

Gene expression profile testing of cancer tissue

Final evidence report

February 16, 2018

Health Technology Assessment Program (HTA)

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

Gene Expression Profile Testing of Cancer Tissue

Final Evidence Report

February 16, 2018

Prepared by:

Center for Evidence-based Policy

Oregon Health & Science University 3030 SW Moody, Suite 250 Portland, OR 97201 Phone: 503.494.2182 Fax: 503.494.3807 <u>http://centerforevidencebasedpolicy.org/</u>



Authors:

Valerie King, MD, MPH, Craig Mosbaek, MPH, Susan Carson, MPH, Brittany Lazur, MPH, Allison Leof, PhD, Robyn Liu, MD, MPH, FAAFP, Curtis Harrod, PhD

The authors would like to acknowledge the contributions of the following:

Galen Gamble, BA, Andrew Hamilton, MLS, MS, Joan Holup, MA, Anitra Ingham, MA, MFA, Chris Kelleher, BA, Adam Obley, MD, those who submitted public comments, and the independent peer reviewers of this report.

This health technology assessment report is based on research conducted by the Center for Evidence-based Policy (Center) under contract to the Washington State Health Care Authority (HCA). This report is an independent assessment of the technology question(s) described based on accepted methodological principles. The findings and conclusions contained herein are those of the authors, who are responsible for the content. These findings and conclusions do not necessarily represent the views of the Washington HCA and thus, no statement in this report shall be construed as an official position or policy of the HCA.

The information in this assessment is intended to assist health care decision makers, clinicians, patients, and policy makers in making evidence-based decisions that may improve the quality and cost-effectiveness of health care services. Information in this report is not a substitute for sound clinical judgment. Those making decisions regarding the provision of health care services should consider this report in a manner similar to any other medical reference, integrating the information with all other pertinent information to make decisions within the context of individual patient circumstances and resource availability.

About the Center for Evidence-based Policy

The Center is recognized as a national leader in evidence-based decision making and policy design. The Center understands the needs of policymakers and supports public organizations by providing reliable information to guide decisions, maximize existing resources, improve health outcomes, and reduce unnecessary costs. The Center specializes in ensuring that diverse and relevant perspectives are considered and appropriate resources are leveraged to strategically address complex policy issues with high-quality evidence and collaboration. The Center is based at Oregon Health & Science University in Portland, Oregon.

<u>Conflict of Interest Disclosures</u>: No authors have conflicts of interest to disclose. All authors have completed and submitted the Oregon Health & Science University form for Disclosure of Potential Conflicts of Interest, and none were reported.

Table of Contents

List of Tables	iii
List of Figures	iv
List of Abbreviations	iv
Executive Summary	1
Structured Abstract	1
Background	
Methods	
Results	
Clinical Practice Guidelines	
Selected Payer Coverage Determinations	
Conclusions	
Technical Report	29
Background	
Technology Description	29
Policy Context	
Washington State Utilization and Cost Data	
Methods	
Search Results	
Evidence Summary	
Clinical Practice Guidelines	
Selected Payer Coverage Determinations	
Conclusions	
References	
Appendix A. Search Strategy	
Appendix B. Additional Methods	
Appendix C. Evidence Tables	111
Appendix D. Risk of Bias Assessments	170
Appendix E. GRADE Profile Ratings	178
Appendix F. Studies Registered at ClinicalTrials.gov	186
Appendix G. Detailed Clinical Practice Guidelines	193
Appendix H. See Attachment for Excluded Studies	196

List of Tables

Table 1. CPT Codes and Descriptions for Gene Expression Profile Tests	38
Table 2. 2015–2017 PEBB/UMP Claims for Gene Expression Profile Testing	40
Table 3. Study Inclusion and Exclusion Criteria	44
Table 4. Recommendations for Lymph Node Status in Guidelines on the Use of Gene Expre Tests in Early-Stage Breast Cancer	
Table 5. Evidence Table: Breast Cancer Systematic Reviews	112
Table 6. Evidence Table: Breast Cancer Randomized Controlled Trials	117
Table 7. Evidence Table: Breast Cancer Observational Studies	121
Table 8. Evidence Table: Breast Cancer Economic Studies	146
Table 9. Evidence Table: Prostate Cancer Observational Studies	149
Table 10. Evidence Table: Prostate Cancer Economic Studies	160
Table 11. Evidence Table: Colon Cancer Observational Studies	164
Table 12. Evidence Table: Colon Cancer Economic Studies	169
Table 13. Risk of Bias: Breast Cancer Systematic Reviews and Meta-analyses	170
Table 14. Risk of Bias: Breast Cancer Randomized Controlled Trial	170
Table 15. Risk of Bias: Breast Cancer Observational Studies	171
Table 16. Risk of Bias: Breast Cancer Economic Studies	172
Table 17. Risk of Bias: Breast Cancer Guidelines	173
Table 18. Risk of Bias: Prostate Cancer Observational Studies	174
Table 19. Risk of Bias: Prostate Cancer Economic Studies	175
Table 20. Risk of Bias: Prostate Cancer Guidelines	175
Table 21. Risk of Bias: Colon Cancer Observational Studies	176
Table 22. Risk of Bias: Colon Cancer Economic Studies	176
Table 23. Risk of Bias: Colon Cancer Guidelines	177
Table 24. Risk of Bias: Multiple Myeloma Guidelines	177
Table 25. GRADE Quality of Evidence for Breast Cancer	178
Table 26. GRADE Quality of Evidence for Prostate Cancer	181
Table 27. GRADE Quality of Evidence for Colon Cancer	183
Table 28. GRADE Quality of Evidence for Multiple Myeloma	185

List of Figures

Figure 1. Analytic Framework	43
Figure 2. Study Flow Diagram: Breast Cancer	49
Figure 3. Study Flow Diagram: Prostate Cancer	50
Figure 4. Study Flow Diagram: Colon Cancer	51
Figure 5. Study Flow Diagram: Multiple Myeloma	52

List of Abbreviations

aOR	adjusted odds ratio
AHRQ	Agency for Healthcare Research and Quality
ASCO	American Society of Clinical Oncology
BCI	Breast Cancer Index
CI	confidence interval
CMS	Centers for Medicare & Medicaid Services
ER	estrogen receptor
FDA	U.S. Food and Drug Administration
HER2	human epidermal growth factor receptor 2
HR	hazard ratio
HTA	health technology assessment
LDT	laboratory-developed test
LN	lymph node
NCCN	National Comprehensive Cancer Network
NR	not reported
OR	odds ratio
PR	progesterone receptor
RCT	randomized controlled trial
RR	risk ratio
PSA	prostate-specific antigen
U.S.	United States

Executive Summary

Structured Abstract

Purpose

The purpose of this evidence report is to review the clinical utility and cost-effectiveness of selected gene expression profile tests used to guide treatment decisions for breast, prostate, and colon cancers and multiple myeloma. There are a growing number of gene expression profile tests for cancers designed to help inform treatment after diagnosis. The potential benefits of these tests could involve more appropriate treatment decisions and better patient outcomes, including avoidance of treatment-related side effects by forgoing unnecessary treatments.

This evidence review will help to inform Washington's independent Health Technology Clinical Committee as it determines coverage regarding the use of the Oncotype DX, MammaPrint, Prosigna, EndoPredict, Breast Cancer Index (BCI), and Mammostrat tests for early invasive breast cancer; the Decipher, Prolaris, and Oncotype DX tests for prostate cancer; the ColoPrint and Oncotype DX tests for stage 2 or 3 colon cancers; and the Myeloma Prognostic Risk Signature (MyPRS) and SKY92 tests for multiple myeloma.

Key Questions

The following key questions guided this review:

- 1. Effectiveness: What is the clinical utility of gene expression profile testing of cancer tissue to inform treatment decisions for patients with breast, prostate, and colon cancers and multiple myeloma?
 - a. Is there evidence that test results affect treatment decisions?
 - b. Do treatment decisions guided by gene expression profile testing of cancer tissue result in clinically meaningful improvements in patient outcomes?
- 2. Harms: What harms are associated with conducting gene expression profile testing of cancer tissue?
- 3. Special populations: Compared with usual care, do treatment decisions, patient outcomes, or harms after gene expression profile testing of cancer tissue vary by:
 - a. Patient demographics (e.g., age, sex, race/ethnicity)?
 - b. Clinical history (e.g., means of diagnosis, stage or grade of cancer, results of other testing, previous treatments, chronicity)?
 - c. Medical comorbidities?
 - d. Provider type or care setting?
- 4. What are the cost-effectiveness and other economic outcomes of gene expression profile testing used to inform treatment management decisions?

Data Sources

Center researchers conducted searches of Ovid MEDLINE, the Cochrane Database of Systematic Reviews, and the Cochrane Central Register of Controlled Trials for English-language studies published from January 2007 to the present (end search dates for the report were in early November 2017). Searches were conducted for eligible health technology assessment (HTA) and evidence reviews from the Agency for Healthcare Research and Quality (AHRQ), the Blue Cross/Blue Shield HTA program, the U.K. National Institute for Health and Care Excellence (NICE), and the Veterans Administration Evidence-based Synthesis Program. Studies were also identified from the reference lists of included studies, test manufacturer websites, and a dossier that had been submitted to the Washington State Agency Medical Directors' Group in December 2016 for one of the tests. In addition, Center researchers searched for ongoing and recently completed registered trials using the ClinicalTrials.gov database. The evidence sources outlined above and the AHRQ National Guideline Clearinghouse were searched for clinical practice guidelines published in the past five years. Center researchers searched the Centers for Medicare & Medicaid Services (CMS) website for the Medicare Coverage Database for National and Local Coverage Determinations applicable to the state of Washington. The Aetna, Cigna, and Regence websites were searched for coverage policies for these private payers.

Study Selection

Two independent Center researchers screened all titles and abstracts for potential inclusion based on pre-specified inclusion and exclusion criteria. Eligible study designs for Key Questions (KQ) 1, 2, and 3 were systematic reviews (with and without meta-analysis), randomized controlled trials (RCTs), and comparative observational studies that reported clinical outcomes, including survival, decision impact, quality of life, and harms. These study designs were also eligible for KQ 4 with the addition of economic modeling studies for cost-effectiveness. Center researchers performed dual full-text review for inclusion criteria. In instances of disagreement, an independent third screener resolved disputes.

Data Extraction

Using standardized processes and forms, one Center researcher extracted data and a second researcher checked the extraction for accuracy. Two researchers independently assessed the risk of bias of included studies and clinical practice guidelines. A rating of high, moderate, or low risk of bias was assigned to each study or review based on adherence to recommended methods and potential for internal and external validity biases.

Data Synthesis

Database searches and other sources yielded 2,949 citations: 2,005 for breast cancer, 266 for prostate cancer, 431 for colon cancer, and 247 for multiple myeloma. A total of 35 studies met inclusion criteria: 22 for breast cancer, 10 for prostate cancer, 3 for colon cancer, and no studies for multiple myeloma. Center researchers applied the Grading of Recommendations, Assessment, Development, and Evaluation Working Group (GRADE) system to rate the overall quality of evidence on key clinical utility outcomes for each genomic test. Clinical utility is

defined for these purposes as the clinical outcomes associated with test use.¹ These outcomes include mortality, morbidity, quality of life, and changes in clinical management decisions, and harms associated with the use of a test because of false reassurance or concern.¹ There was no high-quality evidence of clinical utility to guide decisions about the use of gene expression profile tests for breast, prostate, or colon cancers. There is moderate-quality evidence that women with early invasive breast cancer (breast cancer that has spread outside the milk duct or lobule into adjacent breast tissue, but where there is no or very minimal spread to nearby lymph nodes) who are at high clinical risk based on the Adjuvant! Online risk assessment tool can safely forego adjuvant systemic chemotherapy if their MammaPrint risk score is low. Moderatequality evidence supports the use of Oncotype DX because of its impact on clinical treatment recommendations and chemotherapy use for women with early invasive breast cancer (particularly its ability to identify low-risk women who would not benefit from adjuvant systemic chemotherapy). There is low-quality evidence about the decision impact of using MammaPrint and low-quality evidence that both Oncotype DX and MammaPrint are cost-effective at conventional thresholds when used to guide treatment decisions among women with early invasive breast cancer. Among the remaining conditions and tests, there is a mix of very lowquality evidence or an absence of evidence to support the use of these tests to improve clinical decision making and important patient outcomes.

Limitations

The risk of bias of included studies varied, but was often high. The evidence base was very limited for assessing the clinical utility, harms, and cost-effectiveness of most of the tests. Populations included in studies were generally not diverse in terms of race, ethnicity, and socioeconomic factors. Although a large number of the studies were conducted in the U.S., many were conducted in Europe, which could limit generalizability to the U.S. context. Given limited evidence regarding effectiveness, economic modeling studies did not have adequate quality estimates of effectiveness to include in these modeling exercises. In addition, the economic studies used a variety of modeling techniques and other assumptions, making direct comparisons among them difficult.

Conclusions

There is moderate-quality evidence to support the use of MammaPrint (for clinically high risk individuals based on the Adjuvant! Online assessment tool) and Oncotype DX breast cancer assay for important outcomes related to early-stage invasive breast cancer with negative or positive lymph nodes. Based on limited economic modeling studies, these tests are likely supported at current conventional economic thresholds for use. There is a mix of low-quality, very low-quality, and no evidence to support the other included tests for prostate cancer, colon cancer, and multiple myeloma. Multiple ongoing clinical trials on most of the tests will be reporting results in the next few years and will hopefully improve the evidence base for decision making regarding the clinical usefulness and economic effects of these tests.

Background

The lifetime risk of developing cancer is about 40%, and one in every five Americans will die from cancer.² Strategies for reducing the burden of cancer include prevention, early diagnosis, and appropriate treatments.³ Common treatments for cancer are surgery, chemotherapy, radiation therapy, hormone therapy, and immunotherapy.⁴ The most appropriate treatments for a particular cancer depend on the cancer's characteristics (e.g., cancer stage and grade), the patient's age and health status, response to previous treatments, and other factors.⁵

In recent years, gene expression profile testing of cancer tissue has been used to help inform decisions on appropriate treatments. Gene expression profile testing identifies the genes in a cancer cell or tissue that are making messenger RNA (mRNA), which carries the genetic information that cancer cells need to make proteins. Some gene expression profile tests are designed to increase the accuracy of the prognosis for a patient with cancer. If a test predicts that a cancer is slow growing or is unlikely to metastasize, then active surveillance of the cancer could be the most appropriate course. If a test predicts that a cancer is likely to progress and metastasize, then more aggressive treatments could be warranted.⁶

However, to have usefulness in clinical practice, a gene expression profile test must do more than predict better or worse prognosis. Clinical utility means that a test has an impact on clinical decision making or can help direct therapy in an actionable way to improve patient outcomes. Gene expression profile tests have been developed for patients with breast, prostate, and colon cancers and multiple myeloma, among others. This report examines the clinical utility and costeffectiveness of selected genomic tests for these four cancers.

Technology Description

This report reviews gene expression profile tests for breast, prostate, and colon cancers and multiple myeloma. The six breast cancer tests are indicated for women with early stage, invasive breast cancer: Oncotype DX breast cancer assay, MammaPrint, EndoPredict, Prosigna Breast Cancer Prognostic Gene Signature Assay (PAM50), Breast Cancer Index (BCI), and Mammostrat. These tests are used to predict the cancer's aggressiveness and thus inform treatment decisions.

Decipher, Prolaris, and Oncotype DX prostate cancer assay are gene expression tests for prostate cancer. The Decipher test is used after radical prostatectomy to predict the probability of metastasis and inform clinical decisions on the use of adjuvant prostate cancer treatments. The other two tests, Prolaris and Oncotype DX prostate cancer assay, are used after an initial diagnosis of prostate cancer to predict the cancer's aggressiveness and thus inform treatment decisions.

ColoPrint and Oncotype DX colon cancer assay are gene expression tests for colon cancer. ColoPrint is performed using fresh or frozen tumor samples from stage 2 colon cancer patients who have undergone surgery with curative intent. The Oncotype DX colon cancer assay is indicated for patients with resected anatomic stage 2, mismatch repair proficient (MMR-P) and stage 3 A/B colon cancers.

My Prognostic Risk Signature (MyPRS) and SKY92 (formerly EMC92) are gene expression tests for multiple myeloma. These tests are used to predict the risk of disease relapse and survival outcomes in patients with multiple myeloma.

Molecular diagnostic tests are regulated by the U.S. Food and Drug Administration (FDA) and Centers for Medicare & Medicaid Services (CMS).⁷ The FDA has exercised discretion in its requirements for approval of in vitro diagnostic assays.⁷ In vitro tests developed, validated, and performed in-house by a specific reference laboratory are required to abide by the Clinical Laboratory Improvement Amendments (CLIA), but FDA clearance and approval is currently not required for these laboratory-developed tests (LDTs).⁷ Most of the tests described here are regulated as LDTs. Two of the breast cancer tests, MammaPrint and Prosigna, have received FDA premarket approval. SkylineDx plans to make MMprofiler, based on the SKY92 test for multiple myeloma, available soon as an LDT in the U.S.⁸ MMprofiler is available in Europe, but is only available in the U.S. for research use.⁹

Policy Context

The number of gene expression profile tests for cancer tissue is increasing. Potential benefits of these tests are better patient outcomes and more appropriate treatment decisions, including avoidance of unnecessary treatments and subsequent treatment-related side effects and costs.

This topic was selected for a health technology assessment because of medium concerns for the safety of these tests, medium to high concerns for efficacy, and high concerns for cost. This evidence review will help to inform Washington's independent Health Technology Clinical Committee as the committee determines coverage regarding selected gene expression profile tests for patients with eligible breast, prostate, or colon cancers or multiple myeloma.

Methods

This evidence review is based on the final key questions published on November 14, 2017.

Population: Adults with breast, prostate, or colon cancers or multiple myeloma

Interventions: Gene expression profile testing of cancer tissue to inform treatment decisions, including the following tests by cancer type:

- Breast Cancer—Oncotype DX breast cancer assay, MammaPrint, EndoPredict, Prosigna Breast Cancer Prognostic Gene Signature Assay (PAM50), Breast Cancer Index (BCI), Mammostrat
- Prostate Cancer—Decipher, Prolaris, Oncotype DX prostate cancer assay
- Colon Cancer—ColoPrint, Oncotype DX colon cancer assay
- Multiple Myeloma—Myeloma Prognostic Risk Signature (MyPRS), SKY92-signature (formerly EMC92)

Comparators: Usual care without gene expression profile testing of cancer tissue, alternate gene expression profile tests (i.e., one test intervention listed above versus another)

Outcomes:

- Patient management decisions (including selection of active surveillance rather than active treatment)
- Clinical outcomes (e.g., morbidity, mortality, quality of life)
- Harms, such as consequences of false-positive or false-negative test results
- Cost-effectiveness and other economic outcomes

Time period for literature search: January 2007 to November 2017

Key Questions

- 1. Effectiveness: What is the clinical utility of gene expression profile testing of cancer tissue to inform treatment decisions for patients with breast, prostate, and colon cancers and multiple myeloma?
 - a. Is there evidence that test results affect treatment decisions?
 - b. Do treatment decisions guided by gene expression profile testing of cancer tissue result in clinically meaningful improvements in patient outcomes?
- 2. Harms: What harms are associated with conducting gene expression profile testing of cancer tissue?
- 3. Special populations: Compared with usual care, do treatment decisions, patient outcomes, or harms after gene expression profile testing of cancer tissue vary by:
 - a. Patient demographics (e.g., age, sex, race/ethnicity)?
 - b. Clinical history (e.g., means of diagnosis, stage or grade of cancer, results of other testing, previous treatments, chronicity)?
 - c. Medical comorbidities?
 - d. Provider type or care setting?
- 4. What are the cost-effectiveness and other economic outcomes of gene expression profile testing used to inform treatment management decisions?

Eligible Studies

Randomized controlled trials (RCTs) and nonrandomized comparative studies, and systematic reviews (with and without meta-analysis) of these two types of studies that assess clinical utility were considered for Key Questions 1, 2, and 3. Cost-effectiveness studies and other comparative economic evaluations, along with systematic reviews (with and without meta-analysis) of these types of studies, were considered for Key Question 4.

Data Sources and Searches

Center researchers conducted a search of the peer-reviewed published literature using Ovid MEDLINE. RCTs, nonrandomized comparative studies, and systematic reviews (with and without

meta-analysis) and health technology assessments of these studies that assess clinical utility were considered for Key Questions 1, 2, and 3. Cost-effectiveness studies and other comparative economic evaluations, along with systematic reviews (with and without meta-analysis) reporting economic outcomes, were considered for Key Question 4. The following electronic databases were searched to identify relevant peer-reviewed studies and reviews:

- Ovid MEDLINE
- Cochrane Database of Systematic Reviews
- Cochrane Central Register of Controlled Trials

The full Ovid MEDLINE search strategies for each of the indications (i.e., breast cancer, prostate cancer, colon cancer, and multiple myeloma) are in Appendix A. The search dates were January 1, 2007 through December 1, 2017. Center researchers also screened reference lists of relevant studies and used lateral search functions such as *related articles* and *cited by*. Citations from the Myriad Genetic Laboratories dossier for coverage of EndoPredict, which was submitted to the Washington State Agency Medical Directors' Group in December 2016, were also considered for inclusion. In addition, these core sources were searched:

- Agency for Healthcare Research and Quality (AHRQ)
- Blue Cross/Blue Shield HTA program
- National Institute for Health and Care Excellence (NICE)—Evidence
- Veterans Administration Evidence-based Synthesis Program

Center researchers scanned manufacturer websites and conducted a general Internet search for appropriate published studies and relevant gray literature.

Two independent Center researchers screened titles and abstracts. For studies that could not be excluded by title and abstract screening, a full-text review for inclusion criteria was performed. In instances of disagreement, an independent third screener resolved disputes.

If Center researchers identified a high-quality systematic review (with or without meta-analysis) addressing any of the key questions, a search was conducted to find eligible studies published after the search dates of the systematic review. Center researchers excluded systematic reviews if all of the included studies were also summarized by a more comprehensive systematic review, a systematic review of a higher methodological quality, or a more recently published systematic review.

Center researchers searched the sources listed above for clinical practice guidelines. In addition, searches of the AHRQ's National Guideline Clearinghouse (guidelines.gov) and websites of relevant professional organizations were conducted. Guidelines published in the past five years were considered for inclusion. In addition, Center researchers searched the CMS website for the Medicare Coverage Database for National and Local Coverage Determinations (NCDs and LCDs) applying to the state of Washington. The Aetna, Cigna, and Regence websites were searched for coverage policies for these private payers.

Center researchers searched the online database of clinical trials (clinicaltrials.gov) maintained by the National Library of Medicine at the National Institutes of Health. Information in this database is provided by the sponsor or principal investigator of clinical studies. Studies are generally registered in the database when they begin, with information updated as the study progresses. The search included the names of the gene expression profile tests and other common names for them: Oncotype DX breast (21-gene), MammaPrint (70-gene), EndoPredict (12-gene), Prosigna (PAM50, 50-gene), Breast Cancer Index (BCI), Mammostrat, Decipher (22gene), Prolaris (46-gene), Oncotype DX prostate (17-gene), ColoPrint (18-gene), Oncotype DX colon (12-gene), Myeloma Prognostic Risk Signature (MyPRS), Myeloma (MyPRS, 70-gene), and SKY92 (EMC92, 92-gene).

Data Abstraction and Quality Assessment

One Center researcher used standardized procedures to extract relevant data from each of the included studies and at least one other investigator cross-checked the data for accuracy. Two independent Center researchers evaluated studies for methodological risk of bias, and critical outcomes disagreement among these assessments was settled by a third independent Center researcher. Each study was assessed using Center instruments adapted from international standards and assessments for methodological quality.¹⁰⁻¹⁵ A rating of high, moderate, or low risk of bias was assigned to each study or review based on adherence to recommended methods and potential for bias or other limitations affecting internal and external validity. The risk-of-bias criteria for all of the study types are in Appendix B.

Center researchers assigned each outcome a summary judgment for the overall quality of evidence based on the system developed by the Grading of Recommendations, Assessment, Development, and Evaluation Working Group (GRADE).^{16,17} The GRADE system defines the overall quality of a body of evidence for an outcome in the following manner:

- **High:** Raters are very confident that the estimate of the effect of the intervention on the outcome lies close to the true effect. Typical sets of studies are RCTs with few or no limitations, and the estimate of effect is likely stable.
- **Moderate:** Raters are moderately confident in the estimate of the effect of the intervention on the outcome. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is different. Typical sets of studies are RCTs with some limitations or well-performed nonrandomized studies with additional strengths that guard against potential bias and have large estimates of effects.
- Low: Raters have little confidence in the estimate of the effect of the intervention on the outcome. The true effect may be substantially different from the estimate of the effect. Typical sets of studies are RCTs with serious limitations or nonrandomized studies without special strengths.
- **Very low:** Raters have no confidence in the estimate of the effect of the intervention on the outcome. The true effect is likely to be substantially different from the estimate of

effect. Typical sets of studies are nonrandomized studies with serious limitations or inconsistent results across studies.

• Not applicable: Researchers did not identify any eligible articles.

Results

Across all four cancer indications and 13 tests of interest, searches and other sources generated 2,949 citations and abstracts: 2,005 for breast cancer, 266 for prostate cancer, 431 for colon cancer, and 247 for multiple myeloma. After applying predefined exclusion criteria (see Table 3 for study inclusion and exclusion criteria and rationale), 35 studies met inclusion criteria (breast cancer: 22 studies; prostate cancer: 10 studies; colon cancer: 3 studies; multiple myeloma: no studies).

KQ1: Clinical Utility

Only two included studies provided information about morbidity and mortality outcomes related to tests used for breast cancer.^{18,19} There was no direct information and little indirect information about quality of life related to use of these tests in clinical decision making. A few studies for breast cancer, colon cancer, and prostate cancer reported outcomes such as changes in decisional conflict, confidence in the decision, and anxiety. Most of the evidence about clinical utility outcomes was based on the test affecting treatment recommendations or decisions for breast, prostate, and colon cancer. These results, along with any available information about indirect quality of life findings, are summarized in the following paragraphs by cancer type. No studies met inclusion criteria for multiple myeloma.

Breast Cancer

The majority of research evidence on clinical utility outcomes (22 of 35 studies) included in this report pertains to the use of gene expression profile tests for early-stage invasive breast cancer. These tests have been available for the longest amount of time and simply have a larger research base. Within this group of studies, the majority of evidence relates to the Oncotype DX test (38 primary studies from three systematic reviews and 10 additional studies). The research base for MammaPrint is moderate (seven primary studies from two systematic reviews and four additional studies) and small for Prosigna (one primary study from a systematic review and two additional studies). EndoPredict, BCI, and Mammostrat have very small research bases consisting of one study each.

Systematic Reviews

A systematic review and meta-analysis by Augustovski et al. reported the global pooled decision impact (defined as the proportion of patients whose treatment decision was altered with use of the Oncotype DX test), and global pooled net chemotherapy change (defined as the difference in the number of patients assigned to receive chemotherapy before vs. after the test), from 15 primary studies of women with lymph node (LN)-negative, early-stage invasive breast cancer.²⁰ A group of studies that used universal rather than selective enrollment reported a pooled decision impact of 28.97% (95% confidence interval [CI], 26.65% to 31.34%), $I^2 = 0.00\%$, and a

corresponding net chemotherapy change of 9.00% (95% CI, 4.00% to 14.00%), $I^2 = 89.00\%$, indicating less overall chemotherapy use.²⁰

The Blok et al. systematic review also presented pooled estimates for the proportion of patients who had a change in treatment recommendation to a more or less intensive recommended treatment strategy (i.e., chemotherapy, endocrine therapy, or no treatment).²¹ These pooled estimates combined studies of patients with LN-negative and LN-positive tumors, although the majority of included studies were of patients with LN-negative tumors.²¹ Four included studies with 790 patients used the MammaPrint test; 22 studies (n = 3,743) examined the Oncotype DX test; and one study each involved the use of Prosigna (n = 200) and EndoPredict (n = 167).²¹ Blok et al. reported that lower proportions of patients received recommendations for treatments that were more invasive (MammaPrint: -17%; Oncotype DX: -14.6%; Prosigna: -12.9%; and EndoPredict: -34%).²¹ Correspondingly, the proportion of patients who were recommended to have less intensive treatment increased with use of all the tests (MammaPrint: +32.2%; Oncotype DX: +51.1%; Prosigna: +37.3%; and EndoPredict: +53.2%).²¹ This review did not provide any risk-of-bias assessment for the included studies and did not perform any assessment of statistical heterogeneity across the groups of studies or provide confidence intervals for the estimates.

The Scope et al. systematic review did not present any pooled estimates because the authors were concerned about heterogeneity among studies.²² In a narrative synthesis, the authors reported that the use of Oncotype DX led to changes in treatment recommendations for 21% to 74% of patients enrolled in these studies.²² Change from a recommendation of chemotherapy to no chemotherapy ranged from 6% to 51% of patients after Oncotype DX use, but in one study the proportion of patients who were recommended to receive chemotherapy actually increased after use of the test.²² Similarly, the authors stated that the use of MammaPrint led to treatment recommendation changes of 10% to 40% of patients, and that between 2% and 32% of patients would have a recommendation that changed from chemotherapy to no chemotherapy after the test was used.²²

The search also identified two RCTs and 15 observational studies published after the search dates for the included systematic reviews. One¹⁹ of the RCTs was conducted using the MammaPrint test, and the other¹⁸ involved the Oncotype DX test alone. Nine²³⁻³¹ of the 15 additional observational studies used the Oncotype DX test, three³²⁻³⁴ used MammaPrint, two^{35,36} were on the use of Prosigna/PAM50, and one³⁷ study involved the BCI test.

Additional Studies

Oncotype DX

Bear et al. conducted a small RCT¹⁸ that was assessed by Center researchers as having a high risk of bias because of a very small sample size of 33 subjects with baseline characteristics that varied between groups, study funding from the test manufacturer, and one author who was employed by the test manufacturer. Women with Oncotype DX scores of 11 to 25 were randomized to neoadjuvant hormone therapy (NHT) or neoadjuvant chemotherapy (NCT);

however, two patients crossed over from the NCT to NHT group and no ITT results were presented.¹⁸ (The Oncotype DX Recurrence Score ranges from zero to 100 and represents the longer-term estimated risk of distant cancer recurrence; a score less than 18 represents a low risk of recurrence. See the section Technology Description for a more detailed description.) The randomized groups were not similar at baseline in terms of race, tumor stage, or menopausal status.¹⁸ The authors reported that the clinical and partial response rates by treatment received were significantly different in an ordinal regression that controlled for age, race, menopausal status, and study site.¹⁸ Women who received NHT had a lower clinical response rate than did women who received NCT (22.2% vs. 36.4% (p = .034).¹⁸ However, clinical response rate is a poor surrogate for survival, and this study provides very little evidence about the clinical utility of the test for important patient outcomes.

Friese et al. conducted a retrospective cohort study using the Surveillance, Epidemiology, and End Results (SEER) registries from Los Angeles County and Georgia, and included 1,527 women with estrogen receptor (ER)-positive, human epidermal growth factor receptor 2 (HER2)-negative, early-stage invasive breast cancer (60% were LN-negative), 778 of whom had a treatment recommendation made on the basis of their Oncotype DX test result compared to a group who had treatment recommendations based on other factors.²⁴ The study population was racially and ethnically diverse and included women from a wide range of educational and income levels.²⁴ Women with low-risk Oncotype DX scores were less likely to receive chemotherapy than women who were not tested (odds ratio [OR], 0.1; 95% CI, 0.1 to 0.2); women with high- and medium-risk Oncotype DX scores were more likely to receive chemotherapy compared to women who were not tested (OR for high-risk recurrence score 2.8 (95% CI, 2.0 to 4.0); OR for medium-risk recurrence score 1.4 (95% CI, 1.1 to 1.7). This study reported that 64% of patients who received the Oncotype DX test found it "very" or "extremely" helpful in making their treatment decision.²⁴ Just under 65% of women with a low-risk score reported that it shifted their opinion away from chemotherapy and slightly over 73% of women with a high-risk result reported that they shifted toward wanting chemotherapy.²⁴ Center researchers assessed this study as having a moderate risk of bias because of nonresponse bias and missing data.

Jasem et al. conducted two retrospective cohort studies using data from the U.S. National Cancer Data Base (NCDB).^{25,26} The NCDB captures approximately 70% of newly diagnosed cancers in the nation, from more than 1,500 Commission on Cancer-accredited facilities.^{25,26} Receipt of adjuvant chemotherapy was highly associated with having an intermediate- or high-risk recurrence score (adjusted odds ratios [aORs] of 12 and 83, respectively), although the analysis did not report the proportions of women with low-risk tests or untested women who received adjuvant chemotherapy.²⁵ The second study by Jasem et al. included more than 30,000 patients with a diagnosis of breast cancer in 2010 through 2012, about a third of whom had the Oncotype DX test ordered.²⁶ Patients who had the test ordered received chemotherapy less often than did those who did not have the test ordered (38% vs. 75%), with an aOR of 0.21 (95%)

Cl, 0.20 to 0.22).²⁶ As with the previous study, patients with intermediate- and high-risk scores were substantially more likely to receive chemotherapy (aOR 4.5 and 19.8, respectively) compared to those with low-risk scores.²⁶ Because of potential for miscoding of tests ordered and incomplete control for confounding variables, Center researchers assessed both of these studies as having a high risk of bias.

Parsons et al. conducted a retrospective cohort study that also used the NCDB.²⁹ They included more than 132,000 women diagnosed between 2010 and 2013 who were aged 18 to 70 and had ER-positive (or borderline), HER2-negative (or borderline) early-stage invasive breast cancer.²⁹ Patients who did not receive the test were more likely to have chemotherapy treatment: OR 1.21 (95% CI, 1.17 to 1.25).²⁹ Patients with intermediate- and high-risk Oncotype DX scores were also more likely to receive chemotherapy (OR 12.9 and OR 87.2) than those who had a low-risk result.²⁹ Center researchers assessed this study as having a high risk of bias because of potential for miscoding of tests ordered and incomplete control for confounding variables.

O'Neill et al. conducted a retrospective cohort study using a five-state (CA, GA, KY, NY, OH) commercial insurance claims database with linkage to Oncotype DX test results from the patent holder and provider of the test.²⁸ The study included approximately 5,000 women age 65 and under, and aimed to enroll those diagnosed with stage 1 or 2 hormone receptor-positive, HER2-negative cancers between 2006 and 2010.²⁸ However, the HER2 status was missing for more than half of the subjects, and about a third of them had histologic grade 3 tumors.²⁸ The authors reported the initiation of endocrine therapy within six months of diagnosis and continuation of the medication.²⁸ Oncotype DX test receipt was associated with endocrine therapy initiation (aOR 2.48; 95% CI, 2.03 to 3.04) and was not associated with medication discontinuation (aOR, 0.93; 95% CI, 0.85 to 1.02).²⁸ Center researchers assessed this study as having a high risk of bias because of misclassification due to missing data, potential for coding errors in claims data, and the inability to control for important confounders not recorded in the database.

Ray et al. used Kaiser Permanente's Northern California tumor registry to examine the relationship between use of the Oncotype DX test and receipt of chemotherapy for the treatment of women with ER-positive, HER2-negative, stage 1 and 2 breast cancers who received treatment from 2005 to 2012.³¹ The racial and ethnic composition of the group was fairly diverse: 71% Caucasian, 14% Asian, 9% Hispanic, and 6% African American.³¹ The database included 1,567 women who received Oncotype DX testing and 5,437 who did not.³¹ In a propensity score-matched analysis (n = 2,923), women who had received the Oncotype DX test were less likely to receive chemotherapy (OR, 0.74; 95% Cl, 0.63 to 0.87).³¹ The absolute reduction in chemotherapy use for women who received the test was 6.2% (95% Cl, 5.3% to 13.6%).³¹ Center researchers assessed this study as having a moderate risk of bias because of its design and calculation of some outcome measures using an OR that would likely inflate the effect estimates, although the authors did attempt to control confounding by conducting a propensity-matched analysis.

Two small before and after studies did not contribute new or different decision impact information and are detailed in the main body of this report.^{27,38} The Evans et al. study reported some information that is indirectly related to quality of life, which is discussed below in the section KQ2: Harms.²³

MammaPrint

Cardoso et al. conducted a large European RCT enrolling women aged 18 to 70 with early-stage invasive breast cancer (most women were postmenopausal and had ER-positive, HER2-negative tumors with negative lymph nodes and grade 1 or 2 tumors). All 6,693 enrolled women underwent clinical risk assessment using a modified version of the Adjuvant! Online tool and genomic risk assessment using the 70-gene signature test (MammaPrint). Only women with discordant clinical and genomic risks were randomized to receive or not receive chemotherapy (n = 2,187 randomized and in the intention-to-treat [ITT] analysis).¹⁹ The groups randomized to chemotherapy or no chemotherapy had very similar five-year survival without distant metastasis rates (95.9% in the chemotherapy group, 94.4% in the no chemotherapy group) with an adjusted hazard ratio (HR) of 0.78 (95% CI, 0.50 to 1.21).¹⁹ Among the group of randomized women who had low clinical risk but high genomic risk, the risks of death and distant metastases were very similar in the chemotherapy versus no chemotherapy groups (95.8% vs. 95.0%; aHR, 1.17; 95% CI, 0.59 to 2.28). Because of test changes that caused different risk classification of patients for a period during the trial, Center researchers assessed this RCT as having a moderate risk of bias.¹⁹

The retrospective cohort study by Kuijer et al. included 2,043 women with early-stage invasive breast cancer surgically treated in the Netherlands between November 2011 and October 2013.³³ Nearly 300 of them received the 70-GS (MammaPrint) test to inform treatment decisions, and for 1,745 women, treatment was determined by standard clinicopathological factors.³³ In a mixed-effects linear regression model, use of the test after controlling for age, incidence year, size and grade of tumor, and axillary node involvement was associated with an absolute 9.5% reduction in the use of chemotherapy (95% CI, -15.7% to -3.3%).³³ Because of the design and the authors' concerns about the inability to control for confounding by indication, Center researchers rated this study's risk of bias as high.

Kuijer et al. also published a prospective before-after study involving 660 women in the Netherlands, with similar characteristics to their retrospective cohort study, who were treated at 33 hospitals during calendar years 2013 through 2015.³² Women enrolled had surgically treated ER-positive, early-stage invasive breast cancer and were eligible for adjuvant chemotherapy treatment.³² About 40% of all patients had a recommendation for chemotherapy before the test was performed.³² After the test results were available, treatment recommendations changed for 51% of the women (95% Cl, 46 to 56), and actual treatment received changed for 52%.³² Oncologists' initial treatment recommendations did not predict the test result well (no to slight agreement, based on a kappa of 0.02).³² Although tests were ordered in accord with the national guideline, ordering was selective and not universal. Thus, it is possible that tests were ordered more when adjuvant treatment was uncertain or there were other unmeasured confounding

factors. For this reason and general issues with the before-after design, Center researchers assessed this study as being at high risk of biasCenter researchers assessed this before-after study as being at high risk of bias.

Tsai et al. conducted a before-after study of 840 women with early-stage invasive breast cancer treated between May 2012 and December 2015 at 58 participating U.S. institutions.³⁴ The study was designed to test whether application of the 70-GS (MammaPrint) test affected treatment decisions among women with an intermediate Oncotype DX score (score of 18 to 30).³⁴ The 70-GS result was highly associated with changes in recommendations to add or remove chemotherapy treatment.³⁴ For patients initially recommended to have chemotherapy, but who had a low-risk 70-GS score, the odds for withdrawing the chemotherapy recommendation were high, although with a wide CI (OR, 108; 95% CI, 18.98 to 4304.77).³⁴ Not surprisingly, for women with an initial recommendation of chemotherapy and a high-risk 70-GS score, the odds of having that chemotherapy recommendation withdrawn were very low (OR, 0.01; 95% CI, 0.001 to 0.04).³⁴ Among all patients, the odds of chemotherapy treatment recommendation withdrawal were 0.64 (95% CI, 0.50 to 0.82).³⁴ Overall, 33.6% of treatment recommendations changed after the 70-GS test.³⁴ Physicians were surveyed about how the addition of the 70-GS test influenced their decision, and they reported that it increased their confidence in the final treatment plan in 78.6%, reduced it in 5.8%, and had no influence in 15.6% of cases.³⁴ Center researchers assessed this study as having a high risk of bias because the study was funded by the test manufacturer who had input on the study design, management of the study, analysis, interpretation of the data, and approval of the manuscript.

Prosigna

Two additional observational studies on the Prosigna (PAM50) test met inclusion criteria^{35,36} and were in broad agreement with the findings of the Blok et al. systematic review.²¹ Both were before-after studies using a single group of patients for whom a treatment recommendation was available prior to the test result being provided, and the final treatment recommendation after the additional information from the test was available.^{35,36} Hequet et al. studied 210 postmenopausal women with ER-positive, HER2-negative, LN-negative breast cancer. Overall, treatment recommendation changes occurred for 18% of the women.³⁵ The direction of change was from a recommendation of no adjuvant chemotherapy to adjuvant chemotherapy for 13% of the women, and from adjuvant chemotherapy to no chemotherapy among 5% of them.³⁵ Physicians reported increased confidence in 39%, decreased confidence in 11%, and no change in confidence in 51% of cases.³⁵ The study also gathered information on women's decisional conflict using the Decisional Conflict Scale (DCS) and on function using a functional assessment measure (the Functional Assessment of Cancer Therapy-General, or FACT-G).³⁵ Patients' overall scores on the DCS decreased 3.5 points after the test results, from a mean of 9.8 to 6.2 (p < .001)³⁵ The overall mean FACT-G increased from a baseline level of 79.4 to a posttest mean of 80.2 (p = .264) and then decreased at six months post-diagnosis to 76.77 (no statistical testing reported).³⁵ The changes on both of these scales are minor and are not likely to

represent clinically meaningful differences.³⁵ Center researchers rated the risk of bias for this study as high because of the study design, selective reporting of outcomes, and funding and authorship by the test manufacturer.

Wuerstlein et al. reported treatment recommendations before and after Prosigna study results were available in the West German Study Group (WSG) Breast Cancer Intrinsic Subtype study.³⁶ Overall, for 18% of cases (n = 198) test results were associated with any change in treatment recommendation.³⁶ In 11% of cases, a no adjuvant chemotherapy recommendation changed to an adjuvant chemotherapy recommendation, and for 2% an adjuvant chemotherapy recommendation changed to one against adjuvant chemotherapy.³⁶ For 5% of women, there was a change in the particular type of chemotherapy regimen.³⁶ Physicians reported increased confidence in their treatment recommendations after test results in 89% of cases.³⁶ Patients reported a decrease in "state-anxiety" scores from a mean of 40.5 before the test to 38.5 after the test (p = .082).³⁶ Center researchers assessed this study as being at a high risk of bias because of the study design, the test manufacturer funded the study, and several authors had financial relationships with the company.

Breast Cancer Index (BCI)

Center researchers identified one additional study by Sanft et al. on the BCI test, a small beforeafter study of 96 women from a single U.S. institution who had completed at least 3.5 years of adjuvant endocrine (hormonal therapies such as tamoxifen or aromatase inhibitors) therapy and were eligible for extended endocrine treatment.³⁷ Overall, 26% of women had a change of treatment recommendation after use of the test, with an overall decrease in recommendations for extended adjuvant endocrine therapy (74% before the test vs. 54% after the test); with a statistically significant OR for reduced use of adjuvant endocrine therapy (OR, 0.14; 95% CI, 0.04 to 0.46).³⁷ More physicians felt "strongly confident" in their treatment recommendation after the test (8% vs. 24%; OR, 4.75; 95% CI, 1.62 to 13.96).³⁷ Center researchers rated the risk of bias for this study as high because of inherent methodological flaws with the study design. In addition, the use of an OR to estimate the effect of an outcome that is not rare, particularly with a small sample size as seen in this study, will overstate the estimate of effect.

Prostate Cancer

Eight studies contributed data for this KQ, all of which were before-after designs that reported treatment recommendations before and after the test result was available.³⁹⁻⁴⁶ Four of these studies used a single group of patients and tracked decision outcomes before and after the test results were provided.^{40,44-46} The other four employed a historical comparator group from a time period when treatment decisions were made without the assistance of genomic testing.^{39,41-43} Four studies provided evidence about the Oncotype DX prostate cancer test^{39,40,42,43} and two provided information about the Prolaris test.^{41,46,47} For both of these tests, men were generally categorized as having very low- to intermediate-risk disease based on clinicopathological criteria. All of these studies were conducted in the U.S. and predominantly enrolled Caucasian men over 60 years old on average.

Because of their study designs, Center researchers rated all of these studies as having a high risk of bias. For Oncotype DX and Prolaris, however, there were consistent findings associating the use of the tests with decreased treatment intensity. Two studies on the Prolaris test found that for between 40% and 70% of patients, the recommended or actual treatments were less invasive or intensive with the use of the test than before test results were available.^{41,46,47} Similar results were reported for all four of the Oncotype DX studies, which found that more men had recommendations for watchful waiting or active surveillance rather than more intensive forms of treatment in three of the studies.^{39,42,43} The magnitude of these changes to noninvasive forms of treatment varied by study, but ranged from 21% to 51% of subjects compared to the group without the test. The fourth study reported that treatment intensity decreased for 15.8%, increased for 8.9% and was unchanged for 38.7%.⁴⁰ Two of these studies also reported that physicians found the test useful and that Oncotype DX test results increased their confidence in treatment recommendations.^{40,43} In addition, Eure et al. reported that 96% of patients found the test to be useful in decision making.⁴³

The Decipher test is used differently for men with prostate cancer. It is indicated for men who have been treated with radical prostatectomy and are making a decision about subsequent additional therapies such as hormonal treatment or adjuvant or salvage radiotherapy. Enrolled subjects generally had similar demographics to those in the studies on the Prolaris and Oncotype DX prostate tests. Two before-after studies using a single group of patients were available on this test.^{44,45} Gore et al. used multivariable regression models to estimate that an independent association between Decipher test use and changes in treatment recommendations was present for a group of men considering adjuvant radiotherapy (OR, 1.48; 95% CI, 1.19 to 1.85), and for a group considering salvage radiotherapy (OR, 1.30; 95% CI, 1.03 to 1.65).⁴⁴ The second study reported that 42% of patients who had a recommendation of any active treatment experienced a change to observation only, and nearly 18% with an initial recommendation of observation had a posttest recommendation of an active treatment strategy.⁴⁵ Gore et al. found that median patient DCS scores for both the adjuvant and salvage radiotherapy groups had a statistically significant decrease after the test results were known.⁴⁴

Although the entire group of studies on gene tests to inform treatment of prostate cancer have a high risk of bias, their findings are consistent regarding an association between test use and decreased treatment intensity and increased decision confidence for patients and physicians. The overall quality of evidence for these findings is very low because of substantial limitations, including use of before-after designs and recommended rather than actual treatments, in addition to the lack of important patient outcomes such as survival or treatment-related morbidity.

Colon Cancer

Two included studies reported decreased treatment intensity with the use of the Oncotype DX colon cancer test. One study had outcomes reported in two articles.^{48,49} The Renfro et al. article reported on patient and physician decisional conflict and perceptions of the test.⁴⁸ Regarding

clinical decision-making outcomes, Svrivastava et al. reported that use of the test resulted in no treatment recommendation change for 55.7% of patients, recommendations for increased intensity of therapy for 11.4%, and decreased intensity recommendations for 32.9%.⁴⁹ Brenner et al. reported actual treatment received compared to initial treatment recommended before the test results were known, and found no change in treatment intensity for 62.1% of patients, increased intensity for 9.7%, and decreased intensity for 28.3%.⁵⁰ Thus, across both studies, patients were more likely to receive decreased rather than increased treatments or recommendations for treatments after use of the Oncotype DX test, but the majority did not have a change.^{49,50} Renfro et al. reported that there was a statistically significant decrease in the mean overall DCS score and in each subscale score after receipt of the test (mean overall change, -7.88; 95% CI, -11.25 to -4.50; p < .001).⁴⁸ However, initial scores were fairly low even before the test and below the threshold generally accepted as associated with decisional delay or uncertainty. There were no decision impact studies for the ColoPrint test.

KQ2: Harms

There was no direct information regarding harms arising from false reassurance or false alarms from gene expression profile testing for any of the tests. One study on the Oncotype DX test for women with breast cancer found that levels of pre- and posttest emotional distress were similar and attributed increased levels of distress among particular groups of women to concurrent receipt of chemotherapy.²³ Knowledge about the benefits and harms of chemotherapy increased and perceived risk of recurrence decreased after receipt of the test.²³ This study was assessed as having a high risk of bias because of its design.

KQ3: Special Populations

Clinical utility studies included in the report did not often report findings stratified by subgroups defined by age, gender, race or ethnicity, or clinical history. No study for any cancer reported subgroup results according to medical comorbidity, provider type, or care setting.

Breast Cancer

Although some studies reported differences in subgroups that were more or less likely to receive testing, these studies did not relate differences in who was tested and clinical utility outcomes such as decision making.^{26,31} For example, Jasem et al. reported that older patients were more likely to receive testing than younger patients, that women were more likely to be tested than men, and that African American women and those without insurance were less likely to be tested, but there was no analysis about whether these factors were also associated with subgroup differences for treatment recommendations or actual treatments received.²⁶ The included studies tended to enroll similar populations in terms of clinical characteristics, although some studies enrolled women with only positive or only negative lymph nodes, and some studies had populations with mixed nodal status.

When specific results were reported for a subgroup of interest, the effect of confounding could usually not be ruled out. For example, Jasem et al., in a study that included only LN-negative

women, reported that younger African American women were more likely to receive a recommendation for chemotherapy even with a low Oncotype DX score after controlling for other factors, but the adjusted OR was 1.33 and although statistically significant, might not represent a clinically significant difference given the potential for bias inherent in the study and small difference in the effect.²⁵ Although Jasem et al. reported small differences in use of adjuvant chemotherapy among groups who received and did not receive the Oncotype DX test, these differences are accompanied by overlapping confidence intervals or are not clinically surprising (i.e., it would be expected that women with stage 3 disease would more often be recommended to have adjuvant chemotherapy, even if they had the test).²⁵ Evans et al. reported that younger women were more likely to receive chemotherapy even when controlling for test result.²³

Prostate Cancer

As described above, the Prolaris and Oncotype DX prostate tests are used on biopsy specimens and the Decipher test is used on surgical prostatectomy tissue. The clinical characteristics of patients in these different clinical situations and who have these tests are likely different, but no study describing subgroup differences related to clinical utility met inclusion criteria.

Colon Cancer

Patients included in these studies had resected stage 2 or 2A colon cancer with mismatch repairproficient (MMR-P) tumors. No included study presented an analysis of any outcome of interest by special population or subgroup.

KQ4: Cost-effectiveness and other economic outcomes

Breast Cancer

The Blok et al systematic review included studies with economic outcomes.²¹ The search retrieved one additional economic study by Hall et al.⁵¹ published after the Blok et al. systematic review²¹ and one decision analytic study⁵² published before the systematic review that reported on the BCI test, which Blok et al. did not include. In addition, the Loncaster et al. study, which is included for KQ1, reported some cost outcomes related to the use of the Oncotype DX test.²⁷

Blok et al.²¹ did not conduct individual study risk of bias assessment in their systematic review, but given the descriptions of the modeling studies included in their review, Center researchers would most likely rated them as having a moderate to high risk of bias. There is more and higher-quality information available for the Oncotype DX test than any of the other tests in this study²¹ and the only two included studies^{53,54} using actual (rather than modeled) patient group inputs both studied the economic impact of the Oncotype DX test and estimated that its use would increase costs. However, given that some modeling studies suggest Oncotype DX dominates treatments strategies without genetic testing, it is difficult to say with any certainty what the economic impact in the U.S. would be. The economic study by Hall et al. for the U.K. Health Technology Assessment program included three tests of interest (Oncotype DX, MammaPrint, and Prosigna [subtype]).⁵¹ Center researchers assessed this study as having a

moderate risk of bias because the model assumptions were based on a limited evidence base. The study authors reported that use of Oncotype DX and Prosigna resulted in small cost decreases (-£108 to -£474) and use of MammaPrint was predicted to result in a small cost increase (+£195); however, none of these estimates were statistically significant). This analysis by Hall et al. also reported small increases in quality-adjusted life years (QALYs) of 0.18 to 0.2 that were also not statistically significant.⁵¹ The overall quality of economic evidence about the Prosigna, EndoPredict, and MammoStrat tests is very low, and the quality of economic evidence for the Oncotype DX and MammaPrint tests is low when the Blok et al. systematic review ²¹ and additional economic analysis by Hall et al.⁵¹are considered together.

Prostate Cancer

The search located one cost-effectiveness modeling study⁵⁵ on the use of Decipher for men who have had a prostatectomy and one⁵⁶ on the budget impact of the Prolaris test for men who have received the diagnosis of localized prostate cancer after a prostate biopsy. In addition, the Albala et al. study reported that the aggregate total cost of care related to prostate cancer was \$2,286 less for men who had received the Oncotype DX test compared to historical costs.³⁹

The Ontario Health Technology Advisory Committee commissioned an economic report on adding the Prolaris test to care for men with localized prostate cancer.⁵⁶ The authors found insufficient data to support a primary economic analysis because the test's effect on patient-important outcomes such as survival or need for subsequent radical prostatectomy is not known.⁵⁶ Therefore, the authors conducted a budget impact analysis of adding the test to public coverage in the provincial health service.⁵⁶ Center researchers assessed this study was assessed to be at moderate risk of bias, largely because of the very limited evidence base of the inputs and the small number of sensitivity analyses that can be considered in this type of analysis. Taking into account the costs of the test and physician visits, and the savings from treatment changes, the net budget impact was estimated to add approximately CAD\$8 million in costs per year.⁵⁶

Lobo et al. conducted a cost-effectiveness study of care guided by the Decipher test for men who have had a radical prostatectomy.⁵⁵ Compared to usual care, this cost-effectiveness analysis found that test-based care increased the average per-person cost of care from \$18,370 to \$23,823, but increased the mean QALY per individual by 0.066 (95% CI, 0.016 to 0.117).⁵⁵ The incremental cost-effectiveness ratio (ICER) was \$90,883,⁵⁵ which is in the middle of the range generally proposed by the Institute for Clinical and Economic Review for ICER interpretation in the U.S.⁵⁷

There were no economic studies of Oncotype DX, and the overall quality of evidence regarding economic outcomes for the Decipher and Prolaris tests is very low because of the risk of bias in the evidence that informs the assumptions and limitations associated with these type of modeling studies.

Colon

One study was identified that addressed economic outcomes for the Oncotype DX colon cancer test. The Alberts et al. study included a decision analysis to determine the cost-effectiveness of using the Oncotype DX test to guide therapy for patients with resected stage 2 MMR-P colon cancer.⁵⁸ This cost-effectiveness analysis used information from the Srivastava et al. study to populate the assumptions for the clinical decision-making within the model.⁴⁹ Center researchers rated this study as having moderate risk of bias because of limited information about treatment effectiveness and outcome data from a single source and lack of complete cost modeling.⁵⁸ Alberts et al. reported slightly lower total lifetime costs (\$991 less) with the test (\$103,775) than without it (\$104,767).⁵⁸

The overall quality of evidence for this outcome is very low for Oncotype DX for colon cancer, and there was no evidence included that pertained to the ColoPrint test.

Multiple Myeloma

No studies met inclusion criteria for this key question.

Summary

Table ES-1 summarizes the GRADE quality of evidence for key outcomes by cancer type and for each test considered for that cancer.

Outcome	Breast Cancer	Prostate Cancer	Colon Cancer	Multiple Myeloma
Clinical Utility— mortality or morbidity	Oncotype DX Very low • ං ං MammaPrint Moderate ••• ං	Not applicable (no eligible studies)	Not applicable (no eligible studies)	Not applicable (no eligible studies)
Clinical Utility— patient management decisions	Oncotype DX Breast Moderate ••• MammaPrint Low •• Prosigna, EndoPredict, BCI, and Mammostrat Very low •••	Oncotype DX Prostate, Prolaris, and Decipher Very low • • •	Oncotype DX Colon Cancer Very low • • • • ColoPrint Not applicable (no eligible studies)	Not applicable (no eligible studies)
Clinical Utility— quality of life	Oncotype DX, Prosigna, and BCI Very low • Other tests Not applicable (no eligible studies)	Oncotype DX, Very low • ୦୦୦ Other tests Not applicable (no eligible studies)	Oncotype DX, Very low • ColoPrint Not applicable (no eligible studies)	Not applicable (no eligible studies)
Harms	Oncotype DX Very low • Other tests Not applicable (no eligible studies)	Not applicable (no eligible studies)	Not applicable (no eligible studies)	Not applicable (no eligible studies)

Table ES-1: Summary of GRADE Quality of Evidence for Selected Outcomes

Outcome	Breast Cancer	Prostate Cancer	Colon Cancer	Multiple Myeloma
Cost- effectiveness and other economic outcomes	Oncotype DX and MammaPrint: Low •••• EndoPredict, Mammostrat, Prosigna, BCI: Very low ••••	Oncotype DX Very low • ः Prolaris and Decipher Not applicable (no eligible studies)	Oncotype DX Very low • ः ColoPrint Not applicable (no eligible studies)	Not applicable (no eligible studies)

Clinical Practice Guidelines

Breast Cancer

The most detailed clinical practice guideline, *Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women with Early-Stage Invasive Breast Cancer*, was published by the American Society of Clinical Oncology (ASCO) in 2016.⁵⁹ ASCO published a guideline update in 2017 modifying the recommendations regarding MammaPrint, which draws upon recently published studies.⁶⁰ Center researchers rated both of these guidelines as having good methodological quality.

The ASCO guidelines outline recommendations for when Oncotype DX breast cancer assay, MammaPrint, EndoPredict, Prosigna, BCI, and Mammostrat should or should not be used in patients with early-stage breast cancer. All of these tests, except for Mammostrat, are recommended for use in patients that have ER-positive/progesterone receptor (PR)-positive, HER2-negative, LN-negative breast cancer.^{59,60} The guidelines recommend against the use of Mammostrat for the following categories of breast cancer: ER-positive/PR-positive, HER2negative (LN-positive or negative); HER2-positive; or ER-negative/PR-negative, HER2-negative, LN-negative.⁵⁹

According to the ASCO guidelines, MammaPrint should not be used in patients with low clinical risk (as defined by the Adjuvant! Online tool as used in the MINDACT study)¹⁹ because women in the low clinical risk category had very good outcomes and did not appear to benefit from chemotherapy even with a genomic high-risk cancer.⁶⁰ MammaPrint is recommended for use in patients with ER-positive/PR-positive, HER2-negative, LN-positive breast cancer who have one to three positive nodes and are at high clinical risk per MINDACT categorization.⁶⁰ Still, these patients should be informed that a benefit of chemotherapy cannot be excluded, particularly among patients with more than one involved lymph node.⁶⁰ The ASCO guidelines recommend against using the other tests in patients with ER-positive/PR-positive, HER2-negative, LN-positive, HER2-negative, LN-positive breast cancer; HER2-positive breast cancer; or ER-negative/PR-negative, HER2-negative, LN-negative breast cancer.⁵⁹

The 2017 National Comprehensive Cancer Network (NCCN) clinical practice guidelines on breast cancer discuss the evidence for Oncotype DX (21-gene breast cancer assay), MammaPrint (70-gene assay), and Prosigna (50-gene assay).⁶¹ According to the guidelines, Oncotype DX may be considered for ER-positive/PR-positive, HER2-negative cancers with pT1, pT2, or pT3, and pN0 or pN1mi \leq 2 mm axillary node metastasis and a tumor greater than 0.5 cm.⁶¹ Oncotype DX can also be considered in certain patients with one to three involved ipsilateral axillary lymph nodes to guide the addition of combination chemotherapy to standard hormone therapy.⁶¹ The NCCN guidelines state that the other gene expression profile tests can be considered to assess risk of cancer recurrence, but that they have not been validated to predict response to chemotherapy.⁶¹ Center researchers rated the NCCN guidelines as having fair methodological quality.

NICE published guidelines in 2013 that assessed the use of Oncotype DX breast cancer assay, MammaPrint, Mammostrat, and immunohistochemical 4 (IHC4) score in early-stage breast cancer.⁶² The guidelines recommend Oncotype DX as an option for guiding adjuvant chemotherapy decisions for people with ER-positive, HER2-negative, LN-negative early-stage breast cancer when the patient is assessed as being at intermediate risk.⁶² According to the guidelines, Oncotype DX should only be used when the test results are likely to help in predicting the course of the disease, and therefore in the decision of whether to prescribe chemotherapy.⁶² MammaPrint and Mammostrat are only recommended for use in research in patients with ER-positive, HER2-negative, LN-negative early-stage breast cancer.⁶² Center researchers rated the NICE guidelines as having good methodological quality.

The European Society for Medical Oncology (ESMO) published breast cancer clinical practice guidelines in 2015.⁶³ The ESMO guidelines recommend that gene expression profile tests, such as Oncotype DX breast cancer assay, MammaPrint, EndoPredict, and Prosigna, can be used to complement pathology assessments to predict the benefit of adjuvant chemotherapy.⁶³ Center researchers rated the ESMO guidelines as having poor methodological quality. The European Group on Tumor Markers (EGTM) published a guideline in 2017 on the use of biomarkers in breast cancer.⁶⁴ These guidelines recommend that the Oncotype DX breast cancer assay, MammaPrint, EndoPredict, Prosigna, and BCI can be used to aid in adjuvant therapy decision-making in ER-positive, HER2-negative, LN-negative patients.⁶⁴ In addition, Oncotype DX, MammaPrint, EndoPredict, and Prosigna can be used in patients with one to three metastatic lymph nodes.⁶⁴ Center researchers rated the EGTM guidelines as having poor methodological quality.

Table ES-2 summarizes these five guidelines on breast cancer, indicating whether the gene expression profile tests are recommended for LN-negative and/or LN-positive cancers.

Table ES-2. Recommendations for Lymph Node Status in Guidelines on the
Use of Gene Expression Profile Tests in Early-Stage Breast Cancer

Test	ASCO	NCCN	NICE	ESMO**	EGTM
Oncotype DX	LN-negative	LN-negative LN-positive	LN-negative	LN-negative LN-positive	LN-negative LN-positive
MammaPrint	LN-negative LN-positive	Not recommended*	Not recommended	LN-negative LN-positive	LN-negative LN-positive
EndoPredict	LN-negative	Not recommended*	No guideline recommendation	LN-negative LN-positive	LN-negative LN-positive
Prosigna	LN-negative	Not recommended*	No guideline recommendation	LN-negative LN-positive	LN-negative LN-positive
Breast Cancer Index	LN-negative	Not recommended*	0	No guideline recommendation	LN-negative
Mammostrat	Not recommended	Not recommended*		No guideline recommendation	No guideline recommendation

*NCCN guidelines state that use of 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. **The ESMO guideline authors did not distinguish between LN-negative and LN-positive cancers in their recommendations.

Prostate Cancer

Two clinical practice guidelines were identified that included recommendations on the use of Decipher, Prolaris, and Oncotype DX prostate cancer assay. The 2017 NCCN guidelines on prostate cancer stated that men with clinically localized prostate cancer may consider the use of tumor-based molecular assays and made specific recommendations on the use of Decipher, Prolaris, and Oncotype DX for prostate cancer.⁶⁵ The guidelines recommend Decipher after a radical prostatectomy for patients with pT2 (confined to prostate) with positive margins, any pT3 (extraprostatic extension) disease, and a rising prostate specific antigen (PSA) level.⁶⁵ Prolaris and Oncotype DX are recommended post-biopsy for low- and very low-risk prostate cancer in patients with at least 10 years of life expectancy who have not received other active treatment for prostate cancer and who are candidates for active surveillance or definitive therapy.⁶⁵ Center researchers rated the NCCN guidelines as having fair methodological quality.

A guideline on clinically localized prostate cancer has been published jointly by the American Urological Association, American Society for Radiation Oncology, and Society of Urologic Oncology in 2017.⁶⁶ These guidelines include the following recommendation based on expert opinion: "Tissue-based genomic biomarkers have not shown a clear role in active surveillance for localized prostate cancer and are not necessary for follow up."^{66(p. 4)} Center researchers rated these guidelines as having good methodological quality.

Colon Cancer

No clinical practice guidelines were found that included recommendations for the use of ColoPrint or Oncotype DX for colon cancer. The American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and ASCO published a guideline on molecular biomarkers for colorectal cancer in 2017.⁶⁷ The guideline authors stated, "A problem of quantitative assays, such as gene expression, microRNA expression, and methylation levels, tested in solid tumors, results from the intrinsic mixed nature of the tissue with significant variability of tumor and nontumor tissue content. Another limitation of molecular biomarker discovery approaches that rely on expression levels is that these biomarkers have not been evaluated in the context of complex molecular regulation of individual cancer subtypes."^{67(p. 1482)} Center researchers rated these guidelines as having good methodological quality.

The fair-methodological-quality 2017 NCCN guideline authors discussed multigene assays, including ColoPrint and Oncotype DX colon cancer assay, and concluded that there is no evidence of predictive value in terms of the potential benefit of chemotherapy for any of the multigene assays.⁶⁸ Similarly, the authors of the ESMO 2016 guidelines on metastatic colon cancer concluded that gene expression signatures have failed to accurately predict disease recurrence and prognosis.⁶⁹ Center researchers rated the ESMO guidelines as having poor methodological quality.

Multiple Myeloma

The 2017 NCCN guidelines on multiple myeloma discussed gene expression profiling tests, including MyPRS and SKY92, but did not make any recommendations about the use of these tests.⁷⁰ The NCCN panel unanimously agreed that although gene expression profile tests are not routinely used, they could be helpful in selected patients to estimate the aggressiveness of the disease and to individualize treatment.⁷⁰ Center researchers rated the NCCN guidelines as having fair methodological quality. The 2017 guidelines on multiple myeloma from ESMO stated that gene-expression profiling is not currently used routinely, and more research is needed to identify molecular markers, which could lead to advances in this area.⁷¹ Center researchers rated the ESMO guidelines as having fair methodological quality. No other clinical practice guidelines were found that included recommendations for the use of MyPRS or SKY92 tests.

Selected Payer Coverage Determinations

Breast Cancer

No Medicare National Coverage Determinations (NCDs) were found for any of the gene expression profile tests for breast cancer. Center researchers identified Local Coverage Determinations (LCDs) by Noridian Healthcare Solutions, which apply to Washington, that provide coverage for EndoPredict, Prosigna, and BCI.

The EndoPredict LCD provides coverage for women with T1-3, N0-1 breast cancer when the following criteria are met:

• Patient is postmenopausal

- Pathology reveals invasive carcinoma of the breast that is ER-positive, HER2-negative
- Patient is either LN-negative or has 1 to 3 positive lymph nodes
- Patient has no evidence of distant metastasis
- Test result will be used to determine treatment choice between endocrine therapy alone vs. endocrine therapy plus chemotherapy⁷²

The Prosigna LCD provides coverage for postmenopausal women with either of the following:

- ER-positive, LN-negative, stage 1 or 2 breast cancer or
- ER-positive, LN-positive (one to three positive nodes), stage 2 breast cancer⁷³

The BCI LCD provides coverage for patients who have non-relapsed, ER-positive, LN-negative breast cancer, among other criteria.⁷⁴ No Medicare LCDs covering Washington were found for the Oncotype DX breast cancer assay, MammaPrint, or Mammostrat.

Center researchers assessed private payer policies for Aetna, Cigna, and Regence. The Aetna policy on tumor markers provides coverage for the Oncotype DX breast cancer assay, MammaPrint, EndoPredict, Prosigna, and BCI to assess the necessity of adjuvant chemotherapy in females or males with recently diagnosed breast tumors.⁷⁵ Oncotype DX and MammaPrint are covered for breast cancers that are LN-negative or with one to three involved ipsilateral axillary lymph nodes.⁷⁵ EndoPredict, Prosigna, and BCI are covered for only LN-negative cancers.⁷⁵ Coverage for all of these tests requires that adjuvant chemotherapy is not precluded by any other factor (e.g., advanced age or significant comorbidities) and that the patient and physician have discussed the potential results of the test and agree to use the results to guide therapy, among other criteria.⁷⁵ Aetna does not cover Mammostrat.⁷⁵

The Cigna policy on gene expression assays covers the Oncotype DX breast cancer assay, MammaPrint, and Prosigna under certain conditions, and does not provide coverage for EndoPredict, BCI, or Mammostrat.⁷⁶ Oncotype DX and MammaPrint are covered for LN-negative cancers and for cancers with up to three positive nodes, and Prosigna is covered for only LNnegative cancers.⁷⁶ The Regence policy on gene expression testing for breast cancer provides coverage for the Oncotype DX breast cancer assay, EndoPredict, and BCI under certain conditions, and does not cover MammaPrint, Prosigna, or Mammostrat.⁷⁷ Regence covers the Oncotype DX breast cancer assay, EndoPredict, and BCI for women with primary breast cancer, stages 1, 2, or 3, that are LN-negative, among other criteria.⁷⁷

Prostate Cancer

No Medicare NCDs were found for Decipher, Prolaris, or Oncotype DX for prostate cancer. There are LCDs for Noridian Healthcare Solutions, applying to the state of Washington, that provide coverage for Decipher, Prolaris, and Oncotype DX for prostate cancer under certain conditions. The LCD for Decipher provides coverage after radical prostatectomy when certain conditions are met.⁷⁸ There are two LCDs providing coverage for Prolaris, under certain conditions, one for patients with early stage, needle-biopsy-proven prostate cancer⁷⁹ and the other for patients with

favorable intermediate-risk, needle-biopsy-proven prostate cancer.⁸⁰ The LCD for Oncotype DX for early-stage, needle-biopsy-proven prostate cancer provides coverage with specified condictions.⁸¹

The coverage policies for Aetna⁷⁵ and Regence⁸² consider Decipher, Prolaris, and Oncotype DX prostate cancer assay to be experimental or investigational. Cigna does not include Decipher, Prolaris, or Oncotype DX prostate cancer assay in the list of medically necessary prostate cancer prognostic tests.⁷⁶

Colon Cancer

No Medicare National or Local Coverage Determinations were found for ColoPrint or the Oncotype DX colon cancer assay. The policies for Aetna,⁷⁵ Cigna,⁷⁶ and Regence⁸³ do not cover ColoPrint or Oncotype DX colon cancer assay.

Multiple Myeloma

No Medicare National or Local Coverage Determinations were found for MyPRS or SKY92. The policy for Aetna does not cover MyPRS and does not mention SKY92.⁷⁵ Cigna's coverage policy on tumor markers does not mention MyPRS or SKY92.⁷⁶ The Regence coverage policy states that all microarray-based gene expression profile testing for multiple myeloma is considered investigational.⁸⁴

Conclusions

There was no high-quality evidence of clinical utility to guide decisions about the use of gene expression profile tests for breast, prostate, and colon cancers. The only condition with quality of evidence ratings above very low for any outcome was breast cancer, and only for the MammaPrint and Oncotype DX tests. Based on a single RCT, there is moderate-quality evidence that women with early-stage invasive breast cancer who are considered to be at high clinical risk by the Adjuvant! Online risk assessment tool may safely forego adjuvant systemic chemotherapy if their MammaPrint genomic risk score is low. There was no statistically significant difference in distant metastasis-free survival at five years.

Moderate-quality evidence supports the use of Oncotype DX because of its impact on clinical treatment recommendations and chemotherapy use for women with early-stage invasive breast cancer, particularly its ability to identify low-risk women who would not benefit from adjuvant systemic chemotherapy. There is also low-quality evidence about the decision impact of using MammaPrint for women with early-stage invasive breast cancer.

The review sought information about two distinct types of clinical utility: the impact on patient outcomes such as mortality, morbidity, and quality of life; and the impact on treatment decisions. Most of the evidence found for this review was on the latter. When testing has been found to influence clinical decision making, it could reflect that testing is effective for this purpose or that clinicians are simply more aware of the test. In the absence of patient-important clinical outcomes, there is no assurance that an effect on decision making does, in fact, improve

care and health outcomes. Avoidance of unnecessary treatments, which by definition result in harms without providing benefits, is not inconsequential. However, patients need to be certain that forgoing treatment based on a low-risk genetic test result does not result in worse clinical outcomes later in life.

Based primarily on modeling studies, there is low-quality evidence that both Oncotype DX and MammaPrint are cost-effective at conventional thresholds when used to guide treatment decisions among women with early-stage invasive breast cancer. Among the remaining conditions and tests, there is very low-quality evidence or a complete absence of evidence to support use of these tests to improve clinical decision making and important patient outcomes.

Clinical practice guidelines are generally more liberal in terms of test recommendations, largely because most of these tests can provide information about overall prognosis even if they lack evidence of predictive ability for clinical utility outcomes. For example, ASCO endorses all of the breast cancer tests (with the exception of Mammostrat) for use in LN-negative patients, and also the use of MammaPrint for those with limited lymph node involvement. However, NCCN has adopted a more rigorous standard and advises the use of Oncotype DX for both LN-negative and LN-positive patients and stating that other tests may be considered to help assess risk of recurrence, but have not been validated to predict response to chemotherapy.

There are multiple ongoing clinical trials on most of the tests included in this report (see Appendix F), including some large and important trials that will be reporting results in the next few years. Among these are the TAILORx and RxPONDER breast cancer trials for Oncotype DX and additional results for the MINDACT trial involving MammaPrint for LN-positive women.

Technical Report

Background

The lifetime risk of developing cancer is about 40%, and one in every five Americans will die from cancer.² Strategies for reducing the burden of cancer include prevention, early diagnosis, and appropriate treatments.³ Common treatments for cancer are surgery, chemotherapy, radiation therapy, hormone therapy, and immunotherapy.⁴ The most appropriate treatments for a particular cancer depend on the cancer's characteristics (e.g., cancer stage and grade), the patient's age and health status, response to previous treatments, and other factors.⁵

In recent years, gene expression profile testing of cancer tissue has been used to help inform decisions on appropriate treatments. Gene expression profile testing identifies the genes in a cancer cell or tissue that are making messenger RNA (mRNA), which carries the genetic information that cancer cells need to make proteins. Some gene expression profile tests are designed to increase the accuracy of the prognosis for a patient with cancer. If a test predicts that a cancer is slow growing or is unlikely to metastasize, then active surveillance of the cancer could be the most appropriate course. If a test predicts that a cancer is likely to progress and metastasize, then more aggressive treatments could be warranted.⁶ Other gene expression tests may identify specific mutations for which there are targeted treatments (i.e., the Anaplastic lymphoma kinase [ALK] mutation in lung cancer can be treated with a specific drug). However, to have usefulness in clinical practice, a gene expression profile test must do more than predict better or worse prognosis. Clinical utility means that a test affects clinical decision making or can help direct therapy in an actionable way to improve patient outcomes. Gene expression profile tests have been developed for patients with breast cancer, prostate cancer, colon cancer, and multiple myeloma, among others. This report examines the clinical utility and cost-effectiveness of selected genomic tests for these four cancers.

Technology Description

This section contains descriptions of gene expression profile tests, along with recent studies on their clinical validity, for breast, prostate, and colon cancers and multiple myeloma. Molecular diagnostic tests are regulated by the U.S. Food and Drug Administration (FDA) and the Centers for Medicare & Medicaid Services (CMS).⁷ The FDA has exercised discretion in its requirements for approval of in vitro diagnostic assays.⁷ In vitro tests developed, validated, and performed inhouse by a specific reference laboratory are required to abide by the Clinical Laboratory Improvement Amendments (CLIA), but FDA clearance and approval is currently not required for these laboratory-developed tests (LDTs).⁷ Most of the tests described in this report are regulated as LDTs. Two of the tests, MammaPrint and Prosigna, have received FDA premarket approval.

Breast Cancer

This evidence review examines six tests for women with early-stage invasive breast cancer: Oncotype DX breast cancer assay, MammaPrint, EndoPredict, Prosigna, Breast Cancer Index (BCI), and Mammostrat.

Oncotype DX

The Oncotype DX breast cancer assay uses reverse transcription polymerase chain reaction (RT-PCR, a common technique to detect and quantify mRNA, to measure gene expression in formalin-fixed paraffin-embedded (FFPE) cancer tissue.⁸⁵ The Oncotype DX breast recurrence score test is used to predict the likely benefit of chemotherapy and the risk of distant recurrence among patients newly diagnosed with early-stage estrogen receptor (ER)-positive, human epidermal growth factor receptor 2 (HER2)-negative invasive breast cancer, both lymph node (LN)-negative and LN-positive.⁸⁵ The test measures the expression of 21 genes, including 16 cancer-related genes and five reference genes.⁸⁵ These genes were selected out of 250 rationally selected candidate genes bases on prognostic capacity and consistency in test performance.²¹ The Oncotype DX recurrence score ranges from zero to 100, with categories of low risk (zero to 17), intermediate risk (18 to 30), and high risk (31 to 100).²¹ The Oncotype DX report also provides a quantitative ER score to help assess the magnitude of hormonal therapy benefit.²¹ Oncotype DX is available as an LDT from Genomic Health, licensed under CLIA, and does not require FDA approval.⁸⁶

The 2017 systematic review by Blok et al. included 20 studies assessing the clinical validity of Oncotype DX for predicting recurrence free survival.²¹ All of the included studies in patients with LN-negative disease showed significant differences in outcomes (usually recurrence-free survival) among low-, intermediate-, and high-risk patients.²¹ Most of the studies that included patients with LN-positive cancer showed significant differences in outcomes among the three risk groups.²¹

MammaPrint

Using microarray technology, MammaPrint measures the mRNA expression of 70 genes in fresh tissue or FFPE tissue.⁸⁷ The 70 genes were identified in 2002 from a total of 25,000 genes using supervised clustering.²¹ MammaPrint categorizes patients as having a low or high risk of breast cancer recurrence for women with stage 1 or stage 2 invasive breast cancer that has a tumor size \leq 5.0 cm and is LN-negative, ER-positive or negative, and HER2-positive or negative.⁸⁷ In 2007, Agendia received 510(k) clearance from the FDA for MammaPrint and received clearance for its use in FFPE tissue in 2015.²¹

The Blok et al. systematic review included 21 studies assessing the clinical validity of MammaPrint.⁸⁸ In studies of patients with LN-negative breast cancer, MammaPrint was reported to be of significant prognostic value for risk-of-recurrence outcomes.⁸⁸ For LN-positive disease, the studies reported statistically significant hazard ratios for distant metastases-free survival and breast cancer-specific survival between low- and high-risk MammaPrint groups.⁸⁸ In the remaining studies, without specific classification or with LN-positive and LN-negative groups combined, most of the results showed a significant difference in risk-of-recurrence outcomes between MammaPrint risk groups.⁸⁸
EndoPredict

EndoPredict tests FFPE tissue to estimate the 10-year risk of distant recurrence of early-stage ER-positive, HER2-negative invasive breast cancer, both LN-negative and LN-positive.⁸⁹ EndoPredict produces a 12-gene molecular score using quantitative RT-PCR on eight signature genes, three normalization genes, and one control gene, with the score ranging from zero to 15.²¹ This 12-gene molecular score is combined in an algorithm with tumor size and nodal status to calculate an EPclin score, which has a clinically reportable range of 1.1 to 6.2.⁸⁹ EPclin scores from 1.1 to 3.3 indicate a low 10-year risk for distant recurrence, and scores from 3.4 to 6.1 indicate a high risk for 10-year distant recurrence.⁸⁹ EndoPredict is available as an LDT from Myriad Genetics, licensed under CLIA, and does not require FDA approval.⁹⁰

Four studies included in the Blok et al. systematic review assessed the clinical validity of EndoPredict.²¹ Three of these studies compared low-risk patients to high-risk patients on risk-of-recurrence outcomes, and all three studies found statistically significant hazard ratios ranging from 1.19 to 1.31, indicating that patients with low-risk EndoPredict scores were more likely to have recurrence-free survival.²¹ The fourth study found a low proportion of distant metastases in the low-risk group, but did not compare outcomes between low-risk and high-risk groups.²¹

Prosigna

The Prosigna breast cancer gene signature assay, based on the PAM50 gene signature, uses quantitative PCR performed on RNA isolated from FFPE breast tumor tissue to classify tumors into four intrinsic subtypes: luminal A, luminal B, HER2-negative enriched, and basal-like.^{91,92} Prosigna is indicated for use in postmenopausal women with hormone receptor-positive, LN-negative (stage 1 or 2), or LN-positive (stage 2) breast cancer with a clinicopathological indication to be treated with adjuvant endocrine therapy.⁹² The Prosigna algorithm produces a risk-of-recurrence score using a measure of the expression of 46 genes, intrinsic subtype, tumor size, nodal status, and proliferation score.²¹ The risk-of-recurrence score is on a zero to 100 scale, which correlates with the probability of distant recurrence in a 10-year period.⁹² In 2013, NanoString Technologies received 510(k) clearance from the FDA for Prosigna.⁸⁸

Five studies included in the Blok et al. systematic review assessed the clinical validity of Prosigna.²¹ Four of the five included studies showed that risk-of-recurrence outcomes were significantly different among risk categories as determined by the Prosigna test; the fifth study did not find a statistically significant difference.²¹

Breast Cancer Index

Breast Cancer Index (BCI) predicts the likelihood of benefit from extended endocrine (hormonal) therapy among women with ER-positive, LN-negative or LN-positive (with one to three positive nodes), early-stage invasive breast cancer.⁹³ BCI measures the expression of 11 genes using quantitative RT-PCR performed on FFPE tissue.⁹⁴ BCI was developed from two gene expression-based biomarkers: the HOXB13:IL17BR ratio, which is associated with tumor responsiveness to endocrine therapy in breast cancer, and the Molecular Grade Index, which consists of the

average expression of five cell cycle-associated genes.⁹⁴ The BCI score ranges continuously from zero to 10. The categories for risk of late (post-five year) distant recurrence are low risk (< 5.0825) and high risk (\geq 5.0825).⁹⁴ The categories for risk of overall (zero to 10 year) distant recurrence are low risk (< 5.0825), intermediate risk (\geq 5.0825 to 6.5025), and high risk (\geq 6.5025).⁹⁴ BCI also predicts the likelihood of benefit from extended (> five year) extended endocrine (hormonal) therapy.⁹³ BCI is available as an LDT from Biotheranostics, licensed under CLIA, and does not require FDA approval.⁹⁰

In a 2013 study by Zhang et al., the BCI model was validated by retrospective analyses of tumor samples from tamoxifen-treated patients from a randomized prospective trial (n = 317) and a multi-institutional cohort (n = 358).⁹⁵ In the first cohort, BCI categorized 65% of patients as low risk, and less than 3% of low-risk patients had distant recurrence in years zero to five and in years five to $10.^{95}$ The second cohort generally had larger tumors, and 55% of patients were classified as BCI low risk.⁹⁵ This low-risk group had less than 5% distant recurrence in years zero to five and years five to $10.^{95}$

Mammostrat

The Mammostrat test, launched by Clarient in 2010, is based on the immunohistochemical (IHC) assay of the expression of five genes in hormone receptor positive, stage 1 or 2 breast cancer cells.⁹⁶ An algorithm is used to calculate a risk index score, with patients categorized as low (≤ 0), moderate (0 to ≤ 0.7) or high (> 0.7) risk for cancer recurrence.²² This test is not technically a molecular gene expression profile test but is included in this group because it is used for similar purposes in practice; this evidence review does not make a distinction between Mammostrat and the other tests under discussion. The five genes were selected from 700 gene targets in gene expression assays in three patient cohorts.⁹⁶ Mammostrat was developed and validated using clinical samples from patients enrolled in the National Surgical Adjuvant Breast and Bowel Project B14 and B20 trials.⁹⁷ Mammostrat is available as an LDT from Clarient, licensed under CLIA, and does not require FDA approval.⁹⁸

Prostate Cancer

Decipher is a gene expression profile test used after prostatectomy to predict the probability of metastasis and inform clinical decisions on the use of adjuvant prostate cancer treatments. Two other tests, Prolaris and Oncotype DX for prostate cancer, are used with biopsy tissue after an initial diagnosis of prostate cancer to predict the cancer's aggressiveness and thus inform initial treatment decisions.

Decipher

Decipher is a test designed to predict the probability of metastasis within five years of radical prostatectomy.⁹⁹ Decipher analyzes a tissue sample removed during surgery that is routinely archived or stored by the pathology laboratory.⁹⁹ Decipher measures the expression levels of 22 RNA biomarkers involved in biological pathways across the genome that are associated with aggressive prostate cancer.⁹⁹ The panel is designed to evaluate the expression of genetic

markers associated with specific biological processes, including cell proliferation and differentiation processes, cell structure, adhesion, motility, immune response, and mitosis processes, cell cycle progression (CCP), and other functional processes.¹⁰⁰ The Decipher Genomic Classifier score ranges continuously from zero to one; a higher score indicates a higher probability of clinical metastasis.¹⁰¹ The risk categories for Decipher scores are low (< 0.45), intermediate (0.45 to 0.60), and high (> 0.60).¹⁰² Decipher is available as an LDT from GenomeDx, licensed under CLIA, and does not require FDA approval.⁷

A recent systematic review by Spratt et al. included a meta-analysis of 855 patients who had the Decipher test and were followed-up for a median of eight years.¹⁰³ Patients were classified by Decipher as low, intermediate, or high risk.¹⁰³ The 10-year cumulative incidence metastases rates for the low-, intermediate-, and high-risk groups were 5.5%, 15.0%, and 26.7% (p = .001), respectively (HR per 0.1 unit change in Decipher score, 1.52; 95% CI, 1.39 to 1.67).¹⁰³ After adjusting for clinicopathological variables, Decipher remained a statistically significant predictor of metastasis risk (HR per 0.1 unit, 1.30; 95% CI, 1.14 to 1.47; p = .001)¹⁰³ The authors concluded that Decipher has good clinical validity and can independently and significantly improve prognostic accuracy for patients after a prostatectomy, within nearly all clinicopathologic, demographic, and treatment subgroups.¹⁰³

Prolaris

The Prolaris CCP test uses a quantitative RT-PCR assay on FFPE biopsy tissue to determine cancer aggressiveness and predict the probability of disease progression in patients with lowand intermediate-risk prostate cancer.^{104,105} Prolaris uses a 46-gene expression signature, including 31 CCP genes and 15 housekeeping genes.¹⁰⁶ The CCP score ranges from zero to 10, with each unit increase representing a doubling of risk of disease progression.¹⁰⁴ The CCP report provides an estimate of a patient's 10-year prostate cancer–specific mortality risk, from an algorithm combining the CCP score and clinicopathological variables.⁵⁶ Prolaris is available as an LDT from Myriad Genetic Laboratories, licensed under CLIA, and does not require FDA premarket review.⁷

A 2016 systematic review and meta-analysis by Sommariva et al. assessed the clinical validity of Prolaris.¹⁰⁷ In a meta-analysis of four studies (n = 1,489), the hazard ratio for biochemical recurrence after prostatectomy per one-unit increase in the CCP score was 1.88 (p < .001) in a univariate model and 1.63 (p < .001) in a multivariable model adjusted for clinical and pathological variables, such as extent of disease, age, clinical stage, and use of hormones.¹⁰⁷ For example, a patient with a CCP score of five has a 1.63 times greater likelihood of biochemical recurrence than a patient with a score of four, if all other factors are equal.

Oncotype DX for Prostate Cancer

The Oncotype DX genomic prostate score assay analyzes prostate cancer gene activity to predict disease aggressiveness among men with clinically low-risk prostate cancer.¹⁰⁸ This test was developed by Genomic Health and was based on studies evaluating 727 genes among patients

undergoing radical prostatectomy between 1987 and 2004 in Cleveland, Ohio.¹⁰⁰ Oncotype DX uses a quantitative RT-PCR assay performed on approximately 1 mm FFPE tissue from a prostate needle biopsy.¹⁰⁶ Oncotype DX measures the expression of 17 genes: 12 cancer-related genes across four genetic pathways (stromal response, androgen signaling, proliferation, and cellular organization) and five housekeeping genes.¹⁰⁸ Gene expression in conjunction with clinical risk factors are used to predict the likelihood of adverse pathology, with Genomic Prostate Scores ranging from zero to 100.¹⁰⁸ Oncotype DX is available as an LDT from Genomic Health, licensed under CLIA, and does not require FDA approval.⁷

In 2016, Brand et al. published a patient-level meta-analysis of two independent clinical validation studies (n = 732) of Oncotype DX for prostate cancer, along with the Cancer of the Prostate Risk Assessment (CAPRA) score and National Comprehensive Cancer Network (NCCN) risk group, as predictors of the likelihood of favorable or adverse pathological staging after surgery.¹⁰⁹ In a decision curve analysis, greater net benefit was shown when combining Oncotype DX score with each clinical classifier compared with the classifier alone.¹⁰⁹ By adding the Oncotype DX score to the CAPRA score, the area under the receiver operating characteristic curve (AUROC) improved from 0.68 to 0.73 (p < .001), and adding Oncotype DX to the NCCN risk group improved the AUROC from 0.64 to 0.70 (p < .001).¹⁰⁹ These AUROCs are generally considered by biostatisticians to represent poor to fair test accuracy.

Colon Cancer

Two gene expression profile tests are described below: ColoPrint, which can be used in stage 2 colon cancer patients, and Oncotype DX for stage 2 and 3 colon cancer patients.

ColoPrint

Released by Agendia in 2012, ColoPrint was developed using whole-genome expression data to identify genes that have the highest correlation to a tumor-relapse in stage 2 colon cancer patients.¹¹⁰ ColoPrint's microarray-based 18-gene expression signature is performed using fresh or frozen tumor samples from stage 2 colon cancer patients who have undergone surgery with curative intent.¹¹¹ ColoPrint categorizes patients as being at low or high risk of distant recurrence.¹¹⁰ ColoPrint is available as an LDT from Agendia, licensed under CLIA, and does not require FDA approval.¹¹²

A 2015 review by Kopetz et al. included a pooled analysis of studies assessing the risk of recurrence in 416 stage 2 colon cancer patients who had available ColoPrint results.¹¹¹ High-risk patients, as identified by ColoPrint, had a statistically significantly higher risk of recurrence compared to low-risk patients (HR, 2.16; 95% CI, 1.28 to 3.65; p = .004).¹¹¹ High-risk patients had a five-year risk of recurrence of 20.9% (95% CI, 14.2% to 27.6%) and low-risk ColoPrint patients had a five-year risk of recurrence of 10.3% (95% CI, 6.6% to 14%).¹¹¹ The hazard ratio remained statistically significant in a multivariable model that included the number of lymph nodes retrieved and a measure of microsatellite instability.¹¹¹

Oncotype DX for Colon Cancer

The Oncotype DX colon cancer assay from Genomic Health is indicated for patients with resected anatomic stage 2, mismatch repair proficient (MMR-P), and stage 3 A/B (cancer has not spread to distant organs, but cancer has spread to one to three lymph nodes and grown into the submucosa or into the outermost layer of the colon) colon cancers.¹¹³ The test was developed from four studies that explored the expression of 761 genes in patient tumor samples that were obtained at the time of initial diagnosis of colon cancer.¹¹⁴ The 12-gene test (seven cancer-related genes and five reference genes) produces a score of zero to 100 and then categorizes patients as having low risk (< 30), intermediate risk (30 to 40), or high risk (\geq 41) of recurrence.¹¹³ Oncotype DX is available as an LDT from Genomic Health, licensed under CLIA, and does not require FDA premarket review.¹¹²

A 2016 study by Yamanaka et al. assessed the predictive ability of Oncotype DX in 597 stage 2 and 3 colon cancer patients.¹¹⁵ The continuous Oncotype DX recurrence score was significantly associated with recurrence-free interval after adjustment for disease stage (HR for 25-unit increase in recurrence score, 2.05; 95% CI, 1.47 to 2.86; p = .001).¹¹⁵

Multiple Myeloma

The two tests described below, My Prognostic Risk Signature (MyPRS) and SKY92, are used as prognostic tests in patients with multiple myeloma.

My Prognostic Risk Signature

MyPRS was developed at the University of Arkansas for Medical Science and then licensed to Signal Genetics.¹¹⁶ The MyPRS test includes a 70-gene Prognostic Risk Score (GEP-70), which quantifies the expression of 70 genes commonly altered in multiple myeloma.¹¹⁷ Using a scale of zero to 100, the test predicts the risk of disease relapse and survival outcomes and classifies patients as low or high risk.¹¹⁷ MyPRS was available as an LDT from Signal Genetics, is licensed under CLIA, and does not require FDA approval.¹¹⁸ In November 2016, Signal Genetics sold its MyPRS assets to Quest Diagnostics.¹¹⁸

A 2007 study by Shaughnessy et al. analyzed tumor cells from 532 newly diagnosed patients with multiple myeloma using the 70-gene test.¹¹⁹ The authors found that a high-risk score was an independent predictor of event-free (HR, 3.41; p = .002) and overall survival (HR, 4.75; p < .001).¹¹⁹

SKY92

The SKY92-signature (formerly EMC92) was discovered at the Erasmus Medical Centre in the Netherlands and measures the expression level of 92 genes that are directly or indirectly related to multiple myeloma.⁹ The SKY92 test is used to predict how likely multiple myeloma is to progress and to aid in treatment decision making.⁹ Starting in 2015, SkylineDx began to produce the MMprofiler test, which uses the SKY92 gene signature to classify multiple myeloma patients into high- or standard-risk groups.⁹ MMprofiler is available in Europe, but is only available in the

U.S. for research use and currently not permitted for clinical diagnostic use.⁹ SkylineDx plans to make MMprofiler available soon as an LDT in the U.S.⁸

A 2015 study by Kuiper et al. of 4,750 multiple myeloma patients assessed the value of 20 markers for predicting patients' overall survival.¹²⁰ The study found that the combination of SKY92 and the International Staging System provided the strongest predictor for overall survival among the 20 markers tested.¹²⁰ Median survival was 24 months for the highest risk group, 47 months for the intermediate-high risk group, 61 months for the intermediate-low risk group, and the median was not reached after 96 months for the lowest risk group.¹²⁰ Relative to the lowest risk group, the hazard ratios were significantly higher in the intermediate-low group (HR, 2.6; 95% CI, 1.6 to 4.5), intermediate-high group (HR, 3.2; 95% CI, 1.9 to 5.4), and high group (HR, 6.9; 95% CI, 4.1 to 11.7).¹²⁰

Policy Context

The number of gene expression profile tests for cancer tissue is increasing. Potential benefits of these tests are more appropriate treatment decisions and better patient outcomes, including avoidance of unnecessary treatments and subsequent treatment-related side effects and costs.

This topic was selected for a health technology assessment because of medium concerns for the safety of these tests, medium to high concerns for efficacy, and high concerns for cost. This evidence review will help to inform Washington's independent Health Technology Clinical Committee in determining coverage regarding selected gene expression profile tests for patients with eligible breast, prostate, or colon cancers or multiple myeloma.

Washington State Agency Utilization and Costs

Programs and Populations

The *Gene Expression Profile Testing of Cancer Tissue* review includes utilization and cost data extracted from the Public Employees Benefit Board Uniform Medical Plan (PEBB/UMP) and HCA Medicaid (fee-for-service) and the managed care organization (MCO) Medicaid programs. The Department of Labor and Industries Workers' Compensation Plan had no qualifying claims. The PEBB/UMP populations includes members also enrolled in Medicare. For this analysis, PEBB/UMP Medicare claims were excluded.

The chart below shows the number of people in Washington Medicaid from 2014 to 2017. In 2017, 46% of Medicaid recipients were 18 years or younger, and 54% were 19 years or older.



Among the PEBB/UMP population, 16% were 17 years or younger, and 84% were 18 years or older in 2017. The chart below shows the number of people on PEBB/UMP from 2014 to 2017.



Gene expression profile testing of cancer tissue: final evidence report

Methods

Population utilization was based on claims data; paid claims identified gene expression profile tests. Analysis found that only the Oncotype DX breast cancer assay had a unique Current Procedure Terminology (CPT) code (81519) assigned to the test. The majority of propriety tests reviewed lacked uniquely assigned CPT codes; they utilized "generalized" codes.

Therefore, this analysis utilized a methodology that combined recognized (for billing purposes) non-specific CPT codes with appropriate International Statistical Classification of Diseases and Related Health Problems (ICD) diagnosis codes (version 9 or version 10) as representative, or as surrogate measures of a gene expression profile test. Diagnosis codes for the following cancers were used to categorize nonspecific CPT codes: breast, prostate, and colon cancer, and multiple myeloma. Data included three years of claims; denied claims were excluded.

Limitations

Use of nonspecific representative coding was a limitation of this analysis. Because overall utilization was approximate, performing detailed analyses was impractical. Given these constraints, only data from the Oncotype DX breast cancer test (81519) are reported in detail.

The analysis period included three calendar years, 2015 to 2017. Because of the timing of the data extraction, this analysis lacks the normal 90 days of claims payments following the close of a financial or accounting period for 2017. Claims are included when the following were true: age of enrollee was 18 years or older, at least one of the specific CPT/H codes from Table 1 was used, and claim included ICD 10 diagnosis of breast, prostate, or colon cancers or multiple myeloma. Denied claims were excluded from the analysis.

CPT Code	CPT Description	Note
81519	Oncology (breast), mRNA, gene expression profiling by real-time RT-PCR of 21 genes, utilizing formalin-fixed paraffin-embedded tissue, algorithm reported as recurrence score	Code is specific to Oncotype DX Breast Cancer Assay
81599	Unlisted multianalyte assay with algorithmic analysis	For this analysis, this unlisted code is used in combination with diagnosis codes to estimate utilization of test that do not have specific assigned codes.
88381	88381: <u>Microdissection</u> (i.e., <u>sample</u> <u>preparation</u> of microscopically identified <u>target</u>); <u>manual</u>	Identified via proxy; utilizing two CPT codes and a diagnosis of breast cancer. The test provides a risk category and numerical score to assess a patient's risk of distant recurrence of disease at 10 years in

Table 1. CPT Codes and Descriptions for Gene Expression Profile Tests

CPT Code	CPT Description	Note
		postmenopausal women with node-negative (Stage I or II) or node-positive (Stage II), hormone receptor-positive (HR+) breast cancer.
		NOTE: new specific CPT codes for 2018 include 81520 (Prosigna) and 81521 (MammaPrint) and their definitions are listed below. 81520: Oncology (breast), mRNA gene expression profiling by hybrid capture of 58 genes (50 content and 8 housekeeping), utilizing formalin-fixed paraffin-embedded tissue, and algorithm reported as a recurrence risk score.
		81521: Oncology (breast), mrna, microarray gene expression profiling of 70 content genes and 465 housekeeping genes, utilizing fresh frozen or formalin-fixed paraffin- embedded tissue, algorithm reported as index related to risk of distant metastasis
81525	Oncology (colon), mRNA, gene expression profiling by real-time RT-PCR of 12 genes (7 content and 5 housekeeping), utilizing formalin-fixed paraffin-embedded tissue, algorithm reported as a recurrence score	Identified via proxy utilizing CPT codes and a diagnosis of colon cancer. Code used to identify Oncotype DX for colon cancer
81479	Unlisted molecular pathology procedure	

Table 2 below presents the total number of claims and total allowed charges for three calendar years, 2015 to 2017, for PEBB/UMP utilization.

Condition/Test	Services	Allowed Dollars
Breast cancer	126	\$515,650
Prostate	1	\$165
Colon cancer	4	\$1,859
Multiple myeloma	13	\$5,558

Table 2. 2015–2017 PEBB/UMP Claims for Gene Expression Profile Testing

Note. Surrogate measures used to identify tests; granularity would provide less than robust results.

The table below presents the total number of claims and total allowed charges for three calendar years, 2015 to 2017, for Medicaid (MCO and HCA) utilization.

2015–2017 Medicaid Claims (MCO and HCA) for Gene Expression Profile Testing

Condition/Test	Services	Paid Dollars
Breast cancer	255	\$690,137
Prostate	1	\$560

The next three tables below present claims data for Oncotype DX Breast (CPT Code 81519) by calendar year, 2015 to 2017, for Medicaid (HCA), Medicaid (HCA), and PEBB/UMP.

2015–2017 Medicaid Claims (HCA) for Gene Expression Profile Testing CPT Code 81519 Oncotype DX Breast

	2015	2016	2017
Tests	6	13	14
Individuals	6	13	14
Average paid	\$1,947	\$2,536	\$2,792
Total Paid	\$10,912	\$32,972	\$39,084

Note. The 2017 data does not include the normal 90-days of claims payments following the close of calendar year 2017.

2017 Medicaid Claims (MCO) for Gene Expression Profile Testing CPT code 81519 Oncotype DX Breast

	2015	2016	2017
Tests	33	93	96
Individuals	33	93	96
Average paid	\$1,887	\$2,792	\$2,980
Total Paid	\$64,275	\$259,809	\$283,085

Note. The 2017 data does not include the normal 90-days of claims payments following the close of calendar year 2017.

2017 PEBB/UMP for Gene Expression Profile Testing CPT Code 81519 Oncotype DX Breast

	2015	2016	2017
Individuals	22	55	45
Services	22	55	45
Sum of Amount-Paid	\$88,238	\$225,951	\$190,470
Average of Amount Paid	\$4,011	\$4,108	\$4,233

Note. The 2017 data does not include the normal 90-days of claims payments following the close of calendar year 2017.

Gene expression profile testing of cancer tissue: final evidence report

Methods

This evidence review is based on the final key questions published on November 14, 2017.

Population: Adults with breast, prostate, or colon cancers or multiple myeloma

Interventions: Gene expression profile testing of cancer tissue to inform treatment decisions, including the following tests by cancer type:

- Breast Cancer—Oncotype DX breast cancer assay, MammaPrint, EndoPredict, Prosigna Breast Cancer Prognostic Gene Signature Assay (PAM50), Breast Cancer Index (BCI), Mammostrat
- Prostate Cancer—Decipher, Prolaris, Oncotype DX prostate cancer assay
- Colon Cancer—ColoPrint, Oncotype DX colon cancer assay
- Multiple Myeloma—Myeloma Prognostic Risk Signature (MyPRS), SKY92-signature (formerly EMC92)

Comparators: Usual care without gene expression profile testing of cancer tissue, alternate gene expression profile tests (i.e., one test intervention listed above versus another)

Outcomes:

- Patient management decisions (including selection of active surveillance rather than active treatment)
- Clinical outcomes (e.g., morbidity, mortality, quality of life)
- Harms, such as consequences of false-positive or false-negative test results
- Cost-effectiveness and other economic outcomes

Time period for literature search: January 2007 to November 2017

Key Questions

- 1. Effectiveness: What is the clinical utility of gene expression profile testing of cancer tissue to inform treatment decisions for patients with breast, prostate, and colon cancers and multiple myeloma?
 - a. Is there evidence that test results affect treatment decisions?
 - b. Do treatment decisions guided by gene expression profile testing of cancer tissue result in clinically meaningful improvements in patient outcomes?
- 2. Harms: What harms are associated with conducting gene expression profile testing of cancer tissue?
- 3. Special populations: Compared with usual care, do treatment decisions, patient outcomes, or harms after gene expression profile testing of cancer tissue vary by:
 - a. Patient demographics (e.g., age, sex, race/ethnicity)?
 - b. Clinical history (e.g., means of diagnosis, stage or grade of cancer, results of other testing, previous treatments, chronicity)?

- c. Medical comorbidities?
- d. Provider type or care setting?
- 4. What are the cost-effectiveness and other economic outcomes of gene expression profile testing used to inform treatment management decisions?

Analytic Framework

The analytic framework shown in Figure 1 guided the selection, synthesis, and interpretation of available evidence.



Figure 1. Analytic Framework

Eligible Studies

Randomized controlled trials (RCTs) and nonrandomized comparative studies, and systematic reviews (with and without meta-analysis) of these two types of studies that assess clinical utility were considered for Key Questions 1, 2, and 3. Cost-effectiveness studies and other comparative economic evaluations, along with systematic reviews (with and without meta-analysis) of these types of studies, were considered for Key Question 4. Table 3 summarizes the study inclusion and exclusion criteria.

For clinical utility studies (KQ 1, 2, and 3), include study if all of the following criteria are met:	Rationale
Study population consists of adults with multiple myeloma, or breast, prostate, or colon cancer.	The clinical population of interest for whom the test may be used.
The intervention is one of the selected gene expression profile tests. Breast Cancer: • Oncotype DX breast cancer assay • MammaPrint • EndoPredict • Prosigna (PAM50) • Breast Cancer Index (BCI) • Mammostrat	These tests were of interest to the requester and were vetted by posting and requesting public comment before being finalized. These tests are available and being used in the U.S.
 Prostate Cancer: Decipher Prolaris Oncotype DX prostate cancer assay 	
Colon Cancer: • ColoPrint • Oncotype DX colon cancer assay	
Multiple Myeloma: • Myeloma Prognostic Risk Signature (MyPRS) • SKY92 (EMC92)	
Comparator included usual care without gene expression profiling of cancer tissue or another gene expression profile test listed.	Without a comparison group, a study cannot measure the effect of the intervention test on the outcomes of interest. Concurrent prospective comparison groups offer the most valid comparison because they more accurately reflect the temporal change that occurs in clinical practice and among populations.
At least one outcome is a measure of direct clinical utility, including patient management decision (by care provider and/or patient); clinical outcomes such as mortality, morbidity, or quality of life measures stemming from patient management decisions resulting from the test; harms such as inaccurate test results influencing patient management decisions; and (for KQ4 only) cost-effectiveness or other economic outcomes resulting from use of the test.	Outcomes of interest are defined to inform the requestor's decision-making needs. Patient- oriented end outcomes rather than surrogate or intermediate outcomes are required to inform this decision. Studies that assess only analytic or clinical validity are excluded because they do not yield information about their usefulness to patients and clinicians in practice. Outcomes reported based on hypothetical or scenario-based studies were not eligible for inclusion because

Table 3. Study Inclusion and Exclusion Criteria

For clinical utility studies (KQ 1, 2, and 3), include study if all of the following criteria are met:	Rationale
	they do not provide direct evidence of clinical utility.
Settings for data collection included clinical facilities (inpatient or outpatient) in any country with substantial applicability to the U.S. setting.	The clinical utility of a gene expression profile test could be expected to vary based on the underlying health system and cancer care within that system, and so only settings with direct applicability to the U.S. are included.
Study designs include systematic reviews (with and without meta-analysis) and health technology assessments that meet inclusion criteria for this review; randomized controlled trials, nonrandomized studies with appropriate comparison groups. Systematic reviews, meta- analyses, and health technology assessments must be of moderate or high methodological quality.	Study designs are selected to minimize bias. Randomized controlled trials and systematic reviews and meta-analyses of them generally offer the lowest risk of bias because they are designed to minimize the effects of confounding factors on the outcomes. Prospective nonrandomized studies with appropriate control groups may also give valuable, although generally less biased, information. Retrospective nonrandomized studies often suffer from biases related to subject selection, and exposure and outcome measurement, but may be useful to decision makers when studies with fewer biases are not available. Thus, we include some study designs with high risk of bias that may be useful to decision makers. Cross-sectional studies do not allow evaluation of whether the intervention preceded the outcome occurrence and are not included because of the uncertainty of this temporal relationship. Case reports and case series are not included because they do not include a comparison group.
Other criteria: Publication in English Publication date 2007 or later Publication available for full-text review Data from study publication is extractable	Report authors and report audience use the English language. Scoping indicated that likely eligible studies of gene expression profile tests were not published prior to 2007 or were captured in systematic reviews published after that date. Studies must be available for review. Data included in publication must be reported in a way that can be used and analyzed in the report.

Data Sources and Searches

Center researchers conducted a search of the peer-reviewed published literature using Ovid MEDLINE. RCTs, nonrandomized comparative studies, and systematic reviews (with and without meta-analysis), and health technology assessments of these studies that assess clinical utility

were considered for Key Questions 1, 2, and 3. Cost-effectiveness studies and other comparative economic evaluations, along with systematic reviews (with and without meta-analysis) reporting economic outcomes, were considered for Key Question 4. The following electronic databases were searched to identify relevant peer-reviewed studies:

- Ovid MEDLINE
- Cochrane Database of Systematic Reviews
- Cochrane Central Register of Controlled Trials

The full Ovid MEDLINE search strategies for each of the indications (i.e., breast cancer, prostate cancer, colon cancer, and multiple myeloma) are in Appendix A. The search dates were January 1, 2007, through December 1, 2017. Center researchers also screened reference lists of relevant studies and used lateral search functions such as *related articles* and *cited by*. Citations from the Myriad Genetic Laboratories dossier for coverage of EndoPredict, which was submitted to the Washington State Agency Medical Directors' Group in December 2016, were also considered for inclusion. In addition, these core sources were searched:

- Agency for Healthcare Research and Quality (AHRQ)
- Blue Cross/Blue Shield HTA program
- National Institute for Health and Care Excellence (NICE)—Evidence
- Veterans Administration Evidence-based Synthesis Program

Center researchers scanned manufacturer websites and conducted a general Internet search for appropriate published studies and relevant gray literature.

Two independent Center researchers screened titles and abstracts. For studies that could not be excluded by title and abstract screening, a full-text review for inclusion criteria was performed. In instances of disagreement, an independent third screener resolved disputes.

If Center researchers identified a high-quality systematic review (with or without meta-analysis) addressing any of the key questions, a search was conducted to find eligible studies published after the search dates of the systematic review. Center researchers excluded systematic reviews if all of the included studies were also summarized by a more comprehensive systematic review, a systematic review of a higher methodological quality, or a more recently published systematic review.

Center researchers searched the sources listed above for clinical practice guidelines. In addition, searches of the AHRQ's National Guideline Clearinghouse (guidelines.gov) and websites of relevant professional organizations were conducted. Guidelines published in the past five years were considered for inclusion. In addition, Center researchers searched the CMS website for the Medicare Coverage Database for NCDs and LCDs applying to the state of Washington. The Aetna, Cigna, and Regence websites were searched for coverage policies for these private payers.

Center researchers searched the online database of clinical trials (clinicaltrials.gov) maintained by the National Library of Medicine at the National Institutes of Health. Information in this database is provided by the sponsor or principal investigator of clinical studies. Studies are generally registered in the database when they begin, with information updated as the study progresses. The search included the names of the gene expression profile tests and other common names for them: Oncotype DX breast (21-gene), MammaPrint (70-gene), EndoPredict (12-gene), Prosigna (PAM50, 50-gene), Breast Cancer Index (BCI), Mammostrat, Decipher (22gene), Prolaris (46-gene), Oncotype DX prostate (17-gene), ColoPrint (18-gene), Oncotype DX colon (12-gene), Myeloma Prognostic Risk Signature (MyPRS), Myeloma (MyPRS, 70-gene), and SKY92 (EMC92, 92-gene).

Data Abstraction and Quality Assessment

One Center researcher used standardized procedures to extract relevant data from each of the included studies and at least one other investigator cross-checked the data for accuracy. Two independent Center researchers evaluated studies for methodological risk of bias, and critical outcomes disagreement among these assessments was settled by a third independent Center researcher. Each study was assessed using Center instruments adapted from international standards and assessments for methodological quality.¹⁰⁻¹⁵ A rating of high, moderate, or low risk of bias was assigned to each study or review based on adherence to recommended methods and potential for bias or other limitations affecting internal and external validity. The risk-of-bias criteria for all of the study types are in Appendix B.

Center researchers assigned each outcome a summary judgment for the overall quality of evidence based on the system developed by the Grading of Recommendations, Assessment, Development, and Evaluation Working Group (GRADE).^{16,17} The GRADE system defines the overall quality of a body of evidence for an outcome in the following manner:

- **High:** Raters are very confident that the estimate of the effect of the intervention on the outcome lies close to the true effect. Typical sets of studies are RCTs with few or no limitations, and the estimate of effect is likely stable.
- **Moderate:** Raters are moderately confident in the estimate of the effect of the intervention on the outcome. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is different. Typical sets of studies are RCTs with some limitations or well-performed nonrandomized studies with additional strengths that guard against potential bias and have large estimates of effects.
- Low: Raters have little confidence in the estimate of the effect of the intervention on the outcome. The true effect may be substantially different from the estimate of the effect. Typical sets of studies are RCTs with serious limitations or nonrandomized studies without special strengths.
- **Very low:** Raters have no confidence in the estimate of the effect of the intervention on the outcome. The true effect is likely to be substantially different from the estimate of

effect. Typical sets of studies are nonrandomized studies with serious limitations or inconsistent results across studies.

• Not applicable: Researchers did not identify any eligible articles.

Search Results

The figures on the following pages show the study flow diagrams for each of the indications: breast cancer, prostate cancer, colon cancer, and multiple myeloma. A list of the excluded studies and reasons for exclusion are in Appendix H.



Figure 2. Study Flow Diagram: Breast Cancer



Figure 3. Study Flow Diagram: Prostate Cancer



Figure 4. Study Flow Diagram: Colon Cancer



Figure 5. Study Flow Diagram: Multiple Myeloma

Evidence Summary

The included studies are summarized below by indication and by key question. Detailed evidence tables are in Appendix C. Risk-of-bias assessments for all the included studies are in Appendix D.

Breast Cancer

The search located 1,675 unduplicated records. After excluding 1,454 citations during title and abstract review, 221 were reviewed in full text. A total of 22 studies met final inclusion criteria, including three systematic reviews (all three contributed data for KQ1 and one also contributed to KQ4). Two RCTs and 15 additional observational studies (one with two publications), published after the last search dates of the included systematic reviews were included for KQ1 and two additional economic modeling studies were included for KQ4. Studies on each of the six genetic tests (Oncotype DX for early-stage invasive breast cancer [also known as the Oncotype DX recurrence score], EndoPredict, MammaPrint, Prosigna Breast Cancer Assay [also known as the PAM50], Mammostrat, and the Breast Cancer Index [or BCI]) were available for inclusion in this analysis.

Key Question 1: Effectiveness Systematic Reviews

The three systematic reviews included a high proportion of overlapping primary studies for the Oncotype DX and MammaPrint tests, but each examined different groups of tests. The systematic review and meta-analysis by Augustovski et al.²⁰ only included the Oncotype DX test, while Blok et al.²¹ included Oncotype DX, MammaPrint, Prosigna, and EndoPredict tests, and Scope et al.²² included multiple overlapping studies for Oncotype DX and MammaPrint, and one study for Mammostrat. Overall, the Blok et al. systematic review was the most comprehensive of the three and was assessed by Center researchers as having a moderate risk of bias. The Augustovski et al.²⁰ systematic review was rated as having a low risk of bias, and the Scope et al.²² systematic review was assessed as being at high risk of bias.

Augustovski et al. reported the global pooled decision impact (defined as the proportion of patients whose treatment decision was altered with use of the Oncotype DX test), and global pooled net chemotherapy change (defined as the difference in the number of patients who were assigned to receive chemotherapy before vs. after the test), from 15 primary studies of women with early-stage invasive breast cancer.²⁰ Because of the clinical heterogeneity and risk of bias among these 15 studies, Augustovski et al. also conducted subgroup analyses for the seven studies with universal enrollment of subjects versus the eight with selective enrollment.²⁰ The authors reported a global pooled decision impact of 29.52% (95% Cl, 26.29 to 32.86%), $l^2 = 59.5\%$, and a global pooled net chemotherapy change of 12.00% (95% Cl, 8.00 to 16.00%), $l^2 = 92.00\%$.²⁰ For the subgroup of studies at lower risk of bias because they used universal subject enrollment, the corresponding decision impact estimate was 28.97% (95% Cl, 26.65% to 31.34%), $l^2 = 0.00\%$, and the corresponding net chemotherapy change was 9.00% (95% Cl, 4.00% to 14.00%), $l^2 = 89.00\%$, indicating less overall chemotherapy use.²⁰ Among the subgroup of

studies with selective enrollment, the pooled decision impact was 29.43% (95% CI, 22.88% to 36.45%), $I^2 = 74.09\%$, and pooled net chemotherapy change was 16.00% (95% CI, 7.00% to 24.00%), $I^2 = 93.00\%$.²⁰ The I^2 statistic is a measure of statistical heterogeneity and is commonly interpreted as indicating high, moderate, or low levels of heterogeneity among the pooled studies for cutoff values of 75%, 50%, and 25%, respectively.¹²¹ These analyses indicate that although the estimate for decision impact is quite stable at about 28% to 29%, the pooled estimate for net chemotherapy change of 12% is more variable and subject to high levels of heterogeneity.

The Blok et al. systematic review presented pooled estimates for the proportion of patients who had a change in treatment recommendation to a more intensive recommended treatment strategy (adjuvant systemic chemotherapy), and the proportion with a change to a less intensive treatment regimen (endocrine therapy or no treatment).²¹ These pooled estimates included studies of patients with either LN-negative or LN-positive tumors, although the majority of included studies were of patients with LN-negative tumors.²¹ Four included studies with 790 patients used the MammaPrint test; 22 studies (n = 3,743) examined the Oncotype DX test; and one study each involved the use of the Prosigna (n = 200) and EndoPredict (n = 167) tests.²¹ Blok et al. reported that a lower proportion of patients received higher intensity treatment recommendations after testing for all genomic tests included in the review (MammaPrint: -17%; Oncotype DX: -14.6%; Prosigna: -12.9%; and EndoPredict: -34%).²¹ Correspondingly, the proportions of patients who were recommended to have less intensive treatment increased with use of all the tests (MammaPrint: +32.2%; Oncotype DX: +51.1%; Prosigna: +37.3%; and EndoPredict: +53.2%).²¹ This review did not provide any risk of bias assessment for the included studies and did not perform any assessment of statistical heterogeneity across the groups of studies or provide confidence intervals for the estimates.

The Scope et al. systematic review, which Center researchers rated as having a high risk of bias due to unclear methodological criteria and detail, did not present any pooled estimates because the authors were concerned about heterogeneity among studies.²² The systematic review included 28 studies reporting outcomes on changes in recommended treatment after Oncotype DX testing and six that used the MammaPrint test.²² The review of included studies did not clearly distinguish between treatments recommended and treatments selected or rendered. Although the review searched for studies on the Mammostrat test, none were identified for this outcome.²² In a narrative synthesis, the authors reported that the use of Oncotype DX led to changes in treatment recommendations for 21% to 74% of patients enrolled in these studies.²² Change from a recommendation of chemotherapy to no chemotherapy ranged from 6% to 51% of patients after Oncotype DX use, but in one study the proportion of patients who were recommended to receive chemotherapy increased after use of the test.²² Similarly, the authors stated that the use of MammaPrint led to treatment recommendation changes in 18% to 40% of patients, and that between 2% and 32% of patients would have a recommendation that changed from chemotherapy to no chemotherapy after the test was used.²²

Additional Studies

The search identified two RCTs and 15 observational studies published after the search dates for the included systematic reviews (with the exception of one study³⁷ that was published on the BCI test and not included in any of the systematic reviews). One of the RCTs was conducted using the MammaPrint test¹⁹ and the other used the Oncotype DX test alone.¹⁸ Nine^{23-29,31,38} of the 15 additional observational studies used the Oncotype DX test, three^{32-34,122} used MammaPrint, two^{35,36} were on the use of Prosigna/PAM50, and one used the BCI test.³⁷ One of the additional observational studies for Oncotype DX was reported in two publications.^{24,123}

Randomized Controlled Trials

The additional RCTs that met inclusion criteria are described by test below.

Oncotype DX

Bear et al. reported the results of a small RCT that was assessed as having a high risk of bias because baseline characteristics varied between groups, the test manufacturer funded the study, and one author was employed by the test manufacturer.¹⁸ The 33 women with Oncotype DX scores of 11 to 25 were randomized to neoadjuvant hormone therapy (NHT) or neoadjuvant chemotherapy (NCT), but two patients refused assignment and crossed over from the NCT to NHT group.¹⁸ The randomized groups were dissimilar at baseline in terms of race, tumor stage, and menopausal status.¹⁸ No intention-to-treat (ITT) results were presented.¹⁸ The authors reported that the clinical and partial response rates by treatment received were significantly different in an ordinal regression that controlled for age, race, menopausal status, and study site. Women who received NHT had a lower clinical response rate than did women who received NCT (22.2% vs. 36.4%; p = .034).¹⁸

MammaPrint

Cardoso et al. conducted a large RCT enrolling women aged 18 to 70 years with early-stage invasive breast cancer.¹⁹ This publication is the first report from the international phase 3 RCT called the Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy (EORTC 10041/BIG 3-04 MINDACT) study, also referred to as the MINDACT study.¹⁹ All 6,693 enrolled women underwent clinical risk assessment using a modified version of the Adjuvant! Online tool and genomic risk assessment using the 70-gene signature test (MammaPrint).¹⁹ Only women with discordant clinical and genomic risks were randomized to receive or not receive chemotherapy (n = 2,187 randomized and in the intent-to-treat [ITT] analysis).¹⁹ The specific demographic and clinical characteristics of this randomized group were not reported, but the overall group had an average age of 55 years and most women had ERpositive, HER2-negative, negative lymph nodes, and grade 1 or 2 tumors.¹⁹ A protocol amendment in August 2009 allowed inclusion of women with up to three positive lymph nodes, and 21% of the final total study population had one to four positive nodes.¹⁹ In the randomized groups (with discordant clinical and genomic risk), 35% of subjects had at least one positive lymph node.¹⁹ Racial composition of the study group was not reported, although it was conducted across 112 centers in nine European countries.¹⁹ Among women with high clinical risk but low genomic risk who were randomized to treatment based on clinical risk (adjuvant chemotherapy) or genomic risk (no adjuvant chemotherapy), there were similar rates of five-year survival without distant metastasis (95.9% in the chemotherapy group, 94.4% in the no chemotherapy group), with an adjusted HR of 0.78 (95% Cl, 0.50 to 1.21).¹⁹ Among the group of women with low clinical risk but high genomic risk, the rates of death or distant metastases at five years were also similar for those randomized to treatment on the basis of genomic risk (adjuvant systemic chemotherapy) compared to those randomized to treatment on the basis of clinical risk (no adjuvant systemic chemotherapy) (95.8% vs. 95.0%; aHR 1.17, 95% Cl, 0.59 to 2.28). This RCT was assessed as having moderate risk of bias because of test changes that caused different risk classification of patients for a period during the trial.¹⁹ Other analyses from the MINDACT study are expected to be released in the future,¹²⁴ including an analysis involving women with more than three positive lymph nodes that is due to be completed in 2020, according to the NLM/NIH clinical trials registry from the National Library of Medicine at the National Institutes of Health (ClinicalTrials.gov). A list of the ongoing clinical trials in this registry on the gene expression profile tests is in Appendix F.

Observational Studies

The additional observational studies that met inclusion criteria are described by test below.

Oncotype DX

Nine of the 15 additional observational studies involved the use of the Oncotype DX test.^{23-29,31,38} Of these, one was a prospective cohort study,²⁴ five were retrospective cohort studies,^{25,26,28,29,125} and three^{23,27,38} were before-after studies using a single group of patients. One of the studies³⁸ reported on the decision impact of the test based on treatment recommendations, and six others reported the actual treatment received.^{24-29,31} One study reported women's reported distress, perceived benefits and harms of treatment, and recurrence risk before and after having the test.²³

Friese et al. conducted a retrospective cohort study (using a database with prospective data collection) involving 1,527 women with ER-positive, HER2-negative early-stage invasive breast cancer (60% were LN-negative), 778 of whom had a treatment recommendation made on the basis of their Oncotype DX test result compared to a group who had treatment recommendations based on other factors.²⁴ Some aspects of this study were also reported in another publication,¹²³ but the Friese et al. article²⁴ is more comprehensive and is summarized here. This study used the Surveillance, Epidemiology, and End Results (SEER) registries from Los Angeles County and Georgia for eligible women treated in 2013 and 2014.²⁴ The study population was racially and ethnically diverse (57% Caucasian, 15% African American, 18% Latina, and 7% Asian) and included women from a wide range of educational and income levels.²⁴ Women with low-risk Oncotype DX scores were less likely to receive chemotherapy than women who were not tested (OR, 0.1; 95% CI, 0.1 to 0.2), whereas women with high- and medium-risk Oncotype DX scores were more likely to receive chemotherapy compared to women who were not tested (OR for high-risk recurrence score, 2.8; 95% CI, 2.0 to 4.0; OR for

medium-risk recurrence score, 1.4; 95% CI, 1.1 to 1.7).²⁴ This study also reported that 64% of patients who received the Oncotype DX test found it "very" or "extremely" helpful in making their treatment decision.²⁴ Just under 65% of women with a low-risk score reported that it shifted their opinion away from chemotherapy and slightly over 73% of women with a high-risk result reported that they shifted toward wanting chemotherapy.²⁴ This study was assessed as having a moderate risk of bias because of nonresponse bias and missing data.

Jasem et al. conducted two retrospective cohort studies using data from the U.S. National Cancer Data Base (NCDB).^{25,26} The NCDB captures approximately 70% of newly diagnosed cancers in the nation from more than 1,500 Commission on Cancer-accredited facilities.^{25,26} The first study reported on receipt of adjuvant chemotherapy for more than 120,000 women with ER+, LN-negative, early-stage invasive breast cancer diagnosed from 2004 to 2012, based on whether they received the Oncotype DX test.²⁵ Women who did not receive the test were more likely to be older, Caucasian, have noncommercial insurance coverage, and grade 1 tumors.²⁵ Receipt of adjuvant chemotherapy was highly associated with having an intermediate- or highrisk recurrence score (aORs of 12 and 83, respectively), although the analysis did not report the proportions of women with low-risk tests, or untested women who received adjuvant chemotherapy.²⁵ However, the authors did report the adjusted odds of chemotherapy receipt for women with and without the test by various demographic and clinical characteristics to assess patterns of care associated with test receipt.²⁵ The authors found that younger African American women were more likely to receive adjuvant chemotherapy even when they had low-risk Oncotype DX test results (aOR, 1.33; 95% CI, 1.16 to 1.54).²⁵ Center researchers assessed this study as having a high risk of bias because of potential for miscoding of tests ordered and incomplete control for confounding variables.

The second study by Jasem et al. also used data from the NCDB registry, and included more than 30,000 patients with a diagnosis of breast cancer in 2010 through 2012, about a third of whom had the Oncotype DX test ordered.²⁶ Jasem et al. reported differences in receipt of chemotherapy among women who had or did not have the test ordered.²⁶ Patients who had the test ordered received chemotherapy less often than patients who did not have the test ordered (38% vs. 75%), with an adjusted odds ratio of 0.21 (95% CI, 0.20 to 0.22).²⁶ As with the previous study, patients with intermediate- and high-risk scores were substantially more likely to receive chemotherapy (aOR 4.5 and 19.8, respectively) compared to those with low-risk scores.²⁶ Center researchers assessed this study was also assessed as having a high risk of bias because of potential for miscoding of tests ordered and incomplete control for confounding variables.

O'Neill et al. conducted a retrospective cohort study using a five-state (CA, GA, KY, NY, OH) commercial insurance claims database with linkage to a registry of Oncotype DX test results from the patent holder and test provider.²⁸ The study included approximately 5,000 women age 65 years and under, and aimed to enroll those diagnosed with stage 1 or 2 hormone receptor-positive, HER2-negative cancers from 2006 to 2010.²⁸ However, the HER2 status was missing for more than half of the subjects and about a third of them had histologic grade 3 tumors.²⁸ The

study sample was predominantly Caucasian (85%), and 75% had LN-negative tumors.²⁸ The authors reported the initiation of endocrine therapy within six months of diagnosis and continuation of the medication.²⁸ Oncotype DX test receipt was associated with endocrine therapy initiation (aOR, 2.48; 95% Cl, 2.03 to 3.04) and was not associated with medication discontinuation (aOR, 0.93; 95% Cl, 0.85 to 1.02).²⁸ Center researchers assessed this study as having a high risk of bias because of misclassification due to missing data, potential for coding errors in claims data, and the inability to control for important confounders not recorded in the database.

Parsons et al. ²⁹ conducted a retrospective cohort study using the NCDB described above in the description of the Jasem et al. studies.^{25,26} The study included over 132,000 women diagnosed between 2010 and 2013 who were aged 18 to 70 years and had ER-positive (or borderline), HER2-negative (or borderline) early-stage invasive breast cancer.²⁹ The study compared receipt of chemotherapy among patients who had or did not have the Oncotype DX test performed.²⁹ Patients who did not receive the test were more likely to have chemotherapy treatment (OR, 1.21; 95% CI, 1.17 to 1.25).²⁹ Patients with intermediate- and high-risk Oncotype DX scores were also more likely to receive chemotherapy (OR 12.9 and OR 87.2, respectively) than patients who had a low-risk result.²⁹ These results are, not surprisingly, similar to those reported by Jasem et al.,^{25,26} given that the analyses were performed on substantially overlapping datasets. Accordingly, Center researchers assessed this study as having a high risk of bias because of potential for miscoding of tests ordered and incomplete control for confounding variables.

Ray et al. used Kaiser Permanente's Northern California tumor registry to examine the relationship between use of the Oncotype DX test and receipt of chemotherapy for the treatment of women with ER-positive, HER2-negative, stage 1 and 2 breast cancers who received treatment between 2005 and 2012.³¹ The authors included women with primary tumors larger than 0.5 cm and with either no nodal involvement or only micrometastases.³¹ The racial and ethnic composition of the group was diverse: 71% Caucasian, 14% Asian, 9% Hispanic, and 6% African American.³¹ The database included 1,567 women who received Oncotype DX testing and 5,437 who did not.³¹ In a propensity score-matched analysis (n = 2,923), women who had received the Oncotype DX test were less likely to receive chemotherapy (OR, 0.74; 95% CI, 0.63 to 0.87).³¹ The absolute reduction in chemotherapy use for women who were tested was 6.2% (95% CI, 5.3% to 13.6%).³¹ Center researchers assessed this study as having a moderate risk of bias because of its design and calculation of some outcome measures using an OR that would likely inflate the effect estimates, although the authors did attempt to control confounding by conducting a propensity-matched analysis.

Loncaster et al. performed a small (n = 201) before-after study of a single group of patients with early-stage invasive breast cancer in Manchester, U.K., who were referred for chemotherapy and then had Oncotype DX testing performed before a final decision about the use of adjuvant treatment.²⁷ Women with both LN-negative (n = 136) and LN-positive (n = 65) tumors were included, and many had grade 3 cancers (46% vs. 52% with grade 2 cancers).²⁷ After receipt of

the test, 63% of subjects overall and 69% of women with LN-positive tumors did not receive chemotherapy.²⁷ This study also reported costs associated with chemotherapy avoidance, which are detailed below in the section on economic outcomes for KQ4. Because of the study's design, Center researchers assessed the risk of bias as high.

Pestalozzi et al. also conducted a small (n = 229) before-after study of a single group of patients drawn from 18 sites in Switzerland.³⁸ Patients had completely resected, early-stage invasive breast cancer (ER-positive, HER2-negative, one to three positive lymph nodes).³⁸ The study compared the treatment recommendations of a multidisciplinary tumor board before and after receipt of the Oncotype DX test.³⁸ Overall, 20% of subjects had a treatment recommendation change after the test results were known; 15% of the low-risk group and 32% of the non-low-risk group had a change of recommendation.³⁸ For 44% of the low-risk group and 40% of the non-low-risk group, the recommendation changed from adjuvant chemotherapy to no chemotherapy.³⁸ The addition of adjuvant chemotherapy was recommended for 4% of the low-risk group and 16% of the non-low-risk group after testing.³⁸ Center researchers rated the risk of bias for this study as high because of its study design.

Evans et al. reported women's views and perceptions before and after receipt of the Oncotype DX test.²³ The 193 women included in the results were enrolled from a group of 352 who were eligible.²³ Perceptions about the benefits and harms of chemotherapy increased and perceived risk of recurrence decreased after receipt of the test.²³ Distress levels were not statistically different in the pre- and posttest assessments, although higher levels of posttest distress were associated with increased receipt of chemotherapy.²³ Center researchers assessed this study as having a high risk of bias because of its design.²³

MammaPrint

In addition to the RCT by Cardoso et al.¹⁹ described in the section above, the search located one retrospective cohort study³³ and two before-after studies with single groups of patients.^{32,34} The retrospective cohort study by Kuijer et al. included 2,043 women with breast cancer surgically treated in the Netherlands between November 2011 and October 2013.³³ Nearly 300 of them received the 70-GS (MammaPrint) test to inform treatment decisions, and for 1,745 women treatment was determined by standard clinicopathological factors.³³ All women in this study would have been eligible for MammaPrint testing according to national guidelines at that time.³³ Most women (86% in each group) had grade 2 tumors and most had negative lymph nodes (80% in tested group and 77% in comparator group).³³ In a mixed-effects linear regression model, use of the test (after controlling for age, incidence year, size and grade of tumor, and axillary node involvement) was associated with a 9.5% absolute reduction in the use of chemotherapy (95% CI, -15.7% to -3.3%).³³ In subgroup analyses, this effect was most pronounced among women with grade 1 tumors larger than 2 cm without nodal involvement and for those younger than 50 years old.³³ Given the nature of the study design, the authors noted that confounding by indication was a concern and that no causal relationship between testing and chemotherapy use should be inferred.³³ Center researchers assessed this study's risk

of bias as high because of the design and concerns about the inability to control for other important sources of confounding.

Kuijer et al. also published a prospective before-after study involving 660 women in the Netherlands who were treated at 33 hospitals during calendar years 2013 through 2015.³² Women enrolled had surgically treated ER-positive, early-stage invasive breast cancer and were eligible for adjuvant chemotherapy treatment.³² Women entered into the database most often had grade 2 tumors (73%) and negative axillary lymph nodes (84%), and were HER2-negative (97%).³² During the postoperative multidisciplinary treatment planning meeting, a treatment recommendation was made based on clinicopathological features and informed by use of the U.K. NHS PREDICT tool.³² Overall, 41% of all patients had a recommendation for adjuvant systemic chemotherapy before the test was performed.³² The 70-GS (MammaPrint) test was administered and results sent back to the treating physician who recorded both the treatment recommendation and the actual treatment administered.³² Treatment recommendations changed for 51% (95% CI, 46% to 56%), and actual treatment received changed for 52%.³² Chemotherapy actually administered comported with what was recommended based on the test 90 to 91% of the time.³² Oncologists' initial treatment recommendations did not predict the test result well (no to slight agreement, based on a kappa statistic of 0.02).³² Although tests were ordered in accord with the national guideline, ordering was selective and not universal. Thus, it is possible that tests were ordered more when adjuvant treatment was uncertain or there were other unmeasured confounding factors. For this reason and general issues with the before-after design, Center researchers assessed this study as being at high risk of bias.

Tsai et al. conducted a before-after study of 840 women with early-stage invasive breast cancer treated between May 2012 and December 2015 at 58 participating U.S. institutions.³⁴ This study was a secondary analysis of the Prospective Study of MammaPrint in Breast Cancer Patients with an Intermediate Recurrence Score (PROMIS) study.³⁴ Inclusion criteria required ER-positive, HER2-negative, LN-negative cancers, but a protocol amendment allowing LN-positive patients with one to three positive nodes was introduced midway through the study.³⁴ Among the final enrolled population, 99.9% were ER-positive, 97.7% were HER2-negative, and 88.6% had no nodal involvement.³⁴ The mean age of the study population was 59 and 87% were Caucasian.³⁴ The study was designed to test whether application of the 70-GS (MammaPrint) test affected treatment decisions among women with an intermediate Oncotype DX score (score of 18 to 30).³⁴ The median Oncotype DX was 22 (IQR 20 to 25).³⁴ The 70-GS result was highly associated with changes in recommendations to add or remove chemotherapy treatment.³⁴ For patients initially recommended to have chemotherapy, but who had a low-risk 70-GS score, the odds for withdrawing the chemotherapy recommendation were high, although with a wide CI (OR, 108; 95% CI, 18.98 to 434.77).³⁴ For patients with an initial recommendation of chemotherapy who received a high-risk 70-GS score, the odds of having that chemotherapy recommendation withdrawn were very low (OR, 0.01; 95% CI, 0.001 to 0.04).³⁴ Among all patients, the odds of chemotherapy treatment withdrawal were 0.64 (95% CI, 0.50 to 0.82).³⁴ Overall, 33.6% of

treatment recommendations changed after the 70-GS test was administered.³⁴ Physicians were surveyed about how the addition of the 70-GS test influenced their decision and reported that it increased their confidence in the final treatment plan in 78.6%, reduced it in 5.8%, and had no influence in 15.6% of cases.³⁴ This study is limited by its before-after design and had extensive involvement by the manufacturer of the 70-GS test in the design and conduct of the study and writing of the report.³⁴ Center researchers assessed the study as having a high risk of bias because the study was funded by test manufacturer who had input on the study design, management of the study, analysis and interpretation of the data, and approval of the manuscript.

Prosigna

Two additional observational studies on the Prosigna (PAM50) test met inclusion criteria.^{35,36} Both were before-after studies using a single group of patients for whom a treatment recommendation was available prior to the test result being available; the final treatment recommendation with additional information from the test was also available.^{35,36}

Hequet et al. studied 210 postmenopausal women with ER-positive, HER2-negative, LN-negative breast cancer diagnosed between March 2015 and January 2016 across eight centers in France.³⁵ Seventy-nine percent of enrollees had stage 1 disease.³⁵ Overall, treatment recommendation changes occurred for 18% of the women.³⁵ The direction of change was from a recommendation of no adjuvant chemotherapy to adjuvant chemotherapy for 13% of the women, and from adjuvant chemotherapy to no chemotherapy among 5% of them.³⁵ Physicians reported increased confidence in 39% of cases, decreased confidence in 11% of cases, and no change in confidence in 51% of cases.³⁵ The study also gathered information on women's decisional conflict using the Decisional Conflict Scale (DCS) measure and function using a functional assessment measure (the Functional Assessment of Cancer Therapy-General [FACT-G]).³⁵ Patients' overall scores on the DCS decreased 3.5 points after the test results, from a mean of 9.8 to 6.2 (p < .001).³⁵ The overall mean FACT-G increased from a baseline level of 79.4 to a posttest mean of 80.2 (p = .264) and then decreased at six months post-diagnosis to 76.77 (no statistical testing reported).³⁵ The changes on both of these scales are minor and do not likely represent patient-important differences.³⁵ Center researchers rated the risk of bias for this study as high because of the study design, selective reporting of outcomes, and funding and authorship by the test manufacturer.

Wuerstlein et al. reported treatment recommendations before and after Prosigna study results were available in the West German Study Group (WSG) Breast Cancer Intrinsic Subtype study.³⁶ Women were enrolled consecutively between October 2013 and October 2014 across 11 centers in Germany.³⁶ The study population included 198 postmenopausal women with ER-positive, HER2-negative, predominantly stage 1 (77%) tumors with evaluable test results.³⁶ For 18% of cases, test results were associated with any change in treatment recommendation.³⁶ In 11% of cases, there was a change from a no adjuvant chemotherapy to an adjuvant chemotherapy recommendation, and for 2% there was a change from an adjuvant chemotherapy

recommendation to a recommendation against adjuvant chemotherapy.³⁶ For 5% of women there was a change in the particular type of chemotherapy regimen.³⁶ Physicians reported increased confidence in their treatment recommendation after test results in 89% of cases.³⁶ Patients reported a decrease in "state-anxiety" scores from a mean of 40.5 before the test to 38.5 after the test (p = .082).³⁶ Center researchers assessed this study as being at a high risk of bias because of the study design, the test manufacturer funded the study, and several authors had financial relationships with the company.³⁶

Breast Cancer Index (BCI)

Sanft et al. conducted a before-after study of 96 women with ER- or PR-positive, stage 1 through 3 breast cancer who had completed at least 3.5 years of adjuvant endocrine (hormonal therapies such as tamoxifen or aromatase inhibitors) therapy and were eligible for extended endocrine treatment.³⁷ Subjects were patients and their physicians from a single U.S. institution, enrolled between February and August 2014.³⁷ Enrollees were predominantly Caucasian (91%), most women had stage 1 cancer (60%), and the majority did not have nodal involvement (73%).³⁷ Overall, 26% of women had a change of treatment recommendation after the test, with an overall decrease in recommendations for extended adjuvant endocrine therapy (74% before the test vs. 54% after the test) with a statistically significant OR for reduced use of adjuvant endocrine therapy (OR, 0.14; 95% Cl, 0.04 to 0.46).³⁷ More physicians felt "strongly confident" in their treatment recommendation after the test (8% vs. 24%; OR, 4.75; 95% Cl, 1.62 to 13.96).³⁷ Center researchers rated the risk of bias for this study as high because of the study design. In addition, the use of an OR to estimate the effect of an outcome that is not rare, particularly with a small sample size as seen in this study, will overstate the estimate of effect.

Summary

Although the majority of studies on gene tests to inform treatment of breast cancer have a high risk of bias, their findings are consistent regarding an association between test use and changes in recommended or actual treatment based on the test result. There is evidence regarding usefulness of the tests in terms of patient management decisions, particularly the decision for or against adjuvant systemic chemotherapy. The largest body of evidence pertains to the Oncotype DX test, with a moderate amount for the MammaPrint and Prosigna tests and very little for the BCI, EndoPredict, and Mammostrat tests. The overall quality of evidence for these findings is, however, limited because of the lack of information on important patient outcomes such as survival (with the exception of the MammaPrint test based on the MINDACT study),¹⁹ and the absence of high-quality study designs. The overall quality of evidence varies by test with moderate quality of evidence for the Oncotype DX test, low quality of evidence for the MammaPrint test, and very low quality of evidence for the EndoPredict, Prosigna, Mammostrat, and BCI tests. Based on a single RCT, there is moderate quality of evidence that women with early invasive breast cancer who are considered to be at high clinical risk by the Adjuvant! Online risk assessment tool may safely forego adjuvant systemic chemotherapy if their MammaPrint genomic risk score is low (no statistically significant difference in distant metastasis-free survival at five years). Moderate quality of evidence supports the use of

Oncotype DX because of its impact on clinical treatment recommendations and chemotherapy use for women with early-stage invasive breast cancer, particularly for low-risk women who would not benefit from adjuvant systemic chemotherapy.

Key Question 2: Harms

No studies reported outcomes related to false reassurance or false alarm from these tests. A few studies reported patients' perceptions of or experiences with the test that could be indirectly related to quality of life, such as anxiety. The study on the use of Oncotype DX testing by Evans et al. did report that women had similar levels of distress before and after receiving the test results, but that any observed differences in the levels of distress likely stemmed from differences among women who were receiving chemotherapy as opposed to the test, per se.²³ In general, studies that reported on decisional conflict, anxiety, function, or patient-perceived usefulness of the test found small differences in favor of testing.^{24,35,36} Similarly, physicians reported finding the test useful or that it increased their confidence in treatment recommendations.³⁵⁻³⁷ These details are presented under the results for KQ1 and in the evidence tables in Appendix C. No other studies met specific inclusion criteria for this key question. Overall, there is no compelling evidence of harm from the tests, but the research is sparse and not likely reported in ways that allow for full evaluation of these types of outcomes. For tests where there is evidence regarding harms, the quality of that evidence is very low. For other tests, there is no evidence included in this report to inform KQ2.

Key Question 3: Subpopulations

Few studies reported results stratified by subpopulations of interest such as age, race, or disease characteristics. When available, results are presented above in the results for KQ1 or below in the results for KQ4. Some studies reported differences in groups that received testing.^{26,31} For example, Jasem et al. reported that older patients were more likely to receive testing than younger patients, that women were more likely to be tested than men, and that African American women and patients without insurance were less likely to be tested.²⁶ However, these factors were not analyzed for subgroup differences with respect to the outcomes of interest in this report. The included studies tended to enroll similar populations in terms of clinical characteristics, although some studies did not enroll women with positive lymph nodes or had variable proportions of women with positive nodes or with higher stage disease.

Most studies enrolled mixed groups consisting of LN-negative and LN-positive patients. Two studies on Oncotype DX^{25,31} enrolled exclusively LN-negative subjects and one²⁶ enrolled only LN-positive patients. For the MammaPrint test, the Blok systematic review contained one small study of LN-negative patients, but the remaining four studies in that systematic review, as well as all four of the additional studies, had mixed nodal status enrollment. All of the studies on the Prosigna test enrolled exclusively LN-negative populations.^{21,35,36} The MINDACT study by Cardoso et al. enrolled LN-negative and LN-positive patients, but did not report results stratified by nodal status.¹⁹ They found a mortality benefit from using MammaPrint only among those with high clinical risk (based on the Adjuvant! Online risk assessment tool).¹⁹ In the one study of

an exclusively LN-positive population, the use of Oncotype DX remained associated with lower use of chemotherapy (aOR, 0.21; 95%CI, 0.02 to 0.22).²⁶

When specific results were reported for a subgroup of interest, the effect of confounding could usually not be ruled out. For example, Jasem et al., in a study that included only LN-negative women, reported that younger African American women were more likely to receive a recommendation for chemotherapy even with a low Oncotype DX score after controlling for other factors, but the adjusted OR was 1.33 and although statistically significant, might not represent a clinically significant difference given the potential for bias inherent in the study and the small difference in the effect.¹²⁶ Although Jasem et al. reported small differences in use of adjuvant chemotherapy among groups who received and did not receive the Oncotype DX test, these differences are accompanied by overlapping confidence intervals or potentially reflect decisions based on clinicopathological factors (i.e., it is expected that women with stage 3 disease would more often receive recommendations to have adjuvant chemotherapy, even if they had the gene expression test).¹²⁶ Evans et al. reported that younger women were more likely to receive chemotherapy even when controlling for the test result.²³

Key Question 4: Cost and Cost-Effectiveness Systematic Review

The Blok et al. systematic review, described under the section on KQ 1 results, included 44 studies that reported some type of economic analysis of the four tests included in their review (MammaPrint, Oncotype DX, Prosigna/PAM50, and EndoPredict).²¹ Of these, 26 reported outcomes in costs per QALY and are summarized below, giving the ranges as reported. Studies adopted different perspectives, included different costs, made varying assumptions, used a variety of analytic techniques, and reported results in differing currency units. The remaining 18 economic studies reported outcomes in terms of whether the test strategy was "dominant" (both clinically effective and cost saving compared to current care), "cost increasing," or "cost saving."²¹ The only two studies^{53,54} that compared measured outcomes associated with testing or not testing using actual patient groups found increased costs associated with the use of Oncotype DX compared to no testing.²¹

The remaining 42 studies summarized by Blok et al. used mathematical modeling to estimate economic outcomes of testing strategies.²¹ For better comparability within this report, only the 26 cost-utility analyses that reported results in costs per QALY are summarized. Blok et al. included five studies that evaluated the costs of MammaPrint compared to no testing.²¹ The authors found cost estimates per QALY across these studies ranging from \$10,000 to \$43,044 in U.S. dollars and €4,614 to €134,000 in euros for women with LN-negative cancers.²¹ One study reported costs per QALY of \$13,723 for women with positive lymph nodes. There were 14 studies that compared Oncotype DX to no testing and reported costs per QALY ranges of \$3,843 to \$43,044, CAD\$3,206 to CAD\$63,064, or £29,502 for women with LN-negative tumors.²¹ Five studies reported costs per QALY of \$1,914 to \$49,059, CAD\$464 to CAD\$14,844, and £5,529 for

women with LN-positive tumors (or studies with mixed LN-negative and LN-positive populations).²¹

Two head-to-head economic analyses reported on economic comparisons of Oncotype DX versus either Mammostrat or MammaPrint among women with LN-negative tumors.²¹ Mammostrat was estimated to be more cost-effective than Oncotype DX (although at \$453,600 per QALY, which would not meet conventional cost-effectiveness thresholds), and MammaPrint was also estimated to be more cost-effective than Oncotype DX (€1,457).²¹

Additional economic studies

The search retrieved one additional economic study⁵¹ published after the Blok et al. systematic review²¹ and one decision analytic study⁵² published before the systematic review that reported on the BCI test, which Blok et al. did not include. In addition, one other study described above in the section for KQ1 included cost outcomes related to use of the Oncotype DX test.²⁷

Hall et al. conducted a modified Markov modeling study for the U.K. health technology assessment program to accompany the OPTIMA preliminary feasibility study.⁵¹ In that trial, Oncotype DX testing (using a score cutoff of greater than 25 rather than the more standard intermediate risk cutoff score of greater than 18) was compared to standard risk assessment to guide treatment recommendations.⁵¹ Remaining biopsy tissue was also tested with the MammaPrint and Prosigna (subtype) tests (as well as other tests not included in the scope of this review) to determine the final selection of test(s) to be used in the full trial.⁵¹ This economic study was assessed as being at moderate risk of bias because model assumptions were based on a limited evidence base. In the base case analysis, the lifetime per-patient cost without testing, assuming chemotherapy treatment of all patients, was £13,961 (95% CI, £10,535 to £21, 203) and the lifetime per-patient QALYs were 7.69 (95% Cl, 5.06 to 9.58).⁵¹ The mean incremental per-person cost changes for each of the three tests compared to this base case were reported as follows: Oncotype DX -£180 (95% CI, -£4,610 to £4,592); MammaPrint £195 (95% CI, -£3.206 to £3,430); and Prosigna -£281 (95% CI, -£3,553 to £2,774).⁵¹ Mean incremental QALYs per person compared to the base case for each test were also calculated as follows: Oncotype DX 0.2 (95% Cl, -1.07 to 1.40); MammaPrint 0.18 (95% Cl, -0.87 to 1.1); and Prosigna 0.18 (95% Cl, -0.91 to 1.15).⁵¹ The probabilities that each test would be cost-effective within the U.K. health system compared to the base case were nearly identical, with estimates of 77% for Oncotype DX, 75% for MammaPrint, and 77% for Prosigna.⁵¹

The study authors performed two sensitivity analyses using Monte Carlo simulation to determine the probability of cost-effectiveness under different assumptions; the first with an assumption of a constant relative effect of chemotherapy and the second modeling variable survival after recurrence.⁵¹ When the relative effect of chemotherapy was held constant, the probabilities of cost-effectiveness decreased for all tests, and were estimated at 33% for Oncotype DX, 31% for MammaPrint, and 41% for Prosigna.⁵¹ This means that a strategy of treating all women with chemotherapy was more cost-effective than use of any of the tests

under the assumption that the relative benefit of adjuvant chemotherapy is constant across risk groups.⁵¹ The second Monte Carlo simulation modeled variation in survival after breast cancer recurrence and found increased cost-effectiveness estimates for all tests compared to the base case: 97% for Oncotype DX, 94% for MammaPrint, and 94% for Prosigna.⁵¹ This finding means that under this assumption, each test offered better cost-effectiveness compared to a strategy of treating all women initially diagnosed without using genomic testing.⁵¹

Hall et al. were careful to say that, given the uncertainty of much of the information contributing to the analysis, it would be incorrect to draw conclusions about either the effectiveness or cost-effectiveness of any of the tests included in the analysis.⁵¹ The analysis was performed for the express purpose of guiding selection of tests, based on the value of information to be gained, for a future randomized trial with the primary endpoint of five-year recurrence-free survival.⁵¹ On a value-of-information basis, the Prosigna ROR, IHC4, IHC4 AQUA, and MammaTyper tests were all preferred over MammaPrint, Prosigna (subtype), and Oncotype DX for future research.⁵¹ This finding can likely be attributed to the more mature research bases for these three tests.⁵¹

Gustavsen et al. conducted a decision-analysis evaluation of the BCI test for two scenarios compared to usual care.⁵² The first scenario involved the use of the test for newly diagnosed women with ER-positive, LN-negative breast cancer who were making decisions about the use of adjuvant chemotherapy.⁵² In the second scenario, the test was used for women who were cancer-free at five years after initial treatment and were making a decision about whether to pursue extended endocrine therapy.⁵² The study assumed a U.S. commercial payer perspective and based model treatment assumptions on the 2012 NCCN guidelines.⁵² Center researchers assessed this study as having high risk of bias because of the use of a modeling technique that is no longer considered best practice, questionable model assumptions due to a limited evidence base, and the use of one-way sensitivity analyses that would not account for interrelationships between either the model inputs or outcomes. The authors reported a mean cost difference per patient of \$3,803 (\$41,634 vs. \$45,437) for the first scenario, with sensitivity analyses ranging from \$2,800 to \$4,000, in favor of the BCI test over usual care.⁵² In the second scenario, BCI testing was also associated with a lower mean cost-per-patient difference estimate of \$1,803 (\$20,904 vs. \$22,708), with a sensitivity analysis range of \$300 to \$2,300.

Loncaster et al. described the projected budget impact for a single group of 201 women in the U.K. who received treatment guided by results from the Oncotype DX test compared to what the costs would have been if the test had not been used.²⁷ The estimated overall budget savings from patients not receiving chemotherapy, including the costs associated with testing all patients, was £266,427, or approximately £1,325 per patient.²⁷ Center researchers assessed this study as having a high risk of bias because of its before-after design with no true comparator group. This is a slightly higher estimate of increased cost on a per-patient basis, compared to the two U.S. non-model based studies (\$1,367 and \$400, respectively)^{53,54} described in the Blok et al. systematic review.²¹ However, the systematic review did not provide details about what costs were included in these estimates.²¹
Summary

Blok et al.²¹ did not conduct individual study risk-of-bias assessment, but given the descriptions of the modeling studies included in their systematic review, most studies would likely be rated as having a moderate to high risk of bias. There is more and higher quality information available for the Oncotype DX test than any of the others in this study,²¹ and the only two included studies^{53,54} using actual (rather than modeled) patient group inputs both studied the economic impact of the Oncotype DX test and estimated that its use would increase costs. This is consistent with the findings by Loncaster et al.²⁷ However, given that some modeling studies suggest that Oncotype DX dominates (in cost-effectiveness models "dominant" strategies cost less and are at least as effective as a comparator) treatment strategies without genetic testing, it is difficult to say with any certainty what the economic impact in the U.S. would be.

Similarly, two of the additional studies described above range from a moderate to a low risk of bias. The quality of evidence about the economic outcomes of the BCI test can be rated as very low, given that there was only one study, found to be at high risk of bias, included in this review. The Hall et al. economic study for the U.K. Health Technology Assessment program has a moderate risk of bias in the assessment of Center researchers and included three other tests of interest (Oncotype DX, MammaPrint, and Prosigna [subtype]).⁵¹ The overall quality of economic evidence about the Prosigna, EndoPredict, and Mammostrat tests is very low and the quality of economic evidence for the Oncotype DX and MammaPrint tests is low when the Blok et al.²¹ systematic review and the additional economic analysis by Hall et al.⁵¹ are considered together.

Prostate Cancer

The search located 231 unduplicated records, and after excluding 198 citations on the basis of abstract review, 33 full-text articles were reviewed. Of these, eight additional observational studies (one of which also reported cost outcomes) were included for KQ1, and two economic modeling studies were included for KQ4. No systematic reviews that met inclusion criteria were available. Studies on each of the three genetic tests (Oncotype DX for prostate cancer, Prolaris, and Decipher) were available for inclusion in the analysis.

Key Question 1: Effectiveness

All included studies were a before-after design that reported treatment decisions before and after test results and Center researchers assessed all of these studies as being at high risk of bias because of their study designs, inadequate controlling for confounding, and funding from the test manufacturers.³⁹⁻⁴⁶ Four studies employed a historical comparator group,^{39,41-43} and the remaining four^{40,44-46} used a group of patients for whom test recommendations were available before and after the test was performed.

Oncotype DX and Prolaris are applied to needle biopsy tissue and are indicated for men who have lower-risk disease and are candidates for active surveillance or definitive therapy such as radical prostatectomy or radiotherapy.^{104,108} Four of the eight included studies used Oncotype DX ^{39,40,42,43} and two^{41,46} studied Prolaris. Men enrolled in the studies using Oncotype DX^{39,40,42,43}

and Prolaris^{41,46} were predominantly categorized as having very low-, low-, or intermediate-risk disease, based on clinicopathological considerations such as PSA level, Gleason score, and biopsy result prior to testing. Patients enrolled in these studies had an average age of 60 or greater, and study populations were predominantly Caucasian. All of the studies were conducted in the U.S.

Crawford et al. and Shore et al. reported changes in treatment intensity after Prolaris testing,^{41,46} although Crawford et al. used a historical comparison group. Crawford et al. found that 40% of study subjects had decreased treatment intensity recommendations, including a recommendation away from interventional treatment in favor of active surveillance or watchful waiting for 37% of subjects.⁴¹ Shore et al. reported that treatment options were altered for 47.8% (95% Cl, 45.0 to 50.6%) of patients from the initial to the final treatment recommendation.⁴⁶ For the 576 of 1,206 patients in this group with a treatment change, the treatment intensity decreased for nearly 75%.⁴⁶ Overall, 17.6% of the study population had a change in treatment modality: 24.2% changed from noninterventional treatments.⁴⁶ At the end of the study, nearly the same proportion of patients were recommended to have interventional and noninterventional treatments, but the treatments recommended to individual patients had shifted.⁴⁶

Three^{39,42,43} of the four studies examined Oncotype DX using a historical comparison group; Badani⁴⁰ et al. employed a single group of patients with recommended treatment strategy reported both before and after the test result was available. Albala et al. reported that more men received active surveillance after the test results became available (59% vs. 38%) and that use of any active treatment (radical prostatectomy, intensity-modulated radiation therapy, or other therapies) decreased by a corresponding proportion when the test was added to practice.³⁹ Using another historical comparison study design, Dall-Era⁴² et al. reported that actual use of active surveillance or watchful waiting increased from 43% in the historical group to 67% in the testing group, and use of radical prostatectomy decreased from 27% to 14%. Eure et al. reported recommended treatment changes with the availability of testing, including that 51% of patients switch from interventional treatment to active surveillance, and 14% switch from active surveillance to interventional treatment.⁴³

This study also examined the proportion of men who continued with active surveillance one year after the initial recommendation.⁴³ The authors reported that a higher proportion of men in the testing group (53%) continued with active surveillance at one year compared to the historical group (34%), but that among men who actually initiated active surveillance (as opposed to simply having it recommended) the persistence rates were quite similar (89% and 86%, respectively).⁴³ The Badani et al. study reported changes in recommended treatment intensity, and found that recommended intensity of therapy decreased for 15.8% of men and increased for 8.9%, but that the majority (38.7%) had no or equivocal changes.⁴⁰

Two studies reported physicians' assessments of their confidence in recommended treatment and perceived usefulness of the test.^{40,43} Both studies found that a majority of physicians reported the test to be useful and that it increased their confidence in their treatment recommendation.^{40,43} In addition, Eure et al. reported that 96% of patients found the test to be useful in decision making.⁴³

The Decipher test, studied by Gore et al.⁴⁴ and Michalopoulos et al.,⁴⁵ is used in men who have been treated with radical prostatectomy and are making decisions about subsequent treatments such as hormonal therapy and adjuvant radiotherapy (ART) versus salvage radiotherapy (SRT). Both of these studies were conducted in the U.S. (with the exception of one study site in Canada⁴⁴). The mean age of patients enrolled in these studies was around 60.^{44,45} More than 80% of both the ART and SRT groups in the Gore et al.⁴⁴ study were Caucasian, but race was not reported in the Michalopoulos et al. study.⁴⁵ The spectrum of disease for patients in the Michalopoulos et al. study.⁴⁵ The spectrum of disease for patients in the Michalopoulos et al. study.⁴⁵ Mere et al.⁴⁴ study. This is reflective of the Gore et al. study enrolling men who had already had a biochemical recurrence of their prostate cancer and were therefore considering SRT.⁴⁴ Men enrolled into the ART arm in the Gore et al. study were not eligible if they were receiving adjuvant systemic chemotherapy which is a somewhat unusual clinical scenario.⁴⁴ Both studies enrolled a single group of men with known treatment recommendations before and after testing, rather than actual treatment received, and reported physician decisional conflict scores.^{44,45}

Gore et al. reported that 88.7% of the ART group and 58.3% of the SRT group had observation recommended prior to testing and that there was a change in recommended treatment strategy for 18% of the ART group and 32% of the SRT group.⁴⁴ Gore et al. used multivariable regression models to estimate the independent association between Decipher test use and change in treatment recommendation, controlling for age, pre-prostatectomy PSA level, Gleason score, extraprostatic extension, seminal vesicle invasion, surgical margin (and for the SRT group, the time to biochemical recurrence or PSA rise).⁴⁴ The adjusted OR for the ART group was 1.48 (95% CI, 1.19 to 1.85), and for the SRT group it was 1.30 (95% CI, 1.03 to 1.65).⁴⁴ Michalopoulos et al. found that before the test result was known, 69.9% of patients were recommended to have observation and 27.4% were recommended to have adjuvant treatment.⁴⁵ After the test result, 27.4% were still recommended to have adjuvant treatment, but recommendations changed for individual patients.⁴⁵ Over 42% of patients who had a recommendation of any active treatment had a change to observation only and nearly 18% with an initial recommendation of observation had a posttest recommendation of an active treatment strategy.⁴⁵ Both studies reported that there was a statistically significant decrease in physician's decisional conflict scores after the test results were available.^{44,45} Similarly, Gore et al. found that median patient Decisional Conflict Scale scores in both the ART and SRT groups had statistically significantly decreased after the test results were known.44

Summary

Although the entire group of studies on gene tests to inform treatment of prostate cancer have a high risk of bias, their findings are consistent regarding an association between test use and decreased treatment intensity and decisional conflict for patients and physicians. The overall quality of evidence for these findings is very low because of substantial limitations, including use of before-after designs, recommended rather than actual treatments, uncertainty about the persistence of treatment choices, and lack of information about important patient outcomes such as survival or treatment-related morbidity.

Key Question 2: Harms

No studies met inclusion criteria for this key question.

Key Question 3: Subpopulations

No studies met inclusion criteria for this key question, with the exception that Oncotype DX for prostate cancer and Prolaris are used in different clinical situations than the Decipher test. Those differences are described above under Key Question 1.

Key Question 4: Cost and Cost-Effectiveness

The search located one cost-effectiveness modeling study⁵⁵ on the use of Decipher for men who have had a prostatectomy and one⁵⁶ on the budget impact of the Prolaris test for men who have received the diagnosis of localized prostate cancer after a prostate biopsy. In addition, the Albala et al. study reported that the aggregate total cost of care related to prostate cancer was \$2,286 less for men who had received the Oncotype DX test compared to a historical group of men who had not.³⁹

The Ontario Health Technology Advisory Committee commissioned an economic report on adding the Prolaris test to care for men with localized prostate cancer and to make a recommendation to the Ontario Ministry of Health about public funding of the service.⁵⁶ The authors found that there was insufficient data to support a primary economic analysis because the test's effect on patient-important outcomes such as survival or need for subsequent radical prostatectomy was not known.⁵⁶ Therefore, the authors conducted a budget impact analysis of adding the test to public coverage in the provincial health service.⁵⁶ Center researchers assessed this study to be at moderate risk of bias, largely because of the very limited evidence base of the inputs and the small number of sensitivity analyses that can be considered in this type of analysis.⁵⁶ The authors reported, using units of 2016 Canadian dollars, that the estimated cost of the test itself in the first five years of use would be CAD\$47.9 million and that additional physician visits required to interpret the test would add another CAD\$0.7 million in cost.⁵⁶ The savings attributed to treatment changes (increased use of active surveillance and decreased use of interventional therapies) was estimated to decrease by CAD\$7.3 million in the same five-year time horizon.⁵⁶ The net budget impact was estimated to add CAD\$41.3 million of costs, or approximately CAD\$8 million additional cost per year.⁵⁶ The impact analysis was most sensitive to assumptions regarding uptake of the test, unit cost of the test, and the distribution of

treatments used after test introduction, but was also influenced by the spectrum of tumor risk groups, changes in prostate cancer incidence, and the proportion of patients electing watchful waiting among those receiving noninterventional treatment.⁵⁶

Lobo et al. conducted a cost-effectiveness study of care guided by the Decipher test for men who have had a radical prostatectomy.⁵⁵ The Markov model's effectiveness estimates were provided by two non-comparative before-after studies of treatment recommendations rather than actual treatments, and other model inputs were generally informed by single studies rather than systematic reviews or multiple studies.⁵⁵ The manufacturer of the Decipher test funded the study and several authors had financial relationships with the company.⁵⁵ For these reasons, Center researchers assessed the study to be at high risk of bias. Compared to usual care, this cost-effectiveness analysis found that test-based care increased the average per-person cost of care from \$18,370 to \$23,823, but increased the mean QALY per individual by 0.066 (95% CI, 0.016 to 0.117).⁵⁵ The authors stated that genetic-based care improved clinical outcomes, including 16% lower incidence of metastases at five years after protastectomy.⁵⁵ The ICER was \$90,883,⁵⁵ which is in the middle of the range generally proposed in 2017 by the Institute for Clinical and Economic Review for ICER interpretation in the U.S.⁵⁷ When genetic-based care was compared to a scenario of 100% of patients receiving adjuvant therapy, the average per-person cost of care was lower with genetic testing (\$23,823) compared to adjuvant treatment (\$31,125), and the mean difference in QALYs was somewhat higher at 0.140 (95% CI, 0.080 to 0.201).

Summary

The overall quality of evidence regarding economic outcomes for the Decipher and Prolaris tests is very low because of the risk of bias in the evidence that informs the assumptions and limitations associated with these type of modeling studies.

Colon

The search located 369 unduplicated citations, of which 35 underwent full-text review. There were no systematic reviews that met final inclusion criteria. Three primary studies (one of these with two separate publications that reported different outcomes) were included, two of which contributed to KQ1 and one to KQ4. All of these studies used the Oncotype DX for colon cancer test. No studies using ColoPrint, the other test of interest for the State of Washington, were identified and met final inclusion criteria.

Key Question 1: Effectiveness

One study was reported in two papers:^{48,49} Srivastava et al. reported treatment recommendation changes and Renfro et al. reported patient decisional conflict change and physician ratings of Oncotype DX usefulness. Both these studies^{48,49} and the other included study by Brenner et al.⁵⁰ involved patients with resected stage 2 or 2A colon cancer with mismatch repair-proficient (MMR-P) tumors. Tumors that are MMR-deficient are also said to have microsatellite instability (MSI). MSI is a predisposition to mutation that results from impaired DNA mismatch repair function. MMR-deficient tumors generally have a better prognosis, but do not respond to adjuvant 5-fluorouracil chemotherapy.⁴⁹ A total of 410 patients and 105 physicians were

included in these two studies. Both studies are at high risk of bias, largely because of the noncomparative before-after design.⁴⁸⁻⁵⁰ Additionally, measurement of changes in treatment recommendations rather than actual treatments received, funding by the test manufacturer, and the financial relationships of some authors with the manufacturer increase the risk of bias.

Srivastava et al. reported that use of the test resulted in no treatment recommendation change for 55.7% of patients, increased intensity therapy recommendations for 11.4%, and decreased intensity recommendations for 32.9%. Brenner et al. reported actual treatment received compared to initial treatment recommended before the test results were known and found that there was no change in treatment intensity for 62.1% of patients, increased intensity for 9.7%, and decreased intensity for 28.3%.⁵⁰ Thus, across both studies, patients were more likely to receive lower intensity adjuvant treatments or recommendations for treatments after use of the Oncotype DX test, but the majority did not see a change.^{49,50}

Renfro et al. reported patients' Decisional Conflict Scale scores before and after receiving Oncotype DX. The Decisional Conflict Scale measures five subscales related to patient perception of the effectiveness of the decision; feeling informed, supported, or uncertain about the decision; and whether their values regarding the decision have been clarified. The scale ranges from zero (no conflict) to 100 (extreme conflict).¹²⁷ There was a statistically significant decrease in mean overall and each subscale scores after receipt of the test (mean overall change -7.88; 95% CI, -11.25 to -4.50; p < .001).⁴⁸ However, initial scores were fairly low even before the test and below the threshold generally accepted as associated with decisional delay or uncertainty.¹²⁷

Renfro et al. also surveyed 150 physicians who cared for the patients in the study and found that most agreed (55.3%) or strongly agreed (28.7%) that they felt more confident in their recommendation after ordering the test.⁴⁸ The physicians also reported that they agreed (55.3%) or strongly agreed (30.7%) that the test provided additional clinically relevant information.⁴⁸ Fewer physicians reported that the test influenced their decision (25.4% agreed and 24.0% strongly agreed).⁴⁸ Center researchers rated this study as having high risk of bias because of the study design and because three authors had financial relationships with the test manufacturer, but it provides limited information that patients and physicians feel the Oncotype DX colon cancer test is helpful in making their decision about treatment for stage 2 colon cancer.⁴⁸

The overall quality of evidence for this outcome is very low for Oncotype DX for colon cancer. There was no evidence included that pertained to any other test.

Key Question 2: Harms

No studies met inclusion criteria for this key question.

Key Question 3: Subpopulations

No studies met inclusion criteria for this key question.

Key Question 4: Cost and Cost-Effectiveness

Alberts et al. conducted a decision analysis to determine the cost-effectiveness of using Oncotype DX to guide therapy for patients with resected stage 2 MMR-P colon cancer.⁵⁸ This cost-effectiveness study used information from the Srivastava et al. study that is described above to populate the assumptions for the clinical decision-making for the model.⁴⁹ Center researchers assessed this study as being at moderate risk of bias because of limited information about treatment effectiveness, outcome data from a single source, and lack of complete cost modeling.⁵⁸ This study was also funded by the manufacturer and a minority of its authors had financial relationships with the manufacturer.⁵⁸ Alberts et al. reported slightly lower total lifetime costs (\$991 less) with the test (\$103,775) than without it (\$104,767).⁵⁸ One-way sensitivity analysis found 76.1% probability of cost savings and QALY improvement with use of the test.⁵⁸ Quality-adjusted survival increased by 0.114 QALY, and at a QALY threshold of \$50,000, the probability of cost-effectiveness was reported to be 95.5%.⁵⁸

The overall quality of evidence for this outcome is very low for Oncotype DX for colon cancer. There was no evidence included that pertained to any other test.

Multiple Myeloma

The search located 215 unduplicated studies, of which 197 were eliminated after dual citation and abstract screening. Full-text review was performed for 19 studies, and after that none remained eligible for final inclusion.

Key Question 1: Effectiveness

No studies met inclusion criteria for this key question.

Key Question 2: Harms

No studies met inclusion criteria for this key question.

Key Question 3: Subpopulations

No studies met inclusion criteria for this key question.

Key Question 4: Cost and Cost-Effectiveness No studies met inclusion criteria for this key question.

The overall quality of evidence for outcomes of gene expression profile tests for multiple

myeloma is not rated because no evidence met inclusion criteria.

Clinical Practice Guidelines

Breast Cancer

The most detailed clinical practice guideline, *Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women with Early-Stage Invasive Breast Cancer*, was published by the American Society of Clinical Oncology (ASCO) in 2016.⁵⁹ ASCO published a guideline update in 2017 modifying the recommendations regarding MammaPrint, which draws upon recently published studies.⁶⁰ Both of these guidelines were rated as having good methodological quality. The detailed ASCO recommendations for the use of biomarkers in early-stage breast cancer are in Appendix G.

The ASCO guidelines outlined recommendations for when Oncotype DX breast cancer assay, MammaPrint, EndoPredict, Prosigna, BCI, and Mammostrat should or should not be used in patients with early-stage breast cancer. All of these tests, except for Mammostrat, are recommended for use in patients who have ER-positive/PR-positive, HER2-negative, LNnegative breast cancer.^{59,60} The guidelines recommend against the use of Mammostrat for the following categories of breast cancers: ER-positive/PR-positive, HER2-negative (LN-positive or negative); HER2-positive; or ER-negative/PR-negative, HER2-negative, LN-negative.⁵⁹

According to the ASCO guidelines, MammaPrint should not be used in patients with low clinical risk (as defined by the Adjuvant! Online tool as used in the MINDACT study¹⁹), because women in the low clinical risk category had very good outcomes and did not appear to benefit from chemotherapy even with a genomic high-risk cancer.⁶⁰ MammaPrint can be used in patients with ER-positive/PR-positive, HER2-negative, LN-positive breast cancer who have one to three positive nodes and are at high clinical risk per MINDACT categorization.⁶⁰ Still, these patients should be informed that a benefit of chemotherapy cannot be excluded, particularly among patients with more than one involved lymph node.⁶⁰ The ASCO guidelines recommended against using the other tests in patients with ER-positive/PR-positive, HER2-negative, LN-positive/PR-positive, HER2-negative, LN-positive breast cancer; HER2-positive breast cancer; or ER-negative, HER2-negative, LN-negative breast cancer.⁵⁹

The authors of the 2017 NCCN clinical practice guidelines on breast cancer discussed the evidence for Oncotype DX (21-gene breast cancer assay), MammaPrint (70-gene assay), and Prosigna (50-gene assay).⁶¹ According to the guidelines, Oncotype DX can be considered for ER-positive/PR-positive, HER2-negative cancers with pT1, pT2, or pT3, and pN0 or pN1mi \leq 2 mm axillary node metastasis and a tumor greater than 0.5 cm.⁶¹ Oncotype DX can also be considered in certain patients with one to three involved ipsilateral axillary lymph nodes to guide the addition of combination chemotherapy to standard hormone therapy.⁶¹ The NCCN guideline authors stated that the other gene expression profile tests can be considered to assess risk of cancer recurrence, but that they have not been validated to predict response to chemotherapy.⁶¹ Center researchers rated the NCCN guidelines as having fair methodological quality.

NICE published guidelines in 2013 that assessed the use of Oncotype DX breast cancer assay, MammaPrint, Mammostrat, and immunohistochemical 4 (IHC4) score in early-stage breast cancer.⁶² The guidelines recommend Oncotype DX as an option for guiding adjuvant chemotherapy decisions for people with ER-positive, HER2-negative, LN-negative early-stage breast cancer when the patient is assessed as being at intermediate risk.⁶² According to the guidelines, Oncotype DX should only be used when the test results are likely to help in predicting the course of the disease, and therefore in the decision of whether to prescribe chemotherapy.⁶² MammaPrint and Mammostrat are only recommended for use in research in

patients with ER-positive, HER2-negative, LN-negative early-stage breast cancer.⁶² Center researchers rated the NICE guidelines as having good methodological quality.

The European Society for Medical Oncology (ESMO) published breast cancer clinical practice guidelines in 2015.⁶³ The ESMO guidelines recommend that gene expression profile tests, such as Oncotype DX breast cancer assay, MammaPrint, EndoPredict, and Prosigna, can be used to complement pathology assessments to predict the benefit of adjuvant chemotherapy.⁶³ In cases when decisions might be challenging, such as in luminal B HER2-negative and LN-negative breast cancer, Oncotype DX, EndoPredict, and Prosigna can be used.⁶³ For all types of breast cancer (pN0–1), MammaPrint can be used in conjunction with clinicopathological factors to help in decision making about treatment.⁶³ Center researchers rated the ESMO guidelines as having poor methodological quality.

The European Group on Tumor Markers (EGTM) published a guideline in 2017 on the use of biomarkers in breast cancer.⁶⁴ These guidelines recommend that the Oncotype DX breast cancer assay, MammaPrint, EndoPredict, Prosigna, and BCI can be used to aid in adjuvant therapy decision making in ER-positive, HER2-negative, LN-negative patients.⁶⁴ In addition, Oncotype DX, MammaPrint, EndoPredict, and Prosigna can be used in patients with one to three metastatic lymph nodes.⁶⁴ Center researchers rated the EGTM guidelines as having poor methodological quality. The detailed recommendations from the EGTM are in Appendix G. Table 4 summarizes these five guidelines on breast cancer, indicating whether the gene expression profile tests are recommended for LN-negative and/or LN-positive cancers.

Test	ASCO	NCCN	NICE	ESMO**	EGTM
Oncotype DX	LN-negative	LN-negative LN-positive	0	LN-negative LN-positive	LN-negative LN-positive
MammaPrint	LN-negative LN-positive	Not recommended*		LN-negative LN-positive	LN-negative LN-positive
EndoPredict	LN-negative	Not recommended*	No guideline recommendation	LN-negative LN-positive	LN-negative LN-positive
Prosigna	LN-negative	Not recommended*	No guideline recommendation	LN-negative LN-positive	LN-negative LN-positive
Breast Cancer Index	LN-negative	Not recommended*	No guideline recommendation	No guideline recommendation	LN-negative
Mammostrat	Not recommended	Not recommended*		No guideline recommendation	No guideline recommendation

Table 4. Recommendations for Lymph Node Status in Guidelines on theUse of Gene Expression Tests in Early-Stage Breast Cancer

*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. **The ESMO guideline authors did not distinguish between LN-negative and LN-positive cancers in their recommendations.

Prostate Cancer

Two clinical practice guidelines were identified that included recommendations on the use of Decipher, Prolaris, and Oncotype DX prostate cancer assay. The 2017 NCCN guidelines on prostate cancer stated that men with clinically localized prostate cancer may consider the use of tumor-based molecular assays, and the authors made specific recommendations on the use of Decipher, Prolaris, and Oncotype DX for prostate cancer.⁶⁵ The guidelines recommend Decipher after a radical prostatectomy for patients with pT2 (confined to prostate) with positive margins, any pT3 (extraprostatic extension) disease, and a rising PSA level.⁶⁵ Prolaris and Oncotype DX are recommended post-biopsy for low- and very low-risk prostate cancer in patients with at least 10 years of life expectancy who have not received other active treatment for prostate cancer and who are candidates for active surveillance or definitive therapy.⁶⁵ Center researchers rated the NCCN guidelines as having fair methodological quality.

A guideline on clinically localized prostate cancer has been jointly published by the American Urological Association, American Society for Radiation Oncology, and Society of Urologic Oncology in 2017.⁶⁶ These guidelines include the following recommendation based on expert opinion: "Tissue-based genomic biomarkers have not shown a clear role in active surveillance for localized prostate cancer and are not necessary for follow up."^{66(p. 4)} Center researchers rated these guidelines as having good methodological quality.

Colon Cancer

No clinical practice guidelines were found that included recommendations for the use of ColoPrint or Oncotype DX for colon cancer. The American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and ASCO published a guideline on molecular biomarkers for colorectal cancer in 2017.⁶⁷ This guideline stated, "A problem of quantitative assays, such as gene expression, microRNA expression, and methylation levels, tested in solid tumors, results from the intrinsic mixed nature of the tissue with significant variability of tumor and non-tumor tissue content. Another limitation of molecular biomarker discovery approaches that rely on expression levels is that these biomarkers have not been evaluated in the context of complex molecular regulation of individual cancer subtypes."^{67(p. 1482)} Center researchers rated these guidelines as having good methodological quality.

The fair-methodological-quality 2017 NCCN guidelines on colon cancer discussed multigene assays, including ColoPrint and Oncotype DX colon cancer assay, and concluded that there is no evidence of predictive value in terms of the potential benefit of chemotherapy for any of the multigene assays.⁶⁸ Similarly, the 2016 guidelines on metastatic colon cancer from ESMO concluded that gene expression signatures have failed to accurately predict disease recurrence and prognosis.⁶⁹ Center researchers rated the ESMO guidelines as having poor methodological quality.

Multiple Myeloma

The authors of the 2017 NCCN guidelines on multiple myeloma discussed gene expression profiling tests, including MyPRS and SKY92, but did not make any recommendations about the use of these tests.⁷⁰ The NCCN panel unanimously agreed that although gene expression profile tests are not routinely used, they could be helpful in selected patients to estimate the aggressiveness of the disease and to individualize treatment.⁷⁰ Center researchers rated the NCCN guidelines as having fair methodological quality. The authors of the 2017 guidelines on multiple myeloma from ESMO stated that gene-expression profiling is not currently used routinely, and more research is needed to identify molecular markers, which could lead to advances in this area.⁷¹ Center researchers rated the ESMO guidelines as having fair methodological practice guidelines were found that included recommendations for the use of My Prognostic Risk Signature (MyPRS) or SKY92 tests.

Selected Payer Coverage Determinations

Breast Cancer

Medicare

No Medicare NCDs were found for any of the gene expression profile tests for breast cancer. Center researchers identified LCDs by Noridian Healthcare Solutions, a Medicare contractor in Washington, that provide coverage for EndoPredict, Prosigna, and BCI.

The EndoPredict LCD provides coverage for women with T1-3, N0-1 breast cancer when the following criteria are met:

- Patient is postmenopausal
- Pathology reveals invasive carcinoma of the breast that is ER-positive, HER2-negative
- Patient is either LN-negative or has 1 to 3 positive lymph nodes
- Patient has no evidence of distant metastasis
- Test result will be used to determine treatment choice between endocrine therapy alone vs. endocrine therapy plus chemotherapy⁷²

The Prosigna LCD provides coverage for postmenopausal women with either:

- ER-positive, LN-negative, stage 1 or 2 breast cancer or
- ER-positive, LN-positive (one to three positive nodes), stage 2 breast cancer⁷³

The BCI LCD provides coverage for patients who meet these criteria:

- Postmenopausal female with non-relapsed, ER-positive breast cancer
- LN-negative
- Completing five years of tamoxifen therapy
- Must be eligible for consideration of extended endocrine therapy based on published clinical trial data or practice guidelines
- Physician or patient is concerned about continuing anti-hormonal therapy because of documented meaningful toxicity or possible significant patient-specific side effects

• 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).⁷⁴

No Medicare LCDs covering Washington were found for Oncotype DX breast cancer assay, MammaPrint, or Mammostrat.

Aetna

The Aetna policy on tumor markers provides coverage for the Oncotype DX breast cancer assay, MammaPrint, EndoPredict, Prosigna, and BCI to assess the necessity of adjuvant chemotherapy in females or males with recently diagnosed breast tumors.⁷⁵ Coverage for these tests requires that adjuvant chemotherapy is not precluded by any other factor (e.g., advanced age or significant comorbidities), and that the patient and physician have discussed the potential results of the test and agree to use the results to guide therapy.⁷⁵ Aetna does not cover Mammostrat.⁷⁵

The following are Aetna's additional criteria for coverage of the Oncotype DX breast cancer assay:

- Breast cancer is nonmetastatic (LN-negative) or with one to three involved ipsilateral axillary lymph nodes
- Breast tumor is ER-positive
- Breast tumor is HER2-negative or breast tumor is HER2-positive and less than 1 cm in diameter⁷⁵

The following are Aetna's additional criteria for coverage of MammaPrint:

- Breast cancer is nonmetastatic (LN-negative) or with one to three involved ipsilateral axillary lymph nodes
- Breast tumor is ER-positive or PR-positive
- Breast tumor is HER2-negative
- Member is determined to be at "high clinical risk" of recurrence using Adjuvant! Online⁷⁵

The following are Aetna's additional criteria for coverage of EndoPredict:

- Breast cancer is nonmetastatic (LN-negative)
- Breast tumor is ER-positive
- Breast tumor is HER2-negative⁷⁵

The following are Aetna's additional criteria for coverage of Prosigna:

- Breast cancer is nonmetastatic (LN-negative)
- Breast tumor is ER-positive
- Breast tumor is HER2-negative⁷⁵

The following are Aetna's additional criteria for coverage of BCI:

- Breast cancer is nonmetastatic (LN-negative)
- Breast tumor is ER-positive
- Breast tumor is HER2-negative⁷⁵

Cigna

The Cigna policy on gene expression assays covers the Oncotype DX breast cancer assay, MammaPrint, and Prosigna under certain conditions, and does not provide coverage for EndoPredict, BCI, or Mammostrat.⁷⁶

The following are Cigna's coverage criteria for the Oncotype DX breast cancer assay:

- Recently diagnosed stage 1 or stage 2 breast cancer
- ER-positive
- HER2-negative
- No evidence of distant metastasis
- Either of the following criteria:
 - Axillary-node status is negative (micrometastasis is no greater than 2.0 mm) whether the woman is pre- or postmenopausal
 - Up to three positive axillary nodes in a postmenopausal woman⁷⁶

The following are Cigna's coverage criteria for MammaPrint:

- Stage 1 or stage 2 invasive breast cancer and being considered for adjuvant systemic therapy
- High clinical risk of recurrence
- ER-positive/PR-positive
- HER2-negative
- Up to three positive nodes⁷⁶

The following are Cigna's coverage criteria for Prosigna:

- Recently diagnosed stage 1 or stage 2 breast cancer
- ER-positive
- HER2-negative
- Postmenopausal
- No evidence of distant metastasis
- Axillary node status is negative (micrometastasis is no greater than 2.0 mm)⁷⁶

Regence

The Regence policy on gene expression testing for breast cancer provides coverage for the Oncotype DX breast cancer assay, EndoPredict, and BCI under certain conditions, and does not cover MammaPrint, Prosigna, or Mammostrat.⁷⁷ Regence covers the Oncotype DX breast cancer assay, EndoPredict, and BCI for women with primary breast cancer, stage 1, 2, or 3, to determine

recurrence risk for deciding whether or not to undergo adjuvant chemotherapy when these criteria are met:

- Individual has had excision of breast mass and full pathologic evaluation of the specimen has been completed (i.e., the test should not be ordered on a preliminary core biopsy)
- Primary tumor size 0.6 cm to 1 cm with moderate/poor differentiation or unfavorable features, OR tumor size of 1 cm or greater
- If there are multiple ipsilateral primary tumors, a specimen from the tumor with the most aggressive histological characteristics should be submitted for testing
- ER-positive/PR-positive
- HER2-negative
- Negative lymph nodes (nodes with micrometastases of 2 mm or less in size are considered node negative)
- The test result will aid the patient in making a decision regarding chemotherapy when chemotherapy is a therapeutic option⁷⁷

Prostate Cancer

No Medicare NCDs were found for Decipher, Prolaris, or Oncotype DX for prostate cancer. There are LCDs for Noridian Healthcare Solutions, applying to the state of Washington, that provide coverage for Decipher, Prolaris, and Oncotype DX for prostate cancer under certain conditions.

The LCD for Decipher provides coverage after radical prostatectomy when all these conditions are met:

- Patient with prostate cancer who has undergone a radical prostatectomy within the previous 60 months and is being considered for postoperative secondary therapy due to one or more cancer-recurrence risk factors
- Patient must have achieved initial PSA nadir (defined as undetectable PSA) within 30 days of radical prostatectomy surgery
- Patient must not have any evidence of distant metastasis
- Patient must not have received any neo-adjuvant treatment prior to surgery
- Decipher test is performed on a patient's radical prostatectomy specimen
- Patient's surgical pathology report or medical records must have documented presence of adverse pathology:
 - o Pathological stage T2 disease with a positive surgical margin, or
 - Pathological stage T3 disease (e.g., extraprostatic extension, seminal vesicle invasion, bladder neck invasion), or
 - Rising PSA after initial PSA nadir
- Testing has been ordered by a physician who is certified in the GenomeDx Decipher Certification and Training Registry⁷⁸

There are two LCDs providing coverage for Prolaris, one for patients with early-stage, needlebiopsy-proven prostate cancer⁷⁹ and the other for patients with favorable intermediate-risk, needle-biopsy-proven prostate cancer.⁸⁰ Coverage for Prolaris for early-stage prostate cancer requires these conditions to be met:

- Needle biopsy with localized adenocarcinoma of prostate (no clinical evidence of metastasis or lymph node involvement)
- Formalin-fixed paraffin-embedded prostate biopsy specimen with at least 0.5 mm of cancer length
- Patient stage as defined by one of the following:
 - Very Low Risk Disease (T1c and Gleason Score ≤ 6 and PSA ≤ 10 ng/mL and < 3 prostate cores with tumor and ≤ 50% cancer in any core and PSA density of < 0.15 ng/mL/g) OR
 - Low Risk Disease (T1-T2a and Gleason Score ≤ 6 and PSA ≤ 10 ng/mL)
- Patient has an estimated life expectancy of greater than or equal to 10 years
- Patient is a candidate for and is considering conservative therapy and yet and would be eligible for definitive therapy (radical prostatectomy, radiation therapy, or brachytherapy)
- Result will be used to determine treatment between definitive therapy and conservative management
- Patient has not received pelvic radiation or androgen deprivation therapy prior to the biopsy
- Test is ordered by a physician certified in the Myriad Prolaris™ Certification and Training Registry
- Patient is monitored for disease progression according to established standard of care
- Physician must report the development of prostate cancer metastasis or prostate cancer deaths in patients not treated definitively who were deemed low risk by the assay⁷⁹

Prolaris coverage for favorable, intermediate-risk prostate cancer requires these conditions to be met:

- Needle biopsy with localized adenocarcinoma of prostate (no clinical evidence of metastasis or lymph node involvement)
- Formalin-fixed paraffin-embedded prostate biopsy specimen with at least 0.5 mm of cancer length
- Patients with favorable intermediate-risk disease, defined by the NCCN as follows:
 - Predominant Gleason grade 3 (i.e., Gleason score 3 + 4 = 7), percentage of positive cores < 50%, and no more than one NCCN intermediate-risk factor)
 - NCCN intermediate risk factors include T2b-T2c, Gleason score 7, and PSA 10-20 ng/mL
- Patient has an estimated life expectancy of greater than or equal to 10 years
- Patient is a candidate for and is considering conservative therapy and would be eligible for definitive therapy (radical prostatectomy, radiation therapy, or brachytherapy)
- Result will be used to determine treatment between definitive therapy and conservative management

- Patient has not received pelvic radiation or androgen deprivation therapy prior to the biopsy
- Patient is monitored for disease progression according to established standard of care⁸⁰

The LCD for Oncotype DX for early-stage, needle-biopsy-proven prostate cancer provides coverage when all these conditions are met:

- Needle biopsy with localized adenocarcinoma of prostate (no clinical evidence of metastasis or lymph node involvement)
- Patient stage as defined by the one of the following:
 - Very Low Risk Disease (T1c and Gleason Score = 6 and PSA = 10 ng/mL and < 3 prostate cores with tumor and = 50% cancer in any core and PSA density of < 0.15 ng/mL/g) OR
 - Low Risk Disease (T1-T2a and Gleason Score = 6 and PSA = 10 ng/mL)
- Patient has an estimated life expectancy of \geq 10 years
- Patient has a life expectancy of 10 to 20 years
- Patient is a candidate for and is considering conservative therapy and would be eligible for definitive therapy (radical prostatectomy, radiation therapy, or brachytherapy)
- Patient has not received pelvic radiation or androgen deprivation therapy prior to the biopsy
- Test is ordered by a physician certified in the Oncotype DX prostate cancer assay Certification and Training Registry
- Patient is monitored for disease progression according to active surveillance guidelines as recorded in NCCN guidelines
- Physician must report the development of metastasis or prostate cancer deaths in patients not treated definitively who were deemed low risk by the assay⁸¹

The coverage policies for Aetna⁷⁵ and Regence⁸² consider Decipher, Prolaris, and Oncotype DX prostate cancer assay to be experimental or investigational. Cigna does not include Decipher, Prolaris, or Oncotype DX prostate cancer assay in the list of medically necessary prostate cancer prognostic tests.⁷⁶

Colon Cancer

No Medicare NCDs or LCDs were found for ColoPrint or the Oncotype DX colon cancer assay. The policies for Aetna,⁷⁵ Cigna,⁷⁶ and Regence⁸³ do not provide coverage for ColoPrint or Oncotype DX colon cancer assay.

Multiple Myeloma

No Medicare NCDs or LCDs were found for MyPRS or SKY92. The policy for Aetna does not cover MyPRS, and this coverage policy does not mention SKY92.⁷⁵ Cigna's coverage policy on tumor markers does not mention MyPRS or SKY92.⁷⁶ The Regence coverage policy states that all microarray-based gene expression profile testing for multiple myeloma is considered investigational.⁸⁴

Conclusions

The GRADE quality of evidence tables in Appendix E summarize the strength of evidence for the gene expression profile tests for the outcomes of interest. No high-quality evidence of clinical utility exists to guide decisions about the use of gene expression profile tests for breast, prostate, or colon cancers. The only condition with quality of evidence ratings above very low was breast cancer and only for the MammaPrint and Oncotype DX tests. A single RCT found moderate-quality evidence that women with early-stage invasive breast cancer who are considered to be at high clinical risk by the Adjuvant! Online risk assessment tool may safely forego adjuvant systemic chemotherapy if their MammaPrint genomic risk score is low, given that there was no statistically significant difference in distant metastasis-free survival at five years. Moderate-quality evidence supports the use of Oncotype DX because of its impact on clinical treatment recommendations and chemotherapy use for women with early-stage invasive breast cancer, and particularly because of its ability to identify low-risk women who would not benefit from adjuvant systemic chemotherapy. Moderate-quality evidence supports the use of Oncotype DX because of its impact on clinical treatment recommendations and chemotherapy use for women with early-stage invasive breast cancer. There is also low-quality evidence about the decision impact of using MammaPrint for women with early-stage invasive breast cancer.

Based primarily on modeling studies, low-quality evidence showed that Oncotype DX and MammaPrint are cost-effective at conventional thresholds when used to guide treatment decisions among women with early-stage invasive breast cancer. Among the remaining conditions and tests, there is very low-quality evidence or a complete absence of evidence to support use of these tests to improve clinical decision making and important patient outcomes. Minimal evidence exists on the clinical utility of gene expression profile tests for prostate and colon cancers, and no evidence exists for multiple myeloma.

Yet, there is also little evidence of direct patient harm from the use of these tests. Although some patients might experience increased anxiety and distress from any test result, more patients were likely to find the tests useful and reassuring, based on very low-quality evidence for breast, prostate, and colon cancer tests.

The review sought information about two distinct types of clinical utility: the impact on patient outcomes such as mortality, morbidity, and quality of life; and the impact on treatment decisions. Most of the evidence found for this review was on the latter. When testing has been found to influence clinical decision making, it could reflect that testing is effective for this purpose or that clinicians are simply more aware of the test. In the absence of patient-important clinical outcomes, there is no assurance that an effect on decision making does, in fact, improve care and health outcomes. Avoidance of unnecessary treatments, which by definition result in harms without providing benefits, is not inconsequential. However, patients need to be certain that forgoing treatment based on a low-risk genetic test result does not result in worse clinical outcomes later in life.

Clinical practice guidelines are generally more liberal in their recommendations about test use, largely because most of these tests can provide information about overall prognosis even if they lack evidence of clinical utility. For example, ASCO endorses all of the breast cancer tests (with the exception of Mammostrat) for use in LN-negative patients and the use of MammaPrint for patients with limited lymph node involvement. In contrast, NCCN has adopted a more rigorous standard and advises the exclusive use of Oncotype DX for LN-negative and LN-positive patients, and states that other tests may be considered to help assess risk of recurrence, but have not been validated to predict response to chemotherapy.

For most of the tests included in this report, multiple clinical trials are underway (see Appendix F), including some large and important trials that will be reporting results in the next few years. Among these are the TAILORx and RxPONDER breast cancer trials for Oncotype DX and additional results for the MINDACT trial involving MammaPrint for LN-positive women.

References

- 1. Burke W. Genetic tests: clinical validity and clinical utility. In: *Curr Protoc Hum Genet*. Vol 81.2014:9.15.11-18.
- 2. American Cancer Society. Lifetime risk of developing or dying from cancer. 2016; https://www.cancer.org/cancer/cancer-basics/lifetime-probability-of-developing-or-dying-from-cancer.html.
- 3. Centers for Disease Control and Prevention. Cancer. 2016; https://www.cdc.gov/chronicdisease/resources/publications/aag/dcpc.htm.
- 4. National Cancer Institute. Types of cancer treatment. 2017; https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types.html.
- American Cancer Society. Making treatment decisions. 2017; https://www.cancer.org/treatment/understanding-your-diagnosis/after-diagnosis/making-treatment-decisions.html.
- Meleth S, Reeder-Hayes K, Ashok M, et al. Technology assessment of molecular pathology testing for the estimation of prognosis for common cancers. In. *Technology Assessment of Molecular Pathology Testing for the Estimation of Prognosis for Common Cancers*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014.
- 7. Canfield SE, Kibel AS, Kemeter MJ, Febbo PG, Lawrence HJ, Moul JW. A guide for clinicians in the evaluation of emerging molecular diagnostics for newly diagnosed prostate cancer. *Rev Urol.* 2014;16(4):172-180.
- 8. SkylineDx. About MMprofiler[™] with SKY92. 2017; http://www.skylinedx.com/news-events.html.
- 9. SkylineDx. News and press releases. 2017; http://skylinedx.com/newsevents/93/SKY92_ISS_Superior_to_current_markers_in_risk_classification_for_myeloma_surviva l.html.
- 10. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: Advancing guideline development, reporting and evaluation in health care. *Can Med Assoc J.* 2010;182(18):E839-E842.
- 11. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. 2017; https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada.
- 12. Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. http://handbook.cochrane.org/2017.
- 13. Scottish Intercollegiate Guidelines Network. Methodology checklist 1: Systematic reviews and meta-analyses. 2015; http://www.sign.ac.uk/checklists-and-notes.html.
- 14. Scottish Intercollegiate Guidelines Network. Methodology checklist 2: Randomised controlled trials. 2015; http://www.sign.ac.uk/checklists-and-notes.html.

- 15. Scottish Intercollegiate Guidelines Network. Methodology checklist 5: Diagnostic studies. 2015; http://www.sign.ac.uk/checklists-and-notes.html.
- 16. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926.
- 17. Schünemann H, Brozek J, Guyatt G, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group, 2013. 2014; http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.
- Bear HD, Wan W, Robidoux A, et al. Using the 21-gene assay from core needle biopsies to choose neoadjuvant therapy for breast cancer: a multicenter trial. *J Surg Oncol.* 2017;115(8):917-923.
- 19. 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.
- 20. Augustovski F, Soto N, Caporale J, Gonzalez L, Gibbons L, Ciapponi A. Decision-making impact on adjuvant chemotherapy allocation in early node-negative breast cancer with a 21-gene assay: systematic review and meta-analysis. *Breast Cancer Res Treat*. 2015;152(3):611-625.
- 21. 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;62:74-90.
- 22. Scope A, Essat M, Pandor A, et al. Gene expression profiling and expanded immunohistochemistry tests to guide selection of chemotherapy regimens in breast cancer management: a systematic review. *Int J Technol Assess Health Care*. 2017;33(1):32-45.
- 23. Evans CN, Brewer NT, Vadaparampil ST, et al. Impact of genomic testing and patientreported outcomes on receipt of adjuvant chemotherapy. *Breast Cancer Res Treat*. 2016;156(3):549-555.
- 24. Friese CR, Li Y, Bondarenko I, et al. Chemotherapy decisions and patient experience with the recurrence score assay for early-stage breast cancer. *Cancer*. 2017;123(1):43-51.
- 25. Jasem J, Amini A, Rabinovitch R, et al. 21-gene recurrence score assay as a predictor of adjuvant chemotherapy administration for early-stage breast cancer: an analysis of use, therapeutic implications, and disparity profile. *J Clin Oncol.* 2016;34(17):1995-2002.
- 26. 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*. 2017;15(4):494-503.
- 27. Loncaster J, Armstrong A, Howell S, et al. Impact of Oncotype DX breast recurrence score testing on adjuvant chemotherapy use in early breast cancer: real world experience in Greater Manchester, UK. *Eur J Surg Oncol.* 2017;43(5):931-937.

- 28. O'Neill SC, Isaacs C, Lynce F, et al. Endocrine therapy initiation, discontinuation and adherence and breast imaging among 21-gene recurrence score assay-eligible women under age 65. *Breast Cancer Res.* 2017;19(1):45.
- 29. Parsons BM, Landercasper J, Smith AL, Go RS, Borgert AJ, Dietrich LL. 21-Gene recurrence score decreases receipt of chemotherapy in ER+ early-stage breast cancer: an analysis of the NCDB 2010-2013. *Breast Cancer Res Treat*. 2016;159(2):315-326.
- 30. Potosky AL, O'Neill SC, Isaacs C, et al. Population-based study of the effect of gene expression profiling on adjuvant chemotherapy use in breast cancer patients under the age of 65 years. *Cancer*. 2015;121(22):4062-4070.
- 31. Ray GT, Mandelblatt J, Habel LA, et al. Breast cancer multigene testing trends and impact on chemotherapy use. *Am J Manag Care*. 2016;22(5):e153-160.
- 32. Kuijer A, Straver M, den Dekker B, et al. Impact of 70-gene signature use on adjuvant chemotherapy decisions in patients with estrogen receptor-positive early breast cancer: results of a prospective cohort study. *J Clin Oncol.* 2017;35(24):2814-2819.
- 33. Kuijer A, van Bommel AC, Drukker CA, et al. Using a gene expression signature when controversy exists regarding the indication for adjuvant systemic treatment reduces the proportion of patients receiving adjuvant chemotherapy: a nationwide study. *Genet Med.* 2016;18(7):720-726.
- 34. Tsai M, Lo S, Audeh W, et al. Association of 70-gene signature assay findings with physicians' treatment guidance for patients with early breast cancer classified as intermediate risk by the 21-gene assay. *JAMA Oncol.* 2017:e173470.
- 35. Hequet D, Callens C, Gentien D, et al. Prospective, multicenter French study evaluating the clinical impact of the breast cancer intrinsic subtype-Prosigna(R) Test in the management of early-stage breast cancers. *PLoS One*. 2017;12(10):e0185753.
- 36. Wuerstlein R, Sotlar K, Gluz O, et al. The West German Study Group Breast Cancer Intrinsic Subtype study: a prospective multicenter decision impact study utilizing the Prosigna assay for adjuvant treatment decision-making in estrogen-receptor-positive, HER2-negative earlystage breast cancer. *Curr Med Res Opin.* 2016;32(7):1217-1224.
- 37. Sanft T, Aktas B, Schroeder B, et al. Prospective assessment of the decision-making impact of the Breast Cancer Index in recommending extended adjuvant endocrine therapy for patients with early-stage ER-positive breast cancer. *Breast Cancer Res Treat.* 2015;154(3):533-541.
- 38. Pestalozzi BC, Tausch C, Dedes KJ, et al. Adjuvant treatment recommendations for patients with ER-positive/HER2-negative early breast cancer by Swiss tumor boards using the 21-gene recurrence score (SAKK 26/10). *BMC Cancer*. 2017;17(1):265.
- 39. Albala D, Kemeter MJ, Febbo PG, et al. Health economic impact and prospective clinical utility of Oncotype DX(R) genomic prostate score. *Rev Urol.* 2016;18(3):123-132.
- 40. Badani KK, Kemeter MJ, Febbo FG, et al. The impact of a biopsy based 17-gene genomic prostate score on treatment recommendations in men with newly diagnosed clinically prostate cancer who are candidates for active surveillance. *Urol Pract.* 2015;2(4):181-189.

- 41. 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;30(6):1025-1031.
- 42. Dall'Era MA, Maddala T, Polychronopoulos L, Gallagher JR, Febbo PG, Denes BS. Utility of the Oncotype DX prostate cancer assay in clinical practice for treatment selection in men newly diagnosed with prostate cancer: a retrospective chart review analysis. *Urology Practice*. 2015;2(6):343-348.
- 43. Eure G, Germany R, Given R, et al. Use of a 17-gene prognostic assay in contemporary urologic practice: results of an interim analysis in an observational cohort. *Urology*. 2017.
- 44. Gore JL, du Plessis M, Santiago-Jimenez M, et al. Decipher test impacts decision making among patients considering adjuvant and salvage treatment after radical prostatectomy: interim results from the Multicenter Prospective PRO-IMPACT study. *Cancer.* 2017;123(15).
- 45. Michalopoulos SN, Kella N, Payne R, et al. Influence of a genomic classifier on post-operative treatment decisions in high-risk prostate cancer patients: results from the PRO-ACT study. *Curr Med Res Opin.* 2014;30(8):1547-1556.
- 46. Shore ND, 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;195(3):612-618.
- 47. Biroschak JR, Schwartz GF, Palazzo JP, et al. Impact of Oncotype DX on treatment decisions in ER-positive, node-negative breast cancer with histologic correlation. *Breast J*. 2013;19(3):269-275.
- 48. Renfro LA, Zhang N, Lopatin M, Chao C, Alberts SR. Prospective evaluation of a 12-gene assay on patient treatment decisions and physician confidence in mismatch repair proficient stage IIA colon cancer. *Clin Colorectal Cancer*. 2017;16(1):23-30.
- 49. Srivastava G, Renfro LA, Behrens RJ, et al. Prospective multicenter study of the impact of Oncotype DX colon cancer assay results on treatment recommendations in stage II colon cancer patients. *Oncologist*. 2014;19(5):492-497.
- 50. Brenner B, Geva R, Rothney M, et al. Impact of the 12-gene colon cancer assay on clinical decision making for adjuvant therapy in stage II colon cancer patients. *Value Health*. 2016;19(1):82-87.
- 51. Hall PS, Smith A, Hulme C, et al. Value of information analysis of multiparameter tests for chemotherapy in early breast cancer: the OPTIMA prelim trial. *Value Health*. 2017(pagination).
- 52. Gustavsen G, Schroeder B, Kennedy P, et al. Health economic analysis of Breast Cancer Index in patients with ER+, LN- breast cancer. *Am J Manag Care*. 2014;20(8):e302-310.
- 53. Su KW, Hall J, Soulos PR, et al. Association of 21-gene recurrence score assay and adjuvant chemotherapy use in the Medicare population, 2008-2011. *J Geriatr Oncol*. 2016;7(1):15-23.
- 54. Epstein AJ, Wong YN, Mitra N, et al. Adjuvant chemotherapy use and health care costs after introduction of genomic testing in breast cancer. *J Clin Oncol.* 2015;33(36):4259-4267.

- 55. Lobo JM, Trifiletti DM, Sturz VN, et al. Cost-effectiveness of the Decipher genomic classifier to guide individualized decisions for early radiation therapy after prostatectomy for prostate cancer. *Clin Genitourin Cancer*. 2017;15(3):e299-e309.
- 56. Health Quality Ontario. Prolaris cell cycle progression test for localized prostate cancer: a health technology assessment. *Ont Health Technol Assess Ser.* 2017;17(6):1-75.
- 57. Institute for Clinical and Economic Review. Overview of the ICER value framework and proposals for an update for 2017-2018. 2017; http://icer-review.org/wp-content/uploads/2016/02/ICER-VAF-Update-Proposals-020117.pdf.
- 58. Alberts SR, Yu TM, Behrens RJ, et al. Comparative economics of a 12-gene assay for predicting risk of recurrence in stage II colon cancer. *Pharmacoeconomics*. 2014;32(12):1231-1243.
- 59. 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.* 2016;34(10):1134-1150.
- 60. Krop I, Ismaila N, Andre F, 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 focused update. *J Clin Oncol.* 2017;35(24):2838-2847.
- 61. 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.
- 62. National Institute for Health and Care Excellence. Gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat. 2013; https://www.nice.org.uk/guidance/dg10.
- 63. Senkus E, Kyriakides S, Ohno S, et al. Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2015;26 Suppl 5:v8-30.
- 64. 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.
- 65. 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.
- 66. Sanda MG, Chen RC, Crispino T, et al. Clinically localized prostate cancer: AUA/ASTRO/SUO Guideline. 2017; http://www.auanet.org/guidelines/clinically-localized-prostate-cancer-new-(aua/astro/suo-guideline-2017).
- 67. Sepulveda AR, Hamilton SR, Allegra CJ, et al. Molecular biomarkers for the evaluation of colorectal cancer: guideline from the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and the American Society of Clinical Oncology. *J Clin Oncol.* 2017;35(13):1453-1486.

- National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: colon cancer-version 2.2017. 2017; https://www.nccn.org/professionals/physician_gls/pdf/colon.pdf.
- 69. Van Cutsem E, Cervantes A, Adam R, et al. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol.* 2016;27(8):1386-1422.
- 70. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: multiple myeloma-version 3.2017. 2017; https://www.nccn.org/professionals/physician_gls/pdf/myeloma.pdf.
- 71. Moreau P, San Miguel J, Sonneveld P, et al. Multiple myeloma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2017;28(suppl_4):iv52-iv61.
- 72. Centers for Medicare & Medicaid Services. Local Coverage Determination (LCD): MoIDX: EndoPredict® breast cancer gene expression test (L37311). 2018; https://www.cms.gov/medicare-coverage-database/details/lcddetails.aspx?LCDId=37311&ver=4&CoverageSelection=Both&ArticleType=All&PolicyType=F inal&s=All&KeyWord=endopredict&KeyWordLookUp=Title&KeyWordSearchType=And&art icleId=52974&bc=gAAAACAAAAA&.
- 73. Centers for Medicare & Medicaid Services. Local Coverage Determination (LCD): MolDX: breast cancer assay: Prosigna (L36386). 2016; https://www.cms.gov/medicare-coveragedatabase/details/lcd-

details.aspx?LCDId=36386&ver=5&CoverageSelection=Both&ArticleType=All&PolicyType=F inal&s=All&KeyWord=prosigna&KeyWordLookUp=Title&KeyWordSearchType=And&articleI d=52974&bc=gAAAACAAAAAAAA%3d%3d&.

74. Centers for Medicare & Medicaid Services. Local Coverage Determination (LCD): MoIDX: Breast Cancer Index[™] genetic assay (L36316). 2016; https://www.cms.gov/medicarecoverage-database/details/lcd-

details.aspx?LCDId=36316&ver=6&CoverageSelection=Both&ArticleType=All&PolicyType=F inal&s=All&KeyWord=breast+cancer+index&KeyWordLookUp=Title&KeyWordSearchType= And&articleId=52974&bc=gAAAACAAAAAAAA%3d%3d&.

- 75. Aetna. Tumor markers. 2017; http://www.aetna.com/cpb/medical/data/300_399/0352.html.
- 76. Cigna. Tumor profiling, gene expression assays and molecular diagnostic testing for hematology/oncology indications. 2017; https://cignaforhcp.cigna.com/public/content/pdf/coveragePolicies/medical/mm_0520_cove ragepositioncriteria_tumor_profiling.pdf.
- 77. Regence. Assays of genetic expression in tumor tissue as a technique to determine prognosis in patients with breast cancer. 2017; http://blue.regence.com/trgmedpol/geneticTesting/gt42.pdf.
- 78. Centers for Medicare & Medicaid Services. Local Coverage Determination (LCD): MoIDX-CDD: Decipher® prostate cancer classifier assay (L36345). 2016; https://www.cms.gov/medicare-coverage-database/details/lcddetails.aspx?LCDId=36345&ver=13&CoverageSelection=Both&ArticleType=All&PolicyType=

Final&s=All&KeyWord=decipher&KeyWordLookUp=Title&KeyWordSearchType=And&articled=52974&bc=gAAAACAAAAAAA%3d%3d&.

- 79. Centers for Medicare & Medicaid Services. Local Coverage Determination (LCD): MoIDX-CDD: Prolaris[™] prostate cancer genomic assay (L36350). 2016; https://www.cms.gov/medicare-coverage-database/details/lcddetails.aspx?LCDId=36350&ver=8&CoverageSelection=Both&ArticleType=All&PolicyType=F inal&s=All&KeyWord=prolaris&KeyWordLookUp=Title&KeyWordSearchType=And&articleI d=52974&bc=gAAACAAAAAAA%3d%3d&.
- 80. Centers for Medicare & Medicaid Services. Local Coverage Determination (LCD): MoIDX: Prolaris[™] prostate cancer genomic assay for men with favorable intermediate risk disease (L37082). 2017; https://www.cms.gov/medicare-coverage-database/details/lcddetails.aspx?LCDId=37082&ver=4&CoverageSelection=Both&ArticleType=All&PolicyType=F inal&s=All&KeyWord=prolaris&KeyWordLookUp=Title&KeyWordSearchType=And&articleI d=52974&bc=gAAACAAAAAA%3d%3d&.
- 81. Centers for Medicare & Medicaid Services. Local Coverage Determination (LCD): MolDX-CDD: Genomic Health[™] Oncotype DX[®] prostate cancer assay (L36368). 2016; https://www.cms.gov/medicare-coverage-database/details/lcddetails.aspx?LCDId=36368&ver=3&CoverageSelection=Both&ArticleType=All&PolicyType=F inal&s=All&KeyWord=oncotype+prostate&KeyWordLookUp=Title&KeyWordSearchType=A nd&articleId=52974&bc=gAAAACAAAAAAA%3d%3d&.
- 82. Regence. Gene expression analysis for prostate cancer management. 2017; http://blue.regence.com/trgmedpol/geneticTesting/gt71.pdf.
- 83. Regence. Multigene expression assay for predicting recurrence in colon cancer. 2017; http://blue.regence.com/trgmedpol/geneticTesting/gt22.pdf.
- 84. Regence. Microarray-based gene expression profile testing for multiple myeloma risk stratification. 2017; http://blue.regence.com/trgmedpol/geneticTesting/gt70.pdf.
- 85. Genomic Health. About the Oncotype DX breast recurrence score test. 2017; http://www.oncotypeiq.com/en-US/breast-cancer/healthcare-professionals/oncotype-dxbreast-recurrence-score/about-the-test.
- 86. Genomic Health. Oncotype DX breast recurrence score. 2016; http://www.oncotypeiq.com/-/media/ODX-Breast/Files/Portal_IBC_Node_Negative_Sample_Report.pdf.
- 87. Agendia. MammaPrint® 70-gene breast cancer recurrence assay. 2017; http://www.agendia.com/healthcare-professionals/breast-cancer/mammaprint/.
- 88. NanoString Technologies. NanoString Technologies receives FDA 510(k) clearance for Prosigna[™] breast cancer prognostic gene signature assay. 2013; http://investors.nanostring.com/releasedetail.cfm?releaseid=789662.
- 89. Myriad Genetic Laboratories. Myriad EndoPredict® technical specifications. 2017; https://s3.amazonaws.com/myriad-web/Managed+Care/EndoPredict-TechnicalSpecifications.pdf.

- 90. Premera Blue Cross. Assays of genetic expression in tumor tissue as a technique to determine prognosis in patients with breast cancer. 2017; https://www.premera.com/medicalpolicies/12.04.36.pdf.
- 91. NanoString Technologies. US product data sheet: The Prosigna[™] breast cancer prognostic gene signature. n.d.; http://prosigna.com/docs/Prosigna_Product_Data_Sheet_US.pdf.
- 92. NanoString Technologies. Use PAM50 to generate an individualized Prosigna score for every patient. 2017; https://www.nanostring.com/diagnostics/prosigna/technology/prosigna-algorithm.
- 93. Biotheranostics. Breast Cancer Index (BCI). 2017; https://www.breastcancerindex.com/.
- 94. Biotheranostics. Breast Cancer Index test report information. 2017; https://www.breastcancerindex.com/reporting.
- 95. 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.* 2013;19(15):4196-4205.
- 96. Sinn P, Aulmann S, Wirtz R, et al. Multigene aAssays for classification, prognosis, and prediction in breast cancer: a critical review on the background and clinical utility. *Geburtshilfe Frauenheilkd*. 2013;73(9):932-940.
- 97. Mislick K, Schonfeld W, Bodnar C, Tong KB. Cost-effectiveness analysis of Mammostrat compared with Oncotype DX to inform the treatment of breast cancer. *ClinicoEcon*. 2014;6:37-47.
- 98. Medica. Gene expression profiling assays for breast cancer 2015; http://medicafoundation.org/-/media/documents/provider/coveragepolicies/gene_expression_profiling_breast_cp.pdf?la=en&hash=6FCE20021ED7F026040583B 2D41AE2AF930F7861.
- 99. GenomeDx. What is the Decipher® prostate cancer test? 2017; https://genomedx.com/decipher-test/decipher-prostate-cancer-test/.
- 100. Na R, Wu Y, Ding Q, Xu J. Clinically available RNA profiling tests of prostate tumors: utility and comparison. *Asian J Androl.* 2016;18(4):575-579.
- 101. GenomeDx. Introducing Decipher. 2017; http://deciphertest.com/physicians/.
- 102. Lee HJ, Yousefi K, Haddad Z, et al. Evaluation of a genomic classifier in radical prostatectomy patients with lymph node metastasis. *Res Rep Urol.* 2016;8:77-84.
- 103. Spratt DE, Yousefi K, Deheshi S, et al. Individual patient-level meta-analysis of the performance of the Decipher genomic classifier in high-risk men after prostatectomy to predict development of metastatic disease. *J Clin Oncol.* 2017;35(18):1991-1998.
- 104. Myriad Genetic Laboratories. What is Prolaris®? 2016; https://prolaris.com/prolaris-for-physicians/what-is-prolaris/.
- 105. Myriad Genetic Laboratories. Prolaris® sample requirements. 2017; https://prolaris.com/prolaris-for-physicians/for-pathologists/sample-requirements/.

- 106. Colicchia M, Morlacco A, Cheville JC, Karnes RJ. Genomic tests to guide prostate cancer management following diagnosis. *Expert Rev Mol Diagn*. 2017;17(4):367-377.
- 107. Sommariva S, Tarricone R, Lazzeri M, Ricciardi W, Montorsi F. Prognostic value of the cell cycle progression score in patients with prostate cancer: a systematic review and metaanalysis. *Eur Urol.* 2016;69(1):107-115.
- 108. Genomic Health. About the Oncotype DX genomic prostate score assay. 2017; http://www.oncotypeiq.com/en-US/prostate-cancer/healthcare-professionals/oncotype-dx-genomic-prostate-score/about-the-test.
- 109. Brand TC, Zhang N, Crager MR, et al. Patient-specific meta-analysis of 2 clinical validation studies to predict pathologic outcomes in prostate cancer using the 17-gene genomic prostate score. *Urology*. 2016;89:69-75.
- 110. Agendia. Agendia announces launch of ColoPrint for colon cancer prognosis and prediction. 2012; http://www.agendia.com/agendia-announces-launch-of-coloprint-for-colon-cancer-prognosis-and-prediction/.
- 111. Kopetz S, Tabernero J, Rosenberg R, et al. Genomic classifier ColoPrint predicts recurrence in stage II colorectal cancer patients more accurately than clinical factors. *Oncologist.* 2015;20(2):127-133.
- 112. BlueCross BlueShield of North Carolina. Multigene expression assay for predicting recurrence in colon cancer. 2017; https://www.bcbsnc.com/assets/services/public/pdfs/medicalpolicy/multigene_expression_as say_for_predicting_recurrence_in_colon_cancer.pdf.
- 113. Genomic Health. About the Oncotype DX colon recurrence score test. 2017; http://www.oncotypeiq.com/en-US/colon-cancer/healthcare-professionals/oncotype-dxcolon-recurrence-score/about-the-test.
- 114. Genomic Health. Oncotype DX colon recurrence score: FAQs. n.d.; http://www.genomichealth.com/sitecore/content/ODX-Colon/Home/Professional/Resources/FAQs?sc_lang=en-US#faq5.
- 115. Yamanaka T, Oki E, Yamazaki K, et al. 12-gene recurrence score assay stratifies the recurrence risk in stage II/III colon cancer with surgery alone: the SUNRISE study. *J Clin Oncol.* 2016;34(24):2906-2913.
- 116. Johnson SK, Heuck CJ, Albino AP, et al. The use of molecular-based risk stratification and pharmacogenomics for outcome prediction and personalized therapeutic management of multiple myeloma. *Int J Hematol.* 2011;94(4):321-333.
- 117. van Laar R, Flinchum R, Brown N, et al. Translating a gene expression signature for multiple myeloma prognosis into a robust high-throughput assay for clinical use. *BMC Med Genomics*. 2014;7:25.
- 118. Premera Blue Cross. Microarray-based gene expression profile testing for multiple myeloma risk stratification. 2017; https://www.premera.com/medicalpolicies/12.04.97.pdf.

- 119. Shaughnessy JD, Jr., Zhan F, Burington BE, et al. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*. 2007;109(6):2276-2284.
- 120. Kuiper R, van Duin M, van Vliet MH, et al. Prediction of high- and low-risk multiple myeloma based on gene expression and the International Staging System. *Blood.* 2015;126(17):1996-2004.
- 121. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in metaanalyses. *BMJ*. 2003;327(7414):557-560.
- 122. Gligorov J, Pivot XB, Jacot W, et al. Prospective clinical utility study of the use of the 21gene assay in adjuvant clinical decision making in women with estrogen receptor-positive early invasive breast cancer: results from the SWITCH study. *Oncologist*. 2015;20(8):873-879.
- 123. Li Y, Kurian AW, Bondarenko I, et al. The influence of 21-gene recurrence score assay on chemotherapy use in a population-based sample of breast cancer patients. *Breast Cancer Res Treat*. 2017;161(3):587-595.
- 124. U.S. National Libary of Medicine. Genetic testing or clinical assessment in determining the need for chemotherapy in women with breast cancer that involves no more than 3 lymph nodes (MINDACT). 2017; https://clinicaltrials.gov/ct2/show/NCT00433589?term=70gene&cond=breast+cancer&rank=2.
- 125. Gray RG, Quirke P, Handley K, et al. Validation study of a quantitative multigene reverse transcriptase-polymerase chain reaction assay for assessment of recurrence risk in patients with stage II colon cancer. *J Clin Oncol.* 2011;29(35):4611-4619.
- 126. Jahn B, Rochau U, Kurzthaler C, et al. Cost effectiveness of personalized treatment in women with early breast cancer: the application of OncotypeDX and Adjuvant! Online to guide adjuvant chemotherapy in Austria. *Springerplus*. 2015;4:752.
- 127. O'Connor AM. User manual Decisional Conflict Scale. 2010; https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_decisional_conflict.pdf.

Appendix A. Search Strategy

Breast Cancer

Databases:

- Ovid MEDLINE <1946 to November Week 4 2017>
- Ovid MEDLINE In-Process & Other Non-Indexed Citations < December 01, 2017>
- EBM Reviews—Cochrane Central Register of Controlled Trials <November 2017>
- EBM Reviews—Cochrane Database of Systematic Reviews <2005 to November 29, 2017>
- 1 exp Transcriptome/
- 2 exp Gene Expression Profiling/
- 3 exp Gene Expression/
- 4 exp Biomarkers, tumor/
- 5 3 and 4
- 6 1 or 2 or 5
- 7 exp Breast Neoplasms/
- 8 6 and 7

9 (((gene* adj2 express* adj5 (profil* or identif* or test or tests or testing or tested or assay* or array* or microarray*)) or transcriptom*) adj7 ((breast* or mammar*) adj5 (cancer or neoplas* or tumor* or tumour* or carcinom* or adenocarcinom* or malig* or metasta*))).mp.

10 (Oncotype or 21-gene or EndoPredict or 12-gene or MammaPrint or 70-gene or Mammostrat or Breast Cancer Index or Prosigna or PAM50 or 50-gene).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sy, sh, kw, tx, ct]

11 7 and 10

12 ((breast* or mammar*) adj5 (cancer or neoplas* or tumor* or tumour* or carcinom* or adenocarcinom* or malig* or metasta*)).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sy, sh, kw, tx, ct]

- 13 10 and 12
- 14 11 or 13
- 15 8 or 9 or 14
- 16 limit 15 to humans [Limit not valid in CCTR,CDSR; records were retained]
- 17 exp Decision Making/
- 18 exp Clinical Decision-Making/
- 19 exp Prognosis/
- 20 exp "Outcome and Process Assessment (Health Care)"/
- 21 exp Mortality/
- 22 mo.fs.

- 23 exp survival analysis/
- 24 exp Professional Competence/
- 25 exp Practice Patterns, Physicians'/
- 26 exp Death/
- 27 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26
- 28 exp Economics/
- 29 ec.fs.
- 30 exp Socioeconomic Factors/
- 31 28 or 29 or 30
- 32 16 and 27
- 33 16 and 31
- 34 32 or 33

35 limit 34 to (comparative study or controlled clinical trial or meta analysis or randomized controlled trial or systematic reviews) [Limit not valid in CCTR,CDSR; records were retained]

- 36 exp epidemiologic studies/
- 37 34 and 36
- 38 limit 34 to guideline [Limit not valid in CDSR; records were retained]

39 limit 34 to (evaluation studies or validation studies) [Limit not valid in CCTR,CDSR; records were retained]

- 40 35 or 37 or 38 or 39
- 41 14 or 40
- 42 remove duplicates from 41

Prostate Cancer

Databases:

- Ovid MEDLINE <1946 to November Week 4 2017>
- Ovid MEDLINE In-Process & Other Non-Indexed Citations <November 30, 2017>
- EBM Reviews—Cochrane Central Register of Controlled Trials <November 2017>
- EBM Reviews—Cochrane Database of Systematic Reviews <2005 to November 29, 2017>
- 1 exp Transcriptome/
- 2 exp Gene Expression Profiling/
- 3 exp Gene Expression/
- 4 exp Biomarkers, tumor/
- 5 3 and 4
- 6 1 or 2 or 5

7 exp prostatic neoplasms/

8 6 and 7

9 (((gene* adj2 express* adj5 (profil* or identif* or test or tests or testing or tested or assay* or array* or microarray*)) or transcriptom*) adj7 (prostat* adj5 (cancer or neoplas* or tumor* or tumour* or carcinom* or adenocarcinom* or malig* or metasta*))).mp.

10 (Oncotype or 17-gene or Prolaris or 46-gene or Decipher or 22-gene).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sy, sh, kw, tx, ct]

11 7 and 10

12 (prostat* adj5 (cancer or neoplas* or tumor* or tumour* or carcinom* or adenocarcinom* or malig* or metasta*)).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sy, sh, kw, tx, ct]

- 13 10 and 12
- 14 11 or 13
- 15 8 or 9 or 14
- 16 limit 15 to humans [Limit not valid in CCTR,CDSR; records were retained] (2508)
- 17 exp Decision Making/
- 18 exp Clinical Decision-Making/
- 19 exp Prognosis/
- 20 exp "Outcome and Process Assessment (Health Care)"/
- 21 exp Mortality/
- 22 mo.fs.
- 23 exp survival analysis/
- 24 exp Professional Competence/
- 25 exp Practice Patterns, Physicians'/
- 26 exp Death/
- 27 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 (2697934)
- 28 exp Economics/
- 29 ec.fs.
- 30 exp Socioeconomic Factors/
- 31 28 or 29 or 30
- 32 16 and 27
- 33 16 and 31
- 34 32 or 33

35 limit 34 to (comparative study or controlled clinical trial or meta analysis or randomized controlled trial or systematic reviews) [Limit not valid in CCTR,CDSR; records were retained]

- 36 exp epidemiologic studies/
- 37 34 and 36
- 38 limit 34 to guideline [Limit not valid in CDSR; records were retained]

39 limit 34 to (evaluation studies or validation studies) [Limit not valid in CCTR,CDSR; records were retained]

- 40 35 or 37 or 38 or 39
- 41 14 or 40
- 42 remove duplicates from 41

Colon Cancer

Databases:

- Ovid MEDLINE <1946 to November Week 4 2017>
- Ovid MEDLINE In-Process & Other Non-Indexed Citations < December 01, 2017>
- EBM Reviews—Cochrane Central Register of Controlled Trials <November 2017>
- EBM Reviews—Cochrane Database of Systematic Reviews <2005 to November 29, 2017>
- 1 exp Transcriptome/
- 2 exp Gene Expression Profiling/
- 3 exp Gene Expression/
- 4 exp Biomarkers, tumor/
- 5 3 and 4
- 6 1 or 2 or 5
- 7 exp Colorectal Neoplasms/
- 8 6 and 7

9 (((gene* adj2 express* adj5 (profil* or identif* or test or tests or testing or tested or assay* or array* or microarray*)) or transcriptom*) adj7 ((colon* or colorect* or bowel* or large intestin*) adj5 (cancer or neoplas* or tumor* or tumour* or carcinom* or adenocarcinom* or malig* or metasta*))).mp.

10 (Oncotype or 12-gene or GEP12 or ColoPrint or 18-gene or GEP18).mp.

11 7 and 10

12 ((colon* or colorect* or bowel* or large intestin*) adj5 (cancer or neoplas* or tumor* or tumour* or carcinom* or adenocarcinom* or malig* or metasta*)).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sy, sh, kw, tx, ct]

- 13 10 and 12
- 14 11 or 13
- 15 8 or 9 or 14

- 16 limit 15 to humans [Limit not valid in CCTR,CDSR; records were retained]
- 17 exp Decision Making/
- 18 exp Clinical Decision-Making/
- 19 exp Prognosis/
- 20 exp "Outcome and Process Assessment (Health Care)"/
- 21 exp Mortality/
- 22 mo.fs.
- 23 exp survival analysis/
- 24 exp Professional Competence/
- 25 exp Practice Patterns, Physicians'/
- 26 exp Death/
- 27 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26
- 28 exp Economics/
- 29 ec.fs.
- 30 exp Socioeconomic Factors/
- 31 28 or 29 or 30
- 32 16 and 27
- 33 16 and 31
- 34 32 or 33

35 limit 34 to (comparative study or controlled clinical trial or meta analysis or randomized controlled trial or systematic reviews) [Limit not valid in CCTR,CDSR; records were retained]

- 36 exp epidemiologic studies/
- 37 34 and 36
- 38 limit 34 to guideline [Limit not valid in CDSR; records were retained]
- 39 limit 34 to (evaluation studies or validation studies) [Limit not valid in CCTR,CDSR; records were retained]
- 40 35 or 37 or 38 or 39
- 41 14 or 40
- 42 remove duplicates from 41

Multiple Myeloma

Databases:

- Ovid MEDLINE <1946 to November Week 4 2017>
- Ovid MEDLINE In-Process & Other Non-Indexed Citations < December 01, 2017>
- EBM Reviews—Cochrane Central Register of Controlled Trials <November 2017>

- EBM Reviews—Cochrane Database of Systematic Reviews <2005 to November 29, 2017>
- 1 exp Transcriptome/
- 2 exp Gene Expression Profiling/
- 3 exp Gene Expression/
- 4 exp Biomarkers, tumor/
- 5 3 and 4
- 6 1 or 2 or 5
- 7 exp Multiple Myeloma/
- 8 6 and 7

9 (((gene* adj2 express* adj5 (profil* or identif* or test or tests or testing or tested or assay* or array* or microarray*)) or transcriptom*) adj7 myeloma*).mp. [mp=ti, ab, ot, nm, hw, kf, px, rx, ui, sy, sh, kw, tx, ct]

- 10 8 or 9
- 11 Myeloma Prognostic Risk Signature.mp.
- 12 MyPRS.mp.
- 13 70-gene.mp.
- 14 gep70.mp.
- 15 SKY92.mp.
- 16 EMC92.mp.
- 17 92-gene.mp.
- 18 11 or 12 or 13 or 14 or 15 or 16 or 17
- 19 7 and 18
- 20 myeloma*.mp.
- 21 18 and 20
- 22 19 or 21
- 23 10 or 22
- 24 exp Decision Making/
- 25 exp Clinical Decision-Making/
- 26 exp Prognosis/
- 27 exp "Outcome and Process Assessment (Health Care)"/
- 28 exp Mortality/
- 29 mo.fs.
- 30 exp survival analysis/

- 31 exp Professional Competence/
- 32 exp Practice Patterns, Physicians'/
- 33 exp Death/
- 34 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33
- 35 exp Economics/
- 36 ec.fs.
- 37 exp Socioeconomic Factors/
- 38 35 or 36 or 37
- 39 23 and 34
- 40 23 and 38
- 41 39 or 40
- 42 19 or 41
- 43 remove duplicates from 42

Appendix B. Additional Methods

Risk of Bias Assessment: Systematic Reviews

Domain	Domain Elements The elements included in each domain are assessed and rated as <i>Yes</i> , <i>No</i> , <i>Unclear</i> , or <i>Not Applicable</i> based on performance and documentation of the individual elements in each domain. The overall risk of bias for the study is assessed as <i>High</i> , <i>Moderate</i> , or <i>Low</i> based on assessment of how well the overall study methods and processes were performed to limit bias and ensure validity.						
Protocol Described & Used	 An <i>a priori</i> design or protocol is described or available Adherence to the protocol is documented, or deviations are discussed and appropriate 						
Search Strategy	• The literature search is sufficiently rigorous to identify all relevant studies (i.e., at least two electronic databases should be searched, and the search strategy is provided and determined to be appropriate)						
Study Selection	 The criteria used to select articles for inclusion is clear and appropriate A PRISMA diagram or equivalent description of study selection is included in the review A list of excluded studies is provided 						
Study Characteristics	• A list of included studies and their characteristics are provided (e.g., study design, setting, study participants, interventions, risk factors, or exposures)						
Included Study Quality Assessment	 Process for assessment of methodological quality of included studies is described, including tools used Individual study methodological quality is assessed appropriately 						
Data Extraction	 Methods for data extraction from included studies and related documents (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators are described 						
Data Synthesis	 There are enough similarities among the included studies (e.g., sample characteristics, intervention, comparator) to make combining them reasonable (in a narrative synthesis and/or meta-analysis) A description of how individual study risk of bias is used in data synthesis is provided Overall strength or quality of evidence for key outcomes of interest is provided 						
Statistical Methods (as applicable)	• Any assumptions made and analysis methods (e.g., categorizing exposures or outcomes, calculating standard deviations, or other statistics) are reasonable						
Domain	Domain Elements						
----------------------	--	--	--	--	--	--	--
	The elements included in each domain are assessed and rated as <i>Yes</i> , <i>No</i> , <i>Unclear</i> , or <i>Not Applicable</i> based on performance and documentation of the individual elements in each domain. The overall risk of bias for the study is assessed as <i>High</i> , <i>Moderate</i> , or <i>Low</i> based on assessment of how well the overall study methods and processes were performed to limit bias and ensure validity.						
	 The most appropriate summary estimate (e.g., odds ratio, risk ratio) is used Methods for combining results of studies are described, including measures of consistency (e.g., l² statistic) and model selection (e.g., random effects, fixed effect) for each meta-analysis Methods for additional analyses (e.g., sensitivity or subgroup analyses, meta-regression) are described, including which were pre-specified or determined to be appropriate If performed, network meta-analysis was conducted appropriately (comparison to common therapy or placebo; confounders/effect modifiers taken into consideration; network of evidence, based on both direct and indirect comparisons in underlying RCTs, considered in overall assessment; transitivity assumption met) 						
Other Biases	List others in table footnote and describe, such as: • Publication bias in table						
Interest Disclosures	 Disclosures of interest are provided for authors/funders/commissioners of the study Interests are unlikely to significantly affect study validity 						
Funding Source	There is a description of source(s) of fundingFunding source is unlikely to significantly affect study validity						

Domain	Domain Elements
	The elements included in each domain are assessed and rated as <i>Yes</i> , <i>No</i> , <i>Unclear</i> , or <i>Not Applicable</i> based on performance and documentation of the individual elements in each domain. The overall risk of bias for the study is Assessed as <i>High</i> , <i>Moderate</i> , or <i>Low</i> based on assessment of how well the overall study methods and processes were performed to limit bias and ensure validity.
Randomization	 An appropriate method of randomization is used to allocate participants or clusters to groups, such as a computer random number generator Baseline characteristics between groups or clusters are similar
Allocation Concealment	 An adequate concealment method is used to prevent investigators and participants from influencing enrollment or intervention allocation
Intervention	 Intervention and comparator intervention applied equally to groups Co-interventions appropriate and applied equally to groups Control selected is an appropriate intervention
Outcomes	 Outcomes are measured using valid and reliable measures Investigators use single outcome measures and do not rely on composite outcomes, or the outcome of interest can be calculated from the composite outcome The trial has an appropriate length of follow-up and groups are assessed at the same time points Outcome reporting of entire group or subgroups is not selective
Masking (Blinding) of Investigators and Participants	 Investigators and participants are unaware (masked or blinded) of intervention status
Masking (Blinding) of Outcome Assessors	• Outcome assessors are unaware (masked or blinded) of intervention status
Intention to Treat Analysis	 Participants are analyzed based on random assignment (intention-to-treat analysis)
Statistical Analysis	 Participants lost to follow-up unlikely to significantly bias the results (i.e., complete follow-up of ≥ 80% of the participants overall and non-differential, ≤ 10% difference between groups) The most appropriate summary estimate (e.g., risk ratio, hazard ratio) is used Paired or conditional analysis used for crossover RCT Clustering appropriately accounted for in a cluster-randomized trial (e.g., use of an intraclass correlation coefficient)
Other Biases (as appropriate)	List others in table footnote and describe, such as: Sample size adequacy Interim analysis or early stopping Recruitment bias, including run-in period used inappropriately

Risk of Bias Assessment: Randomized Controlled Trials

Domain	Domain Elements The elements included in each domain are assessed and rated as <i>Yes</i> , <i>No</i> , <i>Unclear</i> , or <i>Not Applicable</i> based on performance and documentation of the individual elements in each domain. The overall risk of bias for the study is Assessed as <i>High</i> , <i>Moderate</i> , or <i>Low</i> based on assessment of how well the overall study methods and processes were performed to limit bias and ensure validity.				
	Use of unsuitable crossover intervention in a crossover RCT				
Interest Disclosure	 Disclosures of interest are provided for authors/funders/commissioners of the study Interests are unlikely to significantly affect study validity 				
Funding	 There is a description of source(s) of funding Funding source is unlikely to have a significant impact on study validity 				

Domain	Domain Elements
	The elements included in each domain are assessed and rated as <i>Yes</i> , <i>No</i> , <i>Unclear</i> , or <i>Not Applicable</i> based on performance and documentation of the individual elements in each domain. The overall risk of bias for the study is assessed as <i>High</i> , <i>Moderate</i> , or <i>Low</i> based on assessment of how well the overall study methods and processes were performed to limit bias and ensure validity.
Subject Selection	For cohort studies:
	 The two groups being studied are selected from source populations that are comparable in all respects other than the factor under investigation, or statistical adjustment is used appropriately to achieve this The study indicates how many of the people asked to take part did so, in each of the groups being studied The likelihood that some eligible subjects might have the outcome at the time of enrollment is assessed and taken into account in the analysis Fewer than 20% of individuals or clusters in each arm of the study dropped out before the study was completed
	For case-control studies:
	 Cases and controls are clearly specified and defined, with the inclusion and exclusion criteria applied appropriately Cases may be selected by meeting inclusion criteria; controls may be selected by meeting inclusion criteria and then being matched to cases Sampling selection (ratio of cases to control) is justified Cases and controls selected from the same population and same timeframe. When not all cases and controls are selected from the same population, they are randomly selected Among cases, investigators confirm that the exposure occurred before the development of the disease being studied and/or the likelihood that some eligible subjects might have the outcome at the time of enrollment is assessed and taken into account in the analysis
Intervention	 The assessment of exposure to the intervention is reliable Exposure level or prognostic factors are assessed at multiple times across the length of the study, if appropriate For case-control studies, assessors of (intervention) exposure status are unaware (masked or blinded) to the case or control status of participants There is a method to limit the effects of recall bias on the assessment of exposure to the intervention
Control	Control condition represents an appropriate comparator
Outcome	 There is a precise definition of the outcomes used Outcomes are measured using valid and reliable measures; evidence from other sources is used to demonstrate that the method of outcome assessment is valid and reliable

Risk of Bias Assessment: Nonrandomized Studies

Domain	Domain Elements							
	The elements included in each domain are assessed and rated as <i>Yes</i> , <i>No</i> , <i>Unclear</i> , or <i>Not Applicable</i> based on performance and documentation of the individual elements in each domain. The overall risk of bias for the study is assessed as <i>High</i> , <i>Moderate</i> , or <i>Low</i> based on assessment of how well the overall study methods and processes were performed to limit bias and ensure validity.							
	 Investigators use single outcome measures and do not rely on composite outcomes, or the outcome of interest can be calculated from the composite outcome The study has an appropriate length of follow-up for the outcome reported and groups are assessed at the same time points Outcome reporting of entire group or subgroups is not selective When patient-reported outcomes are used, there is a method for validating the measure 							
Masked Outcome Assessment	 The assessment of outcome(s) is made blind to exposure status. Where outcome assessment blinding was not possible, there is recognition that knowledge of exposure status could have influenced the assessment of outcome For case-control study: assessors of exposure status are unaware (masked or blinded) of the case or control status of participant) 							
Confounding	• The main potential confounders are identified and taken into account in the design and analysis of the study							
Statistical Analysis	 Comparison is made between full participants and those who dropped out or were lost to follow up, by exposure status If the groups were not followed for an equal length of time, the analysis was adjusted for differences in the length of follow-up All major confounders are adjusted for using multiple variable logistic regression or other appropriate statistical methods Confidence intervals (or information with which to calculate them) are provided For case-control studies that use matching, conditional analysis is conducted or matching factors are adjusted for in the analysis 							
Other Biases (as appropriate)	List others in table footnote and describe, such as: • Sample size adequacy							
Interest Disclosure	 Disclosures of interest are provided for authors/funders/commissioners of the study Interests are unlikely to significantly affect study validity 							
Funding Source	There is a description of source(s) of fundingFunding source is unlikely to have a significant impact on study validity							

Domain	Domain Elements
	The elements included in each domain are assessed and rated as <i>Yes</i> , <i>No</i> , <i>Unclear</i> , or <i>Not Applicable</i> based on performance and documentation of the individual elements in each domain. The overall risk of bias for the study is assessed as <i>High</i> , <i>Moderate</i> , or <i>Low</i> based on assessment of how well the overall study methods and processes were performed to limit bias and ensure validity.
Target Population	 Target population and care setting described Describe and justify basis for any target population stratification, identify any a priori identifiable subgroups If no subgroup analyses were performed, justify why they were not required
Perspective	• State and justify the analytic perspective (e.g., societal, payer, etc.)
Time Horizon	• Describe and justify the time horizon(s) used in the analysis
Discount Rate	• State and justify the discount rate used for costs and outcomes
Comparators	 Describe and justify selected comparators Competing alternatives appropriate and clearly described
Modelling	 Model structure (e.g., scope, assumptions made) is described and justified Model diagram provided, if appropriate Model validation is described (may involve validation of different aspects such as structure, data, assumptions, and coding and different validation models such as comparison with other models) Data sources listed and assumptions for use justified Statistical analyses are described
Effectiveness	 Estimates of efficacy/effectiveness of interventions are described and justified The factors that are likely to have an impact on effectiveness (e.g., adherence, diagnostic accuracy, values, and preferences) are described and an explanation of how they were factored into the analysis is included The strength of evidence for the relationship between the intervention and outcomes, and any necessary links, is described
Outcomes	 All relevant outcomes are identified, measured, and valued appropriately (including harms/adverse events) for each intervention, and the justification for information/assumptions is given Any quality of life measures used in modelling are described and their use justified Any other outcomes that were considered, but rejected, are described with the rationale for rejection Ethical and equity-related outcomes are considered and included when appropriate

Risk of Bias Assessment: Economic Studies

Domain	Domain Elements The elements included in each domain are assessed and rated as <i>Yes</i> , <i>No</i> , <i>Unclear</i> , or <i>Not Applicable</i> based on performance and documentation of the individual elements in each domain. The overall risk of bias for the study is assessed as <i>High</i> , <i>Moderate</i> , or <i>Low</i> based on assessment of how well the overall study methods and processes were performed to limit bias and ensure validity.
Resource Use/Costs	 All resources used are identified, valued appropriately, and included in the analyses Methods for costing are reporting (e.g., patient level) Resource quantities and unit costs are both reported Methods for costing time (e.g., lost time, productivity losses) are appropriate and a justification is provided if time costs are not considered
Uncertainty	 Sources of uncertainty in the analyses are identified and justification for probability distributions used in probabilistic analyses are given For scenario analyses, the values and assumptions tested are provided and justified
Results	 All results are presented in a disaggregated fashion, by component, in addition to an aggregated manner All results are presented with undiscounted totals prior to discounting and aggregation Natural units are presented along with alternative units (e.g., QALYs) The components of the incremental cost-effectiveness ratio (ICER) are shown (e.g., mean costs of each intervention in numerator and mean outcomes of each intervention in denominator) Results of scenario analyses, including variability in factors such as practice patterns and costs, are reported and described in relation to the reference (base) case
Interest Disclosure	 Disclosures of interest are provided for authors/funders/commissioners of the study Interests are unlikely to significantly affect study validity
Funding Source	There is a description of source(s) of fundingFunding source is unlikely to have a significant impact on study validity

Domain Scope and Purpose	Domain ElementsAssessment indicates how well the guideline methodology and development process were performed to limit bias and ensure validity for elements in domain (each domain rated as Good, Fair, or Poor overall based on performance and documentation of elements)• Objectives specifically described
Rigor of Evidence Development	 Systematic literature search Study selection criteria clearly described Strengths and limitations of individual studies and overall strength of the body of evidence assessed
Rigor of Recommendations Development	 Methods for developing recommendations clearly described Explicit link between recommendations and supporting evidence and includes strengths/limitations Balance of benefits and harms (side effects, risks) considered External review conducted Updating procedure specified
Stakeholder Involvement	 Relevant professional groups are represented Views and preferences of target population sought Target users defined
Clarity and Presentation	 Recommendations specific, unambiguous Different management options clearly presented Key recommendations easily identifiable
Applicability and Implementation	 Provides advice and/or tools on how the recommendations can be put into practice Description of facilitators and barriers to application Potential resource implications considered Monitoring/audit/review criteria presented
Editorial Independence	 Views of funding body have not influenced the content of the guideline Competing interests of members have been recorded, monitored, and addressed appropriately

Risk of Bias Assessment: Clinical Practice Guidelines

Appendix C. Evidence Tables

Abbreviations Used in Evidence Tables

aOR: adjusted odds ratio	HR: hazard ratio	pN1mi: micro-metastases
AS: active surveillance	HT: hormone treatment	PR: progesterone receptor
BCI: Breast Cancer Index	IHC: immunohistochemistry	PSA: prostate-specific antigen
Cl: confidence interval	IMRT: intensity-modulated radiation therapy	QALY: quality-adjusted life years
CT: chemotherapy	ITT: intention to treat	RP: radical prostatectomy
DCS: Decisional Conflict Scale	LN: lymph node	ROR: risk of recurrence
DI: decision impact	NCDB: National Cancer Data Base	RR: risk ratio
ECOG: Eastern Cooperative Oncology Group	N0: no nodal involvement	RS: recurrence score
ER: estrogen receptor	N1: 1-3 nodes	RT: radiotherapy
FACT-G v.4: Functional Assessment of Cancer	N2: 4-9 nodes	SD: standard deviation
Therapy-General, version 4	N3: > 9 or IC	STAI: State-Trait Anxiety Inventory
FFPE: formalin-fixed paraffin-embedded	NR: not reported	STO-3: Stockholm Tamoxifen Trial
GEP: gene expression profile	OR: odds ratio	
HER2: human epidermal growth factor receptor 2	pN0 (i-/i+): no axillary lymph node involvement or isolated tumor cells	

excluded	enrollment (lower risk of bias) of subjects and in 8 it was	12.00% (95% Cl, 8.00 to 16.00%) I ² = 92.00%	For the Oncotype DX test, there is a high overlap of	
Outcomes:	selective (higher risk of bias)		included studies with the	
Impact of test on clinical decision making (proportion of patients		Pooled NC for studies with universal enrollment	Blok 2018 ²¹ and Scope 2017 ²² reviews	
whose treatment decision was altered after the test)		9.00% (95% Cl, 4.00 to 14.00%) I ² = 89.00	Research was funded by a	
Net chemotherapy change (difference in the number of patients who were assigned to		Pooled NC for studies with selective enrollment	grant from the test manufacturer	

Design

Outcomes Design:

analysis

Study Inclusion Criteria

Systematic review and meta-

data on decisions regarding

Studies that reported prospectively

chemotherapy treatment before

and after availability of results of

patients with early breast cancer

Studies of hypothetical decisions

and those with fewer than 80% of

LN-negative (LN- or N0) patients or

those not reporting LN status were

the 21-gene assay (Oncotype DX) in

Study inclusion criteria:

February 16, 2018

Risk of Bias

Comments

No specific protocol described or registered

Subgroup analysis of

bias from selective

enrollment and all

No list of excluded studies

studies with higher risk of

estimates for NC are highly

subject to heterogeneity,

making pooled estimates

suspect for their validity

Low

Breast Cancer

Study Citation

Augustovski

Oncotype DX

Tests

2015²⁰

Table 5. Evidence Table: Breast Cancer Systematic Reviews

Outcomes Reported

Global pooled DI:

 $I^2 = 59.50\%$

enrollment

 $I^2 = 0.00\%$

enrollment

 $I^2 = 74.09\%$

Global pooled NC

29.52% (95% Cl, 26.29 to 32.86%)

Pooled DI for studies with universal

28.97% (95% Cl, 26.65 to 31.34%)

Pooled DI for studies with selective

29.43% (95% Cl, 22.88 to 36.45%)

Included Studies

For Decision Impact (DI)

For Net Chemotherapy

Change (NC) outcome: 15

Included studies conducted in

U.S., Canada, U.K., Australia,

France, Germany, Spain,

Mexico, Israel, and Japan

7 studies had universal

outcome: 15 studies

Characteristics

studies

Characteristics:

Total n = 2,229

Study Citation Design			Included Studies			Outcomes Reported				Risk of Bias						
Tests	Study Inclusion Criteria	Characteristics								Comments						
	Outcomes															
receive chemotherapy before vs. after the test)					16.00 (95% CI, 7.00 to 24.00) I ² = 93.00%					No other author financial relationships with test manufacturer						
Blok 2018 ²¹	<i>Design</i> : Systematic review, with pooled	,		% change to CT and HT/no treatment, by test (pooled LN- and LN+ subjects):				ру	Moderate							
Oncotype DX MammaPrint	outcome measures for change in treatment					% change CT	% change HT/no treatment		No specific protocol described or registered							
Prosigna EndoPredict	<i>Study inclusion criteria:</i> Included 1 of 4 GEP tests available			0	MammaPrint (k = 4, n = 79		-17.0%	+32.2%		No list of excluded studies						
	in Europe, covering either test development, clinical validation, clinical utility (CU), or economic evaluation. For CU tests, decision impact (DI) studies had to include a representative cohort and report absolute change in chemotherapy	chemoth	otherapy, total k = 28		Oncotype DX (k = 22, n = 3		-14.6%	+51.1%		Evidence tables have minimal information						
		clinical utility (CU), or economic evaluation. For CU tests, decision	clinical utility (CU), or economic evaluation. For CU tests, decision	clinical utility (CU), or economic evaluation. For CU tests, decision	clinical utility (CU), or economic evaluation. For CU tests, decision	clinical utility (CU), or economic evaluation. For CU tests, decision	clinical utility (CU), or economic evaluation. For CU tests, decision	of studies (k)	enrolled LN- subjects	early breast cancer subjects	Prosigna/PAN (k = 1, n = 20		-12.9%	+37.3%		included
				with either LN- or LN+	EndoPredict (k = 1, n = 16	57)	-34% +53.2%			No quality assessment of included studies						
	use and shifts in treatment	MammaPrint		3	(statistical t	testing r	not perform	ned)	J							
	category. Large, retrospective,	4 Oncotype		2						For the Oncotype DX test,						
	population-based DI studies	22	14 8 Summary of cost/QALY ranges reported,		l, by	, there is a high overlap of included studies with the										
	reporting use of tests and shifts in therapy decisions also met inclusion				test and nodal status:			7	Augustovski 2015 ²⁰ and							
	criteria, but were not included in	1	1	0	MammaPrint vs. no testing LN- \$10,000-\$43,044		_	Scope 2017 ²² reviews; for								
	tables.	EndoPrec	lict	1						the MammaPrint test,						
	Outcomes:	1	0	1	€4,614-€134,000 LN+ \$13,723				there is an overlap of							
	% change to chemotherapy (CT)			Oncotype DX vs. no testing			-	studies with the Scope 2017 ²² review								
	% change to HT or no therapy (no treatment)	Total # economic evaluation studies by comparison evaluated and those reporting			LN-					No description of funding						

Study Citation Design		Included Studies				Outcomes	s Reported	Risk of Bias	
Tests	Study Inclusion Criteria	Characteristics						Comments	
	Outcomes								
	(% change measures were pooled,	outcomes by QALYs by nodal status			al	LN+	£29,502 \$1,914-\$49,059	No author financial	
	weighted by number of subjects, by gene test, but no formal meta- analysis was conducted and no	number LN- LN+ or of mixed studies				CAD\$464-CAD\$14,844 €11,236 £5,529	relationships with test manufacturers		
	confidence intervals were provided. Pooled % change outcomes were not calculated by nodal status	(k) MammaPrin (usual care, Adjuvant! C	, best prac	tice or		Oncotype [LN-	DX vs. Mammostrat \$453,600 (in favor of Mammostrat)		
	group)		4	1		Oncotype [DX vs. MammaPrint		
	Economic evaluation outcomes were reported in tabular form and	Oncotype DX vs. no testing (usual care, best practice, or Adjuvant! Online tool)			LN-	€1,457 (in favor of MammaPrint)			
	not pooled. Various studies	32	14	5					
	reported cost/QALY (in \$USD, CAD\$, €, or £ depending on study)	EndoPredict vs. no testing (best practice)							
	cost change (increased, decreased, savings), or whether the test strategy was dominant in the economic model. The only comparable unit is QALY.	1 0 1 Head-to-head comparisor (Oncotype DX vs. Mammo and MammaPrint vs. Oncotype DX) 4 2 0 (1 each)		arisons mmostrat					
Scope 2017 ²² Oncotype DX MammaPrint Mammostrat	Systematic review, with narrative DX synthesis rint		Studies reporting change in treatment recommendations: k = 34 Oncotype DX: k = 28 MammaPrint: k = 6			reported o	comes were not pooled or quantitatively due to author about heterogeneity among DX	High No specific protocol described or registered	
(eligible, but no studies identified)	Studies published in English including patients with early invasive breast cancer, tested with	English(review did not locate studievith earlyfor other eligible tests			es	"the use	of Oncotype DX leads to changes ent recommendations for between	Inclusion criteria unclear (and included studies, for example, those with	

Study Citation	Design	Included Studies	Outcomes Reported	Risk of Bias
Tests	Study Inclusion Criteria	Characteristics		Comments
	Outcomes			
	Oncotype DX, MammaPrint, IHC4, or Mammostrat compared to usual care, no test, or statistical prognosis scoring using usual clinical data only Outcomes: The extent to which test results are used in treatment decisions	[Note: Table 3 on pp. 38-41 indicates that some studies did not have a recorded pretest treatment recommendation and for others it was inferred according to existing guidelines or prediction models]	 21% and 74% of all patients who underwent Oncotype DX testing."^{22(p. 37)} "change from chemotherapy to no chemotherapy ranged from 6% to 51% of all patients tested. However, in one study more chemotherapy was used after the introduction of Oncotype DX."^{22(p. 37)} "It was not clear in a large number of the studies whether these figures represented actual changes in the treatments patients received."^{22(p. 37)} MammaPrint "the use of MammaPrint in addition to clinicopathological factors led to changes in treatment recommendations for between 18% and 40% of all patients tested, and between 2% and 32% of all patients would be recommended to change from chemotherapy to no chemotherapy."^{22(p. 37)} "Again, in several of these studies it is not clear if actual treatment changes occurred following introduction of the test."^{22(p. 42)} 	hypothetical decisions, that were excluded from MEDLINE search results by this WA HTA report's authors because they did not meet inclusion criteria) No list of excluded studies Evidence tables have minimal information included Limited narrative synthesis of data Article authors stated that quality of prognostic studies was performed, but neither the tables of those studies or ones including studies treatment change outcomes have description of study quality included, but some issues related to methodological quality of included studies are described in narrative text

Study Citation	Design	Included Studies	Outcomes Reported	Risk of Bias
Tests	Study Inclusion Criteria	Characteristics		Comments
	Outcomes			
				No description of funding source
				None of authors had financial relationships with test manufacturers

Study Citation	Design	Included Stue	dies	Outcomes Repo	rted		Risk of Bias
Tests	Study Inclusion Criteria	Characteristic	cs				Comments
	Outcomes						
Bear 2017 ¹⁸ Oncotype DX	Design: Prospective study of one group of patients (n = 59 after 64 enrolled and 5 excluded, with n = 33 randomized) Intervention: Random assignment of patients (n = 17)	10% tumor st immunohisto HER2-negativ cancers ≥ 2 c core needle b amenable to conserving su Patients had I or 1, and had imaging by U	chemistry), ve, invasive m, diagnosed by biopsy, not breast- urgery (BCS). ECOG status of 0 ipsilateral nodal .S. or MRI within	Clinical complete or partial response (cPR, cCR); pathologic complete response (pCR) in the breast and in the breast and nodes; successful breast conserving surgery (BCS). Assessed by as treated analysis (B: n = 18; C: n = 11) "After adjusting for effects from confounders, including age, race, menopausal status, and site, Group B patients had significantly lower clinical response rate, compared with Group C (Ordinal regression: p = .0369)" ¹⁸ (p. 921)		High Of 33 women randomized (17 to NHT and 16 to NCT), 5 (15%) refused NCT; of these, 2 agreed to NHT instead (crossover C to B) Analysis was by treatment received rather than ITT; no description of the dropouts.	
	with Oncotype DX RS	6 weeks of er	rollment	Response Results	•		
	of 11-25 to neoadjuvant hormone	0	negative: n = 24 CCR (%) B: 22.2 C: 36.4	Groups B and C were not similar at baseline by race,			
	therapy (NHT, Group B)	LN-positive: r		cPR (%)	B: 27.8 C: 36.4		tumor stage, nodal stage, or menopausal status; Group C
		Patient Chara	icteristics:	pCR breast (%)	B: 6.0		more likely to be younger,
	<i>Comparator:</i> Random assignment of patients (n = 16) to	Median age, yrs (range) White	B: 64 (41,80) C: 56.5 (30,75) B: 95/5	pCR breast + nodes (%)	C: 0 B: 0 C: 0	-	African American, T3, and premenopausal
	neoadjuvant chemotherapy (NCT,	/African American (%)	C: 79/21	Successful BCS B: 72.2 C: 63.6		Exact NHT or NCT regimens at provider discretion and	
	Group C) (other groups not randomized and assigned with RS < 11 to NHT (Group A) and	Tumor stage T1/T2/T3/T4 (%)	B: 0/79/16/5 C: 8/62/31/0				not accounted for in analysis No explanation for the non- standard cutoffs (11 and 25 rather than 18 and 30, which

Table 6. Evidence Table: Breast Cancer Randomized Controlled Trials

Study Citation	Design	Included Studies	Outcomes Reported	Risk of Bias
Tests	Study Inclusion Criteria	Characteristics		Comments
	Outcomes			
	with RS \geq 26 to NCT (Group D), these data not reported in table)	Pre-/Post- menopausal (%) B: 16/84 C: 39/61		are typical thresholds for low-risk and high-risk RS)
	Setting: Pilot feasibility study within a larger multicenter trial at 7 U.S. and Canadian centers	Nodal stage N0/N1/NX (%) B: 74/31/5 C: 69/31/0		Authors' conclusion that "patients with low RS results derive little or no benefit from adding chemotherapy" has no basis in the data, as there were no subjects with low RS who received chemotherapy Funding from Genomic
				Health & NIH; one author employed by Genomic Health; lead author received speaking and advisory board honoraria from Genomic Health
Cardoso 2016 ¹⁹ MammaPrint (70-gene signature, 70- GS and Adjuvant!	Design: Randomized, phase 3 study (only patients with discordant clinical and genomic risks were randomized)	Study Population: Patients (n = 6,693, ITT n = 2,187): Women 18 to 70 years of age with early stage, histologically confirmed primary invasive breast cancer (stage T1 or T2 or operable T3)	 Patients with high clinical risk and low genomic risk, no adjuvant chemotherapy: Rate of survival without distant metastasis at 5 years: 94.7% (95% CI, 92.5 to 96.2) <i>ITT population:</i> High clinical risk, low genomic risk, randomized to chemotherapy based on clinical risk: 	Moderate Change in RNA-extraction solution for MammaPrint test caused a temporary shift in the risk calculation from May 24, 2009 to January 30, 2010, which led to more patients
	Intervention: Chemotherapy based		5	

Study	Design	Included St	uded Studies Outcomes Reported		Risk of Bias
Citation		Characteristics			Comments
Tests	Study Inclusion Criteria				Comments
10505	Cinteria				
	Outcomes				
	Outcomeson genomic risk (determined by MammaPrint) and clinical risk 	LN-negative LN-positive: Physician na characteristi <i>Patient Char</i> Median age, y (range) % white % ER-, PR-, or both positive % HER2- positive Tumor grade (%) Grade 1	n = 752 and cs not reported	 5-year rate of survival without distant metastasis: 95.9% (95% Cl, 94.0 to 97.2) High clinical risk, low genomic risk, randomized to no chemotherapy based on genomic risk: 5-year rate of survival without distant metastasis: 94.4% (95% Cl, 92.3 to 95.9) Adjusted hazard ratio for distant metastasis or death (chemotherapy vs. no chemotherapy): 0.78 (95% Cl, 0.50 to 1.21) Low clinical risk, high genomic risk, randomized to chemotherapy based on genomic risk: 5-year rate of survival without distant metastasis: 95.8% (95% Cl, 92.9 to 97.6) 	Financial support from Novartis, F. Hoffmann-La Roche, Sanofi-Aventis, Eli Lilly, and Veridex 2 authors financially affiliated with test manufacturer
	European countries	Grade 2 Grade 3 Missing data Lymph node status (%) Negative 1 node 2 nodes 3 nodes \geq 4 nodes	49.1 28.8 0.5 79.0 14.1 4.5 2.3 0.1	Low clinical risk, high genomic risk, randomized to no chemotherapy based on clinical risk: 5-year rate of survival without distant metastasis: 95.0% (95% Cl, 91.8 to 97.0) Adjusted hazard ratio for distant metastasis or death (chemotherapy vs. no chemotherapy): 1.17 (95% Cl, 0.59 to 2.28) "Among all patients at high clinical risk, the use of the 70-gene signature to guide chemotherapy treatment	

Study Citation	Design	Included Studies	Outcomes Reported	Risk of Bias
Tests	Study Inclusion Criteria	Characteristics		Comments
	Outcomes			
			would lead to a reduction in the use of adjuvant chemotherapy in 1,550 of 3,356 patients (46.2%)." ^{19(p.}	
			5-year survival without distant metastasis (clinical-risk strategy alone vs. genomic-risk strategy alone): 95.0% vs. 94.7%	
			<i>Subgroup analyses:</i> LN-negative, 5-year survival without distant metastasis (chemotherapy vs. no chemotherapy): 95.7% (95% CI, 93.0 to 97.4) vs. 93.2% (95% CI, 90.1 to 95.4)	
			LN-positive, 5-year survival without distant metastasis (chemotherapy vs. no chemotherapy): 96.3% (95% Cl, 93.1 to 98.1) vs. 95.6% (95% Cl, 92.7 to 97.4) ER-positive, HER2-negative, LN-negative disease, 5- year survival without distant metastasis (chemotherapy vs. no chemotherapy): 95.5% (95% Cl, 92.5 to 97.3) vs. 93.9% (95% Cl, 90.6 to 96.1)	
			In multivariate analysis, adjusting for chemotherapy use, clinical risk, and patient and tumor characteristics, HR for distant metastasis or death with chemotherapy vs. no chemotherapy in patients with high vs. low genomic risk using MammaPrint test: HR 2.41 (95% CI, 1.79 to 3.26)	

Study Citation	Design	Study Population		Outcomes	Risk of Bias
Test	Intervention	Characteristics			Comments
	Comparator				
	Setting				
Evans 2016 ²³ Oncotype DX	Design: Before-after study, with single group of patients Intervention: Treatment decision based on Oncotype DX test Comparator: Treatment decision based on other factors without test Setting: Women recruited 2011-2015 from 4 sites in Washington, DC	Study Population:Females with recently diastage 1 or stage 2 breastand had Oncotype DX teordered by treating clinic193 who completed pre-post-interview before anreceiving test resultsPhysician n and characternot reportedPatient Characteristics:Mean age (y, SD)57.14 (the state of the s	cian; n = - and - and ad after	Changes in patient-reported outcomes: Perception of pros and cons of chemotherapy both increased (t = 4.27 and t = 8.54, respectively); perceived risk of recurrence decreased. Test did not affect these changes in any category. Distress did not change with test (i.e., patients were not reassured, but not more distressed either) <i>Predictors of chemotherapy use:</i> Recurrence score: High: 88% Intermediate: 57%, OR 0.04 (95% Cl, 0.01 to 0.27) (vs. high) Low: 5%, OR 0.01 (95% Cl, 0.00 to 0.02) (vs. high) Older age: OR 0.89 (0.83 to 0.95) Pretest distress: OR 0.47 (95% Cl, 0.21 to 1.07) Pretest perceived risk: OR 0.47 (95% Cl, 0.98 to 1.06) Pretest chemotherapy pros: OR 1.19 (95% Cl, 0.66 to 2.16)	 High Recruitment limited to clinics in the Eastern U.S. Some patients unreachable for pretest interview; these patients may have differed from the women who were able to be reached to assess for eligibility and consent to a pretest interview Study funded by grants from the American Cancer Society, the National Cancer Institute Authors declared no conflict of interest

Table 7. Evidence Table: Breast Cancer Observational Studies

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
			Pretest chemotherapy cons: OR 0.74 (95% Cl, 0.39 to 1.38)	
			Posttest distress: OR 2.19 (95% Cl, 1.05 to 4.57)	
			Posttest perceived risk: OR 1.01 (95% Cl, 0.98 to 1.03)	
			Posttest chemotherapy pros: OR 1.83 (95% Cl, 0.95 to 3.50)	
			Posttest chemotherapy cons: OR 0.50 (95% Cl, 0.26 to 0.97)	
Friese 2017 ²⁴ Li 2017 ¹²³ Oncotype DX	Design: Retrospective cohort study using registry with prospectively collected data Intervention: Treatment decision based on Oncotype DX test	Study Population: Women (n = 1,527, intervention n = 778) aged 20 to 79 years, diagnosed with early-stage (1/2) breast cancer (ER/PR-positive, HER2-negative), treated in 2013 to 2014 LN-negative: n = 1,214 LN-positive: n = 303 Missing: n = 10	Chemotherapy recommendations in overall study population: Oncologist recommended against chemotherapy: 47.2% Chemotherapy decision left to patient: 12.3% Oncologist recommended chemotherapy: 40.5% Oncotype DX Test Group: Most patients with high RS test results (scores 31-100) received recommendations for chemotherapy (86.9% to 96.5% across clinical	Moderate Results are limited to two geographic regions in the United States Measures of communication and decision making were determined through patient reports and did not necessarily
	Comparator:	Physician n and characteristics not reported	groups).	represent physician perspectives

Study Citation	udy Citation Design Study Population			Outcomes	Risk of Bias	
Test	Intervention	Characteristics			Comments	
	Comparator					
	Setting					
	Treatment decision based on other factors (no Oncotype DX test)	Patient Characteri Mean age (y) at diagnosis	<i>stics:</i> 61.0	65.9% to 78.2% of women with LN-negative disease (majority with low-risk scores of 0-18) received a recommendation against chemotherapy.	Nonresponse bias and missing data are problematic, but were partially mitigated by the analyses	
	Setting: Los Angeles County and Georgia Surveillance, Epidemiology, and End Results registries	% white Clinical group: b LN-negative 1 LN-negative 2 LN-positive Missing Comorbidities (%) No diagnosis 1 condition 2+ conditions Missing Site (%) Georgia Los Angeles County	56.9 60.1 19.4 19.8 0.7 72.2 21.5 5.8 0.6 54.9 45.1	 LN-negative, favorable-risk disease, recommendation against chemotherapy: 78.2% LN-negative, favorable-risk disease, recommendation for chemotherapy: 11.7% <i>No Test Group:</i> LN-negative, more favorable-risk disease, recommendation for chemotherapy: 23.1% LN-negative, less favorable-risk disease, recommendation for chemotherapy: 60.2% LN-positive, recommendation for chemotherapy 83.2% <i>Chemotherapy Receipt:</i> High test score and receipt of chemotherapy: LN-negative, less favorable: 91.1% LN-positive: 100% Low test score and receipt of chemotherapy: LN-negative, more favorable: 2.9% LN-negative, less favorable: 9.5% 	Study funded through grants from the NIH, California Department of Health, and the CDC One author reported research funding from multiple sources including the test manufacturer	

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
			LN-positive: 26.6%	
			Low-risk test vs. no test: OR 0.1 (95% Cl, 0.1 to 0.2) Medium-risk test vs. no test: or 1.4 (95% Cl, 1.1 to 1.7) High-risk test vs. no test: OR 2.8 (95% Cl, 2.0 to 4.0) <i>Patient Perceptions:</i> 63.9% of patients who received test found it "very" or "extremely" helpful 65.0% of women with low-risk results reported the test results shifted their opinion against chemotherapy	
			73.1% of women with high-risk results reported the test results shifted their opinion toward receiving chemotherapy	
Hequet 2017 ³⁵	Design: Before-after study, of a single group of	<i>Study Population:</i> Patients (n = 210): Postmenopausal women with	Treatment recommendation change after receiving test results (before vs. after test): Overall: 18% (34/194)	High
Prosigna (PAM50)	patients	ER-positive, HER2-negative, and LN-negative tumors, stage 1 or	From no adjuvant chemotherapy to adjuvant chemotherapy: 13% (25/194)	Study reports recommended treatments rather than actual treatments
	Intervention: Treatment recommendation after genetic test results	stage 2, diagnosed from March 2015 to January 2016 Physician n and characteristics not stated	From adjuvant chemotherapy to no adjuvant chemotherapy: 5% (9/194) # with change in adjuvant chemotherapy (CT) treatment recommendation after test result (n = 34):	Prior physician experience using test not reported

Study Citation	Design	Study Popu	lation	Outcomes		Risk of Bias		
Test	Intervention	Characterist	tics				Comments	
	Comparator							
	Setting							
	<i>Comparator:</i> Treatment recommendation for same patient submitted before genetic test result available <i>Setting:</i> Multicenter study at 8 hospitals in France	Patient Char Mean age % PR- positive TNM stage (%) 1 2	racteristics: 61.9 (no SD) 86 79 21	Low % (n) Intermediate % (n) High % (n) Change in ph recommenda Increased co No change: 9 Decreased co Patient (n = measure dec mean of 43.3 decrease of $\frac{1}{2}$ Patient (n = decreased af (SD 10.6) to 0 (10.7) (p < .0) Patient (n = functional as after test res 80.2 (15.2), a	ation after nfidence: 51% (98/19 51% (98/19 51% (98/19 51% (98/19 51% (98/19 171) score reased aft 5 (SD 11.6) 1.8 (10.3) (158) overa ter test re 5.2 (7.7), a 01) 151) overa sessment ults from a	test results, 39% (75/192 92) 11% (21/19 son STAI "ster test results) to 41.5 (12. ($p = .02$). all scores on sults from a mean decre all scores on (FACT-G v.4) a mean of 79	% (n): 2) 2) tate-anxiety" ts from a 7), a mean 7), a mean DCS mean of 9.8 base of 3.5 the) increased 9.4 (13.3) to	Data on 6 month follow-up questionnaires not fully reported, including actual treatment received, number of physicians who completed follow-up visit, and DCS measures Study funded by test manufacturer 4 of 23 authors employed by or funded by test manufacturer

Study Citation	Design	Study Popula	tion		Outcomes	Risk of Bias
Test	Intervention	Characteristics				Comments
	Comparator					
	Setting					
					.264), and at 6 months post-diagnosis there was an overall decrease to below baseline level with FACT-G score of 76.77 (n = 161), no statistical testing reported	
					Concordance on Prosigna test subtyping and ROR between central and replication laboratories:	
					Tumor subtyping: 94% (kappa = 0.88)	
					ROR: 91% (kappa = 0.86)	
Jasem 2016 ²⁵	Design:	Study Populat			Use of adjuvant chemotherapy based on test	High
	Retrospective cohort	Women (n = 1			score in patients who received test:	
Oncotype DX	study, using data from the NCDB	intervention n ER-positive, LI		,	Intermediate test score: aOR 12.00 (95% Cl, 11.38 to 12.65)	Coding errors may have occurred since the test is new
	from the NCDD	stage breast c	0	5	High test score: aOR 83.07 (95% Cl, 74.25 to	occurred since the test is new
	Intervention:	from 2004 to		0	92.94)	Test use may be underreported
	Treatment decision	= 121,128				or erroneously reported
	based on Oncotype				Use of adjuvant chemotherapy in patients with	
	DX test results	Physician n an not reported	nd charact	eristics	vs. without testing, aOR (95% CI):	NCDB coverage is lower for
		not reported			Age, 45-64 vs. < 45 y: 0.45 (0.40 to 0.51) vs. 0.31 (0.29 to 0.34)	elderly patients, Hispanics,
	Comparator:	Patient Chara	cteristics:		Age, 65+ vs. < 45 y: 0.23 (0.21 to 0.25) vs. 0.08	Alaskan natives, and American Indians, which limits the
	Treatment decision based on other		Test	No Test	(0.07 to 0.09)	generalizability of these findings
	factors (no test		(n =	(n =	Gender, female vs. male: 1.23 (0.92 to 1.65) vs.	in these populations
	performed)		55,404)	65,724)	0.79 (0.63 to 0.98)	The study did not include data
	Setting:	Age (y) < 45	57.7	42.3	Race, black vs. white: 1.13 (1.03 to 1.23) vs. 1.29	on patients tested after 2012
	NCDB registry	45-64	55.5	44.5	(1.21 to 1.38)	because of a lag in NCDB data; factors associated with use of
	(captures about 70%					

Study Citation	Study Citation Design		ation		Outcomes	Risk of Bias
Test	Intervention	Characteristics				Comments
	Comparator					
	Setting					
	of newly diagnosed cancers in the United States submitted from about 1,500 hospitals)	65+ % female % white % governmental insurance Grade (%) 1 2 3 Unknown	31.5 45.8 34.4 38.6 49.8 49.8 43.3	68.5 54.2 54.1 65.6 61.4 50.2 56.7	Insurance, governmental vs. private: 0.88 (0.82 to 0.94) vs. 0.94 (0.88 to 1.04) Tumor grade, 2 vs. 1: 1.63 (1.54 to 1.74) vs. 2.45 (2.31 to 2.60) Tumor grade, 3 vs. 1: 3.06 (2.83 to 3.31) vs. 9.12 (8.50 to 9.78) Younger African American women more likely to receive adjuvant chemotherapy despite low RS (aOR, 1.33; 95% Cl, 1.16 to 1.54)	the test may have been different in 2012 because of wider availability, better accessibility, and higher adoption rates Lower representation of uninsured patients in NCDB compared to the general population Study funding not stated 4 of 7 study authors affiliated with industry, but none declared affiliations with test manufacturer
Jasem 2017 ²⁶ Oncotype DX	Design: Retrospective cohort using analysis of NCDBIntervention: Oncotype DX testing prior to treatmentComparator: No Oncotype DX testing prior to treatment	Study Populat Women and P positive breas between Janu December 20 pT2, pT2, pN nodes involve HER2-negativ Patients with 10,434	men with st cancer Jary 2010 12. Patier 1 (with 1- ed), HR-po ye.	LN- diagnosed and hts with 3 lymph ositive,	Factors associated with whether the test was ordered: Older patients more likely to get the test than younger patients: Adjusted odds ratio (aOR), 1.52, (95% Cl, 1.36 to 1.71) Women more likely to get the test than men: aOR, 1.65, (95% Cl, 1.29 to 2.12) Patients diagnosed in 2012 were more likely to get the test than those diagnosed in 2010: aOR, 1.72 (95% Cl, 1.61 to 1.83)	High Retrospective analysis; potential for data recording to be inaccurate Study funding not stated Authors stated they had no financial conflicts of interest

Study Citation	Design	Study Population			Outcomes	Risk of Bias
Test	Intervention	Characterist	tics			Comments
	Comparator					
	Setting					
	Setting: The NCDB breast cancer participant use data file from 2010 to 2012. The NCDB includes information on 70% of newly diagnosed cancer cases nationwide annually and collects information submitted from more than 1,500 Commission on Cancer-accredited facilities	Patients with = 21,991 LN-positive: Patient Chan Age, years, % < 45 45 - 64 \geq 65 % white Male Female Charlson co- morbidity score, % 0 \geq 1 Insurance, % None Private Govt.	n = 32	2,425	 African American patients were less likely to get the test than patients of other races: aOR 0.77 (95% Cl, 0.70 to 0.84) Patients with grade 2 (aOR, 0.81; 95% Cl, 0.76 to 0.86) and grade 3 cancer (aOR, 0.47; 95% Cl, 0.76 to 0.86) were less likely to get the test than patients with grade 1 cancer Patients with two nodes (aOR, 0.54; 95% Cl, 0.50 to 0.57) and thee nodes (aOR, 0.29; 95% Cl, 0.26 to 0.32) involved were less likely to get the test than patients with only one node involved Difference in patients receiving chemotherapy based on whether they received test Test ordered: 38% chemotherapy Test not ordered: 75% chemotherapy aOR, 0.21 (95% Cl, 0.20 to 0.22) Relationship of RS to chemotherapy treatment: Patients with an intermediate RS (aOR, 4.51; 95% Cl, 4.51 to 5.01) and high RS (aOR, 19.79; 95% Cl, 15.39 to 92.94) more likely to receive chemotherapy than patients with low RS 	
					chemotherapy than patients with IOW RS	

Study Citation	Design	Study Popu	lation			Outcomes	Risk of Bias
Test	Intervention	Characterist	tics				Comments
	Comparator						
	Setting						
		2010 2011 2012 Grade, % 1 2 3 Tumor size, % pT1 pT2 # of nodes involved, % 1 2 3	26 33 37 41 34 21 38 25 38 24 14	74 67 63 59 66 79 62 75 62 75 62 76 86	-		
Kuijer 2016 ³³ MammaPrint	Design: Retrospective cohort study Intervention: Treatment recommendation based on gene signature (determined by MammaPrint) Comparator:	Study popula Female patie primary brea than 17 year treated betw 2011 and Ou no prior hist neoadjuvan distant meta diagnosis	ents (n ast can rs of ag veen N ctober cory of t treatr astasis :: n = 1	icer, olde ge, surgio lovembe 2013, an maligna ment, or upon ,584	er cally r id had	Administration of chemotherapy in overall study population (n = 2,043), linear regression adjusted absolute difference, % (95% Cl): -9.5% (95% Cl, -15.7 to -3.3) Subgroup analysis (group status): Administration of adjuvant chemotherapy: Group A (N0, grade 1, > 2 cm; n = 19): -24.0% (95% Cl, -46.5 to -1.6) Group B (N0, grade 2, > 1 cm; n = 220): -5.8% (95% Cl, -13.0 to 1.3) Group C (pN1mi, grade 1/2; n = 59): -16.0% (95% Cl, -28.0 to -4.0)	High Patient clustering (by hospital or physician preference) may have influenced the results (authors tried to mitigate this by using linear mixed-effects modeling) Patient age may have influenced a physician's decision to administer chemotherapy (more aggressive treatment in younger

Study Citation	Design	Study Populati	on		Outcomes	Risk of Bias	
Test	Intervention	Characteristics				Comments	
	Comparator						
	Setting						
reco base conv prog	Treatment recommendation based on conventional prognostic clinicopathological	Physician n and not reported Patient characte		teristics	<i>Subgroup analysis (age):</i> Administration of adjuvant chemotherapy: < 50 years: -26.9% (95% Cl, -38.3 to -15.6) 50 to 59 years: -9.2% (95% Cl, -18.7 to 0.3)	women); unable to determine this because of study design Confounding by indication may be an issue (although authors	
	factors		Test n =	No Test n =	60 to 69 years: 3.5% (95% Cl, -5.9 to 12.9)	attempted to mitigate this through IV analyses)	
<i>Setting:</i> Registry enrolling	Mean age (y) at diagnosis (SD)	298 55.6 (8.3)	1,745 56.2 (8.8)		Study funded by the Dutch Cancer Society		
	patients from	% white	NR	NR		Authors reported no financial	
	practices in the Netherlands	Pathological axillary status (%): pN0 (i-/i+) pN1mi	80 20	77 23		conflicts of interest	
		Invasive tumor grade (%):					
		Grade 1 Grade 2	14 86	14 86			
Kuijer 2017 ³² MammaPrint	<i>Design:</i> Before-after study on one group of patients	Study population Patients (n = 66 70 years old who undergone surge positive, early-s	60) youn Io had gery for	ER-	Concordance between pretest chemotherapy recommendation and MammaPrint risk result, low vs. high risk: No chemotherapy: 59% vs. 41% Chemotherapy: 56% vs. 44%	High Pretest recommendation of chemotherapy was an artificial recommendation made with the	
<i>Intervention:</i> Treatment recommendation		cancer Physician n and	charact	teristics	Unsure: 61% vs. 39%	prospect of receiving a MammaPrint test result	
	recommendation	not reported			Oncologist adhered to MammaPrint test result in 96% of patients:		

Study Citation	Citation Design Study Population		ı	Outcomes	Risk of Bias	
Test	Intervention	Characteristics			Comments	
	Comparator					
	Setting					
	Settingbased on MammaPrint testComparator: Treatment recommendation based on clinicopathological results (preliminary recommendation in same patients before MammaPrint test results)Setting: 33 hospitals in the Netherlands	Patient character Mean age, y (SD) % white Invasive tumor grade (%): 1 2 3 % ER-positive % PR-positive % HER2- negative Axillary lymph node involvement (%): pN0 pN1mi pN1a > pN1a Nx	istics: 57.0 (8.1) NR 15 73 12 100 87 97 84 10 5 1 1 1	 MammaPrint low-risk, chemotherapy recommended: 18 patients MammaPrint high-risk, chemotherapy not recommended: 9 patients Administered chemotherapy was same as posttest recommendation in 94% of patients "Thirty patients to whom CT was recommended, of which 29 had a high-risk 70- GS test result, did not receive CT. Conversely, eight patients received CT despite the advice of the oncologists to withhold CT; seven of these patients had a low-risk 70-GS test result. Eventually administered CT was in line with the 70-GS test result in 91% of patients."^{32(p. 2816)} Treatment recommendation change after MammaPrint test, % of patients: 51% (95% CI, 46% to 56%) Actually administered treatment (CT) deviated from preliminary recommendation in 52% of patients Subgroup analyses: Discordance between preliminary chemotherapy recommendation and MammaPrint risk results: 	Possibility that oncologists used the test in a select group of patients, which would overestimate the adherence to test rates or percentages of treatment change Study funded by test manufacturer 1 author of 11 reported financial conflicts of interest with several commercial entities, but not the test manufacturer	

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
			Patients with tumors > 2 cm (n = 146): 56% (k, -0.04; 95% Cl, -0.24 to 0.17) Patients with axillary lymph node involvement: 62% (k, -0.10; 95% Cl, -0.33 to 0.13)	
Loncaster 2017 ²⁷ Oncotype DX	Design: Before-after study of same group of patients Intervention: Treatment given after genetic test results Comparator: Treatment recommendation for same patient by multidisciplinary	Study Population: (n = 201) Women with newly diagnosed breast cancer referred post- surgery to an oncologist for chemotherapy by a breast cancer multidisciplinary team. ER+ (quick score \geq 5/8) and HER2 0, 1 + or non-amplified; LN-negative or LN-positive for postmenopausal women, PREDICT estimation of overall benefit from chemotherapy > 3% LN-negative: n = 136	Reduction in the number of patients recommended for chemotherapy after test results, prior to test, all 201 patients recommended for chemo.After test, 63% (127/201) did not receive chemotherapyReduction in the number of LN-positive women (n = 65) receiving chemotherapy after test results: (Note: at time NICE guidelines did not allow for genetic testing of LN-positive women)	High The study included 201 patients, 82 from a prospective pilot study of the genetic test from May to December 2012 and then a retrospective review of patient files from December 2012 to March 2015 Authors mislabeled data in Table 2 Study designed and funded by
	tumor board submitted before genetic test result availableLN-positive: $n = 65$ Setting:Patient Characteristics:Patients treated for early breast cancer inTumor grade, $\% (n)$ 11% (2/201)		After test, 69% of LN-positive patients did not receive chemotherapy compared to 63% overall % LN-positive women receiving chemotherapy (CT) by RS score: CT No CT Low 8 92	test manufacturer and test manufacturer revised initial manuscript draft 4 of 8 authors reported consulting or serving on the speaker's bureau for the test manufacturer
	Greater Manchester, United Kingdom	1 1% (2/201) 2 52% (104/201) 3 46 % (93/201)	Intermediate6337High8317	

Study Citation	Design	Study Population		Outcomes	Risk of Bias
Test	Intervention	Characteristics			Comments
	Comparator				
	Setting				
		Tumor size, mm, mean (SD) 26.1 (14.3) Node positive, % 32% PREDICT score, mean (SD) 5.8 (2.4) RS result, mean (SD) 20.5 (10.2) RS result, % (n) 43% (86/2) Low 43% (86/2) High 13% (26/2)	01)	Estimated overall budget savings from patients not receiving chemotherapy (including costs of testing all patients) £ 266,427 Estimated budget savings from LN-positive patients not receiving chemotherapy (including costs of testing all patients) £ 110,452	
O'Neill 2017 ²⁸ Oncotype DX	Design:Retrospective cohortstudy of claims dataIntervention:Treatment decisionbased on OncotypeDX testComparator:Treatment decisionbased on otherfactors (no OncotypeDX testing)	Study Population: Commercially insured w = 5,014) age 24 to 63 ye newly diagnosed with s 2 HR-positive, HER2 not positive breast cancer Physician n and charact not reported Patient Characteristics: Age at diagnosis, y (%) 24-39 40-49 50-59 60-63	romen (n ears, tage 1 or n-	 10% of GEP-tested women did not initiate therapy Initiation of endocrine therapy, low vs. high RS women: OR 0.40 (95% Cl, 0.20 to 0.81) Predicted probability of endocrine initiation: Tested patients vs. untested patients with chemotherapy initiated within 6 months: 90.4% vs. 87.6% Tested patients vs. untested patients without chemotherapy initiated within 6 months: 90.2% vs. 74.3% 	High Participants represent a subset of the commercially insured U.S. population (from CA, GA, KY, NY, and OH) The majority of participants were from urban areas and non- white individuals were underrepresented in the data set There is the potential for coding errors in claims data

Study Citation	Design	Study Population		Outcomes	Risk of Bias
Test	Intervention	Characteristics			Comments
	Comparator				
	Setting				
	Setting: Five state cancer registries with linked claims data and GEP results from 2006 to 2010	% white 85. Stage (%) 60. 1 60. 2 39. % HER2-negative 45. % ER- and PR-positive 85. Histological grade (%) 1-2 1-2 74. Missing 21. 4.0 40.	.5 .5 .7 .5 .5	 16% of test-eligible women did not initiate endocrine therapy Receipt of Oncotype DX testing and likelihood of endocrine therapy initiation: 10% vs. 19% non-initiation; OR 2.48 (95% Cl, 2.03 to 3.04) Test status was not significantly related to discontinuation of endocrine therapy: OR 0.93 (95% Cl, 0.85 to 1.02) Test risk category was not associated with non- adherence: OR 0.88 (95% Cl, 0.76 to 1.02) 	High proportion (> 50%) of patients with unknown HER2 status The study was unable to account for variables related to survivorship care (e.g., rates and severity of side effects from endocrine therapy), receipt of care in academic centers vs. community centers, and other unmeasured patient-level variables Study funded by American Cancer Society and National Cancer Institute 3 authors had affiliations with commercial entities at the time of the study, including 2 with the test manufacturer
Parsons 2016 ²⁹	<i>Design:</i> Retrospective analysis of NCDB	Study population: (n = 132,651) Women age 18 to 70 years, ER-positive (or borderline), HER2-negative (or borderline), early-stage breast cancer (T1b – T3 and N0 or		Association of ordering test with recommendation for chemotherapy:	High
Oncotype DX	registry in U.S. Intervention:			Patients who did not receive test more likely to receive chemotherapy than those who did: OR, 1.21 (95% Cl, 1.175 to 1.249; p < .0001)	Conflict of interest: "This research was conducted by employees of the Gundersen Medical Foundation in which no

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
	Oncotype DX testing prior to treatment <i>Comparator:</i> No Oncotype DX testing prior to treatment <i>Setting:</i> The NCDB breast cancer participant use data file from 2010 to 2013. The NCDB includes information on 70% of newly diagnosed cancer cases nationwide annually and collects information from more than 1,500 Commission on Cancer-accredited facilities	N1mi, stage 1 or 2) diagnosed between 2010 and 2013LN-negative: $n = 121,688$ LN-positive: $n = 8,047$ Patients with tubular and mucinous histology or those who did not have surgery excluded; must have record of test being ordered and whether chemotherapy was received.Group A received Oncotype DX testing ($n = 74,778$)Group B did not receive Oncotype DX testing ($n = 54,987$)Patient Characteristics:Image: Compute Data and test being of the state of the	Association of RS result with receiving chemotherapy: Of the 74,778 patients who received test, the RS results were: Low: 60% (44,505/74,778) Intermediate: 32% (23,920/74,778) High: 8% (6353/74,778) Patients with intermediate RS more likely to receive chemotherapy than low RS: OR, 12.9 (95% Cl, 12.2 to 13.6; p < .001) Patients with high RS more likely to receive chemotherapy than low RS: OR 87.2 (95% Cl, 79.6 to 95.6; p < .001)	sponsorship or financial conflict of interest exists." ^{29(p. 325)}

tudy Citation	Design	Study Popula	Study Population		Outcomes	Outcomes	Outcomes Risk of Bias
Test	Intervention	Characteristic	s				Comments
	Comparator						
	Setting						
		% white	87	86			
		Comorbidities					
		%					
		0					
		1	86 12	85 13			
		2	2	3			
		Insurance, %	-	5			
		Private					
		Medicare	70	61			
		Medicaid	21	24			
			6	7			
		Year diagnosis, %					
		2010					
		2011	71	25			
		2012	15	25			
		2013	8 6	25 24			
		AJCC stage,	0	24			
		%					
		1B	23	38			
		1C	54	40			
		2	22	20			
		3	1	2			
		Node status, %					
		⁷ o pN0	00	00			
		pN0⁺	90 4	90 3			
		pN1mi	6	6			

Study Citation	Design	Study Population	Outcomes				Risk of Bias
Test	Intervention	Characteristics					Comments
	Comparator						
	Setting						
Pestalozzi 2017 ³⁸ Oncotype DX	Design:Before-after study ofsame group ofpatientsIntervention:Treatmentrecommendation bymultidisciplinarytumor board aftergenetic test resultsComparator:Treatmentrecommendation forsame patient bymultidisciplinarytumor boardsubmitted beforegenetic test resultavailableSetting:Multicenter study at18 study sites in	Study Population: (n = 229)Patients with completelyresected breast cancer withpathologically confirmednegative margins, $\geq 10\%$ ER-positive invasive malignant cells,HER2-negative, 1-3 positivelymph nodes, suitable to receivechemotherapyWHO performance status 0–1Patients excluded for bilateralinvasive breast cancer, tumorstate cT4, pT4 or nodes pN ≥ 2 ,known metastatic breast cancer,pregnant women, inability togive consentPatients divided into "low risk"(LR) group (n = 154) (pN0 tumorand ≤ 1 risk factor) or "non-lowrisk" (NLR) group (n = 70) (pN0tumor plus ≥ 2 risk factors orpN1a plus ≥ 1 risk factor)Risk factors: ER-positivity < 50%;	receiving tes Overall: 20% LR group: 15 NLR group: 3 From no adj chemothera LR group: 49 NLR group: 49 From adjuva chemothera LR group: 44 NLR group: 44	tresults (b (45/222) (45/222) (23/154) (23/154) (22/68) (22/68) (22/68) (22/68) (22/68) (5/113; 9) 16% (3/19; (3/19; nt chemother (3/13; 9) 16% (3/19; (3/13; 9) 16% (3/19; (18/41; 40% (19/48) hange of ree LR (n = 154) 15 (23/154) ange 100	5% Cl, 1 to 10%) 95% Cl, 3 to 40%) herapy to no adjuv	:): %) %) vant vant	 High Study reported on "reason for change of recommendation" but did not explain method for determining reason Manufacturer provided tests free of charge Study funded in part by Swiss State Secretariat for Education, Research, and Innovation 1 of 20 authors reported consulting fees from manufacturer
	Switzerland	tumor size > 5 cm; "extensive lympho-vascular invasion;" Ki67 > 30%	Opinion of tumor	(23/23) 4 (1/23)	5 (1/22)		

Study Citation	Design	Study Pop	ulation		Outcomes			Risk of Bias	
Test	Intervention	Characteris	stics						Comments
	Comparator								
	Setting								
		LN-negativ		1	board changed Patient	13 (3/23)	5 (1/22)		
		LN-positive	LN-positive: n = 88 Patient Characteristics:						
		Patient Cha				9 (2/23)	5 (1/22)		
			LR	NLR					
		Age, median (range)	58 (35–82)	58 (32– 79)					
		pT stage, %							
		T1 T1a	1 1	3					
		T1b T1c	12 47	7 48					
		Т2	36	33					
		T3 Tis	3 1	9					
		pN state, %							
		pN0 pN1a	76 24	28 72					
		Tumor grade (BRE), %							
		G1 G2 G3	15 78 8	7 39 54					
Ray 2016 ³¹	Design: Retrospective analysis of Kaiser		<i>lation:</i> n	= 7004,	Patient chard	acteristics a	associated with tes	t use:	Moderate
Study Citation	Design	Study Population			Outcomes	Risk of Bias			
----------------	--	---	--	----------------	--	---			
Test	Intervention	Characteristics				Comments			
	Comparator								
	Setting								
Oncotype DX	DettingPermanente's tumor registry recordsIntervention: Oncotype DX testing prior to treatmentComparator: No Oncotype DX testing prior to treatmentSetting: Patients treated for breast cancer by Kaiser Permanente Northern California 2005–2012	HER2-negative, sta stage 2 breast canc primary tumors ≥ 0 either no node invo only $\leq 2mm$ axillary metastasesLN-negative: n = N LN-positive: n = NF Intervention group: n = 1567, received testControl group: n = receive Oncotype DPatient Characteristic < 65 years old: 57%Characteristic $\leq 1 cm$ Tumor size, % $\leq 1 cm$ Same tage, %	cer with 0.51 cm with olvement of y node mid IR R Oncotype 5437, did r DX test stics: % st No test 72 33 45	r ro- DX	Overall, 22% of women received test (1567/7004) Compared to women aged 50 to < 65 years: Women aged 40 to < 50 years more likely to be tested: OR 1.22 (95% Cl, 1.04 to 1.44) Women aged 65 to < 75 years less likely to be tested: OR 0.42 (95% Cl, 0.36 to 0.49) Women aged 75+ years less likely to be tested, OR 0.04 (95%, Cl 0.02 to 0.06) Compared to women with tumors > 2.0 cm: Women with tumors 0.5 cm to \leq 1.0 cm less likely to be tested: OR 0.51 (95% Cl, 0.42 to 0.61) Women with tumors > 1.0 to \leq 2.0 cm more likely to be tested: OR 1.20 (95% Cl, 1.03 to 1.40) Each \$10,000 increase in residential block- group median income was associated with increased odds of testing: OR 1.05 (95% Cl, 1.03 to 1.07) Correlations between RS and chemotherapy use:	Funded by National Cancer Institute, National Institutes of Health Financial conflicts of interest for study authors not stated			

Study Citation	Design	Study Population		Outcomes			Risk of Bias
Test	Intervention	Characteristics					Comments
	Comparator						
	Setting						
		175225Comorbidity score, %79Low79Intermediate20High1Initial treatment, %26Radiation53Hormone therapy70	49	Among patients tests of chemotherapy by high risk: Low 52 Intermediate 39 High 9 Association of test us chemotherapy use: Between 2005 and 20 receiving test rose from Among women tester 2012, % of women red dropped from 26% test "In analyses with indiscore matching (n = gene test testing was decreased odds of ch 95% CI, 0.63 to 0.87) reduction in the percoreceiving chemother or an absolute reduct to 13.6%)" ^{31(p. 6)}	low, intermediate, ents % patients receive chemo 8 40 72 e with reduction in 012, % of eligible om 8% to 25 % ed, between 2005 a eceiving chemothe o 22% ividual-level prope 2924), receipt of t s associated with nemotherapy (OR, corresponding to centage of women apy from 32.7% to	, and , and women and erapy ensity he 21- 0.74; a 0.26.5%,	

Study Citation	Design	Study Population	Outcomes		Risk of Bias
Test	Intervention	Characteristics			Comments
	Comparator				
	Setting				
Sanft 2015 ³⁷ BCI	Design:Before-after study, cohort was compared before and after BCI testIntervention:Treatment 	Study Population:Patients (n = 96): Women with a history of ER- or PR-positive, stage 1 to 3 breast cancer, who had completed at least 3.5 years of adjuvant endocrine therapy and were eligible for extended adjuvant endocrine therapy not reportedPhysician n and characteristics not statedLN-negative: n = 70 LN-positive: n = 26Patient Characteristics:Median age (range)% white91 % ER-, PR- positive% ER-, PR- positive100, 88 positiveGrade 1 Grade 2 (range)31 Grade 2 (range)TNM stage (%)	Treatment recommendareceiving test results (b Overall: 26% Extended adjuvant ther 0.14; 95% CI, 0.04 to 0.4 Post-BCI recommendat adjuvant treatment, bas recommendation: Pre-BCI Yes: 74% No: 26% Physicians who felt stro recommendation (befo 24% (OR, 4.75; 95% CI,	efore vs. after test): apy: 74% vs. 54% (OR, 46) ion for extended sed on pre-BCI <u>Post-BCI</u> Yes: 69% No: 31% Yes: 12% No: 88% ngly confident in re vs. after test): 8% vs.	 High Prior physician experience using test not reported Use of OR for estimate of effect when an outcome that is not rare generally overestimates the effect size Study funder not disclosed 3 of 14 authors had a financial relationship with test manufacturer

Study Citation	Design	Study Population		Outcomes	Risk of Bias
Test	Intervention	Characteristics			Comments
	Comparator				
	Setting				
		Stage 160Stage 235Stage 35Nodal status at diagnosis* (%)	-		
Tsai 2017 ³⁴ MammaPrint (used to further risk stratify after use of the 21- gene (Oncotype DX) assay)	Design:Before-after study on one group of patientsIntervention:Treatment recommendation based on MammaPrint test resultsComparator: Treatment recommendation made prior to MammaPrint test	N30Study population:Patients (n = 840) with Lpositive or LN-negativecancer and an intermedgene assay result at enrolLN-negative: n = 744LN-positive: n = 96Physician n and characternot reportedPatient characteristics:Mean age, y (range)59 (27-93) (73)% white87.3Tumor grade (%):	breast ate 21- ollment	Change in treatment recommendation based on MammaPrint test results: <i>Low-Risk Patients:</i> Removed chemotherapy for MammaPrint low- risk patients who were initially recommended chemotherapy: OR 108.00 (95% CI, 18.98 to 4304.77) <i>High-Risk Patients:</i> Removed chemotherapy for MammaPrint high-risk patients who were initially recommended chemotherapy: OR 0.01 (95% CI, 0.001 to 0.04) <i>All Patients:</i> Removed chemotherapy as a result of the MammaPrint test when chemotherapy was initially recommended (prior to test): OR 0.64 (95% CI, 0.50 to 0.82)	 High Recurrence and survival data were not collected Inclusion criteria limited patients to those with ER-positive, HER2-negative breast cancer and majority were LN-negative Use of OR for estimate of effect for an outcome that is not rare generally overestimates the effect Study was funded by test manufacturer who had had input on the design of the study

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
	(based on 21-gene assay) Setting: U.S. clinical practices	T1 75.7 23.6 96 % ER-positive 99.9 % HER2- 97.7 negative 88.6 negative 88.6	Treatment recommendations changed after MammaPrint test results received: Overall: 33.6% Low-risk result; chemotherapy removed from treatment: 28.9% High-risk result; chemotherapy added to treatment: 36.7% Recommended endocrine therapy before vs. after MammaPrint test results: 94.5% vs. 95.6% <i>Subgroup analyses:</i> <i>Discordant 21-GA and MammaPrint Results:</i> Patients (n = 368) with discordant results between 21-GA and MammaPrint tests, treatment regimen changed: 75.8% MammaPrint low-risk patients recommended no adjuvant chemotherapy: 90.6% MammaPrint high-risk patients recommended adjuvant chemotherapy: 87.8% <i>LN Status:</i> LN-positive, change in treatment decision: 32.1% LN-negative, change in treatment decision: 33.7% Change in chemotherapy treatment decision due to MammaPrint test result: HR 1.85 (95% CI, 1.31 to 2.63)	and the conduct, collection, management of the study, analysis and interpretation of the data, and review and approval of the manuscript 8 of 14 authors reported financial conflicts of interest, but not with the manufacturer of MammaPrint test

Study Citation	Design	Study Population		Outcomes	Risk of Bias
Test	Intervention	Characteristics			Comments
	Comparator				
	Setting				
Wuerstlein 2016 ³⁶ Prosigna (PAM50)	Design: Before-after study of one group of patients Intervention: Treatment recommendation after genetic test results Comparator: Treatment recommendation for same patient submitted before genetic test result available Setting: Multicenter study at 11 sites in Germany	Study PopulationPatients (n = 201):Postmenopausal women wER-positive, HER2-negativeLN-negative tumors, stagestage 2LN-negative: n = 196LN-positive: n = 1Physician n and characteristicsMedian age64 (range 40 to 81)%ER-, PR- positive99, 87positiveECOG status (%) 0090110Tumor Stage (%) pT177 pT223	ve, and e 1 or	Treatment recommendation change after receiving test results (before vs. after test): Overall: 18% (36/198) From no adjuvant chemotherapy to adjuvant chemotherapy: 11% (22/198) From adjuvant chemotherapy to no adjuvant chemotherapy: 2% (5/198) Change in chemotherapy regimen: 5% (9/198) Change in chemotherapy regimen: 5% (9/198) Change in physician confidence in treatment recommendation after test results % (n): Increased confidence: 89% (177/198) No change: 9% (18/198) Decreased confidence: 2% (3/198) Change in physician confidence with patient cancer characteristics (subtype, ROR) % (n): Increased confidence: 88% (174/198) No change: 10% (20/198) Decreased confidence: 2% (4/198) Patient (n = 187) scores on STAI "state-anxiety" measure decreased after test results from a mean of 40.5 (SD 11.1) to 38.5 (11), (p = .082)	 High Study reports recommended treatments rather than actual treatments Study funded by test manufacturer 7 of 17 authors employed by or received funding from test manufacturers

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
			Study did not report patient overall scores on DCS or FACT-G v.4 in supplementary materials table	
			Concordance on Prosigna test subtyping and ROR between central and replication laboratories:	
			Tumor subtyping: 94% (kappa = 0.89)	
			ROR: 93% (kappa not reported; Pearson correlation 0.963)	
			Discordance in intrinsic subtype classification between Prosigna test results and local assessments by treating physician (IHC-based): 28%	

Citation	Design	Population	Main Findings	Risk of Bias
Test	Comparators	Analytic Assumptions		Comments
Gustavsen 2014 ⁵² BCI	Design: Decision-analytic model Comparators: BCI at initial diagnosis and at 5 years post- diagnosis, compared to usual care	Population:Women with ER+, LN- breast cancer at initial diagnosis making decision regarding adjuvant chemotherapy, and women with ER+, LN- breast cancer recurrence- free at 5 years post- diagnosis making decision regarding duration of endocrine therapyAnalytic Assumptions: U.S. commercial payer perspective, patient flow models based on 2012 NCCN guidelines; 2 scenario decision tree representing outcomes with and without use of BCI; health outcomes modeled were recurrence or no recurrence after 5 years and extended endocrine treatment or observation at 5 years post-diagnosis; BCI cost at diagnosis \$4,950 and at 5 years \$3,450	Mean costs per patient (net difference) for payer with BCI used at diagnosis for adjuvant chemotherapy planning vs. usual care using clinicopathological factors to guide treatment: \$41,634 vs.\$45,437 (-\$3,803) Sensitivity analyses ranged from -\$4,000 to -\$2,800 in favor of BCI testing Mean costs per patient (net difference) for payer with BCI used at 5 years post-diagnosis for extended endocrine treatment planning vs. usual care using clinicopathological factors to guide treatment: \$20,904 vs .\$22,708 (-\$1,803) Sensitivity analyses ranged from -\$2,300 vs\$300	 High Discount rate not specified Statistical methods not well described, but employs deterministic decision-analysis modeling, which is generally not considered best practice One-way sensitivity analyses do not account for interrelationships among inputs and outcomes Model parameters based on questionable assumptions (e.g., assumed same decision impact as Oncotype DX because no such studies exist for BCI test) Based on NCCN guideline recommendations, which are now out of date Study funded by the marketer of the test 3 of 7 authors had a financial relationship with test manufacturer

Table 8. Evidence Table: Breast Cancer Economic Studies

Citation	Design	Population	Main Findings	Risk of Bias
Test	Comparators	Analytic Assumptions		Comments
Hall 2017 ⁵¹ Oncotype DX MammaPrint Prosigna	Design: Modified Markov modeling using methods guidance from U.K. NICE Comparators: Oncotype DX testing vs. alternative gene testing with MammaPrint, Prosigna and IHC4 tests	Population: Women age 40 or older with ER-positive, HER2- negative, clinically high risk (1-10 LN+ or LN- with tumor size \geq 30 mm) surgically treated early invasive breast cancer Analytic Assumptions: UK NHS perspective using lifetime horizon; inputs based on OPTIMA preliminary feasibility (preliminary) trial of standard treatment vs. Oncotype DX guided therapy using high risk cutoff score of > 25; tissue samples from OPTIMA preliminary trial were also tested with MammaPrint, Prosigna and IHC4 tests; costs reported in 2012-12 £; Monte Carlo simulation sensitivity analyses based on effectiveness of chemotherapy, and survival after recurrence; test costs £2,580 for Oncotype, £2.207 for	Base-Case Analysis:Lifetime per-patient cost in (no testing), assuming chemotherapy for all patients:£13,961 (95% Cl, £10,535 to £21,203Lifetime per patient QALYs in base-case analysis (no testing), assuming chemotherapy for all patients:7.69 (95% Cl, 5.06 to 9.58)Mean incremental cost (in £) per person (95% Cl):Oncotype DX: -108 (-4,610 to 4,592)MammaPrint: 195 (-3,206 to 3,430)Prosigna Subtype: -281 (-3,553 to 2,774)Prosigna ROR: -474 (-4,078 to 2,955)Mean incremental QALYs (95% Cl) per person for testing:Oncotype DX: 0.2 (-1.07 to 1.4)MammaPrint: 0.18 (-0.87 to 1.1)Prosigna Subtype: 0.18 (-0.91 to 1.15)Probability that test is cost-effective:Oncotype DX: 0.77MammaPrint: 0.75Prosigna ROR: 0.77Prosigna ROR: 0.77Sensitivity Analyses—Constant relative chemotherapy effect	Moderate Effectiveness estimates for alternative tests assumed from the degree of discordance between the test and the Oncotype DX test seen in the OPTIMA preliminary trial because those tests were not directly used in the trial Study funded by U.K. NIHR HTA Programme Authors were all in academic positions, but there was no specific declaration of interests for each author

Citation	Design	Population	Main Findings	Risk of Bias
Test	Comparators	Analytic Assumptions		Comments
		(£1.576-£1,773) for	Oncotype DX: 0.33	
		Prosigna	MammaPrint: 0.31	
			Prosigna Subtype: 0.41	
			Prosigna ROR: 0.35	
			Sensitivity Analyses—Variable survival after recurrence	
			Probability that test is cost-effective:	
			Oncotype DX: 0.97	
			MammaPrint: 0.94	
			Prosigna Subtype: 0.94	
			Prosigna ROR: 0.94	

Prostate Cancer

Table 9. Evidence Table: Prostate Cancer Observational Studies

Study Citation	Design	Study Population	1	Outcomes	Risk of Bias
Test	Intervention	Characteristics			Comments
	Comparator				
	Setting				
Albala 2016 ³⁹ Oncotype DX	SettingDesign:Before-after study, cohortwith historicalcomparison groupIntervention:Oncotype DX testComparator:No genomic testingduring historical periodSetting:Physicians and patientsfrom one large integratedgroup practice inSyracuse, NY withurologists, pathologists,and radiation oncologists,onsite laboratory andIMRT facilities.Patients had to beinsured by the major	Intervention grou very low-risk, low- intermediate-risk enrolled July 2, 20 	risk or favorable prostate cancer, 014 through p, n = 100, ole low-risk working ecember 31, 201 for selection not stics: No test 6 64.9 52 97.5 79	comparator group [Intervention, n = 51 vs. control, n = 71, (difference)]: Active surveillance (AS): 59% vs. 38% (+21%) Radical prostatectomy: 25% vs. 35% (- 10%) B IMRT: 12% vs. 25% (-13%)	 High Historical comparator group with clinical and temporal differences that were not controlled for in analysis Could not mask physicians to use of test or treatment decision Study funder not named, one author employed by test manufacturer

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
	carrier in area (Excellus BCBS).	% NCCN risk group:kVery low28.825Low3546Intermediate36.329	diagnostic biopsy and for 180 days afterward) Untested historical group: \$1,007,434 Tested cohort: \$890,829 Average difference on per patient basis: \$2,286 less for tested patients (n = 51) Cost of test per patient, included in total aggregate costs: \$4,520	
Badani 2015 ⁴⁰	Design:	Patients (n = 158): Men newly	Treatment recommendation change	High
Oncotype DX	Before-after study of one group of patients Intervention: Treatment recommendation after genetic test results Comparator: Treatment recommendation for the same patient submitted before genetic test result available Setting: Patients and urologists from 1 academic and 2	diagnosed (within past 6 months) with very low-, low-, or low- intermediate risk prostate cancer, 50 years and older with greater than 10-year life expectancy Urologist n and characteristics not stated Patient characteristics: Mean age 69.9 (+/- SD) (7.26) % white 76.6 % Gleason: 3+3 70.3 3+4 29.7 Mean PSA 5.8, (2.9) (ng/mL), (+/- SD) \$\$5.8, (2.9) % NCCN risk group:	after receiving test results (before vs. after test): Overall: 18% Active surveillance: 41% vs. 51% Radical prostatectomy (with and without LN dissection): 51% vs. 42% EBRT: 6% vs. 4% Multimodal: 2% vs. 3% 6 of 64 (9%) recommendations for active surveillance were changed to radical prostatectomy Change in recommended treatment intensity, # (%): 25 (15.8%) decreased treatment intensity 14 (8.9%) increased treatment intensity 2 (1.4%) equivocal intensity change 59 (37.3%) no change	Urologists provided with NCCN risk classification training after baseline treatment recommendation, which may have influenced posttest treatment recommendation if not part of clinical practice prior to study Before-after studies do not have an independent control group Prior physician experience using test not reported Up to 6 months could have elapsed between diagnosis and enrollment in study, and time between initial diagnosis or

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
	community urology practices in the U.S.	Very low22.2Low44.9Low-32.9intermediate	Physician confidence in treatment recommendation after test: 84.8% of cases rated as strongly agreed, agreed, or somewhat agreed	treatment recommendation and availability of test result and final treatment recommendation not reported Actual treatment received and
			8.8% of cases rated as disagreed or somewhat agreed	patient attitudes not reported
			Physician assessment of test result usefulness in clinical decision making: 78.5% of cases rated as extremely	Study funded by test manufacturer
			useful, useful, or somewhat useful 15.8% of cases rated as not useful or somewhat not useful	7 of 8 authors had a financial relationship with test manufacturer
Crawford 2014 ⁴¹	Design:	Physicians (n = 150) from 31 U.S.	Overall % with change in recommended	High
Prolaris	Before-after study of one group of patients	states who completed a voluntary survey when ordering the test, 134 (89.3%) specialized in urology	therapy after test: 198/305 (64.9%) Recommendation before vs. after test, # (net % change):	Prior physician experience using test not reported
	Treatment recommendation after test results (options included RP, RT, ART, WW, AS, other, or combinations)	Patients (n = 305), 26 removed from 331 enrolled in registry because they selected "undecided" as treatment either pre- or posttest Patient characteristics:	Radical prostatectomy: 105 vs. 53 (- 49.5%) Radiation: 105 vs. 75 (-28.6%) [note, abstract reports -29.6%] Change from recommended interventional therapy to watchful	Enrollment of patients into registry was voluntary, so sample may represent a group of patients for whom increased prognostic information was desired
	Comparator:	Mean age 67.4 (+/- SD) (7.43)	waiting/active surveillance, # (net % change):	Other factors contributing to treatment recommendation or

Study Citation	Design	Study Population			Outcomes	Risk of Bias
Test	Intervention	Characteristics				Comments
	Comparator					
	Setting					
SettingTreatmentrecommendation for thesame patients before teresults (same categorieas pretest options)Planned interim reportfirst 331 patients enrollin registry, July 19, 2013to December 9, 2013Setting:Registry enrollingpatients from U.S. basepractices		% white76.6% Gleason: \leq \leq 651.4 $3+4$ 28.5 $4+3$ 11.8Mean PSA57.7, (10/mL), (+/- SD)% AUA risk group:28.32)Wath Normalized Low44.3Intermediate High42.9High12.8			61/164 (37.2%) Change in recommended treatment intensity, pre- vs. posttest, # (%): Reduced intensity: 122 (40%) No change: 107 (35.1%) Increased intensity: 76 (24.9%) For 116 patients (38% of subjects), actual treatment was confirmed with chart review, # (%): Concordant with recommendation: 93/116 (80.2%) Discordance with recommendation:	actual treatment initiated, including patient attitudes, were not assessed Decision impact confirmed by chart review for approximately one-third of the sample Study funded by test manufacturer 5 of 7 authors declared a financial interest with the manufacturer Journal peer reviewers received
					23/116 (19.8%), 6/23 had decreased intensity treatment and 17/23 had increased intensity treatment	honoraria for their reviews
Dall-Era 2015 ⁴²	<i>Design:</i> Before-after study, cohort with historical	Physicians (n = ordered at lea the test's laun	st 4 tests,	0	Treatment recommended (test group [n = 114] vs. historical group [n = 60]) Active surveillance/watchful waiting:	High
Oncotype DX	Intervention: Treatment recommendation after test results, collected from medical records by physician/staff, tests	Patients with low- or low- intermediate risk prostate cancer whose cases were identified and enrolled in study by their urologists (n = 124 in test group, n = 87 in historical group)		te cancer fied and ir st group, n	Active survemance/watchidi waiting. 61% vs. 50% Immediate treatment (total): 39% vs. 50% Radical prostatectomy only: 17% vs. 32% Radiation (including IMRT) only: 12% vs. 10% Brachytherapy only: 4% vs. 3%	Only urologists familiar with the test, having ordered at least 4 tests in the May 2013 to February 2014 period, were invited to participate Patients were selected by their urologists for inclusion in the study, with no reporting of how

Study Citation	Design	Study Populatio	on	Outcomes	Risk of Bias
Test	Intervention	Characteristics			Comments
	Comparator				
	Setting				
	Settingordered soon after test became available in May 2013Comparator:Patients diagnosed before test was developed, from May 2012 to April 2013Setting:U.Sbased urologists (93% in large urology groups), with a mean of 20 years of clinical practice experience, in CA, CO, GA, NJ, NY, PA, TX, VA	Patient characterImageTexMean age,65years (+/-(9.100)SD)%% white81% Gleason:≤≤ 3+3843+416% PSA 4-1068(ng/mL)%% NCCN riskgroup:Very Low42Low40Intermediate17	est Historical 5.2 64.4 .9) (10.4) 1 75 4 84 5 16 3 72 2 39 4 3	Multimodality therapy: 1% vs. 2% Other monotherapy: 6% vs. 3% Treatment received (test group [n = 117] vs. historical group [n = 83]) Active surveillance/watchful waiting: 67% vs. 43% Immediate treatment (total): 33% vs. 57% Radical prostatectomy only: 14% vs. 27% Radiation (including IMRT) only: 6% vs. 14% Brachytherapy only: 4% vs. 6% Multimodality therapy: 3% vs. 5% Other monotherapy: 6% vs. 5%	many subjects were evaluated for inclusion Reasons for missing data for recommended treatment in the test group (10/124, 8%) and historical group (27/87, 31%) not reported; lower proportion of missing data for treatment actually received—6% and 5%, respectively; treatment recommendation was not documented in 37 charts and for 11 patients the actual treatment received was not documented Outcomes abstracted from patient charts by urologists or their staff rather than independent entity Timing of data collection relative to chart entry of treatment recommended and actually received not stated;

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
				abstracted at the same time for the test group Patients were not informed or asked to consent to having their data used in the study because a commercial IRB deemed study to be exempt Study funding source not stated All authors had a financial relationship with the test
				manufacturer
Eure, 2017 ⁴³	<i>Design:</i> Before-after study, cohort	Physicians (n = unclear, 26 practices enrolled) who practiced	Test result reclassified risk status for 23% of patients in test group	High
Oncotype DX	with historical comparison group; pre-specified interim analysis of an observational study of test in urology practices <i>Intervention:</i> Treatment recommendation after test results	at a center that had experience in AS and annually treated at least 25 newly diagnosed prostate cancer patients who were NCCN very low, low, or intermediate risk Patients (n = 297) with a valid test result, pretest physician treatment recommendations, and posttest shared decision recorded, enrolled between July 2014 and September 2015 Historical group (n = 247) patients diagnosed between January 2012	Test group patients with a change in shared management decision after test results discussed, #, %: NCCN low risk group: 31/111 (28%) Switch from IT to AS: 21/41 (51%) Switch from AS to IT: 10/70 (14%) % with shared management decision for AS (test group after testing vs. historical group):	Historical group patients had to meet entry criteria and were not individually consented for study, but test group patients were individually consented for study, which may bias toward a group more likely to seek information and have involvement in treatment decisions—this could have led to an overestimate of the influence of the test
		and September 2014, before test	Total: 62% vs. 40%	Pretest treatment recommendations recorded only

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
	<i>Comparator:</i> Patients diagnosed before test was developed, identified by chart review	was available, and who had at least a year of recorded follow-up after diagnosisIntermediate risk: 23% vs. 19% Low risk: 73% vs. 43% Very low risk: 88% vs. 57%Patient characteristics:% with AS persistence rate at one ye (test group vs. historical group)		for the NCCN low-risk group, but the difference in AS recommendation (62% vs. 40%) suggests that the groups may not have been clinically similar in ways that may not be evident
	<i>Setting:</i> 22 community urology practice centers in the	% with age 46.1 42.9 ≥ 65 years 81 75	Overall: AS maintained by 53% vs. 34% of study groups, or 89% vs. 86% of those who chose AS	Fewer practices contributed patients to the historical group
	U.S. contributed patients in the test group, and 9 practices contributed	% Gleason: 75.6 3+3 75.6 3+4 24.4	In 90% of cases, physicians reported the test was useful	than the test group, which may contribute to non-comparability of the groups
	patients to the historical group; practices	% PSA 4.1- 69.8 74.5 9.9 (ng/mL) 74.5	In 92% of cases, physicians reported the test was a source of increased	Study funded by test
	described as "geographically diverse," but locations not specified	% NCCN risk group:43.0Very Low26.4Low43.0Intermediate30.632.4	confidence 96% of patients found the test useful in decision making Low decisional conflict (< 25 on DCS) was reported by 59% of men after, compared with 36% before testing	manufacturer All authors had a financial relationship with the test manufacturer
Gore 2017 ⁴⁴ Decipher	744Design: Before-after study of aPatients (n = 265) with prostate cancer that was previously treated by DD and warms heine sensitive addressed		Treatment recommendation before test, ART group Observation: 88.7% Adjuvant radiotherapy: 11.3%	High This is an interim data analysis that reports on recommended
	radiotherapy (ART) and another group making a	= 115)	Treatment recommendation before test, SRT group	treatments rather than actual treatments received

Study Citation	Design	Study Population		Outcomes	Risk of Bias
Test	Intervention	Characteristics			Comments
	Comparator				
	Setting				
	Settingdecision about salvage radiotherapy (SRT)Intervention: Treatment recommendation after test results, with recruitment into study between May 2014 and February 2016Comparator: Treatment recommendation for the same patients before test resultsSetting: 19 U.S. and one Canadian community and academic	Physicians (n = not treating providers of patients Patient characteristi ART Group Median age at RP, years (range) % non- Hispanic white % Gleason: 3+3 75.6 3+4 24.4 Median 6.3 (0.9- pre-RP PSA (range)	fenrolled	 Observation: 58.3% Either ART, or SRT and ART: 9.6% Change after testing ART arm: 18% (27 patients) SRT arm: 32% (37 patients) Multivariable model OR (95% Cl) for either ART or SRT after test result known, adjusted for age, pre-RP PSA, extraprostatic extension, seminal vesicle invasion, surgical margin, Gleason score, (and for SRT group, years to biochemical recurrence or PSA rise) ART group: aOR 1.48 (1.19 to 1.85) SRT group: aOR 1.30 (1.03 to 1.65) Patient median decisional conflict score (IQR) before and after test, ART group: 	Patients acted as their own control and there was no group enrolled that did not get testing—such a group given an equivalent amount of time to make a decision could have similar levels of decisional conflict and their physicians might make similar types of treatment recommendations Patients in the SRT group had varying amounts of time since RP, which could influence treatment recommendations as well as decisional conflict scores Use of OR for estimate of effect for an outcome that is not rare generally overestimates the effect
	sites participating in the Society of Urologic Oncology Clinical Trials Consortium	(ng/mL) % Gleason score: 6 1.3 3+4 52 4+3 27.3	7.8 42.6 31.3	 25 (8 to 44) and 19 (2 to 30) Patient median decisional conflict score before and after test, SRT group: 27 (16 to 41) and 23 (4 to 30) Physician median decisional conflict score (IQR) before and after test, ART group: 32 (28 to 36) and 28 (23 to 34) 	Study funded by test manufacturer 7 of 22 authors declared a financial interest with the manufacturer

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
			Physician median decisional conflict score before and after test, SRT group: 33 (26 to 36) and 29 (22 to 34) All DCS changes from pre- to posttest, among patients and physicians, reported as statistically significant, p < .001	
Michalopoulos 2014 ⁴⁵ Decipher	Design:Before-after study of urologists caring for a group of patientsIntervention:Treatment recommendation after test resultsComparator: Treatment recommendation for the same patients before test resultsSetting: Community-based practices (community	 Physicians (n = 15 of 18 invited to participate in study) who had ordered the test as part of another study and had performed at least one RP Median time in clinical practice, years (min-max): 11 (3 to 20) Median # (min-max) RP performed per year: 70 (18 to 300) Patients of these 15 urologists (n = 146) who were eligible for adjuvant therapy after radical prostatectomy according to clinical practice guidelines <i>Patient characteristics:</i> 	Treatment recommendation before test results given: Adjuvant therapy: 27.4% Close observation: 69.9% Other (close observation until PSA rises): 2.7% Treatment recommendation after test results: Adjuvant therapy: 27.4% (64.3% for high-risk test result, 4.4% for low-risk test result) Change in treatment recommendation after test, #, % (95% Cl) Overall: 40/146 patients, 30.8% (23 to 39%)	High Study reported recommended treatments rather than actual treatments No patient views or decision input were reported and study presented limited data from a small group of urologists Urologists invited to participate on the basis of having ordered the test, which may select for a group with more interest in test information and information- seeking to inform decisions than a group of urologists who did

Study Citation	Design	Study Population	Outcomes	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
	hospital, large urology groups, or private practice) of a group of 15 U.S. board-certified urologists in KS, MO, NE, OK, TX, and TN	Median age at RP (min-max) 63 (48-79) % white Not reported % biopsy Gleason: 31.5 3+4 30.8 4+3 14.4 % pre-RP PSA < 10 (ng/mL)	Change from any active treatment (adjuvant radiation therapy and/or hormone therapy) to observation: 17/40 patients, 42.5% (27 to 59%) Change from observation to any active treatment (adjuvant radiation therapy and/or hormone therapy): 18/102 patients, 17.6% (11 to 26%) Physicians' DCS score before and after reviewing test results: 34 and 27; p < .0001	not have a history of test ordering Study funded by test manufacturer 5 of 8 authors declared a financial interest with the manufacturer Journal peer reviewers received honoraria for their reviews
Shore 2016 ⁴⁶ Prolaris	Design: Before-after study of a group of patients at 4 time points enrolled in PROCEDE-1000 registry Time periods: A: physician recommendation before test B: physician recommendation after test result available	Patients (n = 1206): newly diagnosed (6 months or less), histologically proven, presumed clinically localized prostate cancer, who had not received any treatment and had sufficient biopsy tissuePhysicians (n = 124) Patient characteristics:Mean age, years (+/- SD) $\%$ white 77 $\%$ Gleason: 6 47.8	Treatment option altered from initial recommendation (time period A) to actual treatment (time period D). #, %, (95% Cl): Overall: 576/1206, 47.8% (45.0 to 50.6) Of 576 patients with treatment changes Decrease in intensity: 72.1% Increase in intensity: 26.9% Treatment recommendations involving multiple treatment options, before test vs. actual treatment: 31.6% vs. 12.9%	High No patient views or decision input reported Urologists participated in registry on the basis of using the test, which may select for a group with more interest in test information and information- seeking to inform decisions than a group of urologists who did not have a history of test ordering

Study Citation	Design	Study Population		Outcomes	Risk of Bias
Test	Intervention	on Characteristics			Comments
	Comparator				
	Setting				
	SettingC: treatmentrecommendation afterconsultation with patientD: actual treatmentreceived, measured 3-6months after consultationIntervention:Treatmentrecommendation andtreatment after testresults (Time periods B, C, D)Comparator:Treatmentrecommendation for the		27.9 11.9 7.8 (8.15) 40.3 42.0 17.7	Change from noninterventional to interventional treatments: 101/417 (24.2%) Change from interventional to noninterventional treatments: 112/789 (14.2%)	Study funded by test manufacturer 5 of 13 authors declared a financial interest with the manufacturer
	same patients before test results (Time period A) <i>Setting:</i> U.Sbased urology practices, with study site selection based on previous investigator experience with the test (no further details given)				

Citation	Design	Population	Main Findings	Risk of Bias
Test	Comparators	Analytic Assumptions		Comments
Albala 2016 ³⁹ Oncotype DX				See observational studies evidence table for details, one economic outcome (costs) reported, but no formal economic analysis provided
Lobo 2017 ⁵⁵ Decipher	Design: Cost-effectiveness analysis Comparators: Genetic testing- based care vs. alternatives of 100% uptake of adjuvant therapy or usual care for outcomes of QALYs, and biochemical recurrence or metastasis at 5- and 10-year posttreatment intervals	Population: Men with prostate cancer in the post- prostatectomy setting Analytic Assumptions: Model included inputs for disease progression, death, treatment complications above those expected for RP alone, probabilities for adjuvant therapy delivery or salvage therapy delivery across usual care and genetic test risk groups, utility values for side effects and health states, and costs	Markov modeling conducted to provide estimates below Genetic test-based care vs. 100% adjuvant therapy Mean difference in life years: 0.024 (95% Cl, -0.012 to 0.059; p = .187) Mean difference in QALYs: 0.140 (95% Cl, 0.080 to 0.201; p < .001) Average per-person cost of care, assuming a test cost of \$4,000: \$23,823 vs. \$31,125 Patients receiving 100% adjuvant therapy had improved clinical outcomes at 5 and 10 years Genetic test-based care vs. usual care (mean difference, 95% Cl) Mean difference in life years: 0.023 (95% Cl, -0.012 to 0.058; p = .191) Mean difference in QALYs: 0.066 (95% Cl, 0.016 to 0.117; p = .0098) Average per-person cost of care, assuming a test cost of \$4,000: \$23,823 vs. \$18,370 Genetic test-based care provided improved clinical outcomes, including reduced incidence of metastases ICER: \$90,883	 High No analytic perspective stated Effectiveness estimates for clinical outcomes based on two non-comparative before-after studies of treatment decisions rather than actual treatment or other clinical outcomes Other model inputs generally based on single studies rather than systematic reviews of multiple studies, and sensitivity analysis ranges were varied by set amounts (e.g., +/-10%), rather than the plausible ranges reported across a comprehensive set of studies or other data sources Study funded by test manufacturer 3 of 11 authors had a financial relationship with test manufacturer

Table 10. Evidence Table: Prostate Cancer Economic Studies

Citation	Design	Population	Main Findings	Risk of Bias
Test	Comparators	Analytic Assumptions		Comments
		Genetic test cost modeled at \$4,000/test and varied from \$0 to 5,000 in sensitivity analyses Discount rate set at 3% and varied from 0% to 5%	Clinical outcomes (%, 95% Cl) Biochemical recurrence-free at 5 years Genetic test-based care: 68.9 (68.0 to 69.8) 100% adjuvant therapy: 78.3 (77.5 to 79.2) Usual care: 64.0 (63.1 to 65.0) Biochemical recurrence-free at 10 years Genetic test-based care: 47.0 (46.1 to 48.0) 100% adjuvant therapy: 58.3 (57.4 to 59.3) Usual care: 42.8 (41.9 to 43.8) Metastasis at 5 years Genetic test-based care: 3.46 (3.10 to 3.82) 100% adjuvant therapy: 2.76 (2.44 to 3.08) Usual care: 4.12 (3.73 to 4.51) Metastasis at 10 years Genetic test-based care: 9.48 (8.91 to 10.1) 100% adjuvant therapy: 7.95 (7.42 to 8.48) Usual care: 10.9 (10.3 to 11.5) p < .001 for all comparisons Sensitivity analysis Results showed that model findings were robust based on outputs not being sensitive to changes in inputs	
Ontario 2017 ⁵⁶ Prolaris	<i>Design:</i> Budget impact analysis <i>Comparators:</i> New scenario vs. reference scenario:	Population: Men with newly diagnosed, low- or intermediate-risk, localized prostate cancer in Ontario, Canada Analytic Assumptions:	 (Note: all costs and estimates are in 2016 Canadian Dollars) Base case: Estimated costs associated with test itself: +\$47.9 million in the first 5 years Estimated costs associated with additional physician visits required to interpret the test result: +\$0.7 million 	Moderate Not a cost-effectiveness analysis per se, but an estimate of budget impact of test adoption in one Canadian province (Ontario) Costs expressed in 2016 Canadian dollars

Citation	Design	Population	Main Findings	Risk of Bias
Test	Comparators	Analytic Assumptions		Comments
	Anticipated clinical practice with vs. without the test	Incidence, 8,500 cases per year, 65% of which would occur in men over 65 years of age, with 34% in the low-risk and 46% in the intermediate- risk group Initial treatment patterns assumed that for low-risk men, 61% would have AS/WW, 22% RT, 17% RP, and for intermediate-risk men 34% would have AS/WW, 38% RT, and 28% RP	 Estimated savings due to treatment changes (increased use of active surveillance and decreased use of interventional treatment): -\$7.3 million Publicly funding the CCP test would result in a net budget impact of +\$41.3 million in the first 5 years Sensitivity analysis: Net budget impact was most sensitive to assumptions regarding the uptake of the test and the extent to which the test altered the distribution of treatment from current practice. Three factors had a moderate impact on the results by affecting the size of the target population: Percentage of low- and intermediate-risk tumors among all patients newly diagnosed with prostate cancer Proportion of patients on watchful waiting among those on noninterventional treatment Annual change in prostate cancer incidence: Budget impact was sensitive to the unit cost of the test, but not to the annual costs of prostate cancer treatments or the number of extra physician visits associated with the test, because the cost of the test is much higher than either the extra physician visits or the savings associated with treatment change 	Study authors stated that there was not sufficient data to support a primary economic evaluation of the test for use in men with localized prostate cancer. "based on the results of the clinical evidence review, the effect of the CCP test on patient-important clinical outcomes (e.g., survival or biochemical recurrence) is currently unknown. No prospective studies have been conducted to evaluate these outcomes." ^{56(p. 32)} No specific comparator or competing alternatives aside from current clinical practice patterns included in budget impact analysis Individual-level and patient- level costs not modeled Limited number of sensitivity analyses considered Authors are contractors or employees of Health Quality Ontario, a governmental agency, which also funded the report to support the work of the Ontario

Citation	Design	Population	Main Findings	Risk of Bias
Test	Comparators	Analytic Assumptions		Comments
				Health Technology Advisory Committee and make recommendations to the Ontario Ministry of Health about public funding of services

Gene expression profile testing of cancer tissue: final evidence report

Colon Cancer

Study Citation	Design	Study Population		Results	Risk of Bias
Test	Intervention	Characteristics			Comments
	Comparator				
	Setting				
Renfro 2017 ⁴⁸ Srivastava 2014 ⁴⁹ Oncotype DX	ava Before-after study of one group of patients and the physicians <i>Intervention:</i> Treatment recommendations	Consecutive patient with stage 2A colon = 141 of whom had T3 mismatch repair (MMR-P) tumors an included in the prim analysis Physicians (n = 105)	n cancer, n d resected -proficient nd were nary	 Posttest treatment recommendations, by pretest recommendation: Observation (n = 68) Observation (no change): 79.4% 5-FU monotherapy: 11.8% 5-FU + oxaliplatin: 11.1% 5-FU monotherapy (n = 34) Observation: 70.6% 	High MMR-P status is associated with higher risk of colon cancer recurrence Study did not enroll targeted number of patients with MMR-P
	recommendations after test results <i>Comparator</i> :	patients were surve Patient characteristi	eyed	 5-FU monotherapy (no change): 23.5% 5-FU + oxaliplatin: 5.9% 5-FU + oxaliplatin (n = 39) Observation: 53.8% 	tumors because of higher number of MMR-D tumors than expected in enrollees
	Treatment recommendations for the same patient before the test <i>Setting:</i> 17 academic and community sites in the U.S. that enrolled patients between May 2010 and May 2012	141):5-FU mMean age (range), years63.2 (27- 87)Mean age (range), years63.2 (27- 87)% male58.9% male58.9% white92.2% adenocarcinoma tumor type93.6% WHO tumor grade93.6Low88.7Low88.7High11.3No cha een May 2010Increase		 5-FU monotherapy: 5.1% 5-FU + oxaliplatin (no change): 41.0% Total (n = 141) Observation: 70.2% 5-FU monotherapy: 12.8% 5-FU + oxaliplatin: 17.0% <i>Treatment recommendation change, by recurrence score group:</i> Low (< 30), n = 100 No change: 54.0% Increased intensity: 7.0% Decreased intensity: 39.0% 	Before-after studies do not have an independent control group Slightly different n included in each paper, data primarily abstracted from Srivastava 2014, ⁴⁹ unless otherwise indicated, because it was a more complete reporting of outcomes <i>Renfro 2017</i> ⁴⁸ Study funded by the U.S. NIH under a Clinical Translational Science Award Grant K12

Table 11. Evidence Table: Colon Cancer Observational Studies

Study Citation	Design	Study Population	Results	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
	Setting		Intermediate (30-40), n = 33 • No change: 60.1% • Increased intensity: 18.2% • Decreased intensity: 21.2% High (\geq 41), n = 7 • No change: 57.1% • Increased intensity: 42.8% • Decreased intensity: 0.0% Total (n = 140) • No change: 55.7% • Increased intensity: 11.4% • Decreased intensity: 32.9% Mean change (95% Cl) in patient DCS score after test (n = 139) (data from Renfro 2017 ⁴⁸) Effective decision: -5.74 (-9.21 to -2.27); p = .001	 3 of 5 authors had a financial relationship with test manufacturer <i>Srivastava 2014⁴⁹</i> Study funded by test manufacturer 3 of 13 authors had a financial relationship with test manufacturer
			Effective decision: -5.74 (-9.21 to -2.27); p = .001 Informed: -9.32 (-13.52 to -5.12); p < .001 Support: -4.63 (-7.85 to -1.41); p = .005 Uncertainty: -12.13 (-17.91 to -6.35); p < .001 Values clarity: -8.27 (-12.45 to -4.09); p < .001 Total: -7.88 (-11.25 to -4.50); p < .001 <i>Physician survey results, % of patient decisions (n = 150)</i> More confident in recommendation after ordering test	

Study Citation	Design	Study Population	Results	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
			 Strongly agree: 28.7% Agree: 55.3% Neither agree nor disagree: 12.7% Disagree: 1.3% Strongly disagree: 2.0% Test results provided additional clinically relevant information Strongly agree: 30.7% Agree: 55.3% Neither agree nor disagree: 10.7% Disagree: 1.3% Strongly disagree: 2.0% Test results influenced my treatment decision Strongly agree: 24.0% Agree: 25.4% Neither agree nor disagree: 10.9% Disagree: 1.4% Strongly disagree: 2.2% 	
			 Patient survey results (n = 138) Test results influenced treatment decision Strongly agree: 60.1% Agree: 25.4% Neither agree nor disagree: 10.9% Disagree: 1.4% Strongly disagree: 2.2% 	

Study Citation	Design	Study Population	Results	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
Brenner 2016 ⁵⁰ Oncotype DX	Design: Before-after study of one group of patients and the physicians Intervention: Treatment recommendations after test results Comparator: Treatment recommendations for the same patient before the test Setting: Registry enrolling patients from 7 institutions in Israel	All patients (n = 312) with stage 2 colon cancer who had testing at 7 institutions between January 2011 and May 2012 (n = 269 with MMR-P tumors, test results analyzed) • 39 excluded due to MMR- deficient tumors • 3 excluded due to no pretest treatment recommendation • 1 excluded based on pathology/laboratory exclusion criteria Patient characteristics: Median age (IQR), 68 (60-75) years % 83.6 adenocarcinoma % tumor grade low 18.2 moderate 72.5 high 7.1	Treatment received after test, by pretest recommendation: Observation (n = 148): • Observation (no change): 83.1% • 5-FU only: 15.5% • 5-FU only: 15.5% • 5-FU + oxaliplatin: 1.4% 5-FU only (n = 105): • Observation: 57.1% • 5-FU only (no change): 41.9% • 5-FU + oxaliplatin: 1.0% 5-FU + oxaliplatin: 1.0% 5-FU + oxaliplatin (n = 16) • Observation: 68.8% • 5-FU + oxaliplatin (no change): 0% Total (n = 269): • Observation: 72.1% • 5-FU + oxaliplatin: 1.1% Treatment change by recurrence score group Low (< 30), n = 157 • No change: 62.4% • Increased intensity: 2.6% • Decreased intensity: 35.0% Intermediate (30-40), n = 85 • No change: 69.4% • Increased intensity: 12.9%	 High Preplanned retrospective analysis from a prospectively designed registry Before-after studies do not have an independent control group and changes in treatment decisions may be due to additional or other factors—no analysis for these potential confounders was undertaken Study reported actual treatment received rather than recommended Study funded by test manufacturer's industrial representative in Israel 5 of 19 authors had a financial relationship with test manufacturer

Study Citation	Design	Study Population	Results	Risk of Bias
Test	Intervention	Characteristics		Comments
	Comparator			
	Setting			
			Decreased intensity: 17.7%	
			 High (≥ 41), n = 27 No change: 37.0% Increased intensity: 40.8% Decreased intensity: 22.2% 	
			Overall (n = 269) • No change: 62.1% • Increased intensity: 9.7% • Decreased intensity: 28.3%	

Citation	Design	Population	Main Findings	Risk of Bias
Test	Comparators	Analytic Assumptions		Comments
Alberts 2014 ⁵⁸ Oncotype DX	Design: Cost-effectiveness using decision analysis Treatment recommendation after gene testing. Treatment recommendation before gene testing	 Patients (n = 141) with stage 2, T3, MMR-P colon cancer in clinical practice who had undergone surgery and were eligible for adjuvant chemotherapy Analytic Assumptions: Perspective adopted was that of U.S. third- party payer; costs standardized to 2014 U.S. dollars Model employed patient and clinical decision impact estimates from the Mayo Clinic Cancer Research Consortium (MCCRC) study of fluoropyrimidine monotherapy vs. combination chemotherapy with oxaliplatin (FOLFOX) According to NCCN guidelines, with and without the availability of genetic testing for patients with stage 2 colon cancer, test cost was set at \$3,640 per patient, economic estimates were determined for the end outcomes of recurrence or not recurrence and subsequent survival or death Sensitivity analyses were conducted over the plausible ranges derived from the MCCRC study, the QUASAR chemotherapy RCT, and U.S. incidence and mortality statistics CMS provided drug costs; a 3% discount rate was applied and varied over 1-5% 	Total estimated lifetime costs with the genetic test vs. without (U.S. 2014 dollars, difference): \$103,775 vs. \$104,767, -\$991 Sensitivity analysis found 76.1% of cost savings and QALY improvement found in 96.1% of scenarios with use of genetic test At a QALY threshold of \$50,000, the probability of cost-effectiveness was 95.5%	Moderate Note: this paper draws decision impact estimates derived from the MCCRC study, ^{48,49} which is abstracted in the accompanying evidence table for colon cancer GEP studies Effectiveness and outcome estimates based largely on data from single RCT Some costs, such as lost time and productivity, are not modeled Study funded by test manufacturer 2 of 12 authors had a financial relationship via a contract with test manufacturer

Table 12. Evidence Table: Colon Cancer Economic Studies

Appendix D. Risk of Bias Assessments

Breast Cancer

Table 13. Risk of Bias: Breast Cancer Systematic Reviews and Meta-analyses

Citation	Protocol Described & Used	Search Strategy	Study Selection	Study Characteristics	Quality Assessment of Included Studies	Data Extraction	Study Synthesis	Statistical Methods	Interest Disclosure	Funding Source	Overall Risk of Bias Assessment
Augustovski 2015 ²⁰	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Low
Blok 2018 ²¹	Unclear	Yes	Yes	Unclear	No	Yes	Yes	Yes	Yes	Unclear	Moderate
Scope 2017 ²²	Unclear	Yes	Unclear	Unclear	Unclear	Yes	Yes	NA	Yes	Unclear	High

Table 14. Risk of Bias: Breast Cancer Randomized Controlled Trial

Citation/ Test	Randomization and Allocation Concealment	Intervention	Outcomes	Masking of Investigators and Participants	Masking of Outcome Assessors	Intention to Treat Analysis	Statistical Analysis	Interest Disclosure	Funding Source	Overall Risk of Bias Assessment
Bear 2017 ¹⁸ Oncotype DX	Yes	Yes	Yes	No	No	No	Yes	Yes	No	High
Cardoso 2016 ¹⁹	Unclear	Yes	Yes	No	No	Yes	No	No	No	Moderate
MammaPrint										

Citation/ Test	Subject Selection	Intervention	Control	Outcome	Masked Outcome Assessment	Confounding	Statistical Analysis	Interest Disclosure	Funding Source	Overall Risk of Bias Assessment
Evans 2016 ²³ Oncotype DX	No	Yes	No	Unclear	No	Unclear	Yes	Yes	Yes	High
Friese 2017 ²⁴ Oncotype DX	Unclear	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Moderate
Hequet 2017 ³⁵ Prosigna	Yes	Yes	No	No	No	No	Yes	No	No	High
Jasem 2016 ²⁵ Oncotype DX	Unclear	Yes	No	Yes	No	Yes	Yes	No	Yes	High
Jasem 2017 ²⁶ Oncotype DX	Yes	No	No	Yes	No	Yes	Yes	Yes	Unclear	High
Kuijer 2016 ³³ MammaPrint	Unclear	Yes	Yes	Yes	No	No	Yes	Yes	Yes	High
Kuijer 2017 ³² MammaPrint	Unclear	Yes	No	Yes	No	No	No	No	No	High
Loncaster 2017 ²⁷ Oncotype DX	No	Yes	No	Yes	No	No	Yes	No	No	High
O'Neill 2017 ²⁸ Oncotype DX	Unclear	Yes	Yes	Yes	No	Yes	Yes	No	Yes	High
Parsons 2016 ²⁹ Oncotype DX	Yes	Yes	No	Yes	No	No	No	Yes	Yes	High
Pestalozzi 2017 ³⁸ Oncotype DX	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	High
Ray 2016 ³¹ Oncotype DX	Yes	Yes	No	Yes	No	Yes	Unclear	No	Yes	Moderate
Sanft 2015 ³⁷ BCI	Unclear	No	Yes	Yes	No	No	No	No	Unclear	High
Tsai 2017 ³⁴ MammaPrint	Unclear	Yes	Yes	Yes	No	No	No	No	No	High

Table 15. Risk of Bias: Breast Cancer Observational Studies

Citation/ Test	Subject Selection	Intervention	Control	Outcome	Masked Outcome Assessment	Confounding	Statistical Analysis	Interest Disclosure	Funding Source	Overall Risk of Bias Assessment
Wuerstlein 2016 ³⁶ Prosigna	Yes	Yes	No	Yes	No	No	No	No	No	High

Table 16. Risk of Bias: Breast Cancer Economic Studies

Part 1

Citation	Target Population	Perspective	Time Horizon	Discount Rate	Comparators	Modeling	Effectiveness
Gustavsen 2014 ⁵²	Yes	Yes	Yes	No	Yes	No	No
Hall 2017 ⁵¹	Yes	Yes	Yes	Yes	Yes	Yes	Unclear

Part 2

Citation	Outcomes	Resource Use/Costs	Uncertainty	Results	Interest Disclosure	Funding Source	Overall Risk of Bias Assessment
Gustavsen 2014	No	Yes	Unclear	No	No	No	High
Hall 2017	Yes	Yes	Yes	Yes	Unclear	Yes	Moderate

Citation	Scope & Purpose	Rigor of Evidence Development	Rigor of Recommendations Development	Stakeholder Involvement	Clarity & Presentation	Applicability & Implementation	Editorial Independence	Overall Assessment Comments
Duffy 2017 ⁶⁴	Good	Poor	Fair	Poor	Good	Fair	Poor	Poor
Harris 2016 ⁵⁹	Good	Good	Good	Good	Good	Good	Fair	Good
Krop 2017 ⁶⁰	Good	Good	Good	Good	Good	Good	Fair	Good
NCCN 2017 ⁶¹	Fair	Fair	Good	Good	Fair	Fair	Fair	Fair
NICE 201362	Good	Good	Good	Good	Good	Good	Fair	Good
Senkus 201563	Fair	Poor	Fair	Poor	Good	Fair	Poor	Poor

Table 17. Risk of Bias: Breast Cancer Guidelines

Prostate Cancer

Table 18. Risk of Bias: Prostate Cancer Observational Studies

Citation/ Test	Subject Selection	Intervention	Control	Outcome	Masked Outcome Assessment	Confounding	Statistical Analysis	Interest Disclosure	Funding Source	Overall Risk of Bias Assessment
Albala 2016 ³⁹ Oncotype DX	Yes	Yes	No	Yes	No	No	N/A	No	No	High
Badani 2015 ⁴⁰ Oncotype DX	No	Yes	Yes	Yes	No	No	N/A	No	No	High
Crawford 2014 ⁴¹ Prolaris	No	Yes	Yes	No	No	No	N/A	No	No	High
Dall'Era, 2015 ⁴² Oncotype DX	No	Yes	Yes	Unclear	Unclear	No	N/A	No	No	High
Eure 2017 ⁴³ Oncotype DX	Yes	Yes	No	Unclear	No	No	N/A	No	No	High
Gore 2017 ⁴⁴ Decipher	Yes	Yes	Yes	Unclear	No	No	N/A	No	No	High
Michalopoulos 2014 ⁴⁵ Decipher	No	Yes	Yes	No	No	No	N/A	No	No	High
Shore 2016 ⁴⁶ Prolaris	No	Yes	Yes	Yes	No	No	N/A	No	No	High
Table 19. Risk of Bias: Prostate Cancer Economic Studies

Part 1

Citation	Target Population	Perspective	Time Horizon	Discount Rate	Comparators	Modeling	Effectiveness
Lobo 2017 ⁵⁵	Yes	No	Yes	Yes	Yes	Yes	No
Ontario 2017 ⁵⁶	Yes	Yes	Yes	N/A	Unclear	Yes	Yes

Part 2

Citation	Outcomes	Resource Use/Costs	Uncertainty	Results	Interest Disclosure	3	Overall Risk of Bias Assessment
Lobo 2017	Yes	Yes	Yes	Yes	No	No	High
Ontario 2017	Yes	No	Yes	Yes	Yes	Yes	Moderate

Table 20. Risk of Bias: Prostate Cancer Guidelines

Citation	Scope & Purpose	Rigor of Evidence Development	Rigor of Recommendations Development	Stakeholder Involvement	Clarity & Presentation	Applicability & Implementation	Editorial Independence	Overall Assessment Comments
NCCN 201765	Fair	Fair	Good	Good	Fair	Fair	Fair	Fair
Sanda 2017 ⁶⁶	Good	Good	Good	Good	Good	Good	Fair	Good

WA – Health Technology Assessment

Colon Cancer

Citation/ Test	Subject Selection	Intervention	Control	Outcome	Masked Outcome Assessment	Confounding	Statistical Analysis	Interest Disclosure	Funding Source	Overall Risk of Bias Assessment
Brenner 2016 ⁵⁰	Yes	Yes	Yes	Yes	No	No	Yes	No	No	High
Oncotype DX										
Renfro 2017 ⁴⁸	Yes	Yes	Yes	No	No	No	No	No	No	High
Srivastava 2014 ⁴⁹										
Oncotype DX										

Table 21. Risk of Bias: Colon Cancer Observational Studies

Table 22. Risk of Bias: Colon Cancer Economic Studies

Part 1

Citation	Target Population	Perspective	Time Horizon	Discount Rate	Comparators	Modeling	Effectiveness
Alberts 2014 ⁵⁸	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Part 2

Citation	Outcomes	Resource Use/Costs	Uncertainty	Results	Interest Disclosure	Funding Source	Overall Risk of Bias Assessment
Alberts 2014	No	No	Yes	Yes	Yes	No	Moderate

Citation	Scope & Purpose	Rigor of Evidence Development	Rigor of Recommendations Development	Stakeholder Involvement	Clarity & Presentation	Applicability & Implementation	Editorial Independence	Overall Assessment Comments		
NCCN 2017 ⁶⁸	Fair	Fair	Good	Good	Fair	Fair	Fair	Fair		
Sepulveda 2017 ⁶⁷	Good	Good	Good	Fair	Good	Good	Fair	Good		
Van Cutsem 2016 ⁶⁹	Fair	Poor	Poor	Fair	Fair	Fair	Poor	Poor		

Table 23. Risk of Bias: Colon Cancer Guidelines

Multiple Myeloma

Table 24. Risk of Bias: Multiple Myeloma Guidelines

Citation	Scope & Purpose	Rigor of Evidence Development	Rigor of Recommendations Development	Stakeholder Involvement	Clarity & Presentation	Applicability & Implementation	Editorial Independence	Overall Assessment Comments
Moreau 2017 ⁷¹	Fair	Poor	Fair	Poor	Fair	Fair	Poor	Fair
NCCN 2017 ⁷⁰	Fair	Poor	Good	Fair	Fair	Fair	Fair	Fair

Appendix E. GRADE Profile Ratings

Table 25. GRADE Quality of Evidence for Breast Cancer

Number of Studies, by Test and Study Design	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Other Factors*	Comments	Overall Quality of Evidence Rating
Outcome: Clinical Utili	ty—mortality o	r morbidity						
Oncotype DX (1 RCT) MammaPrint (1 RCT)	High for Oncotype DX; Moderate for MammaPrint	Not evaluated (1 study for each test)	Indirectness serious for Oncotype DX and not serious for MammaPrint	N/A (only one study for each test)	Not evaluated		Oncotype DX RCT used clinical response as a surrogate/intermediate outcome measure; MammaPrint RCT reported survival	Oncotype DX Very low •০০০ MammaPrint Moderate •••
Outcome: Clinical Utili	ty—patient mar	nagement decisio	ons					
Oncotype DX (3 SRs with largely overlapping primary studies, total k = 38; and k = 10 additional studies, 1 RCT and 9 observational studies) MammaPrint (2 SRs with overlapping primary studies, total k = 7; 4 additional studies, 1 RCT and 3 observational studies) Prosigna	Moderate to High for Oncotype DX; Moderate to High for MammaPrint; High for all other tests	Results overall consistent for these outcomes	Minor indirectness; One meta- analysis for Oncotype DX found high I ² for some comparisons	Precision related to size of study, as expected, but not serious	Evaluated only for Oncotype DX in one meta-analysis and was not detected			Oncotype DX Moderate ••• MammaPrint Low •• Prosigna, EndoPredict, BCI, and Mammostrat Very low •፡ •

Number of Studies, by Test and Study Design	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Other Factors*	Comments	Overall Quality of Evidence Rating
 (1 SR with k = 1 study, and 2 additional observational studies) BCI (0 SRs and 1 additional observational study) EndoPredict (1 SR with k = 1 study, and 0 additional studies) BCI (0 SRs and 1 additional observational study) Mammostrat (1 SR with k = 1 study; and 0 additional study) 								
Outcome: Clinical Utili	ty—quality of li	fe						
Oncotype DX (2 additional observational studies) Prosigna (2 additional observational studies) BCI (1 additional observational study)	High	Results generally consistent across studies that reported the outcomes	Serious	Not evaluated	Not evaluated		No studies reported on true QoL measures; but some studies did report on pre- and posttest levels of decisional conflict, anxiety, and distress	Oncotype DX, Prosigna, and BCI Very low • • • • Other tests Not applicable (no eligible studies)

Number of Studies, by Test and Study Design		Inconsistency	Indirectness	Imprecision	Publication Bias	Other Factors*	Comments	Overall Quality of Evidence Rating
Outcome: Harms								
Oncotype DX (1 additional observational study)	High	N/A	N/A	N/A	Not evaluated		No studies reported harms related to false- positive or false-negative test results. One study reported similar pre- and posttest levels of patient-reported distress with the use of Oncotype DX	Oncotype DX Very low • Other tests Not applicable (no eligible studies)
Outcome: Cost-Effectiv	veness and Econ	omic Outcomes						
1 SR with k = 44 studies (n = NA), and 2 additional studies (n = NA)	Moderate (SR) to High (additional studies)	Serious	Serious	Serious	Not evaluated			Oncotype DX and MammaPrint: Low •• EndoPredict, Mammostrat, Prosigna, BCI: Very low • •

Note. *Other factors for upgrading of findings from observational studies: large effect size, dose-response gradient, plausible residual confounding. Abbreviations. SR: systematic review; N/A: not applicable

Number of Studies (k), and Subjects (n), by Test and Study Design	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Other Factors*	Comments	Overall Quality of Evidence Rating
Outcome: Clinical	Utility—mo	ortality or morbi	dity					
0							No studies reported outcome	Not applicable (no eligible studies)
Outcome: Clinical	Utility—pa	tient manageme	nt decisions					
Oncotype DX k = 4 n = 659 tested, 1,093 total, and 15 physicians Prolaris k = 1 n = 1,206 Decipher k = 2	High	Results generally consistent across studies	Some studies did not report actual treatment received	Not evaluated	Not evaluated		3 studies on Oncotype DX used historical comparator group and 4 used same patient comparator (1 on Oncotype DX, 1 on Prolaris, and 2 on Decipher) and were consistent in finding decreased treatment intensity after testing	Oncotype DX, Prolaris, and Decipher Very low • • •
k = 2 n = 411, and 15 physicians (all studies are before-after designs)								

Table 26. GRADE Quality of Evidence for Prostate Cancer

WA – Health Technology Assessment

Number of Studies (k), and Subjects (n), by Test and Study Design	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Other Factors*	Comments	Overall Quality of Evidence Rating
Outcome: Clinical	Utility—qu	ality of life						
0								Not applicable (no eligible studies)
Outcome: Harms								
0							No studies reported outcome	Not applicable (no eligible studies)
Outcome: Cost-Ef	fectiveness	and Economic O	utcomes				·	·
Oncotype DX k = 3 n = 80 tested in 1 study, and modeling in 2 studies	High	Not serious	Serious (modeling designs and non-U.S. setting)	Not evaluated	Not evaluated		No comparison across studies possible, given that 1 study reported costs only, 1 was a budget impact analysis, and 1 was a cost- effectiveness analysis; however, there was consistency about test increasing costs in the 2 modeling studies There is some indirectness because the modeling study with moderate risk of bias was done from a Canadian perspective	Oncotype DX Very low • ం Prolaris and Decipher Not applicable (no eligible studies)

Note. *Other factors for upgrading of findings from observational studies: large effect size, dose-response gradient, plausible residual confounding.

Number of Studies (subjects), by Test, and Study Design	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Other Factors*	Comments	Overall Quality of Evidence Rating
Outcome:	Clinical Uti	lity—mortality	or morbidity					
0							No studies reported outcome	Not applicable (no eligible studies)
Outcome:	Clinical Uti	lity—patient ma	anagement de	cisions				
Oncotype DX k = 2 n = 410 (both studies are before- after designs)	High	Not serious	Not serious				Both studies used same patient comparator in before-after analyses and were consistent in finding a decrease in recommended or actual treatment intensity after testing, with larger study reporting actual treatment	Oncotype DX Very low • ColoPrint Not applicable (no eligible studies)
Outcome:	Clinical Uti	lity—quality of	life					
Oncotype DX k = 1 n = 139 (before- after design)	High	Not evaluated (1 study)	Serious	Not evaluated	Not evaluated		1 study reported patient decisional conflict, which may relate to quality of life, but is an incomplete and indirect measure	Oncotype DX Very low • ColoPrint Not applicable (no eligible studies)

Table 27. GRADE Quality of Evidence for Colon Cancer

Number of Studies (subjects), by Test, and Study Design	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Other Factors*	Comments	Overall Quality of Evidence Rating
Outcome:	Harms							
0							No studies reported outcome	Not applicable (no eligible studies)
Outcome:	Cost-Effect	iveness and Ecc	onomic Outcom	nes				
Oncotype DX k = 1 (decision analysis modeling study) n = NA	Moderate	Serious	Serious	Not evaluated	Not evaluated		Underlying evidence used in modeling is scant and sensitivity analysis showed decreased cost-effectiveness approximately 25% of the time, raising questions about consistency. Modeling limited with few outcomes, raising questions regarding applicability to current practice.	Oncotype DX Very low • ColoPrint Not applicable (no eligible studies)

Note. *Other factors for upgrading of findings from observational studies: large effect size, dose-response gradient, plausible residual confounding

Number of Studies (subjects), by Study Design	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Other Factors*	Comments	Overall Quality of Evidence Rating
Outcome: 0	Clinical Ut	ility—mortality	or morbidity					
0							No studies reported outcome	Not applicable (no eligible studies)
Outcome: 0	Outcome: Clinical Utility—patient management decisions							
0							No studies reported outcome	Not applicable (no eligible studies)
Outcome: 0	Clinical Ut	ility—quality of	life					
0							No studies reported outcome	Not applicable (no eligible studies)
Outcome: I	Harms							·
0							No studies reported outcome	Not applicable (no eligible studies)
Outcome: 0	Outcome: Cost-Effectiveness and Economic Outcomes							
0							No studies reported outcome	Not applicable (no eligible studies)

Table 28. GRADE Quality of Evidence for Multiple Myeloma

Note. *Other factors for upgrading of findings from observational studies: large effect size, dose-response gradient, plausible residual confounding

Appendix F. Studies Registered at ClinicalTrials.gov

Breast Cancer

Registered Clinical Trial Number	Title of Study	Study Completion Date Indicated on https://clinicaltrials.gov/
Oncotype DX		
NCT01446185 (France)	Treatment Decision Impact of OncotypeDX [™] in HR+, N- Breast Cancer Patients (SWITCH)	April 2013
NCT02627703 (Canada)	(Oncotype DX [®]) in Estrogen Receptor Positive (ER+) HER-2 Negative (HER 2-) 1-3 Node Positive (pN1) Breast Cancer	December 2017
NCT01423890 (Canada)	A Prospective Cohort Study to Evaluate the Oncotype DX® Test in Early Stage Breast Cancer (ONCOTYPEDX)	March 2014
NCT02347449 (Canada)	The Impact of the Oncotype DX® Breast Cancer Assay on Treatment Decisions in a Canadian Population	September 2016
NCT01926964 (Switzerland)	Adjuvant Treatment Recommendation and Oncotype DX [®] in Early Breast Cancer	February 2015
NCT00899639 (U.S.)	Studying Tumor Tissue Samples From Patients With Early- Stage Breast Cancer	Withdrawn (not "applicable clinical trial")
NCT02236572 (U.S.)	Neoadj ph 2 Al Plus Everolimus in Postmenopausal Women w/ ER Pos/HER2 Neg, Low Risk Score	September 2018
NCT01293032 (U.S.)	Hormone Therapy Or Chemotherapy Before Surgery Based on Gene Expression Analysis in Treating Patients With Breast Cancer	March 2016
NCT01272037 (U.S.)	Tamoxifen Citrate, Letrozole, Anastrozole, or Exemestane With or Without Chemotherapy in Treating Patients With Invasive RxPONDER Breast Cancer	February 2022
NCT00310180 (U.S.)	Hormone Therapy With or Without Combination Chemotherapy in Treating Women Who Have Undergone Surgery for Node-Negative Breast Cancer (The TAILORx Trial) (TAILORx)	December 2017
NCT01619306 (Singapore)	Knowledge and Attitudes of Patients and Healthcare Professionals on a Spectrum of Genetic Tests Relevant to Breast Cancer Patients	December 2014

Registered Clinical Trial Number	Title of Study	Study Completion Date Indicated on https://clinicaltrials.gov/
NCT01560663 (Spain)	Predictors of Response to Neoadjuvant Docetaxel- Carboplatin Chemotherapy for Patients With Stage II and III Triple Negative Breast Cancer	June 2017
NCT03183050 (U.S.)	MEND 2: Making Treatment Decisions Using Genomic Testing (MEND2)	December 2018
NCT02400190 (U.S.)	The IDEA Study (Individualized Decisions for Endocrine Therapy Alone)	March 2026
NCT02476786 (U.S.)	Endocrine Treatment Alone for Elderly Patients With Estrogen Receptor Positive Operable Breast Cancer and Low Recurrence Score	July 2023
NCT00921115 (U.S.)	Anti-Hormone Therapy (With Anastrazole and Fulvestrant) Before Surgery to Treat Postmenopausal Women With Breast Cancer	September 2017
NCT02095184 (U.S.)	GCC 1366: Anti-Proliferative Response to NeoAdjuvant Als in Overweight and Obese Patients	March 2019
NCT02766881 (Canada)	Evaluation of the DCIS Score for Decisions on Radiotherapy in Patients With Low/Intermediate Risk DCIS (DUCHESS)	September 2018
NCT01420185 (U.S.)	Identification of Genes That Predict Local Recurrence in Samples From Patients With Breast Cancer Treated on NSABP-B-28	December 2015
NCT00897065 (U.S.)	Biomarkers in Predicting Response to Tamoxifen and Letrozole in Postmenopausal Women With Primary Breast Cancer Treated on Clinical Trial CAN-NCIC-MA17	June 2006
MammaPrint		
NCT02269813 (Germany, Switzerland, Austria)	PRospective Study to Measure the Impact of MammaPrint on Adjuvant Treatment in Hormone Receptor-positive HER2-negative Breast Cancer Patients (PRIMe)	September 2016
NCT02670577 (U.S.)	Measuring the Impact of MammaPrint on Adjuvant and Neoadjuvant Treatment in Breast Cancer Patients: A Prospective Registry (IMPACt)	August 2017

Registered Clinical Trial Number	Title of Study	Study Completion Date Indicated on https://clinicaltrials.gov/
NCT01479101 (U.S.)	NBRST: Prospective Neo-adjuvant REGISTRY Trial (NBRST)	January 2020
NCT03053193 (U.S.)	MammaPrint, BluePrint, and Full-genome Data Linked With Clinical Data to Evaluate New Gene EXpression Profiles (FLEX)	December 2027
NCT02669745 (U.S.)	The Study of Molecular Risk Panels (70-gene Assay) in Chinese Breast Cancer Patients	November 2017
NCT01501487 (U.S.)	MINT I Multi- Institutional Neo-adjuvant Therapy MammaPrint Project I (MINT)	June 2017
NCT00904566 (Japan)	Validation of 70-gene MammaPrint Profile in Japanese Population	January 2009
NCT01617954 (U.S.)	PRospective Study Of MammaPrint in Patients With an Intermediate Recurrence Score (PROMIS)	December 2015
NCT02209857 (Netherlands)	The Symphony Triple A Study: Using Symphony in Treatment Decisions Concerning Adjuvant Systemic Therapy (Symphony)	May 2016
NCT03080428 (N/A)	Assessment of Genomic Test Impact on Shared Decision of Adjuvant Chemotherapy in ER-positive, Her2-negative Early Breast Cancer (OPTIGEN)	May 2023 <i>Withdrawn</i> due to lack of funding
NCT02689921 (U.S.)	NEOADjuvant Aromatase Inhibitor and Pertuzumab/Trastuzumab for Women With Breast Cancer (NEOADAPT)	April 2018
NCT00433589 (Netherlands)	Genetic Testing or Clinical Assessment in Determining the Need for Chemotherapy in Women With Breast Cancer That Involves No More Than 3 Lymph Nodes (MINDACT)	March 2020
EndoPredict		
NCT02773004 (France)	Prospective Study Assessing EndoPredict® Genomic Test Impact on Shared Decision of Adjuvant Chemotherapy in Patients With ER-positive, Her2-negative Early Breast Cancer (ADENDOM)	December 2017

Registered Clinical Trial Number	Title of Study	Study Completion Date Indicated on https://clinicaltrials.gov/
NCT03080428 (N/A)	Assessment of Genomic Test Impact on Shared Decision of Adjuvant Chemotherapy in ER-positive, Her2-negative Early Breast Cancer (OPTIGEN)	May 2023 <i>withdrawn</i> due to lack of funding
NCT01805271 (France)	Safety Study of Adding Everolimus to Adjuvant Hormone Therapy in Women With High Risk of Relapse, ER+ and HER2- Primary Breast Cancer, Free of Disease After Receiving at Least One Year of Adjuvant Hormone Therapy	June 2031
Prosigna (PAM	50)	
NCT01899079 (Spain)	A Prospective Observational Study of Clinical Outcomes for the NanoString® Technologies Prosigna Gene Signature Assay	July 2014
NCT02653755 (U.S.)	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	June 2023
NCT03197805 (France)	Assessment of the Impact of RNA Genomic Profile on Treatment Decision-making in HER2 Equivocal Breast Cancer Patients (EQUIVOK)	April 2019
NCT02889874 (Australia, New Zealand)	EXamining PErsonalised Radiation Therapy for Low-risk Early Breast Cancer (EXPERT)	December 2023
NCT01560663 (Spain)	Predictors of Response to Neoadjuvant Docetaxel- Carboplatin Chemotherapy for Patients With Stage II and III Triple Negative Breast Cancer	June 2017
NCT00991263 (U.S.)	Study of Tissue Samples From Women Treated With Paclitaxel for Breast Cancer on Clinical Trial CALGB-9344 or CALGB-9741	February 2012
Breast Cancer II	ndex (BCI)	
NCT02057029 (U.S.)	Assessment of the Decision-making Impact of the Breast Cancer Index in Recommending Extended Adjuvant Endocrine Therapy for Patients With Early Stage ER- positive Breast Cancer	June 2015

Registered Clinical Trial Number	Title of Study	Study Completion Date Indicated on https://clinicaltrials.gov/
Mammostrat		
NCT02067416 (U.S.)	PREDATOR: Neoadjuvant Gene Prediction for Breast Cancer	June 2014 <i>Terminated</i> when funding agent withdrew funding

Note. N/A = *not available*

Prostate Cancer

Registered Clinical Trial Number	Title of Study	Study Completion Date Indicated on https://clinicaltrials.gov/
Decipher		
NCT02723734 (U.S.)	Validation Study on the Impact of Decipher® Testing - VANDAAM Study (VANDAAM)	May 2021
NCT02080689 (U.S.)	Prospective Clinical Utility Study to Assess the Impact of Decipher on Treatment Decisions After Surgery (PRO-IMPACT)	February 2017
NCT02034812 (U.S.)	Decision Impact Study to Measure the Influence of DECIPHER on Treatment Recommendations for Radiation Oncologists (DECIDE3)	July 2014
NCT02020876 (U.S.)	Decision Impact Study to Measure the Influence of DECIPHER on Treatment Recommendations (DECIDE)	December 2013
NCT02034825 (U.S.)	Multi-site Decision Impact Study for Decipher (ASSESS-D)	January 2015
NCT03237026 (Taiwan)	Decipher the Biology of Lethal Prostate Cancer—Using Urine Metabolomics Profiling to Search for Predictive and Prognostic Markers for Treatment Outcomes in Prostate Cancer Patients	July 2022
NCT03237702 (Taiwan)	Decipher Lethal Prostate Cancer Biology - Urine Metabolomics	April 2020
NCT02609269 (U.S.)	Decipher Genomics Resource Information Database (GRID®)	December 2020
NCT02783950 (U.S.)	Genomics in Michigan Impacting Observation or Radiation (G-MINOR)	November 2021

Registered Clinical Trial Number	Title of Study	Study Completion Date Indicated on https://clinicaltrials.gov/			
NCT02721979 (U.S.)	ARN-509 in Treating Patients With Prostate Cancer Who Are in Active Surveillance	July 2020			
Prolaris					
NCT03290508 (U.S.)	Long-term Study to Evaluate and Clinical Outcomes in Patients With Favorable Intermediate Risk Localized Prostate Cancer	September 2027			
NCT03152448 (U.S. Veterans Affairs Health System)	Prospective Prolaris Value and Efficacy (P-PROVE)	January 2024			
NCT02209584 (U.S.)	Open Registry Measuring Impact of Genomic Testing on Treatment Decision After Biopsy in Newly Diagnosed Prostate Cancer Patients (PROCEDE-2000)	December 2015			
NCT02454595 (U.S.)	Registry to Measure the Impact of Adding Genomic Testing	July 2017 <i>Terminated</i> due to lack of enrollment			
NCT01954004 (U.S.)	Open Registry Measuring Impact of Genomic Testing on Treatment of Prostate Cancer Patients	September 2014			
Oncotype DX prostate					
NCT02668276 (U.S.)	Engaging Newly Diagnosed Men About Cancer Treatment Options (ENACT)	August 2022			

Colon Cancer

Registered Clinical Trial Number	Title of Study	Study Completion Date Indicated on https://clinicaltrials.gov/
ColoPrint		
NCT00903565 (U.S.)	A Prospective Study for the Assessment of Recurrence Risk in Stage II Colon Cancer Patients Using ColoPrint (PARSC)	December 2019
Oncotype DX (Colon	
NCT01479894 (U.S.)	Genetic Analysis to Evaluate the Racial Difference in the Outcome of Patients With Colon Cancer	January 2015

Multiple Myeloma

Registered Clinical Trial Number	Title of Study	Study Completion Date Indicated on https://clinicaltrials.gov/
My Prognostic Risk Signature MyPRS		
NCT01313897 (U.S.)	UARK 2010-35, A Study of Expanded Natural Killer Cell Therapy for Multiple Myeloma (NK2010-35)	October 2016
SKY92-EMC92		
NCT02911571 (U.S.)	PRospective Multiple Myeloma Impact Study (PROMMIS)	December 2024

Appendix G. Detailed Clinical Practice Guidelines

American Society of Clinical Oncology Guidelines for Breast Cancer^{59,60}

Oncotype DX

- If a patient has ER/PR-positive, HER2-negative (LN-negative) breast cancer, the clinician can use Oncotype DX to guide decisions on adjuvant systemic chemotherapy.
- If a patient has ER/PR-positive, HER2-negative (LN-positive) breast cancer, the clinician should not use Oncotype DX to guide decisions on adjuvant systemic chemotherapy.
- If a patient has HER2-positive breast cancer or triple-negative breast cancer, the clinician should not use Oncotype DX to guide decisions on adjuvant systemic therapy.

MammaPrint

- If a patient has ER/PR-positive, HER2-negative, LN-negative breast cancer, the MammaPrint assay may be used in those with high clinical risk per MINDACT categorization to inform decisions on withholding adjuvant systemic chemotherapy because of its ability to identify a good-prognosis population with potentially limited chemotherapy benefit.
- If a patient has ER/PR-positive, HER2-negative, LN-negative breast cancer, the MammaPrint assay should not be used in those with low clinical risk per MINDACT categorization to inform decisions on withholding adjuvant systemic chemotherapy, because women in the low clinical risk category had excellent outcomes and did not appear to benefit from chemotherapy even with a genomic high-risk cancer.
- If a patient has ER/PR-positive, HER2-negative, LN-positive, breast cancer, the MammaPrint assay may be used in patients with one to three positive nodes and at high clinical risk per MINDACT categorization to inform decisions on withholding adjuvant systemic chemotherapy because of its ability to identify a good-prognosis population with potentially limited chemotherapy benefit. However, such patients should be informed that a benefit of chemotherapy cannot be excluded, particularly in patients with greater than one involved lymph node.
- If a patient has ER/PR-positive, HER2-negative, LN-positive, breast cancer, the MammaPrint assay should not be used in patients with one to three positive nodes and at low clinical risk per MINDACT categorization to inform decisions on withholding adjuvant systemic chemotherapy. There are insufficient data on the clinical utility of MammaPrint in this specific patient population.
- If a patient has HER2-positive breast cancer, the clinician should not use the MammaPrint assay to guide decisions on adjuvant systemic therapy. Additional studies are required to address the role of MammaPrint in patients with this tumor subtype who are also receiving HER2-targeted therapy.
- If a patient has ER/PR-negative and HER2-negative (triple negative) breast cancer, the clinician should not use the MammaPrint assay to guide decisions on adjuvant systemic chemotherapy.

EndoPredict

- If a patient has ER/PR-positive, HER2-negative (LN-negative) breast cancer, the clinician can use EndoPredict to guide decisions on adjuvant systemic chemotherapy.
- If a patient has ER/PR-positive, HER2-negative (LN-positive) breast cancer, the clinician should not use EndoPredict to guide decisions on adjuvant systemic chemotherapy.
- If a patient has HER2-positive breast cancer or triple-negative breast cancer, the clinician should not use EndoPredict to guide decisions on adjuvant systemic therapy.

Prosigna (PAM50 risk of recurrence score)

- If a patient has ER/PR-positive, HER2-negative (LN-negative) breast cancer, the clinician can use the Prosigna score, in conjunction with other clinicopathologic variables, to guide decisions on adjuvant systemic therapy.
- If a patient has ER/PR-positive, HER2-negative (LN-positive) breast cancer, the clinician should not use Prosigna to guide decisions on adjuvant systemic therapy.
- If a patient has HER2-positive breast cancer or triple-negative breast cancer, the clinician should not use Prosigna to guide decisions on adjuvant systemic therapy.

Breast Cancer Index

- If a patient has ER/PR-positive, HER2-negative (LN-negative) breast cancer, the clinician can use BCI to guide decisions on adjuvant systemic therapy.
- If a patient has ER/PR-positive, HER2-negative (LN-positive) breast cancer, the clinician should not use BCI to guide decisions on adjuvant systemic therapy.
- If a patient has HER2-positive breast cancer or triple-negative breast cancer, the clinician should not use BCI to guide decisions on adjuvant systemic therapy.

Mammostrat

- If a patient has ER/PR-positive, HER2-negative (LN-positive or LN-negative) breast cancer, the clinician should not use the five-protein assay to guide decisions on adjuvant systemic therapy.
- If a patient has HER2-positive breast cancer or triple-negative breast cancer, the clinician should not use the five-protein assay (Mammostrat) to guide decisions on adjuvant systemic therapy.

European Group on Tumor Markers Guidelines on the Use of Biomarkers in Breast Cancer⁶⁴

Oncotype DX

 Oncotype DX may provide added value to established factors for determining prognosis and aiding decision-making with respect to administration of adjuvant chemotherapy in newly diagnosed breast cancer patients with lymph LN-negative invasive disease that is ER-positive but HER2-negative (Level of evidence [LOE], IB; strength of recommendation [SOR], A). In addition, Oncotype DX may be considered for identifying HER2-negative, ER-positive patients with 1 to 3 involved lymph nodes for treatment with adjuvant chemotherapy (LOE, IB; SOR, A). • Before performing the test, any biopsy cavity in the cancer specimen should be removed by manual dissection.

MammaPrint

 MammaPrint may be used for determining prognosis and guiding decision-making with respect to the administration of adjuvant chemotherapy in patients with newly diagnosed invasive breast cancer that is LN-negative or LN-positive (1 to 3 metastatic nodes). Patients at high risk based on clinical and pathological criteria but at low risk based on MammaPrint may be the candidates for avoiding having to receive adjuvant chemotherapy (LOE, IA; SOR, A).

EndoPredict

 In combination with established clinical and pathological factors, EndoPredict may be used for predicting outcome and aiding adjuvant therapy decision-making in hormone receptor-positive, HER2-negative patients that are either LN-negative or LN-positive (1 to 3 metastatic nodes). (LOE, IB; SOR, A).

Prosigna

 In combination with established clinical and pathological factors, Prosigna may be used for predicting outcome and aiding adjuvant therapy decision-making in hormone receptor-positive, HER2-negative patients that are either lymph LN-negative or lymph LN-positive (1 to 3 metastatic nodes). (LOE IB; SOR, A).

BCI

• In combination with established clinical and pathological factors, BCI may be used for predicting outcome and aiding adjuvant therapy decision-making in lymph LN-negative, hormone receptor-positive and HER2-negative patients. (LOE, IB; SOR, A).

Appendix H. Excluded Studies

See attachment for excluded studies.