

Gene Expression Profiling for Guiding Adjuvant Chemotherapy Decisions in Women with Early Breast Cancer

An Evidence-Based and Economic Analysis

*Presented to the Ontario Health Technology
Advisory Committee in August, 2010*

December 2010



Medical Advisory Secretariat
Ministry of Health and Long-Term Care

Suggested Citation

This report should be cited as follows:

Medical Advisory Secretariat. Gene expression profiling for guiding adjuvant chemotherapy decisions in women with early breast cancer: an evidence-based and economic analysis. *Ont Health Technol Assess Ser* [Internet]. 2010 December [cited YYYY MM DD]; 10(23) 1-57. Available from: http://www.health.gov.on.ca/english/providers/program/mas/tech/reviews/pdf/gep_20101213.pdf

Permission Requests

All inquiries regarding permission to reproduce any content in the *Ontario Health Technology Assessment Series* should be directed to MASinfo.moh@ontario.ca.

How to Obtain Issues in the Ontario Health Technology Assessment Series

All reports in the *Ontario Health Technology Assessment Series* are freely available in PDF format at the following URL: www.health.gov.on.ca/ohtas.

Print copies can be obtained by contacting MASinfo.moh@ontario.ca.

Conflict of Interest Statement

All analyses in the Ontario Health Technology Assessment Series are impartial and subject to a systematic evidence-based assessment process. There are no competing interests or conflicts of interest to declare.

Peer Review

All Medical Advisory Secretariat analyses are subject to external expert peer review. Additionally, the public consultation process is also available to individuals wishing to comment on an analysis prior to finalization. For more information, please visit http://www.health.gov.on.ca/english/providers/program/ohtac/public_engage_overview.html.

Contact Information

The Medical Advisory Secretariat
Ministry of Health and Long-Term Care
20 Dundas Street West, 10th floor
Toronto, Ontario
CANADA
M5G 2C2
Email: MASinfo.moh@ontario.ca
Telephone: 416-314-1092

ISSN 1915-7398 (Online)
ISBN 978-1-4435-4830-4 (PDF)

About the Medical Advisory Secretariat

The Medical Advisory Secretariat is part of the Ontario Ministry of Health and Long-Term Care. The mandate of the Medical Advisory Secretariat is to provide evidence-based policy advice on the coordinated uptake of health services and new health technologies in Ontario to the Ministry of Health and Long-Term Care and to the healthcare system. The aim is to ensure that residents of Ontario have access to the best available new health technologies that will improve patient outcomes.

The Medical Advisory Secretariat also provides a secretariat function and evidence-based health technology policy analysis for review by the Ontario Health Technology Advisory Committee (OHTAC).

The Medical Advisory Secretariat conducts systematic reviews of scientific evidence and consultations with experts in the health care services community to produce the *Ontario Health Technology Assessment Series*.

About the Ontario Health Technology Assessment Series

To conduct its comprehensive analyses, the Medical Advisory Secretariat systematically reviews available scientific literature, collaborates with partners across relevant government branches, and consults with clinical and other external experts and manufacturers, and solicits any necessary advice to gather information. The Medical Advisory Secretariat makes every effort to ensure that all relevant research, nationally and internationally, is included in the systematic literature reviews conducted.

The information gathered is the foundation of the evidence to determine if a technology is effective and safe for use in a particular clinical population or setting. Information is collected to understand how a new technology fits within current practice and treatment alternatives. Details of the technology's diffusion into current practice and input from practising medical experts and industry add important information to the review of the provision and delivery of the health technology in Ontario. Information concerning the health benefits; economic and human resources; and ethical, regulatory, social and legal issues relating to the technology assist policy makers to make timely and relevant decisions to optimize patient outcomes.

If you are aware of any current additional evidence to inform an existing evidence-based analysis, please contact the Medical Advisory Secretariat: MASinfo.moh@ontario.ca. The public consultation process is also available to individuals wishing to comment on an analysis prior to publication. For more information, please visit http://www.health.gov.on.ca/english/providers/program/ohtac/public_engage_overview.html.

Disclaimer

This evidence-based analysis was prepared by the Medical Advisory Secretariat, Ontario Ministry of Health and Long-Term Care, for the Ontario Health Technology Advisory Committee and developed from analysis, interpretation, and comparison of scientific research and/or technology assessments conducted by other organizations. It also incorporates, when available, Ontario data, and information provided by experts and applicants to the Medical Advisory Secretariat to inform the analysis. While every effort has been made to reflect all scientific research available, this document may not fully do so. Additionally, other relevant scientific findings may have been reported since completion of the review. This evidence-based analysis is current to the date of the literature review specified in the methods section. This analysis may be superseded by an updated publication on the same topic. Please check the Medical Advisory Secretariat Website for a list of all evidence-based analyses: <http://www.health.gov.on.ca/ohtas>.

Table of Contents

ABBREVIATIONS	6
EXECUTIVE SUMMARY	7
Objective	7
Clinical Need: Condition and Target Population	7
Technology of Concern	7
Research Questions	8
Research Methods	8
Literature Search	8
<i>Search Strategy</i>	8
<i>Inclusion Criteria</i>	8
<i>Exclusion Criteria</i>	8
Outcomes of Interest	9
Summary of Findings	9
BACKGROUND	11
Objective of Analysis	11
Definitions	11
Clinical Need and Target Population	11
Description of Breast Cancer	12
Prevalence and Incidence	12
Gene Expression Profiling	13
Oncotype-DX	13
Status in Ontario	13
EVIDENCE-BASED ANALYSIS	14
Research Questions	14
Key Questions	14
Research Methods	14
Literature Search	14
<i>Search Strategy</i>	14
<i>Inclusion Criteria</i>	15
<i>Exclusion Criteria</i>	15
<i>Outcomes of Interest</i>	15
Expert Panel	15
Statistical Analysis	15
Quality of Evidence	16
Results of Evidence-Based Analysis	16
Literature Search Results	16
Key Question 1. What is the laboratory performance of Oncotype-DX?	17
<i>a. How reliable is Oncotype-DX (i.e., how repeatable and reproducible is Oncotype-DX)?</i>	17
<i>b. How often does Oncotype-DX fail to give a useable result?</i>	17
Key Question 2. What is the prognostic value of Oncotype-DX?	18
Key Question 3. What is the predictive value of Oncotype-DX?	23
Key Question 4. How does Oncotype-DX compare to other known predictors of risk such as Adjuvant!	26
Key Question 5. How does Oncotype-DX impact patient quality of life and clinical/patient decision-making?	27
Societal and Ethical Considerations	28
Limitations	29

GRADE Quality of Evidence	30
External Levels of Evidence	31
Conclusions	32
ECONOMIC ANALYSIS	34
Study Question	34
Economic Analysis Overview	34
Economic Literature Review	34
Target Population	35
Perspective	35
Comparators	35
Time Horizon	36
Discounting	36
Model Structure	36
Outcomes	36
Resource Use and Costs	36
Parameter Estimates	37
Sensitivity Analysis	37
Results and Discussion	37
Budget Impact Analysis	38
APPENDICES	39
Appendix 1: Literature Search Strategies	39
Appendix 2: Laboratory Performance of Oncotype-DX	41
Appendix 3: Prognostic Value of Oncotype-DX	43
Appendix 4: GRADE Assessment of Quality of Evidence	49
Appendix 5: Subgroup Effect Checklist	51
Criteria	51
Results	52
REFERENCES	53

Abbreviations

ATAC	Arimidex, Tamoxifen, Alone or in Combination trial
CI	Confidence interval
CEA	Cost-effectiveness analysis
CUA	Cost-utility analysis
CAD	Canadian dollars
DFS	Disease-free survival
DR	Disease recurrence
ER	Estrogen receptor
FFPE	Formalin-fixed paraffin-embedded
DCIS	Ductal carcinoma <i>in situ</i>
GEP	Gene expression profile
HER	Human Epidermal growth factor Receptor
HR	Hazard ratio
KM	Kaplan Meier
LN	Lymph node
MAS	Medical Advisory Secretariat
MOHLTC	Ontario Ministry of Health and Long-Term Care
NR	Not reported
NSABP	National Surgical Adjuvant Breast and Bowel Project
OHTAC	Ontario Health Technology Assessment Committee
PR	Progesterone receptor
QALY	Quality-adjusted life year
RCT	Randomized controlled trial
RS	Recurrence score
RT-PCR	Reverse transcription polymerase chain reaction
SD	Standard deviation

Executive Summary

In February 2010, the Medical Advisory Secretariat (MAS) began work on evidence-based reviews of published literature surrounding three pharmacogenomic tests. This project came about when Cancer Care Ontario (CCO) asked MAS to provide evidence-based analyses on the effectiveness and cost-effectiveness of three oncology pharmacogenomic tests currently in use in Ontario.

Evidence-based analyses have been prepared for each of these technologies. These have been completed in conjunction with internal and external stakeholders, including a Provincial Expert Panel on Pharmacogenomics (PEPP). Within the PEPP, subgroup committees were developed for each disease area. For each technology, an economic analysis was also completed by the Toronto Health Economics and Technology Assessment Collaborative (THETA) and is summarized within the reports.

The following reports can be publicly accessed at the MAS website at: www.health.gov.on.ca/mas or at www.health.gov.on.ca/english/providers/program/mas/mas_about.html

1. Gene Expression Profiling for Guiding Adjuvant Chemotherapy Decisions in Women with Early Breast Cancer: An Evidence-Based and Economic Analysis
2. Epidermal Growth Factor Receptor Mutation (EGFR) Testing for Prediction of Response to EGFR-Targeting Tyrosine Kinase Inhibitor (TKI) Drugs in Patients with Advanced Non-Small-Cell Lung Cancer: An Evidence-Based and Economic Analysis
3. K-RAS testing in Treatment Decisions for Advanced Colorectal Cancer: an Evidence-Based and Economic Analysis

Objective

To review and synthesize the available evidence regarding the laboratory performance, prognostic value, and predictive value of Oncotype-DX for the target population.

Clinical Need: Condition and Target Population

The target population of this review is women with newly diagnosed early stage (stage I–IIIa) invasive breast cancer that is estrogen-receptor (ER) positive and/or progesterone-receptor (PR) positive. Much of this review, however, is relevant for women with early stage (I and II) invasive breast cancer that is specifically ER positive, lymph node (LN) negative and human epidermal growth factor receptor 2 (HER-2/*neu*) negative. This refined population represents an estimated incident population of 3,315 new breast cancers in Ontario (according to 2007 data). Currently it is estimated that only 15% of these women will develop a distant metastasis at 10 years; however, a far great proportion currently receive adjuvant chemotherapy, suggesting that more women are being treated with chemotherapy than can benefit. There is therefore a need to develop better prognostic and predictive tools to improve the selection of women that may benefit from adjuvant chemotherapy.

Technology of Concern

The Oncotype-DX Breast Cancer Assay (Genomic Health, Redwood City, CA) quantifies gene expression for 21 genes in breast cancer tissue by performing reverse transcription polymerase chain reaction (RT-PCR) on formalin-fixed paraffin-embedded (FFPE) tumour blocks that are obtained during initial surgery (lumpectomy, mastectomy, or core biopsy) of women with early breast cancer that is newly diagnosed. The panel of 21 genes include genes associated with tumour proliferation and invasion, as well

as other genes related to HER-2/*neu* expression, ER expression, and progesterone receptor (PR) expression.

Research Questions

1. What is the laboratory performance of Oncotype-DX?
 - a. How reliable is Oncotype-DX (i.e., how repeatable and reproducible is Oncotype-DX)?
 - b. How often does Oncotype-DX fail to give a useable result?
2. What is the prognostic value of Oncotype-DX?*
3. What is the predictive value of Oncotype-DX?*
4. How does Oncotype-DX compare to other known predictors of risk such as Adjuvant! Online?
5. How does Oncotype-DX impact patient quality of life and clinical/patient decision-making?

*Key Questions #2 and #3 are the primary questions of this analysis.

Research Methods

Literature Search

Search Strategy

A literature search was performed on March 19th, 2010 using OVID MEDLINE, MEDLINE In-Process and Other Non-Indexed Citations, EMBASE, the Cumulative Index to Nursing & Allied Health Literature (CINAHL), the Cochrane Library, and the International Agency for Health Technology Assessment (INAHTA) for studies published from January 1st, 2006 to March 19th, 2010. A starting search date of January 1st, 2006 was because a comprehensive systematic review of Oncotype-DX was identified in preliminary literature searching. This systematic review, by Marchionni et al. (2008), included literature up to January 1st, 2007. All studies identified in the review by Marchionni et al. as well as those identified in updated literature searching were used to form the evidentiary base of this review. The quality of the overall body of evidence was identified as high, moderate, low or very low according to GRADE methodology.

Inclusion Criteria

- Any observational trial, controlled clinical trial, randomized controlled trial (RCT), meta-analysis or systematic review that reported on the laboratory performance, prognostic value and/or predictive value of Oncotype-DX testing, or other outcome relevant to the Key Questions, specific to the target population was included.

Exclusion Criteria

- Studies that did not report original data or original data analysis,
- Studies published in a language other than English,
- Studies reported only in abstract or as poster presentations (such publications were not sought nor included in this review since the MAS does not generally consider evidence that is not subject to peer review nor does the MAS consider evidence that lacks detailed description of methodology).

Outcomes of Interest

Outcomes of interest varied depending on the Key Question. For the Key Questions of prognostic and predictive value (Key Questions #2 and #3), the prospectively defined *primary outcome* was risk of 10-year distant recurrence. The prospectively defined *secondary outcome* was 10-year death due to any cause (i.e., overall survival). All additional outcomes such as risk of locoregional recurrence or disease-free survival (DFS) were not prospectively determined for this review but were reported as presented in included trials; these outcomes are referenced as *tertiary outcomes* in this review. Outcomes for other Key Questions (i.e., Key Questions #1, #4 and #5) were not prospectively defined due to the variability in endpoints relevant for these questions.

Summary of Findings

A total of 26 studies were included. Of these 26 studies, only five studies were relevant to the primary questions of this review (Key Questions #2 and #3). The following conclusions were drawn from the entire body of evidence:

1. There is a lack of external validation to support the reliability of Oncotype-DX; however, the current available evidence derived from internal industry validation studies suggests that Oncotype-DX is reliable (i.e., Oncotype-DX is repeatable and reproducible).
2. Current available evidence suggests a moderate failure rate of Oncotype-DX testing; however, the failure rate observed across clinical trials included in this review is likely inflated; the current Ontario experience suggests an acceptably lower rate of test failure.
3. In women with newly diagnosed early breast cancer (stage I–II) that is estrogen-receptor positive and/or progesterone-receptor positive and lymph-node negative:
 - a. There is low quality evidence that Oncotype-DX has prognostic value in women who are being treated with adjuvant tamoxifen or anastrozole (the latter for postmenopausal women only),
 - b. There is very low quality evidence that Oncotype-DX can predict which women will benefit from adjuvant CMF/MF chemotherapy in women being treated with adjuvant tamoxifen.
4. In postmenopausal women with newly diagnosed early breast cancer that is estrogen-receptor positive and/or progesterone-receptor positive and lymph-node positive:
 - a. There is low quality evidence that Oncotype-DX has limited prognostic value in women who are being treated with adjuvant tamoxifen or anastrozole,
 - b. There is very low quality evidence that Oncotype-DX has limited predictive value for predicting which women will benefit from adjuvant CAF chemotherapy in women who are being treated with adjuvant tamoxifen.
5. There are methodological and statistical limitations that affect both the generalizability of the current available evidence, as well as the magnitude and statistical strength of the observed effect sizes; in particular:
 - a. Of the major predictive trials, Oncotype-DX scores were only produced for a small subset of women (<40% of the original randomized population) potentially disabling the effects of treatment randomization and opening the possibility of selection bias;
 - b. Data is not specific to HER-2/*neu*-negative women;

- c. There were limitations with multivariate statistical analyses.
6. Additional trials of observational design may provide further validation of the prognostic and predictive value of Oncotype-DX; however, it is unlikely that prospective or randomized data will become available in the near future due to ethical, time and resource considerations.
 7. There is currently insufficient evidence investigating how Oncotype-DX compares to other known prognostic estimators of risk, such as Adjuvant! Online, and there is insufficient evidence investigating how Oncotype-DX would impact clinician/patient decision-making in a setting generalizable to Ontario.

Background

In February 2010, the Medical Advisory Secretariat (MAS) began work on evidence-based reviews of published literature surrounding three pharmacogenomic tests. This project came about when Cancer Care Ontario (CCO) asked MAS to provide evidence-based analyses on the effectiveness and cost-effectiveness of three oncology pharmacogenomic tests currently in use in Ontario.

Evidence-based analyses have been prepared for each of these technologies. These have been completed in conjunction with internal and external stakeholders, including a Provincial Expert Panel on Pharmacogenomics (PEPP). Within the PEPP, subgroup committees were developed for each disease area. For each technology, an economic analysis was also completed by the Toronto Health Economics and Technology Assessment Collaborative (THETA) and is summarized within the reports.

The following reports can be publicly accessed at the MAS website at: www.health.gov.on.ca/mas or at www.health.gov.on.ca/english/providers/program/mas/mas_about.html

1. Gene Expression Profiling for Guiding Adjuvant Chemotherapy Decisions in Women with Early Breast Cancer: An Evidence-Based and Economic Analysis
2. Epidermal Growth Factor Receptor Mutation (EGFR) Testing for Prediction of Response to EGFR-Targeting Tyrosine Kinase Inhibitor (TKI) Drugs in Patients with Advanced Non-Small-Cell Lung Cancer: an Evidence-Based and Economic Analysis
3. K-RAS testing in Treatment Decisions for Advanced Colorectal Cancer: an Evidence-Based and Economic Analysis

Objective of Analysis

To review and synthesize the available evidence regarding the laboratory performance, prognostic value, and predictive value of Oncotype-DX for the target population.

Definitions

The following definitions are important for framing the objectives of this review.

Prognostic value – For a pharmacogenomic test to have prognostic value, its result must be associated with prognostic outcome in women in the absence of therapy or given a standard therapy. A prognostic test helps clinicians understand how a patient progresses (or fares) in the absence of therapy or given a standard therapy. In this context, tests are often considered as being “predictive” of “prognosis”, but this description should not be confused with true predictive value of a pharmacogenomic test (below).

Predictive value – For a pharmacogenomic test to have predictive value, its result must predict significant benefit in terms of improvements in meaningful clinical outcomes for women receiving a particular therapy in comparison to women receiving an alternative therapy, placebo or observation. A predictive test should help clinicians guide a decision between two or more treatment options.

Clinical Need and Target Population

The target population of this review is women with newly diagnosed early stage (stage I–IIIa according to The American Joint Committee on Cancer staging system) invasive breast cancer that is estrogen-receptor (ER) positive and/or progesterone-receptor (PR) positive. Much of this review, however, is relevant for

women with early stage (I and II only) invasive breast cancer that is specifically ER positive, lymph node (LN) negative and human epidermal growth factor receptor 2 (HER-2/*neu*) negative. This refined population represents an estimated incident population of 3,315 new breast cancers in Ontario (according to 2007 data). This would place the refined population second in cancer incidence only to lung cancer (although breast cancer as a whole is by and large the most incident cancer in Ontario) in women in Ontario in 2007. (1)

Description of Breast Cancer

Early breast cancer is subdivided into two major categories, *in situ* disease, mainly in the form of ductal carcinoma in situ (DCIS), and invasive cancer. Breast cancer that is *in situ* has confined itself to the ducts or lobules of the breast and has not spread to the surrounding tissues in the breast or to other parts of the body. Breast cancer that is *in situ*, however, may become invasive. Invasive (infiltrating) breast cancers spread outside the membrane that lines a duct or lobule, invading the surrounding tissues. Invasive cancers may spread cancer to other parts of the body through the bloodstream and lymphatic system. Invasive cancer is far more prevalent than *in situ* disease and is the focus of this review.

For invasive cancer, a variety of adjuvant therapy options are available including chemotherapy, hormonal therapy, combined chemotherapy plus hormonal therapy, or observation alone. Treatment recommendations have traditionally been based upon the patient's risk of recurrence and the estimated or perceived benefits weighted against the potential adverse events of therapy. Factors impacting on risk of tumour recurrence include tumour size, grade, presence of lymphovascular invasion (LVI), proliferation index, and HER-2/*neu* overexpression as well as prognostic risk estimated using risk classification tools such as Adjuvant! Online.

Of women diagnosed with invasive breast cancer, approximately 65% are expected to have lymph node-negative disease at diagnosis. (2) Of these women, only 15% are expected to develop a distant metastasis at 10 years; (3) however, a far great proportion currently receive adjuvant chemotherapy, suggesting that more women are being treated with chemotherapy than can benefit. Current risk classifiers and treatment guidelines -- including Adjuvant! Online, the National Institutes of Health (NIH) Consensus Development criteria, the St. Gallen Expert Opinion Criteria, the National Comprehensive Cancer Network (NCCN) guideline, as well as traditional clinical judgement -- do not accurately identify all those women with invasive breast cancer who should or should not receive chemotherapy. Better estimators of baseline risk (i.e., prognostic risk) and response to chemotherapy (i.e., predictive response) could ensure that more women receive the appropriate treatment. Such tests may also have an economic advantage at a health systems level by potentially reducing the cost of unnecessary chemotherapy and by reducing downstream costs associated with adverse or disease-related events in women receiving improper therapy. The increased cost of testing, however, may offset these economic reductions thereby highlighting the need for formal economic evaluation prior to introducing broad scale testing into dynamic health systems.

Prevalence and Incidence

Breast cancer is the most common cancer among Canadian women. In 2007, breast cancer represented 28.9% of all new cancer cases (first among cancers) as well as 15.5% of all deaths due to cancer (second only to lung cancer) in Canada. In Ontario, this translates to 8,500 new cases of breast cancer, or an age-standardized incidence rate of 104 new cases per 100,000 population, and 2,000 deaths, or an age-standardized mortality rate of 25 new cases per 100,000 in 2007. (1) Of these new cases, a small proportion is diagnosed in advanced stages, where the tumour has spread extensively throughout the breast or to other organs of the body. In addition, many women who have been previously treated will subsequently develop a local or distant recurrence (i.e., metastasis); however, these women would not factor into incident cases.

Gene Expression Profiling

Gene expression profiling is an emerging technology for identifying genes whose activity within tumours may provide insight towards appropriately assessing disease prognosis and guiding therapy. Gene expression profiling examines the composition of cellular messenger ribonucleic acid (mRNA) within tumours providing information about the global activity of genes that give rise to mRNA. As it relates to cancer, these genes may be important for the proliferative, invasive, evasive and angiogenetic properties of tumour cells. Messenger RNA in breast cancer specimens can be measured by two different techniques, real-time reverse transcription polymerase chain reaction (RT-PCR) and deoxyribonucleic acid (DNA) microarray. (4)

Over the past decade, a number of gene expression profiling assays have seen development in breast cancer research; however, very few assays have progressed through development to commercial availability. While this review sought to identify all gene expression profile assays relevant to Ontario, only the Oncotype-DX assay (Genomic Health, Redwood City, CA) was identified as being commercially available and of current relevance to the Ontario breast cancer population and this review therefore focuses on this assay alone.

Oncotype-DX

The Oncotype-DX Breast Cancer Assay quantifies gene expression for 21 genes in breast cancer tissue by RT-PCR using formalin-fixed paraffin-embedded (FFPE) tumour blocks that are obtained during initial surgery (lumpectomy, mastectomy, or core biopsy). The panel of 21 genes include genes associated with tumour proliferation and invasion, as well as other genes related to HER-2/*neu* expression, ER expression, and PR expression. (5)

Oncotype-DX was originally intended as a prognostic risk classifier to estimate the likelihood of recurrence in women of all ages with newly diagnosed stage I or II invasive breast cancer that is LN negative and ER positive, who will be treated with tamoxifen. (5) Since Oncotype-DX's introduction, however, its indications have expanded (according to its manufacturer, Genomic Health). Oncotype-DX is now indicated as above, but also for women who will be treated with aromatase inhibitors instead of tamoxifen. Beyond prognosis, indications for Oncotype-DX have also expanded to accommodate the predictive abilities of the test. Oncotype-DX is now intended to inform adjuvant chemotherapy treatment decisions by predicting the likelihood of adjuvant chemotherapy benefit for the target population. Indications have further expanded such that Oncotype-DX is now intended for use both as a prognostic and predictive tool in postmenopausal women with early, invasive breast cancer that is LN positive. Oncotype-DX is further being evaluated in the neoadjuvant setting (6-9), metastatic setting, (10) and for DCIS (www.oncotypedx.com).

Status in Ontario

Oncotype-DX is currently being reimbursed on a case-by-case basis by the Out of Country Services (OOS) program of the Ontario Health Insurance Plan (OHIP). As an RT-PCR test, Oncotype-DX is considered a routine laboratory service and thus does not require licensing by Health Canada. All samples are currently being shipped to the Genomic Health reference laboratory in Redwood City, CA. The cost of the test for Ontario is currently 4,075 USD.

Evidence-Based Analysis

Research Questions

The overall objective, as stated above, is to review and synthesize the available evidence regarding the laboratory performance, prognostic value, and predictive value of Oncotype-DX for the target population of women with newly diagnosed early breast cancer that is ER positive or ER and/or PR positive (i.e., hormone-receptor positive). This requires several key questions and sub-questions.

Key Questions

1. What is the laboratory performance of Oncotype-DX?
 - a. How reliable is Oncotype-DX (i.e., how repeatable and reproducible is Oncotype-DX)?
 - b. How often does Oncotype-DX fail to give a useable result?
2. What is the prognostic value of Oncotype-DX?*
3. What is the predictive value of Oncotype-DX?*
4. How does Oncotype-DX compare to other known predictors of risk such as Adjuvant! Online?
5. How does Oncotype-DX impact patient quality of life and clinical/patient decision-making?

*Key Questions #2 and #3 are the primary questions of this analysis.

Research Methods

Literature Search

Search Strategy

A literature search was performed on March 19th, 2010 using OVID MEDLINE, MEDLINE In-Process and Other Non-Indexed Citations, EMBASE, the Cumulative Index to Nursing & Allied Health Literature (CINAHL), the Cochrane Library, and the International Agency for Health Technology Assessment (INAHTA) for studies published from January 1st, 2006 to March 19th, 2010. Please note that a previous systematic review (4) had been identified in preliminary literature searching. This systematic review, by Marchionni et al., was deemed comprehensive and complete up until its literature search date of January 9th, 2007. All studies included in the review by Marchionni et al. were therefore included in this review as long as studies met the inclusion/exclusion criteria of this review. A one-year window of overlap was used when updating the literature search of this review to account for any publications that may not have yet been indexed in major science literary databases when Marchionni et al. had conducted their literature search.

Abstracts were reviewed by a single reviewer and, for those studies meeting the eligibility criteria, full-text articles were obtained. Reference lists of included studies and reviews on the topic were also examined for any additional relevant studies not identified through the search. An expert panel, the Oncotype-DX Subgroup Panel (described further below), was established and contacted for knowledge of references that may have been missed. Articles with an unknown eligibility were reviewed with a second clinical epidemiologist and then a group of clinical epidemiologists at the MAS until consensus was established.

Inclusion Criteria

- Any observational trial, controlled clinical trial, randomized controlled trial (RCT), meta-analysis or systematic review that reported on the laboratory performance, prognostic value and/or predictive value of Oncotype-DX testing, or other outcome relevant to the Key Questions, specific to the target population was included.

Exclusion Criteria

- Studies that did not report original data or original data analysis,
- Studies published in a language other than English,
- Studies reported only in abstract or as poster presentations (such publications were not sought nor included in this review since the MAS does not generally consider evidence that is not subject to peer review nor does the MAS consider evidence that lacks detailed description of methodology).

Outcomes of Interest

Outcomes of interest varied depending on the Key Question. For the Key Questions of prognostic and predictive value (Key Questions 2 and 3), the prospectively defined *primary outcome* was risk of 10-year distant recurrence. The prospectively defined *secondary outcome* was 10-year death due to any cause (i.e., overall survival). All additional outcomes such as risk of locoregional recurrence or disease-free survival (DFS) were not prospectively determined for this review but were reported as presented in included trials; these outcomes are referenced as *tertiary* outcomes in this review. Outcomes for other Key Questions (i.e., Key Questions 1, 4 and 5) were not prospectively defined due to the variability in endpoints relevant for these questions.

Expert Panel

A multidisciplinary panel of experts in the field of pharmacogenomics, entitled the Provincial Expert Panel on Pharmacogenomics (PEPP), was established to contextualize and discuss evidence produced from this and other reviews of pharmacogenomic tests. A subgroup of this panel that included additional experts in the field of breast cancer, entitled the Oncotype-DX Subgroup Panel, was established to further contextualize and discuss evidence specific to this review of Oncotype-DX.

Statistical Analysis

Whenever possible, the risk of primary or secondary outcome was presented as the Kaplan-Meier (KM) estimate of risk rather than the absolute proportion of women who developed the outcome in order to account for censoring. If KM estimates were not reported, they were estimated from KM survival curves using the free graphical software package GIMP 2.6 (www.gimp.org). When neither the KM estimates nor the KM survival curves were available, absolute proportions were reported.

For outcomes of clinical sensitivity and specificity, test positivity was determined according to Oncotype-DX recurrence score (RS) risk category. Because Oncotype-DX provides three different risk categories, two different scenarios were entertained. In the first scenario, a positive test referred to all women in the Oncotype-DX high risk group ($RS \geq 31$) whereas a negative test referred to all women in the Oncotype-DX low ($RS 18-31$) and intermediate ($RS \leq 17$) risk groups. In the second scenario, a positive test referred to all women in the Oncotype-DX high and intermediate risk groups whereas a negative test referred to all women in the Oncotype-DX low risk group only. Outcomes of sensitivity and specificity were then calculated by constructing a 2x2 table using the number of women who tested positive/negative on one axis and the number of women who did/did not develop the outcome, e.g., distant recurrence, on the other

axis. Kaplan-Meier estimates of the risk of distant recurrence, for example, were used to calculate the number of women who did/did not develop a distant recurrence.

Quality of Evidence

The quality of the body of evidence was assessed as high, moderate, low, or very low according to the GRADE Working Group criteria (11) as presented below.

- Quality refers to the criteria such as the adequacy of allocation concealment, blinding and follow-up.
- Consistency refers to the similarity of estimates of effect across studies. If there are important and unexplained inconsistencies in the results, our confidence in the estimate of effect for that outcome decreases. Differences in the direction of effect, the magnitude of the difference in effect, and the significance of the differences guide the decision about whether important inconsistency exists.
- Directness refers to the extent to which the interventions and outcome measures are similar to those of interest.

As stated by the GRADE Working Group, the following definitions of quality were used in grading the quality of the evidence:

High	Further research is very unlikely to change confidence in the estimate of effect.
Moderate	Further research is likely to have an important impact on confidence in the estimate of effect and may change the estimate.
Low	Further research is very likely to have an important impact on confidence in the estimate of effect and is likely to change the estimate.
Very Low	Any estimate of effect is very uncertain.

Results of Evidence-Based Analysis

Literature Search Results

As described above, preliminary searching identified a recent comprehensive systematic review by Marchionni et al. that included literature published up until January 9th, 2007. (4) The review by Marchionni et al. included 14 published trials, (5;7-9;12-21) of which, all but one met the inclusion criteria for this review. The excluded study (21) was a poster presentation from the 2005 St. Gallen Breast Cancer Symposium.

The updated literature search yielded 1,463 articles published from January 1st, 2006 until March 19th, 2010. Of these, 88 full texts were retrieved and an additional nine studies were included. (6;10;22-28) Four additional studies were included. (29-32) These trials were not identified in literature searching but were identified through hand-searching or through referral by experts (three of these four studies were published after the literature search date of March 19th, 2010). An additional three studies were identified after the draft version of this document was made available for public comment (33-35). These references were reviewed following draft publication and were not deemed to influence the conclusions or recommendations of this review. These citations are therefore included for completeness only.

A total of 26 studies were therefore included. (5-10;12-20;22-32)

Of the 26 studies included in this review, no trials randomized women to treatment on the basis of Oncotype-DX status nor were there any prospective cohort trials that tracked the primary outcome of 10-year distant recurrence. The majority of trials were retrospective in design and performed using archived tissue samples obtained either from a cohort of consecutive women who were treated at a hospital or from women who had participated in an RCT that originally randomized participants to one of several adjuvant

treatments. Specifically, there was one nested case-control trial, (14) one case-cohort trial, (25) one prospective hospital cohort trial, (24) one retrospective chart review, (22) one retrospective registry cohort, (10) one retrospective multisite cohort trial, (31) one retrospective hospital cohort trial that was used to develop the Oncotype-DX assay, (19) one retrospective analysis of patients from single arms of multiple randomized controlled trials, (32) two technical laboratory reports, (12;20) two prospective multisite cohort trials, (27;30) two retrospective community-based cohort trials, (16;18) two trials whose design was unclassifiable or unclear, (23;28), two retrospective subgroup analyses of randomized controlled trials, (15;26) and two retrospective analyses of patients from single arms of a randomized controlled trial. (5;29) Finally, there were two economic analyses, (13;17) and four neoadjuvant studies. (6-9) These final six trials were considered outside of the scope of this review as they were not relevant for Key Questions #2 –5 and are therefore not discussed in this review in detail; however, data from these trials were used to inform Key Question #1 (the laboratory performance of Oncotype-DX).

There were 22 studies that actively submitted participant samples for testing (see Appendix 1 Table 2). (5-10;14-16;18;19;22-28;28-32) Of these, 19 studies obtained Oncotype-DX scores via Genomic Health, (5-9;14-16;18;19;22;24-32) whereas two studies recreated the 21-gene RS using homebrew microarray techniques, (10;23) and one study recreated the 21-gene RS using homebrew RT-PCR. (28)

Trial characteristics and results of included studies will be summarized by Key Question below.

Key Question 1. What is the laboratory performance of Oncotype-DX?

The laboratory performance of Oncotype-DX was assessed across 15 studies. (6-9;12;14;16;18-20;24-26;29;31). Trial methodology varied considerably between the 15 studies and the majority of studies focused on patient-related outcomes and were not designed with the purpose of evaluating laboratory performance. There were, however, two technical laboratory reports (12;20) but only one of these assessed outcomes relevant to Subquestions 1a and 1b directly below. (12)

a. How reliable is Oncotype-DX (i.e., how repeatable and reproducible is Oncotype-DX)?

Three studies assessed the reliability of Oncotype-DX although not as a primary endpoint. (5;12;14). Only one study assessed the repeatability, or the degree to which Oncotype-DX provides the same result each time the test is performed on a given sample under identical conditions, by measuring the standard deviation of repeat RS measurements from the same FFPE tumour block. (5) The reproducibility, or the degree to which a measurement provides the same result each time the test is performed on a given sample under changing conditions, was assessed across all three studies by measuring the standard deviation of repeat RS measurements from the same or different FFPE tumour blocks at repeat time points and/or across different instruments or operators.

No between-laboratory or external reproducibility studies were performed as all three studies were funded by the manufacturer, Genomic Health, and all testing was conducted at Genomic Health's reference laboratory in Redwood City, California.

The standard deviation on repeat measurements was generally less than or equal to 3.0 RS units across all three studies suggesting reliability of the Oncotype-DX assay; (5;12;14); however, all of the trials examined reliability in very few women or samples and therefore the outcomes observed may be subject to selection bias. Detailed results are summarized in Appendix 2 Table1.

b. How often does Oncotype-DX fail to give a useable result?

Thirteen trials reported on the failure rate of Oncotype-DX but not as a primary endpoint. (6-9;14;16;18;19;24-26;29;31) Reasons for failure included insufficient tumour sample in tissue block, insufficient RNA for RT-PCR, failure of RT-PCR and low signal of reference genes. The failure rate of

Oncotype-DX was often improperly reported, with insufficient tumour sample in tissue block often not being considered as a reason for test failure.

When calculated including insufficient tumour sample in tissue block as a reason for test failure, the failure rate of Oncotype-DX ranged from 2.7–44.9% across studies with a mean of 20.3% and a median of 11.8%. Detailed results are summarized in Appendix 2 Table 2.

The Oncotype-DX Subgroup Panel indicated that the failure rate observed in the above included trials was notably higher than that being observed in samples tested so far in Ontario. The Panel suggested that the increased age of archived tissue samples used in trials may partially explain the high failure rates observed in the above trials (many trials used archived tissue samples that had been stored in excess of ten years prior to submission to Genomic Health for analysis). Insufficient tumour sample in tissue blocks can also be expected when using tissue samples that have been used in multiple translational studies over many years. The failure rate observed across trials was therefore likely inflated.

Key Question 2. What is the prognostic value of Oncotype-DX?

Fourteen studies examined the association between Oncotype-DX score and prognostic outcome in women with early breast cancer. (5;10;14;14;15;18;19;23;25;26;28;29;31;32)

These studies can be further categorized by patient population, therapy, and outcomes assessed:

- Lymph-Node Negative
 - Four studies assessed the association between Oncotype-DX score and *primary* or *secondary* outcome in women with early breast cancer that is ER positive, or ER positive and/or PR positive (i.e., hormone-receptor positive), and LN negative, who received adjuvant tamoxifen (this represents the key population for which Oncotype-DX is currently indicated and is the focus of Subquestion 2a below). (5;15;29;31)
 - Two studies assessed the association between Oncotype-DX score and *tertiary* outcome in women with early breast cancer that is ER positive and LN negative, who received adjuvant tamoxifen. (14;32)
 - Two studies assessed the association between Oncotype-DX score and outcome in women with early breast cancer that is ER positive and LN negative, who received *adjuvant tamoxifen plus adjuvant chemotherapy* (15;32)
- Lymph-Node Positive
 - Two studies assessed the association between Oncotype-DX score and *primary* or *secondary* outcome in postmenopausal women with early breast cancer that is ER positive, or ER positive and/or PR positive (i.e., hormone-receptor positive), and LN positive, who received tamoxifen (this represents a population for which Oncotype-DX is currently indicated and is the focus of Subquestion 2a below). (26;29)
- Mixed
 - Seven studies assessed the association between Oncotype-DX score and outcome in mixed patient populations receiving mixed adjuvant therapy or no adjuvant therapy. (10;14;18;19;23;25;28;32)

Detailed results from each individual trial are categorized and displayed in Appendix 3 Table 1. However, only those trials relevant to Subquestion 2a are discussed in depth in this review as these trials reflect the population for which Oncotype-DX is currently indicated and being used.

a. Is Oncotype-DX recurrence score associated with the risk of distant recurrence or death due to any cause in women with early breast cancer receiving tamoxifen?

Five trials in total assessed the association between Oncotype-DX score and 10-year risk of distant recurrence or death due to any cause in women of the target population with newly diagnosed early breast cancer that is ER positive, or ER positive and/or PR positive, who received adjuvant tamoxifen. (5;15;26;29;31)

Results will be categorized by lymph-node status of study participants.

Lymph-Node Negative

Four trials assessed the prognostic value of Oncotype-DX in women with early breast cancer that is ER positive, or ER positive and/or PR positive, and LN negative, who received adjuvant tamoxifen or anastrozole. (5;15;29;31)

Three of these four trials were retrospective analyses of patients from single arms of RCTs in which study participants were originally randomized to receive either tamoxifen; (5;15) or tamoxifen alone, anastrozole alone or a combination of tamoxifen and anastrozole. (29) Several authors of these three trials declared financial ties to Genomic Health. Two of these three trials also received direct funding from Genomic Health. (5;15) The fourth trial was a retrospective multisite cohort study performed using consecutive women who entered eight oncology centers across Japan. (31) All four trials used archived tissue samples that were sent to Genomic Health to obtain Oncotype-DX scores. (5;15;29;31) All trials also included women with HER-2/*neu*-positive and HER-2/*neu*-negative disease. Lastly, three of four trials included women of most ages above 18 years with ER-positive disease (5;15;31) while one trial included only postmenopausal women with ER-positive and/or PR-positive (i.e., hormone-receptor-positive) disease (29).

Kaplan-Meier estimates showed significant differences in the 10-year risk of distant recurrence between Oncotype-DX risk categories across most trials (Table 1). Kaplan-Meier estimates for death due to any cause at 10 years showed a similar trend (Table 2).

Multivariate Cox proportional hazard models assessing whether the likelihood of 10-year distant recurrence increased relative to a 50-point difference in continuous RS were reported in three trials. (5;29;31) Two of these three studies modeled additional covariates of age at surgery and tumour size (5;31) while the third study also modelled central grade or local grade. (29) It should be noted that continuous clinical covariates such as age were dichotomized across all studies while Oncotype-DX RS was modelled as continuous. This may impact the magnitude and/or statistical strength of the RS.

The pooled hazard ratio (HR) for a 50-point increment in RS relative to the risk of 10-year distant recurrence based on Cox proportional hazard modeling was 3.47 (95% CI 3.24–3.72) (individual HRs were reported across three of four studies). (5;15;29;31) Table 1 below summarizes individual HRs obtained across studies.

Only one study reported results of multivariate Cox proportional hazard modelling to assess whether the likelihood of death due to any cause increased relative to 50-point increments in continuous RS; however, the HR was not significant (*P*-value not reported) (see Table 2). (31)

Table 1: Association of Oncoypte-DX with 10-year risk of distant recurrence in women with lymph-node-negative disease

Study	N (Events)	KM Risk of 10-Year DR (95% CI)			P-value Test for Significance Between KM Groups	Multivariate Cox Proportional HR for 50-Point Increment in Continuous RS
		High	Intermediate	Low		
Dowsett 2010 (ATAC) (29)	872 (72)	24 .0 (17.0–34.0)*	12.0 (8.0–18.0)*	4.0 (3.0–7.0)*	<0.001†	3.92 (2.08 – 7.39)‡ P-value<0.001 5.25 (2.84–9.73) P-value <0.001
Paik 2004 (NSABP B-14) (5)	668 (99)	30.5 (23.6–37.4)	14.3 (8.3–20.3)	6.8 (4.0–9.6)	<0.001¥	2.81 (1.70–4.64) P-value <0.001
Paik 2006 (NSABP B-20) (15)	227 (27)	39.5 (25.2–53.8)	9.1 (0.6–17.5)	3.2 (0.01–16.7)	NR	NR
Toi 2010 (31)	200 (18)	24.8 (15.7–37.8)	0 (N/A)	3.3 (1.1–10.0)	<0.001¥	6.03 (2.17–16.7) P-value<0.001

* Data represent KM risk of distant recurrence at nine years

†Log-rank test of equality between all KM risk groups

‡Hazard ratios pertain to a model that also adjusts for local grade followed by a model that instead adjusts for central grade

¥Log-rank test of difference in KM estimates between high and low risk groups

Abbreviations: ATAC, Arimidex, Tamoxifen, Alone or in Combination trial; CI, confidence interval; HR, hazard ratio; KM, Kaplan-Meier; N, sample size; NR, not reported; NSABP, National Surgical Adjuvant Breast and Bowel Project; RS, recurrence score.

Table 2: Association of Oncoypte-DX with overall survival in women with lymph-node-negative disease

Study	N (Events)	KM Risk of 10-Year Death Due to Any Cause (95% CI)			P-value Test for Significance Between KM Groups	Multivariate Cox Proportional HR for 50-Point Increment in Continuous RS
		High	Intermediate	Low		
Dowsett 2010 (ATAC) (29)	872 (121)	27.0*	16.0*	12.0*	<0.001†	NR
Paik 2004 (NSABP B-14) (5)	668 (NR)	32.1	22.4	10.0	<0.001†	NR
Paik 2006 (NSABP B-20) (15)	227 (30)	38.3	15.6	3.7	NR	NR
Toi 2010 (31)	200 (NR)	19.1 (11.3–31.3)	2.6 (0.4–16.8)	6.4 (2.9–13.6)	NR	2.67 (0.93–7.62) P-value NR

*Data represent KM risk of death due to any cause at nine years

†Log-rank test of equality between all risk groups

Abbreviations: ATAC, Arimidex, Tamoxifen, Alone or in Combination; CI, confidence interval; KM, Kaplan-Meier; N, sample size; NR, not reported; NSABP, National Surgical Adjuvant Breast and Bowel Project

Estimates of clinical sensitivity and specificity for each individual trial are summarized in Appendix 3 Table 1. Across the four trials, Oncotype-DX had a pooled sensitivity of 54% (95% CI 48–61%) and a pooled specificity of 83% (95% CI 81–85%) for estimating 10-year risk of distant recurrence when a positive test was defined as Oncotype-DX high risk only. When a positive test was defined as Oncotype-DX high and intermediate risk, Oncotype-DX had a pooled sensitivity of 77% (95% CI 71–83) and a pooled specificity of 59% (95% CI 57–62%) for estimating 10-year risk of distant recurrence.

Across the four trials, Oncotype-DX had a pooled sensitivity of 40% (95% CI 34–36%) and a pooled specificity of 82% (95% CI 80–84%) for estimating death due to any cause at 10 years when a positive test was defined as Oncotype-DX high risk only. When a positive test was defined as Oncotype-DX high and intermediate risk, Oncotype-DX had a pooled sensitivity of 65% (95% CI 60–71%) and a pooled specificity of 59% (95% CI 56–61) for estimating death due to any cause at 10 years.

The Oncotype-DX Subgroup Panel expressed difficulties in interpreting estimates of clinical sensitivity and specificity, particularly in the absence of comparative accuracy data for other prognostic classification tools or factors. Emphasis should likely not be placed on estimates of clinical sensitivity and specificity for decision-making considering the lack of comparable data and when considering that the Oncotype-DX RS is a continuous measure and was designed to be used as such.

Lymph-Node Positive

Two trials assessed the prognostic value of Oncotype-DX in postmenopausal women with early breast cancer that is ER and/or PR positive (i.e., hormone-receptor positive), and LN positive, who received adjuvant tamoxifen. (26;29).

Both trials were retrospective analyses of women from single arms of RCTs in which participants were originally randomized to receive tamoxifen or placebo; (26) or tamoxifen alone/ anastrozole alone/a combination of tamoxifen and anastrozole or placebo. (29) Several authors in both trials declared financial ties to Genomic Health. One trial received direct funding from Genomic Health. (26) Both trials used archived tissue samples that were sent to Genomic Health to obtain Oncotype-DX scores and both trials included women with HER-2/*neu*-positive and HER-2/*neu*-negative disease. (26;29) Lastly, primary study endpoints differed between trials. In one trial, the primary endpoint was nine-year distant recurrence. (29) In the other, the primary endpoint was 10-year disease-free survival (DFS) because distant recurrence was not recorded in the parent RCT. (26)

Kaplan-Meier estimates showed significant differences in the 9-year risk of distant recurrence or disease event (i.e., DFS) between Oncotype-DX risk categories (Table 3). Similarly, KM estimates showed significant differences in the risk of death due to any cause at 9 or 10 years between Oncotype-DX risk categories (Table 4).

Multivariate Cox proportional hazard models assessing whether the likelihood of primary outcome increased relative to a 50-point difference in continuous RS were reported in both trials (see Table 3). (26;29) One study adjusted only for the number of positive nodes (as a dichotomous variable) (26) while the other study adjusted for tumour size, central grade, age, and number of positive nodes (all as dichotomous variables). (29) It should be noted that continuous clinical covariates such as age, tumour size and number of positive nodes were dichotomized across all models while Oncotype-DX RS was modelled as continuous. This may impact the magnitude and/or statistical strength of the RS.

In both trials, the hazard of the risk of 9-year distant recurrence or 10-year DFS was significantly increased by a 50-point difference in continuous Oncotype-DX RS (Table 3). (26;29)

Only one study reported results of multivariate Cox proportional hazard modelling to assess whether the

likelihood of death due to any cause increased relative to a 50-point difference in continuous Oncotype-DX RS (see Table 4). (26)

In this trial, the hazard of the risk of 10-year death due to any cause was significantly increased by a 50-point increment in continuous Oncotype-DX RS (Table 4). (26)

Table 3: Association of Oncotype-DX with distant recurrence in women with lymph-node-positive disease

Study	N (Events) [Outcome]	KM Risk of Outcome (95% CI)			P-Value for Test of Significance	Multivariate Cox Proportional HR for 50-point Increment in Continuous RS
		High	Intermediate	Low		
Albain 2010 (SWOG 8814) (26)	148 (66) [10-yr DFS]	57.9	51.4	39.6	0.017*	2.64 (1.33–5.27) P-value 0.006
Dowsett 2009 (ATAC) (29)	306 (74) [9-yr DR]	49.0 (35.0–64.0)	28.0 (20.0–39.0)	17.0 (12.0–24.0)	<0.001*	3.47 (1.64 to 7.38) P-value 0.002

* Log-rank test of equality between all KM risk groups

Abbreviations: ATAC, Arimidex, Tamoxifen, Alone or in Combination; CI, confidence interval; DFS, disease-free survival; DR, distant recurrence; HR, hazard ratio; KM, Kaplan-Meier; N, sample size; NR, not reported; RS, recurrence score; SWOG, Southwest Oncology Group; yr, year

Table 4: Association of Oncotype-DX with overall survival in women with lymph-node-positive disease

Study	N (Events) [Outcome]	KM Risk of 10-Year Death Due to Any Cause (95% CI)			P-Value for Test of Significance	Multivariate Cox Proportional HR for 50-point Increment in Continuous RS
		High	Intermediate	Low		
Albain 2010 (SWOG 8814) (26)	148 (66)	48.6	31.9	22.9	0.003*	4.42 (1.96–9.97) P-value 0.0006
Dowsett 2009 (ATAC) (29)	306 (74)	46.0	31.0	26.0	0.002*	NR

* Log-rank test of equality between all KM risk groups

Abbreviations: ATAC, Arimidex, Tamoxifen, Alone or in Combination; CI, confidence interval; DFS, disease-free survival; DR, distant recurrence; HR, hazard ratio; KM, Kaplan-Meier; N, sample size; NR, not reported; RS, recurrence score; SWOG, Southwest Oncology Group; yr, year

Key Question 3. What is the predictive value of Oncotype-DX?

Only two trials assessed the predictive value of Oncotype-DX (15;26). Both trials were retrospective subgroup analyses of RCTs that originally randomized participants to treatment with tamoxifen or tamoxifen plus CMF/MF chemotherapy, (15) or to tamoxifen or CAF chemotherapy followed by tamoxifen. (26)

a. Does Oncotype-DX recurrence score predict significant benefit in terms of improvements in 10-year distant recurrence or death due to any cause for women receiving tamoxifen plus chemotherapy in comparison to women receiving tamoxifen alone?

Results will be categorized by lymph-node status of study participants.

Lymph-Node Negative

One trial assessed the predictive value of Oncotype-DX in women with newly diagnosed breast cancer that is ER positive and LN negative, who received adjuvant tamoxifen therapy. (15)

The trial, by Paik, (15) was designed as a retrospective subgroup analysis of the NSABP B-20 RCT. The B-20 trial originally randomized 2,299 women with newly diagnosed breast cancer that was ER positive and LN negative to treatment with tamoxifen (n=770) or tamoxifen plus CMF/MF chemotherapy (n=1,529). The B-20 trial included women with HER-2/*neu*-positive and HER-2/*neu*-negative disease. (36) Archived tissue samples were collected and used to obtain Oncotype-DX scores via Genomic Health; however, scores were only produced for 651 (28.9%) of the 2,299 women who participated in the parent trial thus opening the possibility of selection bias. (15) Significant differences in baseline tumour grade characteristics were noted between the population with available tissue and the full NSABP B-20 trial population (online appendix). (15) The trial by Paik received funding from Genomic Health and several investigators declared financial ties to Genomic Health. It should be noted that some women from the tamoxifen-only arm of the NSABP B-20 trial were used in the training data sets for the development of the Oncotype-DX assay (37) and thus a subset of the population examined by Paik may have been biased towards optimization of prediction of recurrence. This could translate into an enhanced estimate of chemotherapy benefit although the effect was likely minimal. (4)

For all women in the trial by Paik, the KM risk of 10-year distant recurrence was 12.2% (95% CI 7.7–16.2%) in the tamoxifen-treated arm and 7.8% (5.1–11.6%) in the tamoxifen plus CMF/MF chemotherapy (see Table 5). When grouped by Oncotype-DX risk category, only those women in the high risk category showed significant improvements in distant recurrence for chemotherapy vs. tamoxifen (RR of 0.26 [95% CI 0.13–0.53] for Oncotype-DX high risk women). Conversely, the risk of recurrence in the Oncotype-DX low risk group actually increased (RR 1.31 [95% CI 0.46–3.78]) with chemotherapy. However, it should be noted that small event sizes within the individual Oncotype-DX risk groups, particularly within the low risk group (which experienced only five events), increased the uncertainty in the point estimate (reflected by large confidence intervals that included chemotherapy benefit in the Oncotype-DX low and intermediate risk groups). Similar results were observed for death due to any cause (i.e., overall survival) with only the Oncotype-DX high risk group showing significant (P -value <0.001) reductions in death due to any cause for chemotherapy vs. tamoxifen (only survival curve data was available; reported in the online appendix of the Paik publication). (15)

Multivariate Cox proportional hazard modeling was performed to test the statistical strength of the relationship between the magnitude of chemotherapy benefit and continuous Oncotype-DX RS by modelling a formal test of statistical interaction between a 50-point increment in continuous RS and chemotherapy treatment (this is the strongest test of predictive benefit). This tests whether the difference in outcome from randomised treatment depended on increasing recurrence score. The likelihood ratio test

for interaction was significant (P -value of 0.038) with a HR of 0.32 (95% CI 0.11–0.94); however, the model was not adjusted for additional covariates. The authors indicated that individual multivariate models that adjusted for single covariates (i.e., one covariate per model) such as patient age, tumour size, quantitative ER and quantitative PR, and grade (all modelled as dichotomous variable) demonstrated persistence in the strength of the interaction term despite that the reported P -value range was no longer below 0.05 (P -value 0.035–0.068 for individual models; HRs were not reported). (15) It is unclear what the statistical strength of the interaction between RS and chemotherapy benefit would have been had the model incorporated multiple covariates. It is also unclear how the addition of HER-2/*neu* (as measured by fluorescent in situ hybridization [FISH] or immunohistochemistry [IHC]) would have influenced the model results.

Table 5: Predictive value of Oncotype-DX for distant recurrence in women with lymph-node-negative disease in a trial by Paik (15)

Arm	N (Events)	KM Risk of 10-Year Distant Recurrence				Relative Risk of Chemotherapy Benefit (95% CI) by Oncotype-DX Risk Group	Cox Proportional HR For Continuous RS*
		High	IM	Low	All Women		
Tamoxifen	227 (30)	39.5 (29.2–53.8)	9.1 (0.6–17.5)	3.2 (0.1–6.3)	12.2 (7.7–16.2)	High 0.26 (0.13–0.53)	0.32 (0.11–0.94) P -value 0.038
						Intermediate 0.61 (0.24–1.59)	
Tamoxifen + CMF/MF	424 (32)	11.9 (5.8–18.0)	10.9 (4.1–17.6)	4.4 (1.4–7.3)	7.8 (5.1–11.6)	Low 1.31 (0.46–3.78)	P -values NR

*Tests the interaction between 50-point increment in continuous RS and chemotherapy benefit
Abbreviations: CI, confidence interval; CMF, cyclophosphamide, methotrexate and fluorouracil 5-FU; HR, hazard ratio; MF, methotrexate and fluorouracil 5-FU; N, sample size; NR, not reported; RS, recurrence score.

Lymph-Node Positive

One trial assessed the predictive value of Oncotype-DX in postmenopausal women with newly diagnosed breast cancer that is ER and/or PR positive and LN positive, who received adjuvant tamoxifen therapy. (26)

The trial, by Albain, (26) was designed as a retrospective subgroup analysis of the SWOG-8814 RCT. The SWOG-8814 trial originally randomized 1,477 women with newly diagnosed breast cancer that was ER and/or PR positive and LN positive to treatment with tamoxifen ($n=361$) or CAF chemotherapy followed by tamoxifen (CAF→T) ($n=566$) or concurrent CAF chemotherapy plus tamoxifen (CAFT) ($n=550$). The SWOG-8814 trial included women with HER-2/*neu*-positive and HER-2/*neu*-negative disease. The authors of the Albain study report that the concurrent treatment (CAFT) arm was excluded from analysis (despite the authors having obtained Oncotype-DX results for this group) because of inferior efficacy of this arm compared to the sequential CAF→T arm in the parent trial. Archived tissue samples were collected and used to obtain Oncotype-DX scores via Genomic Health; however, scores were only produced for 367 (39.6%) of the 927 women in the tamoxifen and CAF→T arms. The trial by Albain received funding from Genomic Health and several investigators declared financial ties to Genomic Health. (38)

Individual multivariate Cox proportional hazard models adjusting for the number of positive nodes (as a dichotomous variable) were analyzed in order to assess the HR for DFS for chemotherapy vs. tamoxifen

across Oncotype-DX risk groups (see Table 6). Both the parent trial SWOG-8814 (HR 0.69, 95% CI 0.56–0.84, *P*-value 0.0003) and the cohort that underwent Oncotype-DX testing in the trial by Albain (HR 0.72, 95% CI 0.51–1.00, *P*-value 0.048) showed improved DFS in favour of chemotherapy when chemotherapy (Arm B) was compared to tamoxifen (Arm A) in all women. When women were stratified according to Oncotype-DX risk category, this chemotherapy benefit was observed only in the high risk group (HR 0.59, 95% CI 0.35–1.01, *P*-value 0.033) (Table 6). (38) Similar results were observed for death due to any cause (Table 7). The authors indicate that the interaction of treatment benefit and recurrence score remained significant after adjustment for age, ethnic origin, tumour size, progesterone status, grade, P53, and HER2 by TAB250 but results are not shown. (38)

Multivariate Cox proportional hazard modeling was performed to test the statistical strength of the relationship between the magnitude of chemotherapy benefit and continuous Oncotype-DX RS by modelling a formal test of statistical interaction between a 50-point increment in continuous RS and chemotherapy treatment (this is the strongest test of predictive benefit). This tests whether the difference in outcome from randomised treatment depended on increasing recurrence score. The likelihood ratio test for interaction was not significant (*P*-value of 0.053) with a HR of 0.43 (95% CI 0.18–1.01) when adjusting for the number of positive nodes (as a dichotomous variable) (see Table 6). Unplanned, additional analysis assessed whether the effect of RS on treatment was constant over time. The likelihood ratio test for interaction was not significant (*P*-value of 0.58) with a HR of 0.66 (95% CI 0.16–2.82) when adjusting for the number of positive nodes (as a dichotomous variable) after 5 years of treatment. Conversely, the likelihood ratio test for interaction was significant (*P*-value of 0.029) with a HR of 0.30 (95% CI 0.10–0.89) when adjusting for the number of positive nodes (as a dichotomous variable) over the first 5 years of treatment (Table 6). The trial authors conclude that RS is predictive of chemotherapy benefit over the first five years of treatment. Similar results were observed for death due to any cause (Table 7).

Table 6: Predictive value of Oncotype-DX for DFS in women with lymph-node-positive disease in a trial by Albain (38)

Arm	Sample Size (Events)	Cox Proportional HR for Death Due to Any Cause for Chemotherapy vs. Tamoxifen by Oncotype-DX Risk Group (95% CI)	Cox Proportional HR For Continuous RS*
Tamoxifen	148 (66)	Parent Trial (SWOG-8814) All Women 0.69 (0.56–0.84) <i>P</i> -value 0.0003	All years 0.43 (0.18–1.01) <i>P</i> -value 0.053
		All Women 0.72 (0.51–1.00) <i>P</i> -value 0.048	First 5 years 0.30 (0.10–0.89) <i>P</i> -value 0.029
CAF→T	219 (77)	High 0.59 (0.35–1.01) <i>P</i> -value 0.033	After 5 years 0.66 (0.16–2.82) <i>P</i> -value 0.58
		Intermediate 0.72 (0.39–1.31) <i>P</i> -value 0.48	
		Low 1.02 (0.54–1.93) <i>P</i> -value 0.97	

*Tests the interaction between 50-point increment in continuous RS and chemotherapy benefit
Abbreviations: CI, confidence interval; CAF, cyclophosphamide, doxorubicin, fluorouracil 5-FU; DFS, disease-free survival; HR, hazard ratio; RS, recurrence score.

Table 7: Predictive value of Oncotype-DX for overall survival in women with lymph-node-positive disease in a trial by Albain (38)

Arm	Sample Size (Events)	Cox Proportional HR for Death Due to Any Cause for Chemotherapy vs. Tamoxifen by Oncotype-DX Risk Group (95% CI)	Cox Proportional HR For Continuous RS*
Tamoxifen	148 (43)	High 0.56 (0.31–1.02) <i>P</i> -value 0.057	All years HR NR <i>P</i> -value 0.026
		Intermediate 0.84 (0.40–1.78) <i>P</i> -value 0.65	First 5 years HR NR <i>P</i> -value 0.016
CAF→T	219 (51)	Low 1.18 (0.55–2.54) <i>P</i> -value 0.68	After 5 years HR NR <i>P</i> -value 0.87

*Tests the interaction between 50-point increment in continuous recurrence score and chemotherapy benefit
Abbreviations: CI, confidence interval; CAF, cyclophosphamide, doxorubicin, fluorouracil 5-FU; HR, hazard ratio; RS, recurrence score.

Key Question 4. How does Oncotype-DX compare to other known predictors of risk such as Adjuvant! Online?

Six trials compared Oncotype-DX to Adjuvant! Online. (16;21;25;28-30) However, in no trial was this comparison a primary endpoint of the respective study.

Five trials assessed the correlation between projected 10-year risk of distant recurrence by Oncotype-DX and projected 10-year risk of any recurrence by Adjuvant! Online (Table 8). (16;21;25;29;30) Oncotype-DX generally showed weak correlation with Adjuvant! Online suggesting that the prognostic information provided by the RS may be independent from the prognostic information provided by Adjuvant! Online.

Table 8: Correlation between Oncotype-DX projected risk of distant recurrence and Adjuvant! Online projected risk of any recurrence

Trial	Year	Correlation Coefficient (<i>P</i> -value)
Bryant (21)	2005	$\rho=0.37$ (<i>P</i> -value<0.0001)
Dowsett (29)	2010	$\rho=0.23$ (<i>P</i> -value<0.001)
Goldstein (25)	2008	$\rho=0.19$ (<i>P</i> -value=NR)
Oratz (16)	2007	$\rho=0.43$ (<i>P</i> -value<0.01)
Wolf (30)	2007	$\rho=0.32$ (<i>P</i> -value=0.0001)

Abbreviations: NR, not reported

Two trials assessed the prognostic value of Oncotype-DX combined with Adjuvant! Online. (25;28)

In a trial by Espinosa, (28) a retrospective analysis was performed using archived tissue samples on 153 women newly diagnosed with early breast cancer between February 1995 and March 2003 at a hospital in Spain. All women were diagnosed with early breast cancer that was hormone-receptor positive; however, the study population was mixed with respect to lymph-node positivity and whether women received adjuvant chemotherapy in addition to tamoxifen. Several GEP tools and prognostic indicators were examined including Oncotype-DX and Adjuvant! Online. Analyses were performed to examine the gain in predictive accuracy when combining prognostic indicators/tools. When combined with Adjuvant! Online, Oncotype-DX score explained more variation (25.8%±1.4) in distant metastasis-free survival (DMFS) than when used in isolation (18.1%±0.9). (28)

In a trial by Goldstein, (25) retrospective analysis was performed using archived tissue samples on 465 women with hormone-receptor-positive breast cancer with zero to three positive axillary nodes who did or did not have recurrence after chemohormonal therapy (case-cohort design). The patient population had originally been enrolled in the Eastern Cooperative Oncology Group (ECOG) trial E2917 that randomized 2,885 women to AC or AT chemotherapy plus endocrine therapy (if hormone-receptor positive). Several authors of this trial declared financial ties to Genomic Health. An integrator of clinicopathologic information was used for comparison with Oncotype-DX. The “Integrator” was modeled after Adjuvant! Online but adjusted to 5-year outcomes rather than 10-year outcomes. Cox proportional hazard modelling and ROC analysis indicated that Oncotype-DX was a stronger predictor of the 5-year risk of any (i.e., distant or local/regional) recurrence. In addition, the absolute 5-year risk of recurrence of trial participants increased with increasing Oncotype-DX and Adjuvant! Online risk category when both tools were combined (see Table 9). (25)

Table 9: Absolute 5-year risk of any recurrence by Oncotype-DX and Adjuvant! Online risk category in a trial by Goldstein (25)

		Oncotype-DX		
		Low	Intermediate	High
Adjuvant! Online	Low	0.03 (0.02-0.06)	0.10 (0.07-0.20)	0.14 (0.18-0.34)
	Intermediate	0.03 (0.01-0.06)	0.23 (0.17-0.31)	0.14 (0.09-0.21)
	High	0.09 (0.05-0.17)	0.10 (0.06-0.16)	0.25 (0.18-0.34)

Overall, the available evidence comparing Oncotype-DX with Adjuvant! Online is weak but hypothesis-generating. Further studies with this comparison as a primary endpoint are required to fully illustrate the relationship between Oncotype-DX and Adjuvant! Online and whether these tools can be combined to improve prognostic or predictive utility and subsequent patient outcomes.

Key Question 5. How does Oncotype-DX impact patient quality of life and clinical/patient decision-making?

Only three studies addressed aspects of quality of life and clinical/patient decision-making. (22;24;27)

In a trial by Asad, (22) a retrospective chart review was completed on all ER positive, LN negative women (n=85) who had an Oncotype-DX score obtained at two hospitals in the U.S. between February 2006 and January 2008. Overall, Oncotype-DX influenced the treatment recommendation to provide or withhold chemotherapy in 44% (n=37) of women. (22)

In a trial by Lo, (27) a prospective, multi-centre trial was completed involving seventeen medical

oncologists at one community and three academic practices across the U.S. between December 2005 and August 2006. The trial consecutively enrolled 89 women with LN–negative, ER positive breast cancer who were medically fit to receive adjuvant chemotherapy. Several authors declared financial ties to Genomic Health and the trial received direct funding from Genomic Health. Each medical oncologist provided their treatment recommendation before and after Oncotype-DX testing. Patient’s treatment decision before and after Oncotype-DX testing was also recorded. Questionnaires were used to gather all relevant data. The medical oncologist treatment recommendation changed for 28 women (31.5%) (see Table 10). Similarly, 24 women (27%) changed their treatment decision. The largest change after Oncotype-DX testing was the conversion from the medical oncologist’s pre-test recommendation for chemotherapy plus hormonal therapy (CHT) to a post-test recommendation for hormone therapy only (HT) for 20 of 42 women originally recommended for CHT. Patient anxiety and decisional conflict were significantly lower after Oncotype-DX result. Meanwhile, quality of life remained stable according to Functional Assessment of Cancer Therapy (FACT-B and FACT-G) surveys. (27)

Overall, the available evidence examining how Oncotype-DX compares to Adjuvant! Online is limited. Further independent studies conducted in an Ontario or Canadian setting will improve the generalizability of this evidence.

Table 10: Medical oncologist treatment recommendation before and after Oncotype-DX testing in a trial by Lo (27)

Pre-Oncotype-DX Post-Oncotype-DX Recommendation	N
HT to HT	40
HT to CHT	3
CHT to CHT	20
CHT to HT	20
HT to equipoise	3
CHT to equipoise	2
Equipoise to equipoise	1

Abbreviations: CHT, chemotherapy plus hormonal therapy; HT, hormonal therapy; N, number of study participants.

The third trial by Geffen (24) only briefly reported on the prospectively planned therapy of 328 consecutive women in a regional oncology practice in Israel between November 2002 and December 2006. Oncotype-DX became available only near the end of the study period and was therefore performed for only 25 women. In nine of these women (36%), treatment recommendations were changed based on the scores, six from chemotherapy to no chemotherapy. (24)

Societal and Ethical Considerations

As a corollary to MAS’ systematic and economic review, MAS commissioned a qualitative review of the ethical and societal issues and considerations surrounding GEP in early breast cancer. The objectives of this qualitative review were as follows:

1. To systematically review published qualitative social science (empirical) and ethics research (both empirical and scholarly) relevant to GEP for guiding chemotherapy in women with early stage breast cancer, and

2. To describe social values and ethics issues potentially relevant to the formulation of OHTAC policy recommendations regarding this technology.

Utilizing a previous framework for defining the scope of a qualitative health technology assessment (39), the current topic was stratified into three relevant subtopics: a) the technology (i.e., gene expression profiling), b) the condition (i.e., early breast cancer), and c) the technology for condition (i.e., gene expression profiling for guiding chemotherapy decisions for women with early breast cancer). For each of these topics, a systematic search was completed of three bibliographic databases -- Medline, CINAHL, and Web of Science (Social Sciences, Humanities) -- over the period January 1, 2005 through June 30, 2010. A total of 104,763 citations were identified from searching that resulted in 70 publications that met the inclusion/exclusion criteria (full details are available in the full report [publication pending]).

Summary findings from the qualitative review are found below:

- Prognostic uncertainty contributes to suffering of breast cancer patients. Information is important, and is sought from many lay and clinical sources. Information can be useful or overwhelming. This depends on the complexity, clarity and volume of information, as well as the patient's preferences, language fluency, health literacy, culture, and other features. Not all cancer clinicians possess adequate genetic literacy to interpret and discuss a GEP test result and few patients possess adequate genetic literacy to fully contextualize discussions of GEP test results.
- Clinician communication affects the usefulness of test information. GEP may be only one of numerous prognostic/predictive assessors, and it is unclear how best to share GEP tests results in the broader context of diagnosis and prognosis. Clinicians may deliberately withhold the fact that a GEP test has been performed or withhold the results, or may communicate patients' prognoses in a deliberately vague way, to manage emotions and uncertainty in the clinical encounter.
- Breast cancer patients vary widely, both as individuals and as members of demographic groups, in their preference for aggressive treatment. The availability of treatment choices, driven by patient preferences, is widely valued (although clinicians and patients construct these preferences in many different ways). Some patients prefer aggressive intervention even when benefit is uncertain or small, or side effects are substantial while other patients may forego aggressive intervention even when perceived benefits are high.
- Groups who are ruled out of access to standard treatment, by virtue of a GEP test, may experience themselves as "orphaned" disease populations.
- Economic incentives may lead to clinically sub-optimal use of pharmacogenomic tests or the implicated drugs. On the supply side, drug manufacturers face incentives to control impact that new tests may have on the demand for profitable drugs (e.g., through strategic marketing or pricing, commercial control of the test, etc.). On the demand side, payers face incentives to control drug expenditures by limiting drug utilization on the basis of pharmacogenomic rationales. These possibilities suggest value in sensitivity analyses, and possible post-market re-assessment, in the HTA of GEP tests.
- Discrimination on the basis of genetic information is a pervasive concern, but may be less problematic with regard to tests of non-heritable tumour genetics than tests of heritable genetics. However, if the future research relates patterns of tumour genetics to population genetics, GEP test information may be used to type – and potentially to discriminate in favour of or against – particular subpopulations.

Limitations

There are a number of limitations across the available body of evidence that affect both the generalizability of this evidence, as well as the magnitude and statistical strength of the observed effect sizes. The majority of limitations discuss here pertain to those studies which addressed the primary

questions of this review, Key Questions #2 and #3.

Firstly, the design of trials was suboptimal, particularly with respect to those trials evaluating the predictive value of Oncotype-DX. These trials, by Paik (15) and Albain, (26) were designed as retrospective subgroup analyses of RCTs. As such, these studies may be evaluated for the credibility of subgroup analysis according to published criteria. (40) When applied to the studies of predictive value included in this review, the published checklist suggested a moderate level of credibility of the subgroup effects reported in these studies (Appendix 5). Further discussion regarding study design is discussed below (see External Levels of Evidence section).

Secondly, as indicated in the Results section above, multivariate statistical analyses performed across key prognostic and predictive trials were limited, thereby creating uncertainty in the estimates of effect presented. Briefly, some models failed to adjust for additional covariates outside of the RS alone, continuous covariates were often modeled as dichotomous, descriptions of why covariates were chosen was lacking, and the majority of models assessed a 50-point difference in continuous RS. Analyses employing a smaller difference in RS could have been included across trials to improve the understanding of the continuous nature of the RS, particularly since the predetermined Oncotype-DX risk categories employ rather narrow RS intervals. The above limitations may have affected the magnitude and/or statistical strength of the observed effect sizes for continuous RS in multivariate analysis.

Finally, none of the trials appropriately controlled for the potential confounding effects of HER-2/*neu* positivity. With the introduction of trastuzumab, HER-2/*neu*-positive women currently undertake a treatment pathway that is distinct from HER-2/*neu*-negative women. HER-2/*neu*-positive women would therefore not currently be candidates for Oncotype-DX testing in Ontario. This can be problematic since HER-2/*neu*-positivity is known to increase the risk of recurrence for those women with early breast cancer (both for women with LN-negative disease) (5) and LN-positive disease (26)). It is also known that HER-2/*neu* status is associated with Oncotype-DX risk score. Of 55 women who were HER-2/*neu*-positive in the prognostic study by Paik, (5) 50 of these women were high risk according to Oncotype-DX (this represents 28% of the Oncotype-DX high risk population in this study). (5) It is unclear what affect removing HER-2/*neu*-positive women would have on the prognostic and predictive value of Oncotype-DX, particularly for the high risk Oncotype-DX group. One trial assessed the impact of HER-2/*neu* status on the prognostic value of Oncotype-DX and found that the hazard of the risk of 5-year recurrence was no longer significantly associated with a 50-point increment in continuous Oncotype-DX RS (HR *P*-value >0.05) in a HER-2/*neu*-negative only population despite being significant (HR 3.13, 95% CI 1.60–6.14, *P*-value 0.0009) for the entire study population that included both HER-2/*neu*-positive and -negative women. (25) One trial assessed the impact of HER-2/*neu* status on the predictive value of Oncotype-DX, in a LN-positive population. (38) The trial, reported above, by Albain, the interaction between treatment benefit and RS remained significant after adjustment for HER-2/*neu* by TAB250 but results were not shown. (38) Full results, including similar adjustments for a LN-negative population would reduce uncertainty with respect to potential confounding by HER-2/*neu* status.

Failure to account for HER-2/*neu* confounding may affect both the generalizability and the strength of the effect sizes observed for the RS in multivariate and KM analyses. Future studies should attempt to control confounding by excluding HER-2/*neu*-positive women in trial entry criteria, by stratifying women by HER-2/*neu* status, or by including HER-2/*neu* status as a covariate in multivariate modeling.

GRADE Quality of Evidence

Please note that all GRADE evaluations were updated following comments obtained regarding this review during a public comment period. GRADE evaluations were updated with guidance of members of the GRADE Working Group, the group responsible for the development of the GRADE tool; however, the overall GRADE level remained unchanged.

The quality of the body of evidence was assessed according to GRADE criteria. GRADE scoring was conducted only for the primary questions of the review, Key Questions #2 and #3, and was carried out independently for the major patient populations under study: women with LN-negative disease and those with LN-positive disease (see Appendix 4). All GRADE evaluations relate to the outcome of distant recurrence.

Overall, the quality of the body of evidence for the prognostic value of Oncotype-DX was low according to GRADE for both patient populations (Appendix 4 Table 1). The quality of the body of evidence was downgraded twice due to study design (because the data available is either observational/descriptive or the design of the studies generated data akin to observational/descriptive data). Limitations in generalizability were also noted but did not contribute to the downgrading of the overall quality of the body of evidence.

Overall, the quality of the body of evidence for the predictive value of Oncotype-DX was very low according to GRADE for both patient populations (Appendix 4 Table 2). The quality of the body of evidence once due to methodological quality (limitations in statistical analyses and potential for selection bias [see Limitations section above or External Levels of Evidence section below]), as well as once due to consistency (due to a lack of confidence in subgroup effect [see Limitations section above]) and once due to imprecision.

In regards to imprecision, confidence intervals for chemotherapy benefit spanned both benefit (i.e., reduction in risk of distant recurrence) and harm (i.e., increase in risk of distant recurrence) for the low and intermediate Oncotype-DX risk groups across both predictive trials, largely due to small effect sizes in these risk groups, indicating uncertainty in how to clinically manage these risk groups. Despite that such analyses reflect secondary endpoints of their respective studies, these risk groups are still relevant to current clinical management of patients tested with Oncotype-DX in Ontario and elsewhere and commentary by Paik (15) (see Discussion section of the study in question), for example, provide interpretations of the anticipated benefit of chemotherapy for women of low RS that may be misleading. Further imprecision is present in the primary endpoint, the HR of the interaction of RS and chemotherapy. As it stands, the confidence interval is sufficiently wide that it spans minimal chemotherapy benefit for a 50-point difference in RS to large clinical benefit. The arbitrary choice of a 50-point continuous interval compared to smaller continuous intervals also influences the magnitude of the effect size further generating uncertainty in the interaction. The lack of data specific to HER-2/*neu*-negative patients also contributes to sparse data. Taken together, this imprecision generates uncertainty into the clinical management of patients when using Oncotype-DX. A downgrading on the basis of imprecision was therefore warranted as there exists “confidence intervals that are sufficiently wide that the estimate is consistent with conflicting recommendations should be considered as imprecise or sparse data”. (41)

Limitations in generalizability were also noted but did not contribute to the downgrading of the overall quality of the body of evidence.

External Levels of Evidence

The validation of biomarker assays remains a considerable challenge due to the multitude of marker assessment methods (e.g., IHC, FISH, RT-PCR, proteomics-based classifiers, and so on); the reliability and reproducibility of the assay (including issues of central versus local testing); the length of time needed to assess relevant outcomes for diseases such as breast cancer; the availability of tissue samples; and the additional costs involved with assessing the marker status of each study participant. As a result, prospective observational or randomized trials are lacking.

Recent publications have explored issues relating to study quality for those studies designed to assess

prognostic or predictive biomarker assays. Publications by Simon (42) as well as by Mandrekar and Sargent, (43–46) have suggested that well-designed retrospective subgroup analyses from well-conducted prospective RCTs may be of sufficient methodological quality to appropriately validate biomarker assays so that these assays may be made available for clinical use in a timely manner. Simon has even proposed revisions to the American Society of Clinical Oncology's (ASCO's) Levels of Evidence (LOE) scale. (42)

Even if such proposed scales were to be adopted for this review, the evidence of predictive value of Oncotype-DX would still be of low quality. Both of the key predictive analyses by Paik (in regards to a LN negative population) (15) and by Albain (in regards to a LN positive population) (26) produced Oncotype-DX scores for a small percentage of the parent population thus opening the possibility of selection bias (Oncotype-DX scores were only produced for less than 40% of the original, relevant RCT arms across both trials, potentially disabling the effects of treatment randomization). Additional limitations as discussed above in the Limitations section are also apparent and contribute to the downgrading of evidence. Subsequent independent validation studies of similar design should be conducted to better understand the predictive value of Oncotype-DX.

Conclusions

1. There is a lack of external validation to support the reliability of Oncotype-DX; however, the current available evidence derived from internal industry validation studies suggests that Oncotype-DX is reliable (i.e., Oncotype-DX is repeatable and reproducible).
2. Current available evidence suggests a moderate failure rate of Oncotype-DX testing; however, the failure rate observed across clinical trials included in this review is likely inflated; the current Ontario experience suggests an acceptably lower rate of test failure.
3. In women with newly diagnosed early breast cancer (stage I–II) that is estrogen-receptor positive and/or progesterone-receptor positive and lymph-node negative:
 - a. There is low quality evidence that Oncotype-DX has prognostic value in women who are being treated with adjuvant tamoxifen or anastrozole (the latter for postmenopausal women only),
 - b. There is very low quality evidence that Oncotype-DX can predict which women will benefit from adjuvant CMF/MF chemotherapy in women being treated with adjuvant tamoxifen.
4. In postmenopausal women with newly diagnosed early breast cancer that is estrogen-receptor positive and/or progesterone-receptor positive and lymph-node positive:
 - a. There is low quality evidence that Oncotype-DX has limited prognostic value in women who are being treated with adjuvant tamoxifen or anastrozole,
 - b. There is very low quality evidence that Oncotype-DX has limited predictive value for predicting which women will benefit from adjuvant CAF chemotherapy in women who are being treated with adjuvant tamoxifen.
5. There are methodological and statistical limitations that affect both the generalizability of the current available evidence, as well as the magnitude and statistical strength of the observed effect sizes; in particular:
 - a. Of the major predictive trials, Oncotype-DX scores were only produced for a small subset of women (<40% of the original randomized population) potentially disabling the effects of treatment randomization and opening the possibility of selection bias;
 - b. Data is not specific to HER-2/*neu*-negative women;

- c. There were limitations with multivariate statistical analyses.
6. Additional trials of observational design may provide further validation of the prognostic and predictive value of Oncotype-DX; however, it is unlikely that prospective or randomized data will become available in the near future due to ethical, time and resource considerations.
 7. There is currently insufficient evidence investigating how Oncotype-DX compares to other known prognostic estimators of risk, such as Adjuvant! Online, and there is insufficient evidence investigating how Oncotype-DX would impact clinician/patient decision-making in a setting generalizable to Ontario.

Economic Analysis

DISCLAIMER: The Medical Advisory Secretariat uses a standardized costing method for its economic analyses of interventions. The main cost categories and the associated methods from the province's perspective are as follows:

Hospital: Ontario Case Costing Initiative cost data are used for in-hospital stay, emergency visit and day procedure costs for the designated International Classification of Diseases (ICD) diagnosis codes and Canadian Classification of Health Interventions procedure codes. Adjustments may be required to reflect accuracy in estimated costs of the diagnoses and procedures under consideration. Due to the difficulties of estimating indirect costs in hospitals associated with a particular diagnosis or procedure, the secretariat normally defaults to considering direct treatment costs only.

Nonhospital: These include physician services costs obtained from the Ontario Schedule of Benefits, laboratory fees from the Ontario Schedule of Laboratory Fees, drug costs from the Ontario Drug Benefit Formulary, and device costs from the perspective of local health care institutions whenever possible or its manufacturer.

Discounting: For cost-effectiveness analyses, a discount rate of 5% is applied as recommended by economic guidelines.

Downstream costs: All numbers reported are based on assumptions on population trends (i.e. incidence, prevalence and mortality rates), time horizon, resource utilization, patient compliance, healthcare patterns, market trends (i.e. rates of intervention uptake or trends in current programs in place in the Province), and estimates on funding and prices. These may or may not be realized by the system or individual institutions and are often based on evidence from the medical literature, standard listing references and educated hypotheses from expert panels. In cases where a deviation from this standard is used, an explanation is offered as to the reasons, the assumptions, and the revised approach. The economic analysis represents *an estimate only*, based on the assumptions and costing methods that have been explicitly stated above. These estimates will change if different assumptions and costing methods are applied to the analysis.

Study Question

The objective of the current economic analysis was to determine the cost effectiveness of Oncotype-DX when used in addition to Adjuvant! Online to guide chemotherapy provision for Ontario women diagnosed with LN-negative ER-positive HER-2/*neu*-negative early-stage breast cancer.

Economic Analysis Overview

A cost-utility analysis (CUA) was performed. This compared all feasible strategies for the provision of Oncotype-DX and chemotherapy, subject to a number of assumptions (see Comparators section). The target population was Ontario women diagnosed with LN-negative ER-positive HER-2/*neu*-negative early-stage breast cancer who are possible candidates for adjuvant chemotherapy. The analysis was conducted from the perspective of the Ontario Ministry of Health and Long-Term Care (MOHLTC) and adopted a lifetime time horizon.

Economic Literature Review

A search was conducted for existing cost-effectiveness analyses of Oncotype-DX. Three studies were identified at the time of the review.

Hornberger et al. (2005) found Oncotype-DX-guided treatment to be cost-saving for patients reclassified as low risk and cost-effective (\$31,000 US per QALY) for patients reclassified as intermediate or high risk. (47) However, this study was funded by the manufacturer and was heavily criticized by the recent AHRQ systematic review study. (48)

Lyman et al. (2007) found Oncotype-DX-guided treatment to be cost-saving versus a strategy of providing all patients with chemotherapy and tamoxifen and cost-effective (\$2,000 US per life year) versus a strategy of providing all patients with tamoxifen alone. (49) However, this study was also manufacturer funded and was similarly criticized by the AHRQ systematic review. (48)

Tsoi et al. (2010) found Oncotype-DX-guided treatment to be cost-effective (\$63,000 CAD per QALY) versus Adjuvant! Online (AOL)-guided treatment. (50) The study authors had no apparent conflicts of interest and the overall quality of the analysis was good. Since the study was conducted from an Ontario public-payer perspective, this is by far the most relevant existing study for consideration by the MOHLTC. However, the study has some limitations, in particular: it did not consider intermediate risk for either AOL or Oncotype-DX; it only considered two strategies (Oncotype-DX is assumed to be provided either to all patients or to no patients, rather than being provided conditional upon AOL score); it considered only a single chemotherapy regimen in the base case; it did not account for parameter uncertainty; it did not consider local recurrence; and it did not account for long term adverse events resulting from chemotherapy use.

Since the Tsoi analysis was both recent and relevant, we decided to build upon this analysis and address as many of its limitations as possible rather than start from scratch. We are grateful to the authors for providing their TreeAge model and a draft of the manuscript prior to publication. A number of limitations were addressed: the analysis now considers intermediate risk groups for both AOL and Oncotype-DX; we considered all the possible strategies for the provision of Oncotype-DX and chemotherapy, rather than considering just two strategies (see Comparators section); we modelled different chemotherapy regimens for patients of different risk of distant recurrence; and we used probabilistic sensitivity analysis to propagate parameter uncertainty throughout the model.

Target Population

The target population was Ontario women diagnosed with LN-negative ER-positive HER-2/*neu*-negative early-stage breast cancer who are possible candidates for adjuvant chemotherapy. The base case analysis considered 50-year-old women, while scenario analyses considered 35-year-old and 65-year-old women.

Perspective

The analytic perspective was that of the MOHLTC.

Comparators

The analysis considered all possible permutations for the provision of Oncotype-DX and chemotherapy, assuming that:

- (a) all patients are first identified as low, intermediate or high risk using AOL (or equivalent clinical judgement);
- (b) Oncotype-DX may be targeted at specific AOL risk groups;

- (c) the recurrence score (RS) provided by Oncotype-DX is used only to identify patients as low, intermediate or high risk (rather than considered on a continuous scale);
- (d) chemotherapy may be targeted at specific AOL risk groups and (where applicable) combined AOL / Oncotype-DX risk groups; and
- (e) only a single chemotherapy regimen is considered for any particular AOL or combined AOL / Oncotype-DX risk group.

Time Horizon

The analysis considered a lifetime time horizon.

Discounting

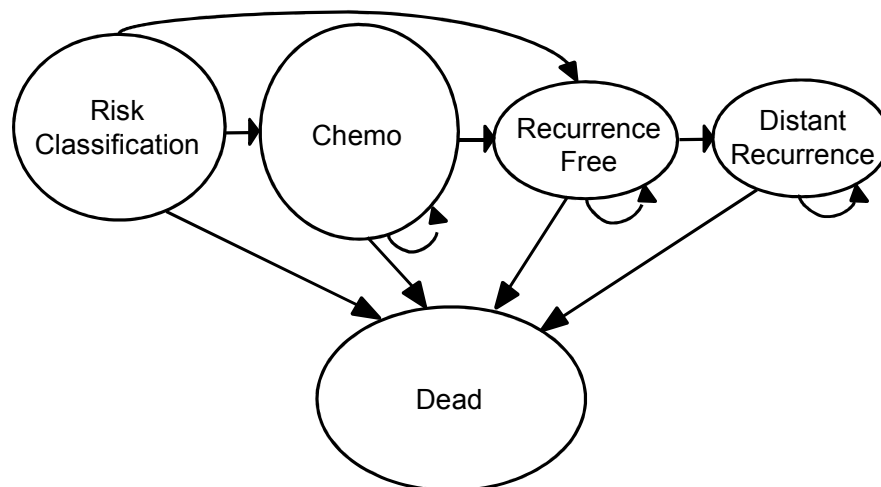
In the base case all costs and QALYs were discounted at a common rate of 5% per annum, following the most recent CADTH guidelines. (51) A sensitivity analysis was conducted in which costs and QALYs were discounted at a common rate of 1.5% per annum.

Model Structure

A schematic of the Markov model is given in Figure 1.

In the first cycle, each patient is classified as low, intermediate or high risk using AOL and (if applicable) Oncotype-DX, and a decision is made as to whether to give the patient chemotherapy. If chemotherapy is given, the patient enters the 'Chemo' state for six months, before entering the 'Recurrence Free' state; if chemotherapy is not given, the patient immediately enters the 'Recurrence Free' state. If the patient has a distant recurrence she immediately enters the 'Distant Recurrence' state. At any time the patient may die, at which point she enters the terminal 'Dead' state.

Figure 1: Schematic of Markov model, adapted from Tsoi et al. (2010)



Outcomes

The outcome of interest was the per-patient lifetime QALYs associated with each strategy.

Resource Use and Costs

The majority of the cost estimates were adapted from Table 2 of Tsoi et al. (2010). Many of these costs

had been inflated to 2008 Canadian dollars (CAD) using an unreferenced 5% inflation rate. We deflated these costs back to their base year and then reflatd all costs to 2010 CAD using a more appropriate inflation rate of 1.6%. (52)

The cost of Oncotype-DX (\$4075 US) was provided by the manufacturer. As of 19 August 2010, this was equivalent to 4191 CAD.

Recent cost estimates for each of the chemotherapy regimens were obtained from the pharmacy of the Sunnybrook Odette Cancer Centre in Toronto.

Parameter Estimates

The probability of distant recurrence was derived from Bryant (2005) and Paik (2006). (53;54) This was assumed to vary depending upon a patient’s AOL and (if applicable) Oncotype-DX risk group(s) and whether or not chemotherapy was given.

The probability of death—whether from distant metastasis, chemotherapy, or other causes—and the utility estimates remained unchanged from those given in Table 2 of Tsoi et al. (2010). (50)

Sensitivity Analysis

Probabilistic sensitivity analysis using Monte Carlo simulation (10,000 iterations) was conducted on the base case analysis.

In addition to the base case analysis, a number of scenario analyses were conducted. These considered the impact of changing a particular model assumption from that adopted in the base case. Scenario analyses separately considered: 35 year old women (base case considered 50 year old women); 65 year old women; and a 1.5% discount rate (base case adopted a 5% discount rate).

Results and Discussion

The results of the base case analysis are summarized in Table 1. The most costly strategy here is to provide Oncotype-DX to all patients, although this is also the most effective. The ICER of this strategy versus the most expensive strategy in which Oncotype-DX is *not* provided to all patients is \$23,983 per QALY. This implies that it is cost-effective to provide Oncotype-DX to all patients at any typical willingness-to-pay for a QALY. All scenario analyses resulted in the same finding.

Table 1: Summarized results of base case analysis

Patients receiving Oncotype-DX	Cost	QALYs	ICER (per QALY)
No patients	\$13,298	13.34	N/A
AOL high risk only	\$13,660	14.04	\$518
AOL intermediate/high risk only	\$13,961	14.42	\$795
All patients	\$17,466	14.64	\$23,983

At a willingness-to-pay of \$75,000 per QALY (the mid-point of the typical range of \$50,000 to \$100,000 per QALY), the probabilistic sensitivity analysis found that the probability that Oncotype-DX is cost-effective is 83.5% for patients identified as AOL low risk, 99.8% for patients identified as AOL intermediate risk, and 65.8% for patients identified as AOL high risk.

Limitations remain with the analysis, in particular with respect to the probability of distant recurrence and utility estimates. The model also does not account for local recurrence or long term adverse events from

chemotherapy. These limitations notwithstanding, Oncotype-DX appears to be cost-effective for all patients, irrespective of AOL risk group.

Budget Impact Analysis

A budget impact analysis was conducted from the perspective of the MOHLTC. This considered the additional cost to the MOHLTC of providing Oncotype-DX to patients in Ontario, along with any corresponding costs associated with changes in chemotherapy provision.

Each year there are approximately 8,500 new cases of breast cancer in Ontario. (1) Assuming 75% are LN negative, of which 60% are ER positive and HER-2/*neu* negative, there are an estimated 3,825 cases for which Oncotype-DX might be considered. If the uptake of Oncotype-DX in such patients was 25%, the total budget impact would be \$4.14M per annum. By far the greater proportion of this (\$4.01M) represents the cost of providing Oncotype-DX itself, while the remaining \$0.13M represents additional chemotherapy costs. All of these costs may be scaled if uptake is expected to be greater than 25%.

Appendices

Appendix 1: Literature Search Strategies

Database: Ovid MEDLINE(R) <1996 to March Week 2 2010>

Search Strategy:

-
- 1 exp Breast Neoplasms/ (98362)
 - 2 ((breast* or mammar*) adj2 (cancer* or adenocarcinoma* or neoplas* or tumo?r* or carcinoma*)).ti,ab. (8172)
 - 3 1 or 2 (103989)
 - 4 exp Early Diagnosis/ (6450)
 - 5 (early or primary or stage I or stage 1 or stage II* or stage 2* or stage III* or stage 3*).ti,ab. (875308)
 - 6 4 or 5 (877524)
 - 7 3 and 6 (23770)
 - 8 exp Carcinoma, Intraductal, Noninfiltrating/ (2586)
 - 9 (dcis or ductal carcinoma in situ).ti,ab. (3024)
 - 10 8 or 9 (4365)
 - 11 7 or 10 (27099)
 - 12 exp Gene Expression Profiling/ (44957)
 - 13 (expression profil* or prognos* profil* or predict* profil* or mRNA expression or real-time polymerase chain reaction or reverse transcriptase polymerase chain reaction or RT-PCR or qRT-PCR or microarray* or predict* assay or prognos* assay or expression assay or predict* signature or prognos* signature or expression signature or gene signature or prognos* expression or predict* expression or gene classifier or molecular signature).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (229272)
 - 14 (oncotype or oncotypedx or nuvoselect or rotterdam signature or metastasis score or two gene ratio or 2 gene ratio or h?i ratio or h?i test or h?i test or h?i ratio or mammaprint or 21 gene assay or 14 gene signature or 76 gene assay or 70 gene profile or two-gene expression ratio or 76 panel or breast cancer gene expression ratio or HOXB13?IL17BR or bioclassifier or invasiveness gene signature or IGS or Sorlie-Perou classifier or theros or breast cancer index).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier] (1378)
 - 15 or/12-14 (230521)
 - 16 11 and 15 (2019)
 - 17 limit 16 to (english language and humans and yr="2006 -Current") (962)

Database: EMBASE <1980 to 2010 Week 10>

Search Strategy:

-
- 1 exp breast tumor/ (173087)
 - 2 ((breast* or mammar*) adj2 (cancer* or neoplas* or tumo?r* or adenocarcinoma* or carcinoma*)).ti,ab. (135317)
 - 3 1 or 2 (190093)
 - 4 exp early diagnosis/ (40409)
 - 5 (early or primary or stage I or stage 1 or stage II* or stage 2* or stage III* or stage 3*).ti,ab. (1215302)
 - 6 4 or 5 (1227627)
 - 7 3 and 6 (39903)
 - 8 exp intraductal carcinoma/ (2759)
 - 9 (dcis or ductal carcinoma in situ or intraductal carcinoma*).mp. [mp=title, abstract, subject headings,

- heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (5545)
- 10 8 or 9 (5545)
- 11 7 or 10 (44095)
- 12 exp gene expression profiling/ (19153)
- 13 (expression profil* or prognos* profil* or predict* profil* or mRNA expression or real-time polymerase chain reaction or reverse transcriptase polymerase chain reaction or RT-PCR or qRT-PCR or microarray* or predict* assay or prognos* assay or expression assay or predict* signature or prognos* signature or expression signature or gene signature or prognos* expression or predict* expression or gene classifier or molecular signature).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (189437)
- 14 (oncotype or oncotypedx or nuvoselect or rotterdam signature or metastasis score or two gene ratio or 2 gene ratio or h?i ratio or h?i test or h?i test or h?i ratio or mammaprint or 21 gene assay or 14 gene signature or 76 gene assay or 70 gene profile or two-gene expression ratio or 76 panel or breast cancer gene expression ratio or HOXB13?IL17BR or bioclassifier or invasiveness gene signature or IGS or Sorlie-Perou classifier or theros or breast cancer index).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name] (1728)
- 15 or/12-14 (191054)
- 16 11 and 15 (2287)
- 17 limit 16 to (human and english language and yr="2006 -Current") (1161)

Appendix 2: Laboratory Performance of Oncotype-DX

Table 1: Reliability of Oncotype-DX

Study	Year	Methods	Results
Cronin (12)	2007	Repeat analyses were conducted across multiple days, operators, RT-PCR plates, 7900HT instruments, and liquid-handling robots. Two aliquots of a single RNA sample were used for repeat testing. Mixed-effect ANOVA was used to estimate components of variance. REML of the components of assay variance were obtained.	Total SD was 0.792 RS units. Between-day SD was 0 RS units. Between-plate SD was 0 RS units. Within-plate SD was 0.792 RS units.
Habel (14)	2006	Pearson's correlation and ANOVA were used to assess between-block reproducibility. Repeat analyses were performed on 60 FFPE blocks that did not undergo macro-dissection from 20 women (2 to 5 blocks per patient).	Overall between-blocks SD was 3.0 RS units. For 16 of the 20 women, the between-blocks SD was less than 2.5 RS units.
Paik (5)	2004	Repeatability within blocks and reproducibility between blocks was assessed by repeating the assay in five serial sections from six FFPE blocks in two women.	Within-block RS SD was 0.72 RS unit (95% CI = 0.55 to 1.04) Total within-patient SD (including between and within-block SD) = 2.2 RS units.

Abbreviations: ANOVA, analysis of variance; FFPE, formalin-fixed paraffin-embedded; REML, restricted estimates of maximum likelihood; RS, recurrence score; SD, standard deviation.

Table 2: Success rates of Oncotype-DX and 21-gene home brew assays

Study		Protocol	Success
Akashi-Tanaka (6)	2009	qRT-PCR (Genomic Health)	43/78 (55.1)
Albain (26)	2010	qRT-PCR (Genomic Health)	367/430 (85.3)
Asad (22)	2008	qRT-PCR (Genomic Health)	Unable to estimate
Chang (9)	2008	qRT-PCR (Genomic Health)	80/97 (82.4)
Cobleigh(19)	2005	qRT-PCR (Genomic Health)	78/85 (91.8)
Dowsett (29)	2010	qRT-PCR (Genomic Health)	1308/2006 (65.2)
Espinosa (28)	2009	qRT-PCR (Home Brew)	Unable to estimate
Esteva (18)	2005	qRT-PCR (Genomic Health)	149/220 (67.7)
Fan (23)	2006	Microarray (Home Brew)	Unable to estimate
Geffen (24)	2009	qRT-PCR (Genomic Health)	Unable to estimate
Gianni (8)	2010	qRT-PCR (Genomic Health)	89/95 (93.7)
Goldstein (25)	2008	qRT-PCR (Genomic Health)	776/880 (88.2)
Habel (14)	2006	qRT-PCR (Genomic Health)	790/865 (91.3)
Kok (10)	2009	Microarray (Home Brew)	Unable to estimate
Lo (27)	2010	qRT-PCR (Genomic Health)	Unable to estimate
Mamounas (32)	2010	qRT-PCR (Genomic Health)	Unable to estimate
Mina (7)	2006	qRT-PCR (Genomic Health)	45/57 (78.9)
Oratz (16)	2007	qRT-PCR (Genomic Health)	72/74 (97.3)
Paik (5)	2004	qRT-PCR (Genomic Health)	668/754 (88.6)
Paik (15)	2006	qRT-PCR (Genomic Health)	Unable to estimate
Toi (31)	2010	qRT-PCR (Genomic Health)	280/313 (89.5)
Wolf (30)	2007	qRT-PCR (Genomic Health)	Unable to estimate

Appendix 3: Prognostic Value of Oncotype-DX

Table 1: Association of RS with specified outcome in women with early breast cancer receiving adjuvant hormone therapy alone

Author	Design	Population	Treatment	N (Event)	Primary Outcome	Risk of Outcome (95% CI)		Positive Test	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Multivariate Cox Proportional HR (95% CI)			
Lymph-Node Negative														
Paik 2004 (5)	Retro. analysis of pts from a single arm of an RCT	ER+ LN-HER2+/- ≤4 cm 18–70 yrs No prior chemo	Tamoxifen	668 (99)	Distant recurrence at 10 years	High	30.5 (23.6–37.4)	High	56 (45–66)	78 (74–81)	Multivariate Cox proportional HR (95% CI) for 50-point increment in continuous RS: 2.81 (1.70–4.64) P-value <0.001			
						IM	14.3 (8.3–20.3)	High + IM	77 (67–85)	55 (51–59)				
						Low	6.8 (4.0–9.6)							
						668 (NR)	Death due to any cause at 10 years	High	32.1	High		46 (38–56)	77 (74–81)	NR
						IM	22.4	High + IM	73 (64–80)	56 (52–60)				
						Low	10.0							
Paik 2006 (15)	Retro. subgroup analysis of an RCT	ER+ LN-HER2+/- ≤5 cm 18+ yrs No prior chemo	Tamoxifen	227 (27)	Distant recurrence at 10 years	High	39.5 (25.2–53.8)	High	70 (50–86)	86 (80–90)	NR			
						IM	9.1 (0.6–17.5)	High + IM	85 (66–96)	66 (58–72)				
						Low	3.2 (0.01–16.7)							
						Death due to any cause at 10 years	High	38.3	High	60 (41–77)		85 (79–90)	NR	
						IM	15.6	High + IM	83 (65–94)	66 (59–72)				
						Low	3.7							

Author	Design	Population	Treatment	N (Event)	Primary Outcome	Risk of Outcome (95% CI)		Positive Test	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Multivariate Cox Proportional HR (95% CI)		
Toi 2010 (31)	Retro. Cohort	ER+ LN-HER2+/- ≤5 cm Any age	Tamoxifen	200 (18)	Distant recurrence at 10 years	High	24.8 (15.7–37.8)	High	84 (60–96)	73 (66–79)	Multivariate Cox proportional HR (95% CI) for 50-point increment in continuous RS: 6.03 (2.17–16.7) P-value<0.001		
						IM	0 (N/A)	High + IM	84 (60–96)	51 (43–58)			
						Low	3.3 (1.1–10.0)						
						200 (NR)	Death due to any cause at 10 years	High	19.1 (11.3–31.3)	High	63 (39–83)	71 (63–77)	Multivariate Cox proportional HR (95% CI) for 50-point increment in continuous RS: 2.67 (0.93–7.62) P-value NR
								IM	2.6 (0.4–16.8)	High + IM	68 (43–86)	49 (42–57)	
								Low	6.4 (2.9–13.6)				
Dowsett 2009 (29)	Retro. analysis of pts from a single arm of an RCT	ER+ and/or PR+ LN-HER2+/- Postmeno. Any size	Tamoxifen or anastrozole or tamoxifen + anastrozole	872 (72)	Distant recurrence at 9 years	High	24.0 (17.0–34.0)	High	41 (30–52)	88 (85–90)	Multivariate Cox proportional HR (95% CI) for 50-point increment in continuous RS: With local grade 3.92 (2.08 – 7.39) P-value<0.001 With central grade 5.25 (2.84–9.73) P-value <0.001		
						IM	12.0 (8.0–18.0)	High + IM	74 (63–83)	62 (59–66)			
						Low	4.0 (3.0–7.0)						
						872 (121)	Death due to any cause at 9 years	High	27.0	High	26 (19–35)	87 (84–89)	NR
								IM	16.0	High + IM	54 (45–62)	61 (57–65)	
								Low	12.0				
Habel 2006 (14)	Nested case-control	ER+ LN-HER2+/- Any size <75 yrs No prio chemo	Tamoxifen	205 (55)	Death due to breast cancer at 10 years	High	15.5 (7.6–22.8)	High	31 (19–45)	87 (80–82)	Multivariate conditional logistic regression OR (95% CI): 5.3 (1.6–17.2) P-value <0.003		
						IM	10.7 (6.3–14.9)	High + IM	71 (57–82)	63 (55–71)			
						Low	2.8 (1.7–3.9)						

Author	Design	Population	Treatment	N (Event)	Primary Outcome	Risk of Outcome (95% CI)		Positive Test	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Multivariate Cox Proportional HR (95% CI)
Mamounas 2010 (32)	Retro. analysis of pts from single arms of multiple RCTs	ER+ LN- HER2+/- ≤5 cm 18+ yrs No prior chemo	Tamoxifen	895 (73)	Locoregional recurrence at 10 years	High	15.8 (10.4–21.2)	High	51 (39–63)	77 (74–80)	Multivariate Cox proportional HR (95% CI) for 50-point increment in continuous RS: 2.16 (1.26–3.69) P-value 0.007-
						IM	7.2 (3.4–11.0)	High + IM	71 (59–81)	55 (51–58)	
						Low	4.3 (2.3–6.3)				
Lymph-Node Positive											
Albain 2010 (26)	Retro. subgroup analysis of an RCT	ER+ and/or PR+ LN+ HER2+/- Any size Any age Postmen. No prior chemo	Tamoxifen	148 (66)	Disease event at 10 years (i.e., DFS)	High	57.9	High	47 (32–62)	75 (65–83)	Multivariate Cox proportional HR (95% CI) for 50-point increment in continuous RS: 2.64 (1.33–5.27) P-value 0.006
						IM	51.4	High + IM	79 (64–89)	45 (35–55)	
						Low	39.6				
				148 (47)	Death due to any cause at 10 years	High	48.6	High	45 (32–59)	75 (66–83)	Multivariate Cox proportional HR (95% CI) for 50-point increment in continuous RS: 4.42 (1.96–9.97) P-value 0.0006
				IM	31.9	High + IM	75 (61–85)	43 (34–53)			
						Low	22.9				
Dowsett 2009 (29)	Retro. analysis of pts from a single arm of an RCT	ER+ and/or PR+ LN+ HER2+/- Postmeno. Any size	Tamoxifen or anastrozole or tamoxifen + anastrozole	306 (74)	Distant recurrence at 9 years	High	49.0 (35.0–64.0)	High	32 (22–44)	88 (83–92)	Multivariate Cox proportional HR (95% CI) for 50-point increment in continuous RS: 3.47 (1.64 to 7.38) P-value 0.002
						IM	28.0 (20.0–39.0)	High + IM	65 (54–76)	58 (52–65)	
						Low	17.0 (12.0–24.0)				
				306 (90)	Death due to any cause at 9 years	High	46.0	High	25 (18–35)	87 (82–91)	NR
						IM	31.0	High + IM	56 (46–65)	56 (50–63)	
						Low	26.0				

Abbreviations: CI, confidence interval; DFS, disease-free survival; ER, estrogen receptor; HR, hazard ratio; IM, intermediate; LN, lymph node; NR, not reported; postmeno, postmenopausal; PR, progesterone receptor; pts, patients; retro, retrospective; RCT, randomized controlled trial; RS, recurrence score.

Table 2: Association of RS with specified outcome in women with early breast cancer receiving adjuvant tamoxifen plus chemotherapy

Author	Design	Population	Treatment	N (Event)	Primary Outcome	Risk of Outcome (95% CI)	Positive Test	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Multivariate Cox Proportional HR (95% CI)
Paik 2006 (15)	Retro. subgroup analysis of an RCT	ER+ LN- HER2+/- ≤5 cm 18+ yrs No prior chemo	Tamoxifen plus CMF or MF chemo	227 (27)	Distant recurrence at 10 years	High	High	41 (25–59)	74 (69–78)	NR
						IM	High + IM	71 (52–84)	53 (48–58)	
						Low		4.4 (1.4–7.3)		
Mamounas 2010 (32)	Retro. analysis of pts from single arms of multiple RCTs	ER+ LN- HER2+/- ≤5 cm 18+ yrs No prior chemo	Tamoxifen + CMF or MF chemo	424 (14)	Locoregional recurrence at 10 years	High	High	64 (36–86)	74 (69–78)	NR
						IM	High + IM	79 (49–94)	52 (47–57)	
						Low		1.6 (0–3.5)		
Goldstein 2008 (25)	Nested case-control	ER+ and/or PR+ LN+/- HER2+/- Any age ≥1.1 cm	Tamoxifen + docetaxel	465 (99)	Any recurrence at 5 years	NR	N/A	Unable to estimate	Unable to estimate	Multivariate Cox proportional HR (95% CI): Central grade 2.12 (0.97–4.65) <i>P</i> -value 0.06 Local grade 3.13 (1.60–6.14) <i>P</i> -value 0.0009 RS not predictive in HER-2/ <i>neu</i> -negative women (<i>P</i> -value NR but not significant)

Abbreviations: CI, confidence interval; ER, estrogen receptor; HR, hormone receptor; IM, intermediate; LN, lymph node; PR, progesterone receptor; RS, recurrence score.

Table 3: Association of RS with specified outcome in women with early breast cancer receiving mixed adjuvant therapies or no adjuvant therapy

Author	Design	Population	Treatment	N (Event)	Primary Outcome	Risk of Outcome (95% CI)	Positive Test	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Multivariate Cox Proportional HR (95% CI)	
Habel 2006 (14)	Nested case-control	ER+ LN- HER2+/- Any size <75 yrs No prio chemo	Observation only	361 (110)	Death due to breast cancer at 10 years	High	19.9 (14.2–25.2)	High	35 (26–44)	82 (77–87)	Multivariate conditional logistic regression OR (95% CI) 2.4 (1.1–5.2) P-value 0.025
						IM	17.8 (11.8–23.3)	High + IM	64 (54–72)	64 (57–70)	
						Low	6.2 (4.5–7.9)				
Mamounas 2010 (32)	Retro. subgroup analysis of multiple RCTs	ER+ LN- HER2+/- ≤5 cm 18+ yrs No prior chemo	Placebo	355 (53)	Locoregional recurrence at 10 years	High	18.4 (9.5–27.4)	High	34 (22–48)	73 (68–78)	NR
						IM	20.0 (9.9–30.0)	High + IM	66 (52–78)	50 (45–56)	
						Low	10.8 (5.8–15.8)				
Cobleigh 2005 (19)	Retro. cohort	ER+/- PR+/- LN+ ≥10 positive nodes HER2+/- Any size Any age	Observation only or tamoxifen and/or various adjuvant chemo regimens (proportions not clearly reported)	78 (55)	Distant recurrence or death at 10 years	High	80 (63–89)	High	69 (56–78)	57 (37–74)	NR
						IM	72 (38–88)	High + IM	95 (85–98)	35 (19–55)	
						Low	29 (0–53)				
Esteva 2005 (18)	Retro. cohort	ER+/- PR+/- LN- HER2+/- ≤5 cm Any age	Observation only	149 (NR)	Distant recurrence at 10 years	NR	N/A	Unable to estimate	Unable to estimate	Multivariate cox proportional HR found no significant correlation between RS and distant recurrence. Log-rank test found no significant difference between the Oncotype-DX risk groups in terms of distant recurrence.	

Author	Design	Population	Treatment	N (Event)	Primary Outcome	Risk of Outcome (95% CI)	Positive Test	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Multivariate Cox Proportional HR (95% CI)		
Kok 2008 (10)	Retro. cohort	ER+ LN+/- HER2 +/- Any size Any age	Observation only	69 (69)	10-year disease-free survival	NR	N/A	Unable to estimate	Unable to estimate	The distribution of DFS between RS risk classes was significant P-value 0.002		
Espinosa 2009 (28)	Retro. cohort	ER+ LN+/- ≤5 cm Any age	Tamoxifen (100%)	153 (34)	Distant recurrence at 5 years	High	31%	High	58 (41–73)	57 (47–66)	Multivariate Cox proportional HR (95% CI): High risk vs. Low risk 8.18 (1.77–37.86) P-value 0.007	
			No chemo (37%) CMF (27%) Anthracyclines (36%)			IM	19%	High + IM	68 (51–82)	42 (33–51)		
						Low	2%					
Fan 2006 (23)	Unclass.	ER+/- LN+/- HER2+/- <5 cm ≤52 yrs	No adjuvant therapy (56%)	295 (NR)	Recurrence-free survival	NR		N/A	Unable to estimate	Unable to estimate	Multivariate Cox proportional HR entire population: High risk vs. low risk 4.27 (2.05–8.92) P-value <0.001 ER+ only: High risk vs. low risk: 2.59 (1.44–4.65) P-value 0.001	
			Tam only (7%)			ER+ 225 (NR)						
			Tam + chemo (7%) Chemo only (30%)									
				295 (NR)	Overall survival	NR		N/A	Unable to estimate	Unable to estimate	Multivariate Cox proportional HR entire population: High risk vs. low risk: 6.14 (1.84–20.4) P-value 0.003 ER+ only: High risk vs. low risk: 4.95 (1.82–13.4) P-value 0.002	

Abbreviations: CI, confidence interval; DFS, disease-free survival; ER, estrogen receptor; HR, hazard ratio; IM, intermediate; LN, lymph node; NR, not reported; postmeno, postmenopausal; PR, progesterone receptor; retro, retrospective; RS, recurrence score; unclass, unclassified.

Appendix 4: GRADE Assessment of Quality of Evidence

Table 1: GRADE Assessment of Quality of Evidence Regarding Prognostic Value of Oncotype-DX

Studies	Quality Assessment					Summary of Findings	
	Design	Quality	Consistency	Directness	Other Modifying Factors	Effect Size	Overall Quality
Lymph-Node Negative							
Toi 2010 (31) Dowsett 2009 (29) Paik 2004 (5) Paik 2006 (15) (n=1,967) (# events=216)	Retrospective observational cohort study following women through the arms of a multisite hospital cohort / Retrospective analysis of a single arm of an RCT (-2) LOW	No major limitations LOW	No important inconsistency LOW	Generalizability issues: - HER-2/ <i>neu</i> LOW	No important imprecision or sparse data LOW	HRs for the association of distant recurrence and 50-point increment in RS: 6.03 (2.17–16.7) 3.92 (2.08–7.39) 2.81 (1.70–4.64) LOW	LOW
Lymph-Node Positive							
Albain 2010 (26) (n=148) (# events=66) Dowset 2009 (29) (n=306) (# events=74)	Retrospective analysis of a single arm of an RCT (-2) LOW	No major limitations LOW	No important inconsistency LOW	Generalizability issues: - HER-2/ <i>neu</i> - Postmenopausal women only LOW	No important imprecision or sparse data LOW	HR for the association of distant recurrence and 50-point increment in RS: 2.64 (1.33–5.27) 3.47 (1.64–7.38) LOW	LOW

Abbreviations: HER-2, Human Epidermal growth factor Receptor 2; HR, hazard ratio; RCT, randomized controlled trial; RS, recurrence score

Table 2: GRADE Assessment of Quality of Evidence Regarding Predictive Value of Oncotype-DX

Studies	Quality Assessment					Summary of Findings	
	Design	Quality	Consistency	Directness	Other Modifying Factors	Effect Size	Overall Quality
Lymph-Node Negative							
Paik 2006 (15) (n=651) (# events=62)	Retrospective subgroup analysis of an RCT	Limitations in statistical analysis and potential for selection bias (-1)	Moderate uncertainty in subgroup effect (-1)	Generalizability issues - HER-2/neu - Older chemo regimens	Imprecise or sparse data (-1)	HRs for interaction b/w chemo treatment and 50-point increment in RS: 0.32 (95% CI 0.11–0.94)	
	HIGH	MODERATE	LOW	LOW	VERY LOW	VERY LOW	VERY LOW
Lymph-Node Positive							
Albain 2010 (26;40) (n=367) (# events=143)	Retrospective subgroup analysis of an RCT	Limitations in statistical analysis and potential for selection bias (-1)	Moderate uncertainty in subgroup effect (-1)	Generalizability issues: - HER-2/neu - Postmenopausal women only - Older chemo regimens	Imprecise or sparse data (-1)	HRs for interaction b/w chemo treatment and 50-point increment in RS: 0.43 (95% CI 0.18–1.01)	
	HIGH	MODERATE	LOW	LOW	VERY LOW	VERY LOW	VERY LOW

Abbreviations: HER-2, Human Epidermal growth factor Receptor 2; HR, hazard ratio; RCT, randomized controlled trial; RS, recurrence score

Appendix 5: Subgroup Effect Checklist

Criteria

Design

1. Is the subgroup variable a characteristic measured at baseline or after randomization?
2. Is the effect suggested by comparisons within rather than between studies?
3. Was the hypothesis specified a priori?
4. Was the direction of the subgroup effect specified a priori?
5. Was the subgroup effect one of a small number of hypothesized effects tested?

Analysis

6. Does the interaction test suggest a low likelihood that chance explains the apparent subgroup effect?
7. Is the significant subgroup effect independent?

Context

8. Is the size of the subgroup effect large?
9. Is the interaction consistent across studies?
10. Is the interaction consistent across closely related outcomes?
11. Is there indirect evidence that supports the hypothesized interaction (biological rationale)?

Source: Sun et al. (40)

Results

Criteria	Study	
	Albain 2010	Paik 2006
1. Is the subgroup variable a characteristic measured at baseline or after randomization?	Yes (baseline)	Yes (baseline)
2. Is the effect suggested by comparisons within rather than between studies?	Yes	Yes
3. Was the hypothesis specified a priori?	Yes	Yes
4. Was the direction of the subgroup effect specified a priori?	Yes	Yes
5. Was the subgroup effect one of a small number of hypothesized effects tested?	Yes	Yes
6. Does the interaction test suggest a low likelihood that chance explains the apparent subgroup effect?*	No	No
7. Is the significant subgroup effect independent?	No	No
8. Is the size of the subgroup effect large?	No	Yes
9. Is the interaction consistent across studies?	No	No
10. Is the interaction consistent across closely related outcomes?	No	No
11. Is there indirect evidence that supports the hypothesized interaction (biological rationale)?	Yes	Yes

Note: Where data for a particular question of the checklist was unclear or not reported or when additional studies were not available, the study was considered not to meet the criteria in question.

*Rather than adopt a threshold for the *P*-value of the interaction term, a preferable way of assessing the *P*-value is that as it gets smaller, the subgroup hypothesis becomes increasingly credible such that we can begin to consider a *P*-value between 0.1 and 0.01 whereas we can take the hypothesis seriously when *P* values reach 0.001 or less. This becomes important when considering additional limitations that may affect the statistical strength of the interaction.

Source: Sun *et al.* (40)

References

- (1) Canadian Cancer Society's Steering Committee. Canadian cancer statistics 2009 [Internet]. Toronto: Canadian Cancer Society. 2009 Apr [cited: 2010 Mar 3]. 124 p. Available from: <http://tinyurl.com/d885qh>
- (2) Ghafoor A, Jemal A, Ward E, Cokkinides V, Smith R, Thun M. Trends in breast cancer by race and ethnicity. *CA Cancer J Clin* 2003; 53(6):342-55.
- (3) Fisher B, Jeong JH, Bryant J, Anderson S, Dignam J, Fisher ER et al. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. *Lancet* 2004; 364(9437):858-68.
- (4) Marchionni L, Wilson RF, Marinopoulos SS, Wolff AC, Parmigiani G, Bass EB et al. Impact of gene expression profiling tests on breast cancer outcomes. *Evid Rep Technol Assess (Full Rep)* 2007;(160):1-105.
- (5) Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004; 351(27):2817-26.
- (6) Akashi TS, Shimizu C, Ando M, Shibata T, Katsumata N, Kouno T et al. 21-Gene expression profile assay on core needle biopsies predicts responses to neoadjuvant endocrine therapy in breast cancer patients. *Breast* 2009; 18:171-4.
- (7) Mina L, Soule SE, Badve S, Baehner FL, Baker J, Cronin M et al. Predicting response to primary chemotherapy: gene expression profiling of paraffin-embedded core biopsy tissue. *Breast Cancer Res Treat* 2007; 103(2):197-208.
- (8) Gianni L, Zambetti M, Clark K, Baker J, Cronin M, Wu J et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. *J Clin Oncol* 2005; 23(29):7265-77.
- (9) Chang JC, Makris A, Gutierrez MC, Hilsenbeck SG, Hackett JR, Jeong J et al. Gene expression patterns in formalin-fixed, paraffin-embedded core biopsies predict docetaxel chemosensitivity in breast cancer patients. *Breast Cancer Res Treat* 2008; 108(2):233-40.
- (10) Kok M, Linn SC, Van Laar RK, Jansen MP, van den Berg TM, Delahaye LJ et al. Comparison of gene expression profiles predicting progression in breast cancer patients treated with tamoxifen. *Breast Cancer Res Treat* 2009; 113(2):275-83.
- (11) Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y,onso-Coello P et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336(7650):924-6.
- (12) Cronin M, Sangli C, Liu ML, Pho M, Dutta D, Nguyen A et al. Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer. *Clin Chem* 2007; 53(6):1084-91.
- (13) Lyman GH, Cosler LE, Kuderer NM, Hornberger J. Impact of a 21-gene RT-PCR assay on treatment decisions in early-stage breast cancer: an economic analysis based on prognostic and

predictive validation studies. *Cancer* 2007; 109(6):1011-8.

- (14) Habel LA, Shak S, Jacobs MK, Capra A, Alexander C, Pho M et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* 2006; 8(3):R25.
- (15) Paik S, Tang G, Shak S, Kim C, Baker J, Kim W et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006; 24(23):3726-34.
- (16) Oratz R, Paul D, Cohn AL, Sedlacek SM. Impact of a commercial reference laboratory test recurrence score on decision making in early-stage breast cancer. *J Oncol Pract* 2007; 3(4):182-8.
- (17) Hornberger J, Cosler LE, Lyman GH. Economic analysis of targeting chemotherapy using a 21-gene RT-PCR assay in lymph-node-negative, estrogen-receptor-positive, early-stage breast cancer. *Am J Manag Care* 2005; 11(5):313-24.
- (18) Esteva FJ, Sahin AA, Cristofanilli M, Coombes K, Lee SJ, Baker J et al. Prognostic role of a multigene reverse transcriptase-PCR assay in patients with node-negative breast cancer not receiving adjuvant systemic therapy. *Clin Cancer Res* 2005; 11(9):3315-9.
- (19) Cobleigh MA, Tabesh B, Bitterman P, Baker J, Cronin M, Liu ML et al. Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes. *Clin Cancer Res* 2005; 11(24 Pt 1):8623-31.
- (20) Cronin M, Pho M, Dutta D, Stephans JC, Shak S, Kiefer MC et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am J Pathol* 2004; 164(1):35-42.
- (21) Bryant J. Toward a more rational selection of tailored adjuvant therapy data from the National Surgical Adjuvant Breast and Bowel Project. 2005 St. Gallen Breast Cancer Symposium. [Complete slide presentation via Genomic Health]. 2005
- (22) Asad J, Jacobson AF, Estabrook A, Smith SR, Boolbol SK, Feldman SM et al. Does oncotype DX recurrence score affect the management of patients with early-stage breast cancer? *Am J Surg* 2008; 196(4):527-9.
- (23) Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006; 355(6):560-9.
- (24) Geffen DB, Amir N, Sion-Vardy N, Ariad S, Kachko L, Bayme M et al. Stage I breast cancer in a regional oncology practice in Israel 2002-2006: clinicopathologic features, risk estimation and planned therapy of 328 consecutive patients. *Breast* 2009; 18(5):316-21.
- (25) Goldstein LJ, Gray R, Badve S, Childs BH, Yoshizawa C, Rowley S et al. Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features. *J Clin Oncol* 2008; 26:4063-71.
- (26) Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* 2010; 11(1):55-65.

- (27) Lo SS, Mumby PB, Norton J, Rychlik K, Smerage J, Kash J et al. Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. *J Clin Oncol* 2010; 28(10):1671-6.
- (28) Espinosa E, Sanchez-Navarro I, Gamez-Pozo A, Marin AP, Hardisson D, Madero R et al. Comparison of prognostic gene profiles using qRT-PCR in paraffin samples: a retrospective study in patients with early breast cancer. *PLoS One* 2009; 4(6):e5911.
- (29) Dowsett M, Cuzick J, Wale C, Forbes J, Mallon EA, Salter J et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. *J Clin Oncol* 2010; 28(11):1829-34.
- (30) Wolf I, Ben-Baruch N, Shapira-Frommer R, Rizel S, Goldberg H, Yaal-Hahoshen N et al. Association between standard clinical and pathologic characteristics and the 21-gene recurrence score in breast cancer patients: a population-based study. *Cancer* 2008; 112(4):731-6.
- (31) Toi M, Iwata H, Yamanaka T, Masuda N, Ohno S, Nakamura S et al. Clinical significance of the 21-gene signature (Oncotype DX) in hormone receptor-positive early stage primary breast cancer in the Japanese population. *Cancer* 2010; 116(13):3112-8.
- (32) Mamounas EP, Tang G, Fisher B, Paik S, Shak S, Costantino JP et al. Association between the 21-gene recurrence score assay and risk of locoregional recurrence in node-negative, estrogen receptor-positive breast cancer: results from NSABP B-14 and NSABP B-20. *J Clin Oncol* 2010; 28(10):1677-83.
- (33) Flanagan MB, Dabbs DJ, Brufsky AM, Beriwal S, Bhargava R. Histopathologic variables predict Oncotype DX recurrence score. *Mod Pathol* 2008; 21(10):1255-61.
- (34) Geradts J, Bean SM, Bentley RC, Barry WT. The oncotype DX recurrence score is correlated with a composite index including routinely reported pathobiologic features. *Cancer Invest* 2010; 28(9):969-77.
- (35) Gwin K, Pinto M, Tavassoli FA. Complementary value of the Ki-67 proliferation index to the oncotype DX recurrence score. *Int J Surg Pathol* 2009; 17(4):303-10.
- (36) Fisher B, Dignam J, Wolmark N, DeCillis A, Emir B, Wickerham DL et al. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. *J Natl Cancer Inst* 1997; 89(22):1673-82.
- (37) Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M et al. Multie-gene RT-PCR assay for predicting recurrence in node negative breast cancer patients - NSABP studies B-20 and B-14 [Abstract]. *Breast Cancer Res Treat* 2003; 82:A16.
- (38) Albain KS, Barlow WE, Ravdin PM, Farrar WB, Burton GV, Ketchel SJ et al. Adjuvant chemotherapy and timing of tamoxifen in postmenopausal patients with endocrine-responsive, node-positive breast cancer: a phase 3, open-label, randomised controlled trial. *Lancet* 2009; 374(9707):2055-63.
- (39) Giacomini M, Abelson J, Goldsmith L, Levin L, DeJean D, Smith A. Social values and health technology policy analysis: the role for qualitative social science and humanities research evidence. Canadian Agency for Drugs and Technology in Health (CADTH) Annual Symposium;

April 19, 2010; Halifax, Nova Scotia. Giacomini M, Abelson J, Goldsmith L, Levin L, DeJean D, Smith A. Social Values and Health Technology Policy Analysis: The Role for Qualitative Social Science and Humanities Research Evidence. Canadian Agency for Drugs and Technology in Health (CADTH) Annual Symposium; April 19, 2010; Halifax, Nova Scotia.

- (40) Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010; 340:c117.
- (41) Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328(7454):1490.
- (42) Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 2009; 101(21):1446-52.
- (43) Mandrekar SJ, Sargent DJ. Predictive biomarker validation in practice: lessons from real trials. *Clin Trials* 2010; 7(5):567-73.
- (44) Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *J Clin Oncol* 2009; 27(24):4027-34.
- (45) Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: one size does not fit all. *J Biopharm Stat* 2009; 19(3):530-42.
- (46) Mandrekar SJ, Grothey A, Goetz MP, Sargent DJ. Clinical trial designs for prospective validation of biomarkers. *Am J Pharmacogenomics* 2005; 5(5):317-25.
- (47) Hornberger J, Cosler LE, Lyman GH. Economic analysis of targeting chemotherapy using a 21-gene RT-PCR assay in lymph-node-negative, estrogen-receptor-positive, early-stage breast cancer. *Am J Manag Care* 2005; 11(5):313-24.
- (48) Marchionni L, Wilson RF, Wolff AC, Marinopoulos S, Parmigiani G, Bass EB et al. Systematic review: gene expression profiling assays in early-stage breast cancer. *Ann Intern Med* 2008; 148(5):358-69.
- (49) Lyman GH, Cosler LE, Kuderer NM, Hornberger J. Impact of a 21-gene RT-PCR assay on treatment decisions in early-stage breast cancer: an economic analysis based on prognostic and predictive validation studies. *Cancer* 2007; 109(6):1011-8.
- (50) Tsoi DT, Inoue M, Kelly CM, Verma S, Pritchard KI. Cost-effectiveness analysis of recurrence score-guided treatment using a 21-gene assay in early breast cancer. *Oncologist* 2010; 15(5):457-65.
- (51) Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada [3rd Edition]. Ottawa: Canadian Agency for Drugs and Technologies in Health. 2006 Mar 74 p. Available from: http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf
- (52) Statistics Canada. Consumer Price Index, health and personal care, by province [Internet]. [updated 2010 Sep 27; cited 2010 Nov 29]. Available from: <http://www40.statcan.gc.ca/101/cst01/econ161a-eng.htm>
- (53) John B. Bryant J. Toward a more rational selection of tailored adjuvant therapy [Poster presentation]. Poster presented at the 9th International St. Gallen Oncology Conference on

Primary Therapy of Early Breast Cancer, 26-29 January 2005, St. Gallen, Switzerland.

- (54) Paik S, Tang G, Shak S, Kim C, Baker J, Kim W et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006; 24(23):3726-34.