

Health Quality Ontario

The provincial advisor on the quality of health care in Ontario

Ontario Health Technology Assessment Series

Ultrasound as an Adjunct to Mammography for Breast Cancer Screening: A Health Technology Assessment

KEY MESSAGES

Screening for breast cancer—the most common cancer in Canadian women—is the process of looking for the disease before symptoms appear so it can be treated early. Many factors can affect a woman’s risk of breast cancer, including age, a strong family history of the disease, and inherited genetic mutations. Women at average risk generally have a less than 15% chance of developing the disease over a lifetime; women at high risk have either an inherited genetic mutation or a greater than 25% (1 in 4) chance (using common risk assessment tools).

In Ontario, mammography (x-ray of the breast) is used to screen average-risk women. Women at high risk get two tests: mammography and magnetic resonance imaging (MRI), although some women are not able to have the MRI test. Ultrasound, an imaging method that uses sound waves, can be used to look for breast cancer missed by mammography alone.

This review looked at the impact of adding ultrasound to mammography for screening both average-risk and high-risk women. We wanted to see if doing both tests catches more breast cancers and saves lives, compared with mammography alone. We also wanted to know if ultrasound produces more false-positives (test results that show a woman has breast cancer when she does not), because false-positives can lead to unnecessary follow-up testing, treatment, and anxiety.

We found no studies on mammography plus ultrasound to screen average-risk women. Studies of high-risk women showed that screening with both tests found more breast cancers than mammography alone, but the combined screening also led to more false-positives. We found no studies looking at whether doing both tests reduces deaths from breast cancer.

We also looked at the costs of using ultrasound plus mammography for the small number of Ontario women who have high risk for breast cancer and cannot have an MRI. Publicly funding the combined screening in this population would add a very small cost to the provincial budget, from \$15,500 to \$30,250 each year for the next 5 years.

JULY 2016
VOL. 16, NO. 15

Let’s make our health system healthier

HEALTH TECHNOLOGY ASSESSMENT AT HEALTH QUALITY ONTARIO

This report was developed by a multidisciplinary team from Health Quality Ontario. The lead clinical epidemiologist was Milica Nikitovic-Jokic, the lead health economist was Hong Anh Tu with assistance from Stefan Palimaka, the medical librarians were Caroline Higgins and Corinne Holubowich, and the medical editor was Amy Zierler. Others involved in the development and production of this report were Irfan Dhalla, Nancy Sikich, Andree Mitchell, Claude Soulodre, Jessica Verhey, and Merissa Mohamed.

We are grateful to our expert advisors: Dr. Anna Chiarelli, Senior Scientist, Prevention and Cancer Control, and Provincial Scientific Lead, Ontario Breast Screening Program, Cancer Care Ontario, and Professor, Dalla Lana School of Public Health, University of Toronto; and Dr. Derek Muradali, Radiologist-in-Chief, Ontario Breast Screening Program, Cancer Care Ontario.

Citation

Health Quality Ontario. Ultrasound as an Adjunct to Mammography for Breast Cancer Screening: a health technology assessment. Ont Health Technol Assess Ser [Internet]. 2016 July;16(5):1-71. Available from: <http://www.hqontario.ca/Evidence-to-Improve-Care/Journal-Ontario-Health-Technology-Assessment-Series>.

ABSTRACT

Background

Screening with mammography can detect breast cancer early, before clinical symptoms appear. Some cancers, however, are not captured with mammography screening alone. Ultrasound has been suggested as a safe adjunct screening tool that can detect breast cancers missed on mammography. We investigated the benefits, harms, cost-effectiveness, and cost burden of ultrasound as an adjunct to mammography compared with mammography alone for screening women at average risk and at high risk for breast cancer.

Methods

We searched Ovid MEDLINE, Ovid Embase, EBM Reviews, and the NHS Economic Evaluation Database, from January 1998 to June 2015, for evidence of effectiveness, harms, diagnostic accuracy, and cost-effectiveness. Only studies evaluating the use of ultrasound as an adjunct to mammography in the specified populations were included.

We also conducted a cost analysis to estimate the costs in Ontario over the next 5 years to fund ultrasound as an adjunct to mammography in breast cancer screening for high-risk women who are contraindicated for MRI, the current standard of care to supplement mammography.

Results

No studies in average-risk women met the inclusion criteria of the clinical review.

We included 5 prospective, paired cohort studies in high-risk women, 4 of which were relevant to the Ontario context. Adjunct ultrasound identified between 2.3 and 5.9 additional breast cancers per 1,000 screens. The average pooled sensitivity of mammography and ultrasound was 53%, a statistically significant increase relative to mammography alone (absolute increase 13%; $P < .05$). The average pooled specificity of the combined test was 96%, an absolute increase in the false-positive rate of 2% relative to mammography screening alone. The GRADE for this body of evidence was low.

Additional annual costs of using breast ultrasound as an adjunct to mammography for high-risk women in Ontario contraindicated for MRI would range from \$15,500 to \$30,250 in the next 5 years.

Conclusions

We found no evidence that evaluated the comparative effectiveness or diagnostic accuracy of screening breast ultrasound as an adjunct to mammography among average-risk women aged 50 years and over.

In women at high risk of developing breast cancer, there is low-quality evidence that screening with mammography and adjunct ultrasound detects additional cases of disease, with improved sensitivity compared to mammography alone. Screening with adjunct ultrasound also increases the number of false-positive findings and subsequent biopsy recommendations. It is unclear if the use of screening breast ultrasound as an adjunct to mammography will reduce breast cancer-related mortality among high-risk women. The annual cost burden of using adjunct ultrasound to screen high-risk women who cannot receive MRI in Ontario would be small.

TABLE OF CONTENTS

LIST OF TABLES	6
LIST OF FIGURES	7
BACKGROUND	8
Clinical Need and Target Population	8
<i>Breast Cancer</i>	8
<i>Classifying Risk of Breast Cancer</i>	8
<i>Screening for Breast Cancer</i>	9
<i>Screening of High-Risk Women</i>	10
<i>Breast Cancer Screening in Ontario</i>	10
Technology/Technique	11
<i>Regulatory Status</i>	12
<i>Ontario Context</i>	12
Research Questions.....	13
CLINICAL EVIDENCE REVIEW	14
Objective of Analysis.....	14
Methods.....	15
<i>Sources</i>	15
<i>Literature Screening</i>	15
<i>Inclusion Criteria</i>	15
<i>Exclusion Criteria</i>	15
<i>Outcomes of Interest</i>	16
<i>Study Designs</i>	16
<i>Quality of Evidence</i>	18
<i>Expert Consultation</i>	18
Results	20
<i>Studies of Women at Average Risk for Breast Cancer</i>	21
<i>Studies of Women at High Risk for Breast Cancer</i>	21
<i>Excluded Studies</i>	21
<i>Findings From Studies of Average-Risk Women</i>	22
<i>Findings From Studies of High-Risk Women</i>	22
Conclusions.....	37
Discussion	37
<i>Study Limitations</i>	38
Existing Guidelines for Technology.....	40
ECONOMIC EVIDENCE REVIEW	41
Objectives.....	41
Methods.....	41
<i>Sources</i>	41
<i>Literature Screening</i>	41
<i>Inclusion Criteria</i>	41
<i>Exclusion Criteria</i>	41
<i>Outcomes of Interest</i>	42

<i>Data Extraction</i>	42
Results	42
<i>Literature Search</i>	42
Identification	43
Screening	43
Eligibility	43
Included.....	43
<i>Critical Review</i>	44
Conclusions.....	44
BUDGET IMPACT ANALYSIS	46
Objectives.....	46
Methods.....	46
<i>Target Population</i>	46
<i>Resource</i>	46
<i>Canadian Costs</i>	48
<i>Scenario Analyses</i>	49
<i>Cost Analysis</i>	49
Results	50
<i>Base Case</i>	50
<i>Scenarios Results</i>	50
Limitations	51
Discussion and Conclusions	51
ABBREVIATIONS	53
GLOSSARY	54
APPENDICES.....	55
Appendix 1: Literature Search Strategies	55
<i>Clinical Evidence Review</i>	55
<i>Economic Evidence Review</i>	56
Appendix 2: Diagnostic Accuracy Definitions and Calculations.....	59
Appendix 3: Excluded Systematic Reviews of Studies of Women With Dense Breasts and no Risk Specified.....	60
Appendix 4: Evidence Quality Assessment	61
Appendix 5: Supplemental Results	63
REFERENCES	66

LIST OF TABLES

Table 1: Summary of Ontario Breast Screening Program Guidelines	11
Table 2: Automated Breast Ultrasound Devices Approved by Health Canada	12
Table 3: Summary of Included Systematic Review for Average-Risk Women	22
Table 4: Summary of Included Systematic Review for High-Risk Women	23
Table 5: Summary of High-Risk Populations in Included Primary Studies	25
Table 6: Summary of Screening and Diagnosis Methods in Included Primary Studies	26
Table 7: Diagnostic Yield of Adjunct Ultrasound Compared With Mammography Alone	28
Table 8: Characteristics of Tumours Identified by Mammography or Adjunct Ultrasound Alone	29
Table 9: Sensitivity and Specificity of Mammography and Adjunct Ultrasound Compared With Mammography Alone in Women at High Risk for Breast Cancer	31
Table 10: Pooled Sensitivity and Specificity of Mammography and Adjunct Ultrasound Compared With Mammography Alone in Women at High Risk for Breast Cancer	32
Table 11: Positive Predictive Value of Screening by Mammography and Ultrasound Compared With Mammography Alone in Women at High Risk for Breast Cancer	33
Table 12: Recall and Biopsy Rates for Screening Mammography and Adjunct Ultrasound Compared With Mammography Alone in Women at High Risk for Breast Cancer	35
Table 13: Summary of Findings for Studies Using BI-RADS Score ≥ 4	36
Table 14: Summary of Findings for Study Using BI-RADS Score 3, 4, or 5	36
Table 15: Clinical Guideline Recommendations on the Use of Screening Breast Ultrasound ...	40
Table 16: Results of Economic Literature Review—Summary	45
Table 17: Number of High-Risk Women Screened by Mammography Plus Ultrasound, Actual and Forecast, 2011/12 to 2019/20	47
Table 18: Expected Number of Abnormal Ultrasound Results After Screening by Mammography Plus Ultrasound, Ontario Breast Screening Program, 2015/16 to 2019/20	47
Table 19: Estimated Number of Follow-Up Procedures for High-Risk Women Screened by Mammography Plus Ultrasound, 2015/16 to 2019/20	48
Table 20: Breast Screening Costs	48
Table 21: Costs of Follow-Up Procedures After Abnormal Results of Screening Breast Ultrasound	48
Table 22: Estimated Breast Cancer Screening Costs for High-Risk Women Contraindicated for MRI, Ontario, 2015 to 2019/20	50
Table 23: Additional Budget Required to Fund Breast Cancer Screening for High-Risk Women Contraindicated for MRI, Assuming Abnormal Results Are Detected by Ultrasound Only	50
Table 24: Additional Budget Required to Fund Breast Cancer Screening for High-Risk Women Contraindicated for MRI, Assuming Abnormal Results Are Detected by Both Mammography and Ultrasound	51
Table A1: Summary of Calculations Used to Assess Diagnostic Performance	59
Table A2: Summary of Systematic Reviews of Screening Breast Ultrasound as an Adjunct to Mammography in Women with Dense Breasts	60
Table A3: AMSTAR Scores of Included Systematic Reviews	61
Table A4: Risk of Bias for Studies of Screening Breast Ultrasound (QUADAS-2)	61
Table A5: GRADE Evidence Profile for Comparative Diagnostic Accuracy of Mammography and Mammography With Adjunct Breast Ultrasound	62
Table A6: Summary of Screening Values in Primary Studies	63
Table A7: Diagnostic Outcomes of Adjunct Ultrasound Compared With Mammography Alone for High-Risk Women, Subgrouped by Personal History of Breast Cancer	64

Table A8: Area Under the Receiver Operating Characteristic Curve for Mammography Plus
 Ultrasound Screening Compared With Mammography Alone for Women at High Risk
 for Breast Cancer65

LIST OF FIGURES

Figure 1: Framework for Screening with Mammography and Adjunct Breast Ultrasound 14
 Figure 2: PRISMA Flow Diagram for Clinical Evidence Review20
 Figure 3: Plot of Change in Sensitivity and Specificity with Addition of Screening Ultrasound
 Relative to Mammography Alone in Women at High Risk for Breast Cancer32
 Figure 4: PRISMA Flow Diagram for Economic Evidence Review.....43
 Figure A1: Forest Plots for Individual Study Sensitivity and Specificity63

BACKGROUND

Clinical Need and Target Population

Breast Cancer

Breast cancer is the most common cancer among Canadian women, with an estimated 1 in 9 women expected to develop the disease during their lifetime.¹ In Ontario, an estimated 9,500 women are diagnosed and 1,950 will die from breast cancer annually.¹

Most breast cancers are invasive, meaning the cancer invades the surrounding tissue of the breast. Invasive breast cancers can metastasize (spread) to the lymph nodes and other parts of the body. Some women will be diagnosed with a non-invasive breast cancer, meaning abnormal cells have not spread to neighbouring breast tissue. The most common non-invasive breast cancer is ductal carcinoma in situ (DCIS, lesions in the milk ducts). The natural history of DCIS is poorly understood, and it not known which lesions could become invasive.^{2,3}

Treatment options for breast cancer vary depending on the stage of the disease and the cancer pathology. Treatment often involves a combination of surgery, hormone therapy, chemotherapy, and/or radiation therapy.

Classifying Risk of Breast Cancer

The risk of developing breast cancer increases with age, and more than half of breast cancer cases occur in average-risk women between the ages of 50 and 74 years. Various other factors can increase the risk of developing breast cancer. These include a family or personal history of breast or ovarian cancer, extremely dense breast tissue, age at menopause, and lifestyle factors. The strongest known risk factor for breast cancer is hereditary, resulting from gene mutations (changes) inherited from a parent. The most common hereditary breast cancers are due to mutations in the *BRCA* genes. Women with *BRCA1* or *BRCA2* mutations are estimated to have a 40% to 80% lifetime risk of developing breast cancer.^{5,6}

Lifetime risk of breast cancer can be determined based on genetic assessment and common risk assessment tools such as the International Breast Cancer Intervention Study tool and the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm.

The classifications of breast cancer risk are multifaceted and are not standardized across countries or the literature. Women are often considered to be at *average risk* for breast cancer if they have less than a 15% risk of developing the disease over their lifetime.⁴ Women with a risk between 15% and 25% are considered to be at *higher than average* or *intermediate risk* for developing the disease.

Women at *high risk* for breast cancer are generally defined as having one or more of the following risk factors: known mutation carrier, untested first-degree relative of a mutation carrier, a strong degree of high-risk family histories, or a 25% or greater lifetime risk for breast cancer.⁴ Women who had radiation therapy to the chest before the age of 30 and more than 8 years ago are also categorized as high risk.⁴ It is estimated that less than 1% of the general population are at high risk for breast cancer and that about 5% of all breast cancers are due to inherited genetic mutations.⁴ Women at high risk for breast cancer often develop the disease at a younger age, and their cancers tend to grow faster and be more aggressive.^{7,8} Some women at high risk may choose preventative options to reduce their risk of breast cancer through

chemoprevention or prophylactic mastectomy (surgical removal of all or part of the breasts) or oophorectomy (surgical removal of ovaries).

Screening for Breast Cancer

Breast cancer screening is the regular examination of healthy, asymptomatic women. The intent of breast screening programs is to identify breast cancer early so that women can receive timely and effective treatment. Cancers identified and treated at earlier stages tend to have better prognosis than those that have progressed or metastasized.¹ The ultimate goal of breast cancer screening is to reduce breast cancer–related deaths as well as the morbidity associated with advanced stages of the disease.

A successful screening program must also aim to minimize any adverse consequences associated with the screening itself. Screening for breast cancer can pose many challenges. Because no test is perfect, all screening tests have the potential to produce false test results, both false-negative as well as false-positive. False-negative tests—tests that indicate a person does not have the disease when they actually do—may delay necessary treatment. False-positive tests—tests that indicate a person has the disease when they do not—will lead to additional unnecessary testing to confirm the diagnosis. This may include diagnostic mammography, ultrasound, and surgical biopsy, all of which pose their own risks. False-positive tests can also lead to serious distress, anxiety, and uncertainty for patients.^{9,10} Overdiagnosis and overtreatment are other potential risks in breast cancer screening. Some cancers detected by screening may never cause symptoms or become life-threatening. There is currently no definitive way of determining which of these screening-detected cancers will progress, meaning that some women may undergo treatment with surgery, radiation therapy, or chemotherapy that may not be needed.

The primary method used for breast cancer screening is mammography, which uses low-dose x-rays to image the breast, either on film or digitally. Mammography is currently the only screening tool for breast cancer that reduces breast cancer–related deaths through early detection for average-risk women aged 50 to 74 years.¹¹ However, recent reviews have suggested that screening with mammography may not be as effective for this population as originally thought and can result in significant overdiagnosis and overtreatment.¹² For younger average-risk women (aged 40 to 49 years), several reviews have found that mammography is not an effective tool for breast cancer screening.^{11,13}

Mammography is not a perfect test, and several factors such as younger age and increased breast density can decrease its diagnostic accuracy. A high proportion of dense breast tissue (fibrous and glandular tissue) can make it more difficult to detect cancer on mammography. Approximately 40% of all women are estimated to have *heterogeneously dense* breasts (50% to 74% dense tissue) and 10% have *extremely dense* breasts ($\geq 75\%$ dense tissue).¹⁴ Extremely high breast density has also been suggested to increase the risk of breast cancer, although there is considerable debate about the potential correlation between breast density and rates of interval cancer (cancers that are diagnosed between screening rounds).^{15,16} Increased breast density is directly related to younger age: approximately 53% of premenopausal and 23% of postmenopausal women have at least 50% dense tissue. A review by Health Quality Ontario found that digital mammography is more sensitive than film mammography among women with heterogeneously or extremely dense breast tissue.¹⁷

Screening of High-Risk Women

For women at high risk for breast cancer, prior reviews have identified no known published research evaluating the impact of screening on breast cancer–related mortality.^{17,18} Nonetheless, early screening is recommended for women at high risk, based on the high rates of breast cancer in this population and the potential benefits of detecting tumours while they are small and have not become invasive.

Screening with mammography alone has been shown to have significantly poorer diagnostic performance among high-risk women than in the general population. For high-risk women, mammography has lower sensitivity (the rate of true-positive test results) and a higher rate of interval cancers that have often spread to lymph nodes. This has been attributed to many factors among high-risk women including younger age at onset of cancer, higher breast density, and increased tumour growth. Evidence has suggested that breast cancers associated with specific genetic mutations are also more often not visible on the mammogram or misclassified as benign compared with sporadic cancers (cancers not linked to high-risk mutations) because the two types of cancer have both histological and biological features that differ.^{7,8} Cancers associated with *BRCA1* and *BRCA2* mutations are also more likely to present as invasive cancer rather than as DCIS.¹⁹

Supplemental screening with magnetic resonance imaging (MRI) has been shown to significantly improve the detection of breast cancer in women at high risk, compared to mammography alone.^{17,20,21} Sensitivity improves from 32% with mammography alone to 84% with the combination of mammography and MRI.²¹ Although MRI is associated with a higher false-positive rate and no direct evidence of an improvement in mortality, MRI plus mammography has become the standard practice for breast cancer screening in high-risk women in several jurisdictions, including Ontario.

Some women, however, are contraindicated for MRI due to factors such as having severe claustrophobia or anxiety, a metallic implant (e.g., pacemaker), or allergies to the contrast agents that are injected into patients. Women who have high risk of breast cancer and cannot use MRI may benefit from other screening technologies such as ultrasound to compensate for the limitations of mammography screening alone.

Breast Cancer Screening in Ontario

Screening for breast cancer can be done either as part of an organized program or opportunistically (when requested by the patient or offered by a health care provider at a routine or unrelated health care visit). The Ontario Breast Screening Program is a province-wide, organized screening program for breast cancer.⁴ Table 1 summarizes the program's current recommendations for screening mammography.⁴ Average-risk women are offered screening with mammography every two years, while women with higher than average risk (but not at high risk), including those with extremely dense breasts, are offered screening with mammography annually. High-risk women are currently screened with both mammography and MRI when possible.

Table 1: Summary of Ontario Breast Screening Program Guidelines

Age, Years	Risk	Screening Tests and Frequency
Women < 50	Average	Screening is not recommended
Women 50 to 74	Average	Mammography every 2 years
Women 50 to 74	Higher than average ^a	Annual mammography
Women 30 to 69	High	Annual mammography and breast MRI, or screening breast ultrasound if MRI is contraindicated ^b

Abbreviation: MRI, magnetic resonance imaging.

^aDocumented pathology of high-risk lesions, a personal history of ovarian cancer, 2 or more first-degree female relatives with breast cancer at any age, 1 first-degree relative with breast cancer under age 50, 1 first-degree relative with ovarian cancer at any age, breast density greater than 75% as seen on mammogram (reassessed annually by a screening radiologist).

^bContraindications include metallic implants (e.g., pacemakers or aneurysm clips), contrast allergies, and claustrophobia. There are also size and weight restrictions to using the MRI machines.

Approximately 1.15 million women in Ontario aged 50 to 74 years (59% of those eligible) were screened for breast cancer with mammography between 2012 and 2013. Of these women, 76% were screened through the Ontario Breast Screening Program.²² In 2014/15, nearly 6,000 Ontario women at high risk for breast cancer between the ages of 30 and 69 years were screened with a combination of mammography, MRI, and/or ultrasound (data provided by Cancer Care Ontario, 2015).

Technology/Technique

Breast ultrasound (also known as sonography or ultrasonography) is a non-invasive test that assesses the breast tissue through the use of high-frequency sound waves bounced off the breast and converted to images on a screen.

Breast ultrasound can be done using either a hand-held or an automated device. Hand-held ultrasonography involves the manual use of a small transducer and ultrasound gel placed directly on the skin, with representative images obtained by the operator. Newer automated breast ultrasound systems, called whole-breast ultrasound, separate the imaging process from interpretation.¹⁵ Unlike hand-held ultrasound systems, automated systems are not dependent on the operator for image selection and allow radiologists to review the entire dataset for interpretation. Various automated ultrasound systems exist, with several designs, image acquisition approaches, and workstation setup and features. All automated systems allow imaging of the whole breast, with some systems providing both 2-dimensional and 3-dimensional images. The primary drawback to automated breast ultrasound systems is the volume of data acquired during scanning and the corresponding time required to read the scans. Women with large breasts are also more difficult to image with either type of ultrasound system.

This technology is used both for screening and for diagnosis of breast cancer—for example, to evaluate breast lumps or abnormalities found by mammography, breast MRI, or clinical breast exam, or to guide a biopsy procedure. As a breast cancer screening tool, ultrasound can be done either as a sequential screening (a follow-up test when women have had a negative or inconclusive mammogram) or as a simultaneous test (a test done in parallel with mammography).

Adjunct (supplemental) screening with ultrasound has the potential to detect breast cancers that may not be visible on mammography. The use of adjunct ultrasound is thought to be a safe and

inexpensive approach to improve the sensitivity of breast cancer screening with mammography alone, although potentially at the expense of increasing the rate of false-positive findings and the subsequent risk of increased patient anxiety, overdiagnosis, and overtreatment.

Regulatory Status

Ultrasound systems are approved by Health Canada as Class II medical devices. Numerous hand-held ultrasound systems are available in Canada, most of which are licensed as general systems that can be used on the breast but are not specifically indicated for breast cancer diagnosis or screening.

Three automated breast ultrasound systems are currently approved by Health Canada (Table 2). Two have limited approval and are intended only for use as an adjunct to mammography, rather than a replacement for mammography. These devices are marketed primarily for imaging dense breast tissue.

Table 2: Automated Breast Ultrasound Devices Approved by Health Canada

Technology	Device Number	Health Canada Approved Indication
Sofia Automated Tomographic Ultrasound Device	79608	Indicated for use as a B-mode ultrasonic imaging system for imaging of a patient's breast when used with an automatic scanning linear array transducer
Somo-v Automated Breast Ultrasound System	74905	Intended for use as an adjunct to mammography to provide physicians with an increase in the sensitivity of breast cancer detection in diagnosing symptomatic and screening asymptomatic women; the device is not intended to be used as a replacement for screening mammography
SonoCiné Automated Whole Breast Acquisition Screening System	87616	Indicated as an adjunct to mammography for screening asymptomatic women for breast cancer

Ontario Context

The Schedule of Benefits for Physician Services in Ontario includes a fee code for breast ultrasound only for diagnostic imaging.²³ The use of this code for breast cancer screening in average-risk women is not considered appropriate, but it is acceptable when the screening is for high-risk women for whom MRI is contraindicated (Ministry of Health and Long-Term Care, personal communication, April 2015). Some private facilities in Ontario currently provide screening with automated breast ultrasound devices outside of the Ontario Health Insurance Plan.

According to experts, the use of screening breast ultrasound for women who are *not* at high risk of breast cancer is increasing in Ontario. An assessment of the effectiveness of screening breast ultrasound as an adjunct to mammography is needed to either support or discourage the use of this technology in women who are not at high risk of developing breast cancer.

For high-risk women unable to have an MRI, the effectiveness of screening with adjunct ultrasound instead of MRI is unknown. A review of data from the Ontario Breast Screening Program from July 2011 to June 2012 found that 2.2% of high-risk women screened had received an ultrasound without the use of screening breast MRI.²⁴

Research Questions

- 1) What are the effectiveness, safety, and diagnostic accuracy of breast ultrasound as an adjunct to mammography for breast cancer screening compared with mammography alone for women at *average risk* for breast cancer?
- 2) What are the effectiveness, safety, and diagnostic accuracy of breast ultrasound as an adjunct to mammography for breast cancer screening compared with mammography alone for women at *high risk* for breast cancer?
- 3) What is the cost-effectiveness of ultrasound as an adjunct to mammography compared with mammography alone in breast cancer screening for women of both average and high risk?
- 4) What is the estimated cost in Ontario of publicly funding ultrasound as an adjunct to mammography in breast cancer screening for high-risk women contraindicated for MRI?

CLINICAL EVIDENCE REVIEW

Objective of Analysis

The objective of this analysis was to determine the effectiveness, harms and diagnostic accuracy of screening breast ultrasound as an adjunct to screening breast mammography in women at average or high risk for developing breast cancer.

Figure 1 displays the general screening pathway and framework for our research questions. The overarching question is whether adjunct screening with ultrasound improves patient-important outcomes relative to screening with mammography alone. Improvement in patient outcomes can be associated with harms related to screening tests, diagnostic tests, or treatment. False-positive results can lead to unnecessary testing, surgery, or treatment, and false-negative tests can result in more aggressive and difficult to treat cancers.

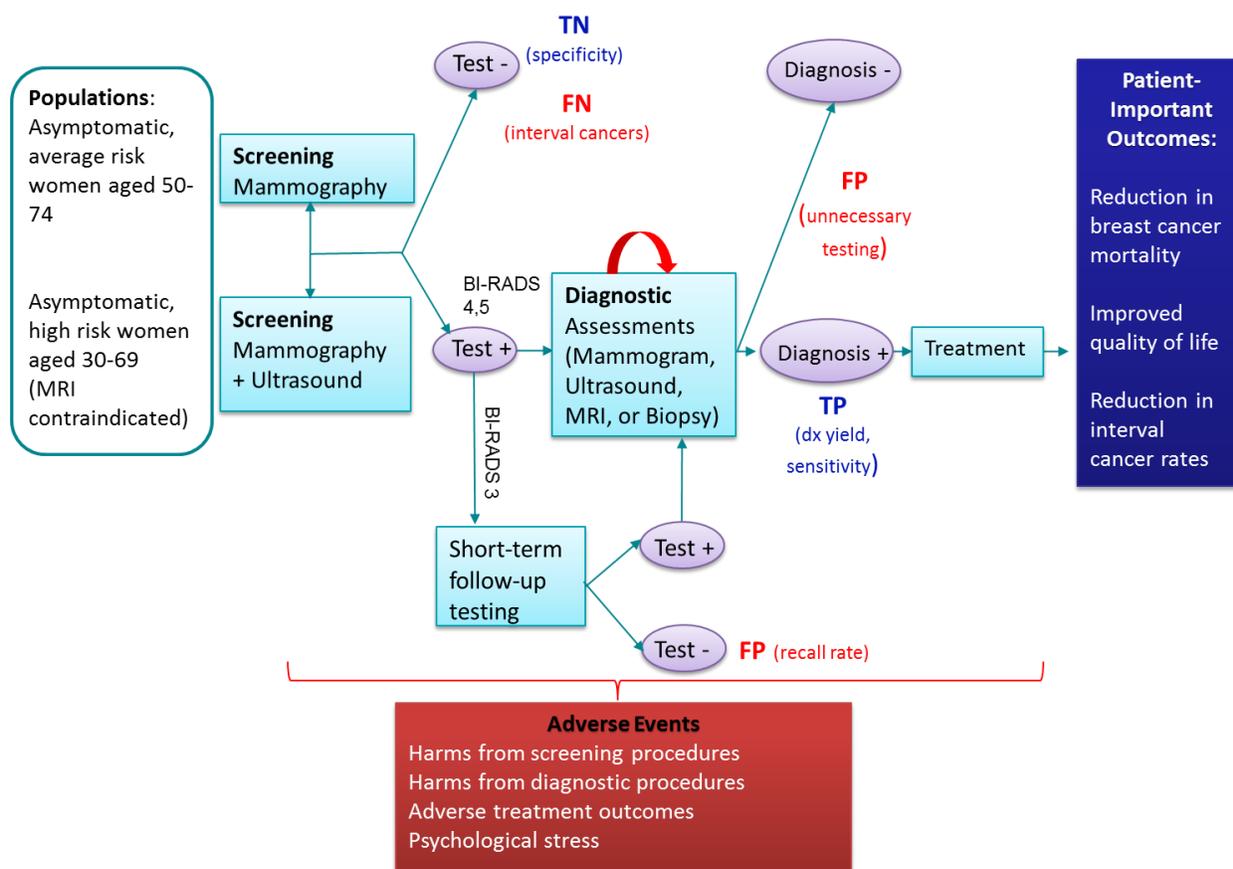


Figure 1: Framework for Screening with Mammography and Adjunct Breast Ultrasound

Abbreviations: BI-RADS, Breast Imaging Reporting and Data System; Dx, diagnostic; FN, false-negative; FP, false-positive; MRI, magnetic resonance imaging; TN, true-negative; TP, true-positive; +, positive; -, negative.

Methods

Sources

We performed a literature search on June 18, 2015, using Ovid MEDLINE, Ovid MEDLINE In-Process and Other Non-Indexed Citations, Ovid Embase, and EBM Reviews, for studies published from January 1, 1998, to June 18, 2015. (Appendix 1 provides details of the search strategies.)

Literature Screening

A single reviewer reviewed the abstracts and, for those studies meeting the eligibility criteria, we obtained full-text articles. We also examined reference lists for any additional relevant studies not identified through the search.

Inclusion Criteria

- English-language full-text publications
- Published between January 1, 1998, and June 15, 2015
- Studies of asymptomatic women aged 50 years and older at *average lifetime risk* for breast cancer
 - defined as less than 15% lifetime risk of breast cancer, or studies excluding high-risk women as defined below, or studies of women with dense breasts and no additional high-risk factors were included;
- Or studies of asymptomatic women at *high lifetime risk* for breast cancer
 - defined as carriers (or first-degree relatives) of the breast cancer mutation genes, chest radiation prior to age 30, $\geq 25\%$ lifetime risk of breast cancer, or as defined in research articles
- Studies assessing screening breast ultrasonography as an adjunct to screening mammography (provided simultaneously or sequentially to mammography)
- Studies using pathology results from biopsy as a reference standard for true-positive tests and a minimum of clinical follow-up for women with negative imaging results
- Studies reporting on one or more outcomes of interest
 - for studies reporting only on diagnostic performance (yield or accuracy), sufficient information to construct a 2 x 2 table (true-positives, true-negatives, false-positives, false-negatives) was required

Exclusion Criteria

- Studies among symptomatic women (e.g., clinical symptoms or palpable breast mass prior to enrollment)
- Studies only among women with a personal history of breast cancer
- Studies where population risk for breast cancer is not specified or results are not stratified by included population risk groups
- Studies comparing ultrasound alone to mammography alone as a primary screening modality
- Case series, case reports, abstracts, editorials, non-systematic reviews

Outcomes of Interest

Effectiveness and Harms

- Breast cancer mortality
- All-cause mortality
- Number needed to screen to prevent one additional death
- Health-related quality of life
- Screening-related harms

Diagnostic Performance

- Incremental diagnostic yield (incremental cancer detection rate)
 - Cancer and tumour characteristics: tumour size, invasiveness, lymph-node status
- Sensitivity (true-positive rate)
- Specificity (true-negative rate)
- False-negative rate
- False-positive rate
- Positive predictive value (the proportion of all positive results that were true-positives) among women who tested positive for disease and among women who received a follow-up biopsy
- Biopsy rate and recall rate

Study Designs

- Systematic reviews and health technology assessments
- For primary studies, we used the following hierarchical approach based on study design:
 - For effectiveness outcomes*
 - 1) Randomized controlled trials and prospective, comparative studies
 - For diagnostic performance*
 - 1) Randomized controlled trials and prospective, comparative studies; paired study designs were considered the ideal design for observational studies²⁵
 - 2) Prospective, non-comparative studies (including studies of ultrasound among women with negative mammography) and retrospective, comparative studies
 - 3) Retrospective, non-comparative studies

We contacted authors via email where there were missing or incomplete data reported, or where clarification was needed regarding study populations or outcome definitions.

Statistical Analysis

Effectiveness Outcomes

For effectiveness outcomes, we planned to calculate the risk ratio or odds ratio with 95% confidence intervals for each outcome. For similar studies with minimal clinical heterogeneity, we planned to pool outcomes using Review Manager 5.3.

Diagnostic Performance

A summary of the definitions and formulas we used to calculate diagnostic performance is provided in Appendix 2. We calculated diagnostic yield as the total number of true-positive cancers identified over the total number of screens in the study. The number needed to screen to identify an additional case of cancer was calculated as the inverse of the incremental diagnostic yield between tests at a given prevalence.

To assess the diagnostic accuracy of each test, we constructed 2 x 2 tables (true-positives, false-positives, true-negatives, false-negatives). We reported calculations of incremental diagnostic yield, sensitivity, specificity, positive predictive value, biopsy rates, and recall rates as provided in the research articles. When the study did not directly report results of interest, we calculated outcomes for each intervention based on data provided in the articles. Confidence intervals around estimates for individual tests were calculated using the binomial Clopper-Pearson exact method based on a beta distribution. We did not calculate the statistical differences between tests within individual studies because the studies did not provide enough data to account for the paired nature of the data and repeated measures among individual study participants.

Given that sensitivity and specificity are dependent on the threshold used to define a positive test, we stratified the results of imaging based on the study's assignment of Breast Imaging Reporting and Data System (BI-RADS) categories, as defined by the American College of Radiology. The current BI-RADS definitions are²⁶:

- 0 Incomplete assessment; additional imaging or review of prior images is needed
- 1 Negative
- 2 Benign finding
- 3 Probably benign finding; short interval follow-up is suggested
- 4 Suspicious abnormality; biopsy should be considered
- 5 Highly suggestive of malignancy; appropriate action should be taken

We accepted the BI-RADS categorization of positive and negative tests for the detection of cancer as reported in the studies. When it was not reported but the study provided sufficient primary data, we classified data as a positive test for BI-RADS categories 0, 4, and 5 and negative for categories 1, 2, and 3. We did not classify BI-RADS category 3 tests as positive for the detection of cancer, but summarized and reported these tests as a harm related to unnecessary follow-up procedures when disease was not present. Where possible, we calculated positive predictive value for women recalled on testing (BI-RADS 3, 4, 5) and for women recommended for biopsy (BI-RADS 4, 5). If the study authors did not directly report combined results for mammography plus adjunct ultrasound but did provide sufficient primary data, we calculated the combined outcome as test-positive if either test was positive, and test-negative only when both tests were negative.

Using Review Manager 5.3, we plotted sensitivity and specificity for each test within each paired study, in the receiver operating characteristic space as well as on forest plots to explore study variations and heterogeneity. Where we found sufficient clinical and methodological homogeneity, R version 3.0.2 software was used to pool studies and calculate the summary estimates of sensitivity and specificity, and their 95% confidence intervals, using the bivariate

model. The bivariate model allows sensitivity and specificity to be jointly analyzed and incorporates correlations between the two measures using a random effects approach.^{25,27}

We compared tests to one another by adding a covariate for test type to the bivariate model to assess whether average sensitivity and/or specificity differed between the tests.²⁷ To assess statistical significance of differences in sensitivity and specificity between tests, we used the Wald test.

We evaluated the impact of screening with mammography alone in comparison with mammography and adjunct ultrasound by assessing the incremental diagnostic yield, the number needed to screen to detect 1 additional cancer, and the number of additional false-positives, based on the summary estimates obtained from the meta-analysis and on the range of prevalence data within the individual studies.

Subgroup Analyses

Where possible, we analyzed results by subgroup based on the following factors:

- screening round—first round (prevalence screen) versus subsequent rounds (incident screens)
- breast density—high density (extremely or heterogeneously dense) versus less than high density (scattered density or fatty breasts)
- personal history of breast cancer—no prior history versus personal history
- type of mammography—digital versus film

Publication Bias

We planned to assess publication bias using funnel plot methodology or statistical tests (e.g., Egger's, Begg's); however these tests could not be performed given the insufficient number of studies.

Quality of Evidence

We used the Assessment of Multiple Systematic Reviews (AMSTAR) measurement tool to assess the methodologic quality of systematic reviews.²⁸

The risk of bias for each included study evaluating diagnostic accuracy was examined using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.²⁹ QUADAS-2 consists of four domains: patient selection, index test, reference standard, and flow and timing.

The quality of the body of evidence for each outcome was examined according to the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group criteria.³⁰⁻³² The overall quality was determined to be high, moderate, low, or very low using a step-wise, structural methodology.

Expert Consultation

In April 2015, we solicited expert consultation on the use of ultrasound as an adjunct to mammography for breast cancer screening. Experts consulted were physicians and researchers in the specialty areas of oncology, radiology, and breast cancer screening. The role of the expert advisors was to contextualize the evidence, provide research guidelines, and provide

advice on screening for breast cancer. However, the statements, conclusions, and views expressed in this report do not necessarily represent the views of the consulted experts.

Results

The database search yielded 2,705 citations published between January 1, 1998, and June 15, 2015 (with duplicates removed). Articles were excluded based on information in the title and abstract. The full texts of potentially relevant articles were obtained for further assessment. Figure 2 presents the flow diagram for the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA).

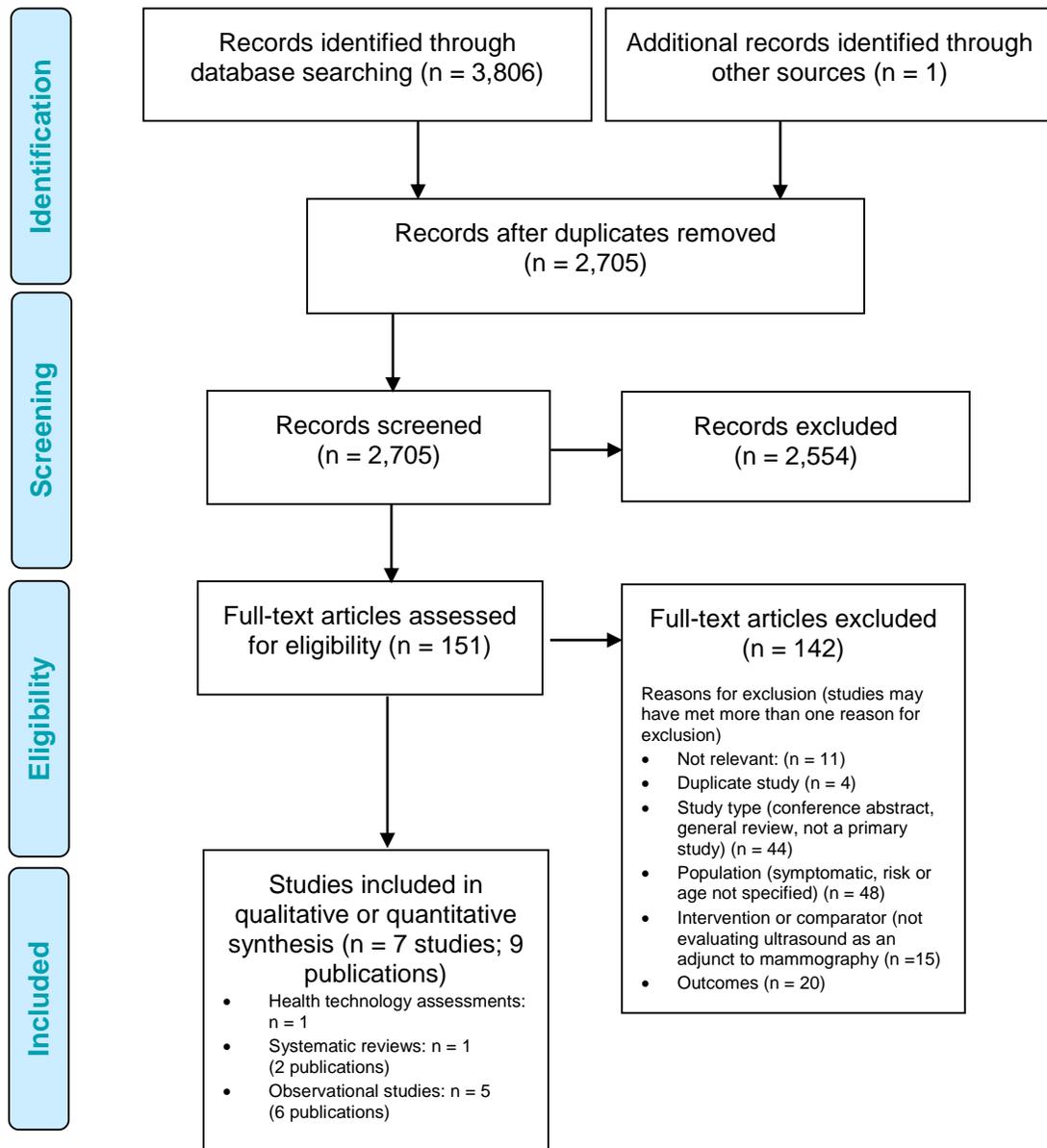


Figure 2: PRISMA Flow Diagram for Clinical Evidence Review

Source: Adapted from Moher et al.³³

Authors of studies were contacted for clarification on the eligibility of the populations included as well as the methods used to evaluate the adjunct use of ultrasound. No authors responded with additional information, and therefore these studies were excluded.

Studies of Women at Average Risk for Breast Cancer

One systematic review, published in two articles, met our inclusion criteria.^{34,35} We used the article published in the Cochrane Database of Systematic Reviews to summarize study results because it provided greater depth of methodology.³⁴

No additional primary studies that met our inclusion criteria were identified.

Studies of Women at High Risk for Breast Cancer

Six studies of women at high risk for breast cancer met the inclusion criteria.

One study used a broader definition of high-risk women than that used by the Ontario Breast Screening Program; however, we included it as the authors self-defined the study population as including only women at elevated risk.³⁶ Results for this study were reported in two articles, with the most recent article including longer follow-up data and updated values.^{36,37} We counted this study as one paper, using values from the most recent publication.

Kuhl et al^{38,39} published two included studies in separate years. We contacted the lead author to determine if study populations or screening events overlapped in the two studies, but we did not receive a response. We therefore assumed these articles were independent and report them as two separate cohorts.

We searched the reference lists of the included studies and relevant health technology assessment websites to identify other relevant studies, and one additional health technology assessment was included.⁴⁰

Two of the included studies^{41,42} published additional data on the methodology used in preceding publications.^{43,44} We used these articles only for the purposes of supplementing data on study methodologies used.

Excluded Studies

Numerous studies and systematic reviews evaluated the use of breast ultrasound as an adjunct screening method for women whose mammograms showed they had dense breasts. These studies, however, did not specify the risk criteria they used, or they included both average- and high-risk women without stratifying results by risk. We contacted authors to confirm the population risk criteria but did not receive responses during our review period. Appendix 3 provides a summary of the excluded systematic reviews evaluating the use of adjunct ultrasound in women with dense breasts.

Several other studies that met our population risk criteria were excluded due to other patient factors. We excluded two studies that did not meet our age criteria for women at average risk: in one, the median age was 47 years (range 27 to 79)⁴⁵; the second included women less than 50 years old (range and inclusion criteria were not provided), and the results stratified by age were not sufficient to calculate 2 x 2 tables.⁴⁶ An additional study evaluated the use of adjunct ultrasound among high-risk Chinese women, but we excluded it as it did not clearly report the factors used in the risk assessment model and more than 90% of the women did not have a

family history of breast cancer.⁴⁷ This study also did not directly assess outcomes related to the accuracy of ultrasound as an adjunct to mammography.

Findings From Studies of Average-Risk Women

One high-quality 2013 systematic review directly evaluated the use of screening breast ultrasound as an adjunct to mammography among women at average risk of breast cancer aged 40 years and onward.³⁴ Table 3 summarizes the study quality and inclusion criteria of the review.

The review identified no methodologically sound studies. It did identify one protocol for a randomized controlled trial comparing mammography to mammography plus ultrasound in women aged 40 to 49 years; however, that patient population does not meet our inclusion criteria.

Inclusion criteria for the systematic review were more restrictive than ours: they required comparative, prospective studies published since 2008, with a low risk of bias, involving women with no personal history of breast cancer, and a sample size of at least 500 people.

Table 3: Summary of Included Systematic Review for Average-Risk Women

Author, Year	Search Dates	AMSTAR Score ^a	Included Populations	Additional Selection Criteria	Studies Included
Gartlehner et al, 2013 ³⁴	2008 to 2012	11	Asymptomatic women aged 40–75 years at average risk of breast cancer (< 15% lifetime risk or dense breasts without additional risk factors), with no personal history of the disease	Intervention: mammography plus ultrasound Comparator: mammography alone Reference standard for diagnostic accuracy: biopsy and minimum follow-up period Outcomes: mortality, harms, false-positive rate, false-negative rate, tumour characteristics Study designs: RCTs or prospective controlled non-randomized studies with low risk of bias; sample size at least 500	0

Abbreviations: AMSTAR, Assessment of Multiple Systematic Reviews; RCT, randomized controlled trial

^aSee Appendix 4 for details of AMSTAR scores.

Findings From Studies of High-Risk Women

Included Systematic Reviews

We identified one systematic review evaluating the use of screening breast ultrasound as an adjunct to mammography as part of a health technology assessment by the National Institute of Health and Care Excellence (NICE) on surveillance of women at high risk for breast cancer.⁴⁰ Table 4 summarizes the study quality and selection criteria.

The NICE review evaluated the effectiveness of various methods of breast cancer screening, including breast ultrasound, in comparison with each other and in combination. The authors included one study that we have included as a primary study (described below).³⁹ However, the

NICE review primarily focused on the use of adjunct MRI and did not reach conclusions related to screening breast ultrasound as an adjunct to mammography.

Table 4: Summary of Included Systematic Review for High-Risk Women

Author, Year	Search Dates	AMSTAR Score ^a	Included Populations	Additional Selection Criteria	Studies Included ^b
NICE, 2013 ⁴⁰	2003 to 2011	6	Women with a family history of breast cancer, with no personal history of breast cancer, aged 18 years and over	Intervention: mammography, MRI, ultrasound, clinical breast exam, any combination of tests, no screening Comparator: each other	1 systematic review of 11 studies; 4 diagnostic studies

Abbreviations: AMSTAR, Assessment of Multiple Systematic Reviews; NICE, National Institute for Health and Care Excellence.

^a See Appendix 4 for summary of AMSTAR scores.

^b Total number of studies included, regardless of interventions evaluated.

Included Primary Studies

Table 5 and Table 6 summarize the characteristics of the five primary studies we identified. All were prospective, paired cohort studies evaluating the use of adjunct ultrasound in comparison with mammography alone to screen women at high risk for breast cancer.

Study Populations

All studies screened women aged 25 or 30 years and over, unless they had family members with an earlier diagnosis of the disease. The mean ages were 41 to 55 years and ranged from 22 to 91 years (Table 5).

All studies included women with a personal history of breast cancer. Most studies indicated that the breast where cancer was found was not included in subsequent screening rounds.

The definition and assessment of high risk for breast cancer varied across studies (Table 5). Four studies used a *BRCA1* or *BRCA2* mutation or lifetime risk greater than 20% to 25% to define risk criteria. None of these four studies included women solely based on a personal history of breast cancer. The number of women with a confirmed *BRCA1* or *BRCA2* mutation ranged from 8% to 65.8%.

Berg et al³⁶ varied from the other studies in their definition of high-risk women and in limiting the study population to women with heterogeneously or extremely dense breasts. This study included a personal history of breast cancer (53% of women) or a personal history of atypical ductal hyperplasia, atypical lobular hyperplasia, or lobular carcinoma in situ (3% of women) as independent factors for including women in the study. Less than 1% of women had a *BRCA1* or *BRCA2* mutation.³⁶ We report results for this study separately.

Interventions and Comparators

Table 6 summarizes the characteristics of study interventions, comparators, and methodology used to assess each. The study by Berg et al³⁶ was the only one designed to directly evaluate the diagnostic yield and performance of screening with adjunct ultrasound compared with mammography alone, with the addition of MRI in the final year of screening. All other studies were designed to evaluate the adjunctive or primary use of screening breast MRI, while comparing to the individual and combined use of mammography and ultrasound.

All included studies used hand-held ultrasound systems. Only one study exclusively used film mammography, and the remainder used both film and digital mammography.

Ultrasound was used as a simultaneous test in each study, with results of mammography and other imaging tests blinded at the time of imaging and assessment. Berg et al³⁶ conducted an integrated assessment of the combined results for mammography and ultrasound: results were evaluated together and results from one test could be downgraded based on results from the other. Sardanelli et al⁴² used a worst-case approach, with a positive test on either ultrasound or mammography considered a positive test. The remaining studies did not clearly describe how they combined the assessment of tests.^{38,39,41}

While all studies required patients to be asymptomatic at study entry, three studies included annual or semi-annual clinical breast exams (physical exams).^{38,39,42} One of these studies did not blind assessors to the results from the clinical breast exam prior to imaging with mammography or ultrasound,³⁹ and one study did not specify if assessors knew the results of the clinical exam at the time of imaging.³⁸ The number of positive clinical breast exams was not reported.

Each study conducted full screening rounds yearly with each intervention under examination, and three studies provided additional semi-annual ultrasound exams to all or a subgroup of patients. It is unclear how results from these screens were incorporated into the annual screening results for all studies.

Classification of Positive Test and Reference Standard

Four studies used a BI-RADS score of 4 or greater to indicate a positive test. Only the study by Berg et al³⁶ used a BI-RADS score of 3, 4, or 5 to indicate a positive screen (Table 6). No study included or reported on tests classified as BI-RADS 0 (incomplete assessment), which should be considered a positive test.

All studies used histology (evaluation of biopsy specimens) to confirm a true-positive test, with cancer defined as either invasive cancer or DCIS. Confirmation of false-positive tests varied across studies. Biopsy confirmed false-positives in most cases; however, given that some women in these studies also received MRI, some ultrasound cases were confirmed as false-positive through negative mammography and MRI, without a biopsy. Clinical follow-up for one year was used in all studies to confirm true-negative tests, with some patients receiving biopsy due to preference or positive diagnoses from alternative screening tests such as MRI (Table 6). The extent of clinical follow-up varied from a clinical visit to a telephone call.

Table 5: Summary of High-Risk Populations in Included Primary Studies

Author, Year,	Country, Number of Sites,	Women, N ^a (completed screens) ^b	Risk Classification	Additional Inclusion/Exclusion	History of Breast Cancer, %	BRCA Mutation, %	Mean Age, Years (Range)	Mean Screens per Woman, N, or Length of Follow-Up
Riedl et al, 2015 ⁴¹	Austria, 1	559 (1,365)	<ul style="list-style-type: none"> • <i>BRCA1</i> or <i>BRCA2</i> mutation • Lifetime risk > 20% 	Exclusion: bilateral mastectomy, stage IV breast cancer, pacemaker, pregnancy, clinical symptoms	Unclear	28	Median 44 (22–83)	Mean 2.45 rounds per woman (range 1–11 rounds)
Berg et al, 2012 ^{36,37}	United States, 21	2,662 (7,473)	<ul style="list-style-type: none"> • <i>BRCA1</i> or <i>BRCA2</i> mutation • Lifetime risk > 25% (Gail/Claus model) or high 5-year risk (Gail model \geq 2.5% or \geq 1.7% and extremely dense breasts) • History of chest irradiation • History of ADH/ALH/LCIS or atypical papilloma • Personal history of breast cancer 	Inclusion: heterogeneously or extremely dense breasts Exclusion: pregnant, lactating, signs or symptoms of breast disease, breast surgery within 12 months, breast implants	53.1	0.9	55 (25–91)	3-year follow-up (93.8% completed second round; 87.3% completed third round)
Sardanelli et al, 2011 ^{42,44}	Italy, 18	501 (1,121)	<ul style="list-style-type: none"> • <i>BRCA1</i> or <i>BRCA2</i> mutation or first-degree relative of carrier • High familial risk (criteria not specified; can include personal history events) 	Exclusion: personal history and bilateral total mastectomy, pregnancy, breast-feeding, current chemotherapy, terminal illness, contraindications to MRI	43.5 ^c	65.8	46 (22–79)	Mean 3.17 rounds per woman (range 1–7 rounds)
Kuhl et al, 2010 ³⁸	Germany, 4	687 (1,679)	<ul style="list-style-type: none"> • Lifetime risk > 20% based on family history (<i>BRCAPRO</i> model), <i>BRCA1</i> or <i>BRCA2</i> mutation, familial and personal history of breast cancer 	Exclusion: current signs or symptoms of breast cancer, bilateral mastectomy, chemotherapy in last 12 months, diagnosis of distant metastases	30.9	9.5	44.6 (25–71)	Mean follow-up 29 months (range 12.8–40 months)
Kuhl et al, 2005 ³⁹	Germany, 4	529 (1,452)	<ul style="list-style-type: none"> • Lifetime risk > 20% based on family history 	Exclusion: current signs or symptoms of breast cancer, bilateral mastectomy, chemotherapy in last 12 months, diagnosis of distant metastases	26	8.1	41.7 (27–59)	Mean follow-up 5.3 years (range 2–7 years)

Abbreviations: ADH, atypical ductal hyperplasia; ALH, atypical lobular hyperplasia; LCIS, lobular carcinoma in situ; MRI, magnetic resonance imaging.

^aTotal number of women enrolled in the screening study.

^bTotal number of annual screening rounds with data for all imaging modalities under investigation.

^cPersonal history of breast cancer and/or ovarian cancer.

Table 6: Summary of Screening and Diagnosis Methods in Included Primary Studies

Author, Year	Ultrasound Description, Performer	Mammography Description, Performer	Additional Annual Screening Tests	Individual and Combined Test Interpretation	Maximum Time Between Tests	Classification of Positive Test	Reference Standard
Riedl et al, 2015 ⁴¹	Various systems, radiologist	Various systems, radiologist	MRI, semi-annual ultrasound for <i>BRCA</i> mutation carriers	Simultaneous, blinded individual test assessment; combined review unclear	1 month	Positive: BI-RADS 4, 5 Negative: BI-RADS 1, 2, 3	Positive: Histology Negative: 1-year clinical follow-up and/or interview Indeterminate: 6-month follow-up screen
Berg et al, 2012 ³⁶	Maximum frequency at least 12 MHz, physician	Film screen or digital, radiologist	MRI substudy for last round of screening only	Simultaneous, blinded individual assessment; integrated review of combined tests	2 weeks	Positive: BI-RADS 3, 4, 5 Negative: BI-RADS 1, 2	Positive: Histology Negative: 1-year clinical follow-up and/or interview
Sardaneilli et al, 2011 ⁴²	Frequency of 7.5 MHz or greater, radiologist	Film screen, phosphor plate digital and full-field digital, radiologist	MRI, CBE	Simultaneous, blinded individual assessment; worst-scenario approach for combined review	Attempted same day (maximum 1 to 2 months)	Positive: BI-RADS 4, 5 Negative: BI-RADS 1, 2, 3	Positive: Histology Negative: 1-year follow-up Indeterminate: 4-month follow-up testing
Kuhl et al, 2010 ³⁸	Maximum frequency at least 12 MHz, specialized physician	Film screen or full field digital, radiologist	MRI, semi-annual CBE and ultrasound for subset of women	Simultaneous, blinded individual assessment; combined review unclear	6 weeks	Positive: BI-RADS 4, 5 Negative: BI-RADS 1, 2, 3	Positive: Histology Negative: 12-month clinical or telephone follow-up Indeterminate: follow-up based on guidelines
Kuhl et al, 2005 ³⁹	7.5 MHz to 13 Mhz probes, radiologist	Film screen, radiologist	MRI, semi-annual CBE and ultrasound	Simultaneous, blinded interpretation of tests, results of CBE known; combined review unclear	8 weeks	Positive: BI-RADS 4, 5 Negative: BI-RADS 1, 2, 3	Positive: Histology Negative: 1-year clinical follow-up or 6-month CBE/ultrasound Indeterminate: 6-month clinical visit

Abbreviations BI-RADS, Breast Imaging Reporting and Data System; CBE, clinical breast exam; MRI, magnetic resonance imaging.

Methodological Quality of the Included Studies

Complete results of the QUADAS-2 assessment for risk of bias of included studies are presented in Appendix 4. Four studies were deemed directly applicable or partially applicable to the research questions.^{38,39,41,42} We assessed one study as being not directly applicable to the Ontario setting.³⁶ All studies had limitations associated with patient selection as well as flow and timing. Overall, there were serious limitations and a high risk of bias for each study assessed.

Effectiveness Outcomes

We did not identify any studies that evaluated the effectiveness of adjunct screening breast ultrasound relative to mammography alone on all-cause mortality, breast cancer–related mortality, or patients' quality of life.

Diagnostic Performance Outcomes

The population rate of breast cancer diagnosed across the five studies ranged from 1.5% to 3.3%. The lowest prevalence of disease was observed by Berg et al³⁶ who used a lower-risk population to define their inclusion criteria, and the highest prevalence was observed in the study by Sardanelli et al,⁴² which included the largest proportion of women with *BRCA1* or *BRCA2* mutations.

All studies used the total number of complete screening rounds or person-years as the denominator for diagnostic outcomes, but methods used to count the number of women and the number of cancers varied across studies. Kuhl et al (2010)³⁸ and Sardanelli et al⁴² calculated diagnostic accuracy on a per-patient basis, not a per-breast or per-lesion basis, for all screening rounds. When accounting for the number of cancers per screening round, Berg et al³⁶ used the participant as the primary unit of analysis for each screening round, and only allowed analysis of the breast without cancer in the subsequent annual screen for women diagnosed with breast cancer during a previous round. Kuhl et al (2005)³⁹ counted diagnostic indices per breast with cancer. Riedl et al⁴¹ considered malignant lesions as independent cancers; however, it was unclear if more than one cancer could be identified per patient, per round, in that study.

Diagnostic Yield

Table 7 summarizes the diagnostic yields for mammography alone and for mammography plus adjunct ultrasound in each study, as well as the incremental diagnostic yield (additional cancers detected) with adjunct ultrasound.

In all studies, screening with adjunct ultrasound increased the breast cancer detection rate compared to mammography alone. The overall incremental diagnostic yield of imaging with adjunct ultrasound ranged from 2.3 per 1,000 screens to 5.9 per 1,000 screens. Only Berg et al³⁶ statistically compared the diagnostic yield between groups, and that study identified a significant increase in cancers detected with adjunct ultrasound in comparison to mammography alone (incremental yield 4.3 per 1,000; $P < .001$).

Based on these diagnostic yields, the number needed to screen with adjunct ultrasound to detect 1 additional breast cancer ranged from 169 to 435.

Table 7: Diagnostic Yield of Adjunct Ultrasound Compared With Mammography Alone

Author, Year	Round	Screens, N	Cancers, N			Diagnostic Yield, Cancers Per 1,000 Screens (95% CI)		
			Total ^a	M	M+US	M	M+US	Incremental Yield
Riedl et al, 2015 ⁴¹	All	1,365	40 ^b	15	20	11.0 (6.2, 18.1) ^c	14.6 (9.0, 22.5) ^c	3.6 (NR)
	1	559	19	8	11	14.3 (6.2, 28.0) ^c	19.7 (9.9, 34.9) ^c	5.4 (NR)
	> 1	806	21	7	9	8.7 (3.5, 17.8) ^c	11.2 (5.1, 21.1) ^c	2.5 (NR)
Sardanelli et al, 2011 ⁴²	All	1,592 (M: 1,095; M+US: 1,047)	52	25	30	22.8 (14.8, 33.5) ^c	28.6 (19.4, 40.6) ^c	5.9 ^d (NR)
Kuhl et al, 2010 ³⁸	All	1,679	27	9	13	5.4 (2.4, 10.1) ^c	7.7 (4.1, 13.2) ^c	2.3 (NR)
Kuhl et al, 2005 ³⁹	All	1,452	43 ^e	14	21	9.6 (5.3, 16.1) ^c	14.5 (9.0, 22.0) ^c	4.9 (NR)
Berg et al, 2012 ³⁶	All	7,473	111	59	91	7.89 (6, 10.2)	12.2 (9.8, 14.9)	4.3 (NR)
	1	2,659	36	20	34	7.5 (4.6, 11.6)	12.8 (8.9, 17.8)	5.3 (2.1, 8.4); <i>P</i> < .001
	2,3	4,814	75	39	57	8.1 (5.8, 11.1)	11.8 (9.0, 15.3)	3.7 (2.1, 5.8); <i>P</i> < .001

Abbreviations: CI, confidence interval; M, mammography; N, number; US, ultrasound; NR, not reported

^aIncludes cancers identified by other screening tests as well as interval breast cancers.

^b38 women with 40 cancers.

^cCalculated using data provided in studies.

^dNot all women had both tests; value does not necessarily represent number of cancers missed by mammography.

^e41 women with 43 cancers.

Across the five studies, between 18% and 52% of breast cancers identified were missed by both mammography and ultrasound. They were either interval cancers (cancers diagnosed in the interval between screenings) or were captured by MRI in the screening rounds. Because no studies randomized women to the two interventions, we could not assess a change in interval cancer rates.

Subgroup: prevalent versus incident screens. Two studies stratified results for incremental diagnostic yield by screening round (Table 7). Both identified a higher incremental diagnostic yield in the first round of screening (prevalent screen) in comparison to subsequent screens (incident screens). Neither study evaluated the statistical significance of these findings.

Subgroup: personal history versus no personal history of breast cancer. Two studies stratified results based on whether or not women had a personal history of breast cancer (Appendix 5, Table A7).^{36,39}

In the 2005 study by Kuhl et al³⁹ there was minimal variation in the incremental diagnostic yield with adjunct ultrasound between women with a personal history and no personal history of the disease. No between-group statistical comparison was conducted.

Berg et al³⁶ found a statistically significant increase in diagnostic yield with adjunct ultrasound compared with mammography alone for both groups: women with and without a personal history of breast cancer. The difference between these two groups was not statistically significant.

Subgroup: digital versus film mammography. Berg et al³⁶ found the diagnostic yield with adjunct ultrasound to increase significantly with both digital mammography ($P < .001$) and film mammography ($P = .003$).

Cancer and Tumour Characteristics

Three of the five studies reported characteristics of the cancers identified by mammography alone compared with those identified with the addition of ultrasound (Table 8).

Of the total number of cancers identified by mammography in the three studies ($n = 88$), approximately 66% were invasive cancers, of which 36% were node-positive (i.e., cancer cells from the breast tumour had spread to the lymph nodes). Among the 44 cancers identified only by ultrasound, most were invasive (89%), of which approximately 11% were node-positive. DCIS represented nearly 30% of all cancers identified by either test among the three studies and about 9% of cancers identified by ultrasound only.

The remaining two studies did not provide tumour characteristics based on screening method. Overall, Kuhl et al (2010)³⁸ noted that DCIS accounted for 53% of all cancers diagnosed during incidence screens (using mammography, ultrasound or MRI). Sardanelli et al⁴² stated that 15.4% of all breast cancers diagnosed in their study were classified as DCIS.

Table 8: Characteristics of Tumours Identified by Mammography or Adjunct Ultrasound Alone

Author, Year	Berg, 2012 ³⁶		Riedl, 2015 ^{41,a}		Kuhl, 2005 ³⁹	
Tumour Characteristic	Mammography	Ultrasound Alone	Mammography	Ultrasound Alone	Mammography	Ultrasound Alone
Cancers detected, N	59	32	15	5	14	7
Invasive, % (n)	69.5 (41)	93.7 (30)	46.7 (7)	60 (3)	71.4 (10) ^b	85.7 (6) ^b
Mean size, mm, (range)	M alone: 1.5 (1–55) M+US: 6 (3–40)	10 (2–40)	NR	NR	13.2 (SD 7.8)	M+US: 13.9 (SD 6.4)
Node-positive, % (n)	40 (12) ^c	3.7 (1) ^d	14.3 (1)	66.7 (2)	40 (4) ^b	16.7 (1)
DCIS, % (n)	30.5 (18)	6 (2)	33.3 (5)	40 (2)	21.4 (3)	0 (0)
Other, % (n)	0	0	20 (3) ^e	0	7.1 (1) ^f	14.3 (1) ^f

Abbreviations: DCIS, ductal carcinoma in situ; M, mammography; N, number; US, ultrasound; SD, standard deviation.

^aAll values calculated from data provided in article.

^bDefined as primary invasive breast cancers only; number of secondary invasive cancers was not clearly reported.

^c30 of the 41 invasive cancers had staging data available.

^d27 of the 30 invasive cancers had staging data available.

^e1 metastasis from ovarian cancer, 1 mucinous invasive ductal carcinoma recurrence, 1 medullary carcinoma.

^fInsufficient detail provided.

Comparative Diagnostic Accuracy

Tables 9 to 11 summarize the results for the diagnostic accuracy of screening breast ultrasound as an adjunct to screening mammography, compared with mammography alone. A summary of the screening test results, corresponding forest plots, and individual study measures of the area under the receiver operating characteristic curve is provided in Appendix 3 (Tables A2 to A4 and Figure A1).

Berg et al³⁶ was the only study to use a BI-RADS score of 3, 4 or 5 as test-positive; in other words, a recommendation for short-term follow-up, additional testing, or biopsy classified as a positive test. The remaining four studies used a BI-RADS score of 4 or 5 to indicate a positive test: in those studies, recommendations for short-term follow-up were classified as test-negative and only recommendations for biopsy (with additional testing) were classified as test-positive.

Sensitivity and Specificity

In all studies, the sensitivity of screening improved with the addition of adjunct ultrasound, with a consequent decrease in specificity (Tables 9 and 10 and Figure 3).

Positive test threshold: BI-RADS 4 or 5. The pooled sensitivity and specificity of screening breast ultrasound and mammography, using the bivariate model among the four studies that used BI-RADS 4 and 5 as their positive test thresholds, are depicted in Figure 3 and Table 10.

The average pooled sensitivity for mammography plus adjunct ultrasound was 53.3% (95% CI 44.9, 61.6), an absolute increase of 13.7% compared with mammography screening alone. The average pooled specificity of mammography plus ultrasound was 96.2% (95% CI 98.4, 91.5), which corresponds to a 2.1% absolute increase in the false-positive rate (1 – specificity) with adjunct ultrasound relative to mammography screening alone.

When we applied a covariate for test type to the model, the difference between mammography alone and mammography plus ultrasound was statistically significant for sensitivity ($P < .05$) and not statistically significant for specificity ($P > .05$). Evaluation of the confidence regions on the receiver operating characteristic space shows a large increase in sensitivity, with little overlap in the confidence regions, and a smaller decrease with greater overlap in specificity (Figure 3). Given the small number of studies included and the limitations of the bivariate model for paired study designs, this represents a conservative estimate.

The quality of this body of evidence on sensitivity and specificity was low, according to our GRADE assessment (Appendix 4, Table A5).

Positive test threshold: BI-RADS 3, 4, or 5. Berg et al³⁶ was the only study to directly evaluate the incremental diagnostic accuracy of screening breast ultrasound as an adjunct to mammography. As noted above, women in this study were defined as having a positive test if they were recommended for short-term follow-up, additional testing, or biopsy.

The sensitivity of mammography plus adjunct ultrasound was 94.4% in the first-year screen (prevalence screen) and 76% for the second- and third-year screens combined (incidence screens). This is a statistically significant increase from mammography screening alone (increase in true-positive rate of 38.9% in year 1 and 24% in years 2 and 3; $P < .001$). The specificity of mammography and adjunct ultrasound decreased significantly relative to mammography alone for each screening round. The increases in the false-positive rate for the prevalence and incidence rounds were 14.8% and 7.1%, respectively. The area under the receiver operating characteristic curve was statistically significantly higher for each individual screening round (Appendix 5, Table A8).

The quality of this body of evidence on sensitivity and specificity was very low, according to our GRADE assessment (Appendix 4, Table A5).

Table 9: Sensitivity and Specificity of Mammography and Adjunct Ultrasound Compared With Mammography Alone in Women at High Risk for Breast Cancer

Author, Year	Round	Screens, N	Sensitivity, % (95% CI) ^a			Specificity, % (95% CI) ^a		
			Mammography	Mammography + Ultrasound	Absolute Change in Sensitivity	Mammography	Mammography + Ultrasound	Absolute Change in Specificity
Positive Test: BI-RADS 4, 5								
Riedl et al, 2015 ⁴¹	All	1,365	37.5 (24.2, 53.0)	50.0 (35.2, 64.8)	12.5 (NR)	97.1 (96.1, 97.9)	95.7 (94.5, 96.7)	-1.4 (NR)
Sardanelli et al, 2011 ⁴²	All	M: 1,095 M+US: 1,047	50.0 (35.5, 64.5)	62.5 (47.3, 76.0)	12.5 (NR)	99.0 (98.2, 99.5)	97.6 (96.4, 98.4)	-1.45 (NR) NS ^b
Kuhl et al, 2010 ³⁸	All	1,679	33.3 (17.2, 53.9)	48.1 (29.1, 67.6)	14.8 (NR); NS ^c	99.1 (98.5, 99.5)	98.4 (97.5, 98.8)	-0.79 (NR)
Kuhl et al, 2005 ³⁹	All	1,452	32.6 (19.0, 48.5) ^a	48.8 (33.3, 64.5) ^a	16.28 (NR)	96.8 (95.7, 97.7) ^a	89.0 (87.2, 90.6) ^a	-7.81 (NR)
Positive Test: BI-RADS 3, 4, 5								
Berg et al, 2012 ³⁶	1	2,659	55.6 (38.1, 72.1)	94.4 (81.3, 99.3)	38.9 (20.2, 57.5); <i>P</i> < .001	89.1 (87.8, 90.3)	74.3 (72.6, 76.0)	-14.8 (-16.3, -13.2); <i>P</i> < .001
	2,3	4,814	52.0 (40.1, 63.7)	76.0 (64.7, 85.1)	24.0 (14.7, 33.3); <i>P</i> < .001	91.3 (90.4, 92.0)	84.1 (83.1, 85.2)	-7.1 (-8.0, -6.3); <i>P</i> < .001

Abbreviations: BI-RADS, Breast Imaging Reporting and Data System; CI, confidence interval; M, mammography; N, number; NR, not reported; NS, not significant; US, ultrasound.

^aCalculated based on data provided in study.

^bAuthors stated that the specificity ranged from 96% to 99% across imaging modalities and combinations without significant differences; however it is unclear if this applies to comparison between mammography alone and mammography with adjunct ultrasound.

^cAuthors stated that the combination of ultrasound and mammography was not statistically significantly higher (*P* < .12) than mammography alone or ultrasound alone. It is unclear if the *P* value represents the comparison to mammography alone, ultrasound alone, or both.

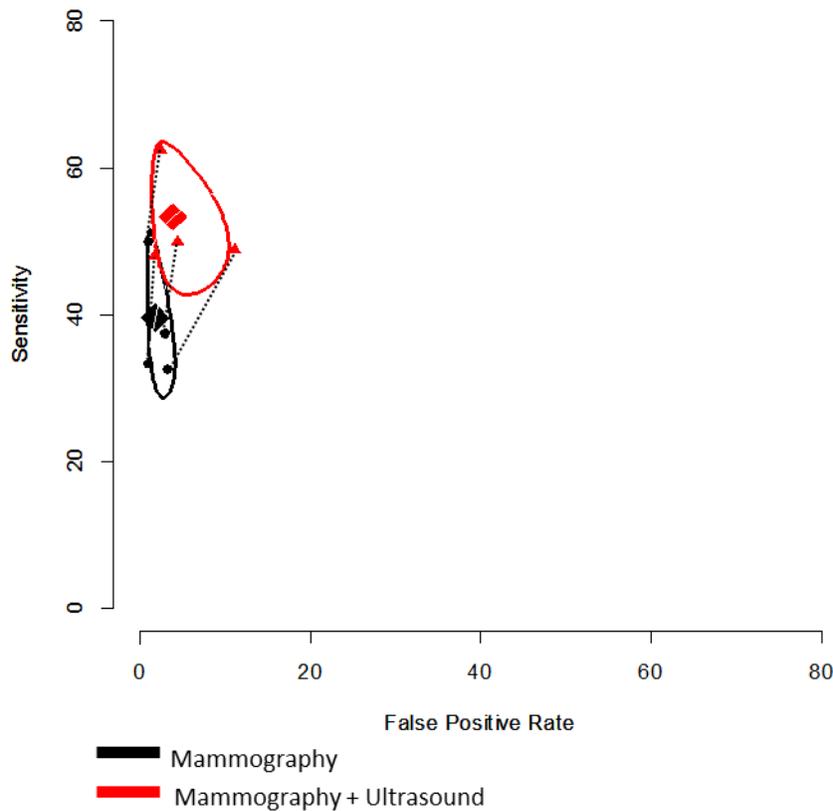


Figure 3: Plot of Change in Sensitivity and Specificity with Addition of Screening Ultrasound Relative to Mammography Alone in Women at High Risk for Breast Cancer

Table 10: Pooled Sensitivity and Specificity of Mammography and Adjunct Ultrasound Compared With Mammography Alone in Women at High Risk for Breast Cancer

Test	Round	Number of Studies	Pooled Sensitivity, % (95% CI)	Pooled Specificity, % (95% CI)
Positive Test: BI-RADS 4, 5				
Mammography	All	4	39.6 (30.8, 49.2)	98.3 (96.6, 99.2)
Mammography + ultrasound	All	4	53.3 (44.9, 61.6)	96.2 (91.5, 98.4)
<i>P</i>			< .05	NS

Abbreviations: BI-RADS, Breast Imaging Reporting and Data System; CI, confidence interval; NS, not significant.

Subgroup: personal history versus no personal history of breast cancer. Two studies reported results separately for women with and without a personal history of disease (Appendix 5, Table A7).

Berg et al³⁶ found a statistically significant increase in sensitivity and a corresponding decrease in specificity for both groups when women were screened with adjunct ultrasound, relative to

mammography alone. There was no statistically significant difference in the incremental sensitivity between the two groups; however, incremental specificity among women with a personal history of the disease was statistically significantly lower than those without the disease. This corresponds to an increase in the false-positive rate (1 – specificity) of 11.6% for women without a personal history and 8.3% for women with a personal history of the disease.

Kuhl et al (2005)³⁹ did not statistically compare results between these groups, but they did find that, with adjunct ultrasound, incremental sensitivity increased from 8.4% among women with no personal history of breast cancer to 19.3% among women with a personal history of the disease. Little variation in the incremental specificity was observed between the two groups.

Subgroup: digital versus film mammography. Sardanelli et al⁴² identified no significant difference ($P = .560$) in the sensitivity of mammography among women screened with film (sensitivity = 55%) or digitally (sensitivity = 42%).

Positive Predictive Value

Table 11 summarizes the positive predictive value (PPV) reported for each study. We have stratified results by the positive test threshold that each study used.

Table 11: Positive Predictive Value of Screening by Mammography and Ultrasound Compared With Mammography Alone in Women at High Risk for Breast Cancer

Author, Year	Round	PPV, % (95% CI)		
		Mammography	Mammography + Ultrasound	Absolute Difference Between Groups
Positive Test: BI-RADS 4, 5				
Riedl et al, 2015 ⁴¹	All	28.3 (18, 41.6)	26 (17.5, 36.7)	-2.3 (NR)
Kuhl et al, 2010 ³⁸	All	39.1 (20.4, 61.2)	32.5 (19.1, 49.2)	-6.6 (NR)
Sardanelli et al, 2011 ⁴²	All	71.4 (53.7, 85.4)	55.6 (41.4, 69.1)	-15.8 (NR) NS ^a
Kuhl et al, 2005 ³⁹	All	23.7 (13.6, 36.6) ^b	11.9 (7.5, 17.6) ^b	-11.8 (NR)
Positive Test: BI-RADS 3, 4, 5				
Berg et al, 2012 ³⁶ .	1	6.5 ^c (4.0, 9.9)	4.8 ^c (3.4 to 6.7)	-1.7 ^c (-3.7, 0.1); $P = .07$
	2,3	8.6 ^c (6.2, 11.6)	7.0 ^c (5.4, 9.0)	-1.6 ^c (-3.1, -0.2); $P = .04$
Berg et al, 2012 ³⁶ .	1	29.2 ^d (18.6, 41.8)	11.4 ^d (7.9 to 15.8)	-17.8 ^d (-26.7, -9.3); $P < .001$
	2,3	38.1 ^d (28.5, 48.6)	16.2 ^d (12.5, 20.6)	-21.9 ^d (-28.7, -14.7); $P < .001$

Abbreviations: BI-RADS, Breast Imaging Reporting and Data System, CI, confidence interval; NR, not reported; NS, not significant; PPV, positive predictive value.

^a The authors stated that the PPV for all included screening tests ranged from 53% to 71%, without statistically significant differences. It is unclear if all test combinations were considered and compared.

^b Calculated based on data provided in article

^c Calculated as the malignancy rate among cases that test positive (recommended for further testing, short-interval follow-up or biopsy) on screening.

^d Defined by authors as the malignancy rate among women with a positive screening test who underwent biopsy of the same lesion. These values could include biopsy resulting from a BI-RADS 3 diagnosis.

Positive test threshold: BI-RADS 4 or 5. Four studies assessed the positive predictive value for tests resulting in biopsy recommendations (BI-RADS 4 or 5) (Table 11). With mammography and adjunct ultrasound, the PPV ranged from 11.9% to 55.6%. This is lower than with mammography alone by 2.3% to 15.8%—the percentage of additional unnecessary follow-up testing or biopsy recommendations resulting from screening with adjunct ultrasound. None of the individual studies reported statistical comparisons.

Positive test threshold: BI-RADS 3, 4, or 5. Berg et al³⁶ provided two measures of positive predictive value: PPV1 as the rate of cancers detected among all women who tested positive (recommended for short-term follow-up, additional testing, or biopsy), and PPV2 as the cancer rate among women with a positive screening test who then underwent biopsy of the same lesion.

In the first round of screening, the PPV1 with mammography plus adjunct ultrasound did not significantly differ from mammography alone ($P = .07$). Among the subsequent two screening rounds, the PPV1 with adjunct ultrasound (7.1%) was statistically significantly lower than with mammography alone (difference of 1.6%; $P = .04$).

The PPV2 with mammography and adjunct breast ultrasound was 11.4% in the first round of screening and 16.2% in subsequent rounds, a statistically significant decrease from mammography screening alone ($P < .001$). This corresponded to increases in the number of unnecessary biopsies of 17.8% and 21.9% for the first and subsequent screening rounds, respectively.

Subgroup: personal history versus no personal history of breast cancer. Appendix 5, Table A7 summarizes the results for positive predictive value of mammography alone and mammography plus adjunct ultrasound in the subgroups of patients with and without a personal history of breast cancer.

Berg et al³⁶ found that adjunct ultrasound decreased the PPV when compared with mammography alone in both women with and without a personal history of breast cancer. However, there was no significant difference in the change in PPV between the two groups ($P \geq .71$ for each comparison). The 2005 study by Kuhl et al³⁹ observed only a very minor difference between the groups: women with no personal history of breast cancer had a slightly smaller decrease in PPV with adjunct ultrasound over mammography alone, compared to women with a history of the disease.

Recall Rates and Biopsy Rates

Only Berg et al³⁶ directly reported comparative recall rates and biopsy rates (Table 12). The recall rate was defined as the proportion of screens that resulted in a recommendation for additional testing, follow-up, or biopsy. Adjunct ultrasound resulted in a statistically significant increase in the number of recalls: an additional 15.1% of screens were recalled in the first year of screening, and an additional 7.4% were recalled in years 2 and 3 combined. Similarly, significantly more women were biopsied following mammography plus adjunct ultrasound compared with mammography alone (7.8% more in the first year of screening and 5.0% in the second year).

Table 12: Recall and Biopsy Rates for Screening Mammography and Adjunct Ultrasound Compared With Mammography Alone in Women at High Risk for Breast Cancer

Author, Year	Round	Recall Rate, ^a % of Screens (95% CI)			Biopsy Rate, ^b % of Screens (95% CI)		
		M	M+US	Difference Between Groups	M	M+US	Difference Between Groups
Berg et al, 2012 ³⁶	1	11.5 (10.3, 12.8)	26.6 (24.9, 28.3)	15.1 (13.5, 16.6); <i>P</i> < .001	2.4 (1.9, 3.1)	10.2 (9.1, 11.4)	7.8 (6.7, 8.8) <i>P</i> < .001
	2, 3	9.4 (8.6, 10.3)	16.8 (15.8, 17.9)	7.4 (6.6, 8.2); <i>P</i> < .001	2.0 (1.6, 2.5)	7.0 (6.3, 7.8)	5.0 (4.4, 5.7) <i>P</i> < .001

Abbreviations: CI, confidence interval; M, mammography; US, ultrasound.

^aCalculated as the number of women recalled from screening for additional follow-up or biopsy over the total number of screens.

^bCalculated as the number of biopsies performed among women who were not diagnosed with cancer over the total number of screens.

Summary of Diagnostic Performance Outcomes

Table 13 provides a summary of all diagnostic performance measures for the studies involving women at high risk of breast cancer and using a BI-RADS threshold of 4 or 5 as a positive test. Table 14 provides the summary of estimates for the single study that used an intermediate- to high-risk population with dense breasts and a BI-RADS threshold of 3, 4, or 5 as a positive test. We calculated theoretical estimates for every 1,000 women screened using the average prevalence estimates of disease among the four studies (2.5%, with a range of 1.5% to 3.3%) with high risk defined similarly as it is in Ontario.

Assuming a 2.5% prevalence of disease among the high-risk population screened, 25 cancers would be expected among 1,000 screens. Using a BI-RADS score of 4 or 5 as a positive test, screening with mammography alone would identify 10 of the 25 cancers, and 20 screens would be falsely classified as positive. With the addition of ultrasound screening, 3 of the cancers missed by mammography would be identified along with an additional 19 false-positive screens. Overall, 308 screens with adjunct ultrasound would be required to identify 1 additional case of breast cancer, with 6 additional false-positive tests (Table 13).

Table 13: Summary of Findings for Studies Using BI-RADS Score ≥ 4

Summary of Estimates for Included Studies ^a					Number Need to Screen and Additional False Positives at Varied Prevalence of Breast Cancer ^b					
	Diagnostic Yield, Cancers per 1,000 Screens	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, %	Prevalence = 1.5%		Prevalence = 2.5%		Prevalence = 3.3%	
					NNS	AFP	NNS	AFP	NNS	AFP
M	Range: 5.4 to 22.8	Range: 33 to 50 Pooled: 40 (31, 49)	Range: 97 to 99 Pooled: 98 (99, 97)	Range: 24 to 71						
M+US	Range: 7.7 to 28.6	Range: 48 to 62 Pooled: 53 (45, 62)	Range: 89 to 98 Pooled: 96 (98, 91)	Range: 12 to 56	513	10	308	6	233	5
Difference	Range: 2.3 to 5.9 NNS: 169 to 435	Range: 12 to 16 Pooled: 13	Range: -1 to -8 Pooled: -2	Range: -2 to -16						

Abbreviations: AFP, additional false-positives; M, mammography; NNS, number needed to screen; PPV, positive predictive value; US, ultrasound.

^a Studies compared screening with mammography alone and with adjunct ultrasound in women at high risk for breast cancer.

^b Calculated using average pooled estimates from meta-analysis. Values represent the number needed to screen to identify 1 additional cancer and the additional false-positives that would result.

Table 14: Summary of Findings for Study Using BI-RADS Score 3, 4, or 5

Summary of Estimates for Included Study ^a								Number Need to Screen and Additional False Positives at Varied Prevalence of Breast Cancer ^b					
	Diagnostic Yield, Cancers per 1,000 Screens	Sensitivity, %		Specificity, %		PPV, %		Prevalence = 1.5%		Prevalence = 2.5%		Prevalence = 3.3%	
		R1	R2/3	R1	R2/3	R1	R2/3	NNS	AFP	NNS	AFP	NNS	AFP
M	Overall 7.9	56	52	89	92	29	38						
M+US	12.2	94	76	74	84	11	16	R1: 175 R2/3: 278	R1: 26 R2/3: 22	R1: 105 R2/3: 167	R1: 15 R2/3: 13	R1: 80 R2/3: 126	R1:12 R2/3: 10
Difference	4.3 NNS: 232	39	24	-15	-7	-18	-22						

Abbreviations: AFP, additional false-positives; M, mammography; NNS, number needed to screen; PPV, positive predictive value; R, round; US, ultrasound.

^a Berg et al³⁶ compared screening with mammography alone and with adjunct ultrasound in women with dense breasts and at high risk for breast cancer.

^b Calculated using summary estimates as reported in Berg et al.³⁶ Values represent the number needed to screen and the additional false-positives to identify an additional cancer.

Conclusions

The use of screening breast ultrasound as an adjunct to mammography has been suggested to improve the detection of breast cancer in asymptomatic women. Among studies of women aged 50 years and over and at average risk for developing breast cancer, we found no evidence on the effectiveness or diagnostic accuracy of screening breast ultrasound as an adjunct to screening mammography.

In women at high risk of developing breast cancer, screening with mammography and adjunct ultrasound detects additional cases of disease with improved sensitivity in comparison to mammography screening alone. Screening with adjunct ultrasound also increases the number of false-positive findings and consequently the number of patient recalls and biopsies. The quality of this body of evidence is low.

Due to a lack of evidence, it is unclear if screening with ultrasound as an adjunct to mammography for high-risk women will reduce breast cancer–related mortality, reduce rates of interval or advanced breast cancer, or detect disease at earlier stages.

Discussion

In Ontario, women aged 50 to 74 years at average risk of breast cancer currently receive mammography screening every two years and are not recommended to receive screening breast ultrasound. Our review identified no primary evidence evaluating the comparative effectiveness or diagnostic accuracy of screening breast ultrasound as an adjunct to mammography for this population. Our findings are in line with the Cochrane review on screening breast ultrasound in women at less than high risk for breast cancer, which identified no methodologically sound studies comparing adjunct ultrasound and mammography with mammography alone in the average-risk population.⁴⁸

Among Ontario women at high risk for breast cancer, annual screening breast mammography and MRI are recommended as the most effective screening tools. Women for whom MRI is contraindicated are funded to receive ultrasound as an adjunct to mammography. We found that the adjunct use of ultrasound can identify breast cancers not visible on mammography, improving the true-positive rate of screening. This improved detection is associated with a corresponding increase in false-positive tests that may lead to unnecessary diagnostic testing and distress for patients.

We did not find any evidence on the impact of these findings on patient-important outcomes such as survival rates. Lead-time bias may therefore be a concern: this occurs when cancer is identified earlier but the progression of disease and survival rates remain the same. The total number of cancers identified across the studies was low, and we could not reach conclusions about the potential of adjunct ultrasound to detect breast cancer at earlier stages. We do not know if the cancers identified by ultrasound were overdiagnosed or overtreated, or if they would have been caught in a later screening. However, additional randomized controlled trials comparing these interventions are unlikely to be feasible or ethical, particularly given that MRI and mammography are currently the gold standard for screening high-risk women and the number of women contraindicated for MRI is likely to remain small.

Other systematic reviews have concluded that screening breast MRI as an adjunct to mammography is superior to screening with mammography alone, to ultrasound alone, or to the

combination of the two with clinical breast exam.^{20,21} Assuming a 2% prevalence of breast cancer, it has been estimated that screening 99 women with adjunct MRI would be needed to identify 1 additional cancer, with 3 additional false-positives, relative to mammography alone.¹⁸ Although the population of high-risk women likely to be contraindicated for MRI is small, these prior systematic reviews established that screening with mammography alone in this population is unlikely to be sufficient.

The prior systematic reviews, however, did not directly compare mammography alone with mammography plus adjunct ultrasound. The sensitivity and specificity of mammography alone in our review is slightly higher than in the earlier reviews. The false-positive rate with MRI and mammography has been estimated at 23% when the threshold for a positive test is a BI-RADS score of 3 or greater, and 5% for a threshold of 4 or greater.²¹ Our review found similar false-positive rates for the same subgroups.

Our review is unique in directly comparing screening breast ultrasound as an adjunct to mammography with mammography screening alone among women at high risk for breast cancer. All studies included in the review used paired study designs, providing direct comparative evidence of effect estimates.

Study Limitations

Several limitations related to the methodology used in the included studies, evidence available, and generalizability of results are discussed below.

Patient Populations

While we identified no studies among average-risk women for inclusion in the present review, limitations related to the definitions used to define breast cancer risk should be noted. This review used a focused definition of breast cancer risk classification, which aligned with current Ontario breast screening definitions. As such, all possible permutations and combinations of breast cancer risk factors could not be thoroughly assessed. This review provides a focused view of literature within the context of Ontario. Numerous excluded studies did not specify the population risk for breast cancer or state if high-risk women were excluded, and therefore could not be included in the analysis.

Although the studies we reviewed among high-risk women used similar inclusion criteria, there were large variations in terms of the percentage of women with a personal history of breast cancer, genetic mutation of the *BRCA1* or *BRCA2* genes, and other high-risk factors. Few studies stratified by risk-factor subgroups, and therefore we could not make meaningful conclusions for these subgroups. If the effectiveness of ultrasound or mammography screening vary depending on the specific factors that classify women as high risk, then studies may be measuring several different population effects.

The study by Berg et al³⁶ included women at elevated but not high risk for breast cancer according to definitions used in Ontario. The higher false-positive rate in this study may therefore reflect the use of a lower BI-RADS threshold as well as the inclusion of women at lower breast cancer risk and cannot be generalized to the high-risk population in Ontario.

Classification of Positive Tests

The interpretation of combined mammography and ultrasound results varied between studies, with some combining the two tests in an integrated assessment and others using a worst-case

approach. It is therefore difficult to ascertain if results from these studies reflect the use of ultrasound provided only to high-risk women with negative mammograms or the use of ultrasound when offered to all high-risk women as a routine part of their breast cancer screening.

In addition to the measured difference in BI-RADS classifications, other non-measurable differences may exist: variation between radiologists, equipment used, and the interpretation of the BI-RADS classification system. While a predefined BI-RADS classification was used in each study, different thresholds may be used by radiologists to move from a score of 3 (probably benign) to 4 (suspicious). As a result, the sensitivity and specificity among the studies may not be based on homogeneous threshold definitions.

The studies that used a BI-RADS threshold of 4 or 5 did not provide a rate of additional short-term follow-up recommendations. The rate of false-positive tests is therefore underestimated if any additional follow-up or testing is to be classified as a potential harm.

Additional Tests

All studies included other screening techniques (clinical breast exam and/or MRI) in addition to mammography and ultrasound. Cancers detected by these additional tests may have developed into interval cancers in the absence of these tests, and therefore true interval cancer rates cannot be established using these study designs. Similarly, cancers detected at a very early stage with MRI may have been detected at an appropriate stage in subsequent screening rounds with mammography and ultrasound. In all the studies, clinical management was based on the results of all the tests that patients received; therefore, we could not determine if the screening methods were associated with differences in actual clinical management.

Meta-analysis

Comparative meta-analytic methods for diagnostic test accuracy studies with paired study designs have not been well validated. We therefore used a simplified approach, assuming independence in estimates. Given the method used and the small number of studies included, we have analyzed the results conservatively.

Existing Guidelines for Technology

Few international guidelines on screening for breast cancer have addressed the adjunct use of ultrasonography. Table 15 provides a summary of guideline recommendations surrounding the use of screening breast ultrasound. In summary, ultrasound has only been recommended in cases where women are at high risk for breast cancer and are contraindicated for MRI. We did not identify any guidelines that recommend the use of screening breast ultrasound for women at average risk for breast cancer.

Table 15: Clinical Guideline Recommendations on the Use of Screening Breast Ultrasound

Guidance, Year	Recommendation
Ontario Breast Screening Program, 2015 ⁴	<ul style="list-style-type: none"> • Offer mammogram and breast MRI every year for women at high risk (screening breast ultrasound when MRI is contraindicated)
National Institute for Health and Care Excellence, 2013 ⁴⁰	<ul style="list-style-type: none"> • Do not routinely offer ultrasound surveillance to women at moderate (greater than 17% to less than 30%) or high risk (30% or greater) of breast cancer but consider it: <ul style="list-style-type: none"> – When MRI surveillance would normally be offered but is not suitable (for example, because of claustrophobia) – When results of mammography or MRI are difficult to interpret
American College of Radiology, 2014 ⁴⁹	<ul style="list-style-type: none"> • Breast ultrasound as a supplement to mammography is appropriately indicated for: <ul style="list-style-type: none"> – Screening for occult cancers in certain populations of women (such as those with dense fibroglandular breasts who are also at elevated risk of breast cancer or with newly suspected breast cancer) who are not candidates for MRI or have no easy access to MRI
International Agency for Research on Cancer, 2015 ⁵⁰	<ul style="list-style-type: none"> • Inadequate to limited evidence of a benefit and sufficient evidence of false-positives with screening breast ultrasonography as an adjunct to mammography in women with dense breasts and negative mammography results • Inadequate evidence of ultrasonography as an adjunct to mammography in screening women with high risk

Abbreviations: MRI, magnetic resonance imaging.

ECONOMIC EVIDENCE REVIEW

Objectives

The objective of this analysis was to review the literature on the cost-effectiveness of ultrasound as an adjunct to mammography compared with mammography alone in breast cancer screening for women at average risk of developing breast cancer and for those at high risk.

Methods

Sources

We performed an economic literature search on June 18, 2015, using Ovid MEDLINE, Ovid MEDLINE In-Process and Other Non-Indexed Citations, Ovid Embase, EBM Reviews, and National Health Service Economic Evaluation Database for studies published from January 1, 1998, to June 18, 2015. We also extracted economic evaluation reports developed by health technology assessment agencies, by searching the websites of agencies such as the Canadian Agency for Drugs and Technologies in Health, Institute of Health Economics, Institut national d'excellence en sante et en services, McGill University Health Centre Health Technology Assessment Unit, and the Cost-Effectiveness Analysis Registry (available at <https://research.tufts-nemc.org/cear4/>). Finally, we reviewed reference lists of included economic literature for any additional relevant studies not identified through the systematic search.

Literature Screening

We based our search terms on those used in the clinical evidence review in this report and applied economic filters to the search results. Study eligibility criteria for the literature search are listed below. Two health economists were involved in the literature review. One reviewed titles and abstracts and, for those studies meeting the inclusion/exclusion criteria, obtained full-text articles. The second health economist reviewed the final full-text articles and abstracted data.

Inclusion Criteria

- English-language full-text publications
- Studies published between January 1, 1998, and June 18, 2015
- Studies on asymptomatic women being screened for breast cancer
- Studies reporting on ultrasound as an adjunct screening method for breast cancer (whether combined with mammography or performed after a negative mammography result)
- Cost-utility analyses, cost-effectiveness analyses, cost-benefit analyses, budget impact analyses, and cost analyses

Exclusion Criteria

- Abstracts, posters, reviews, letters/editorials, foreign language publications, and unpublished studies
- Studies involving ultrasound as a diagnostic intervention after an abnormal mammography result
- Studies in symptomatic women or in breast cancer survivors

Outcomes of Interest

- Full economic evaluations: cost-utility analyses, cost-effectiveness analyses, cost-benefit analyses, cost analyses

Data Extraction

We extracted relevant data on the following:

- Study characteristics (i.e., authors, year of publication)
- Population and comparator
- Interventions
- Outcomes (i.e., health outcomes, costs, and cost-effectiveness)

Results

Literature Search

The database search yielded 408 citations published between January 1, 1998, and June 18, 2015 (with duplicates removed). We excluded a total of 391 articles based on information in the title and abstract. We then obtained the full texts of 16 potentially relevant articles for further assessment. Figure 4 presents the flow diagram for the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA).

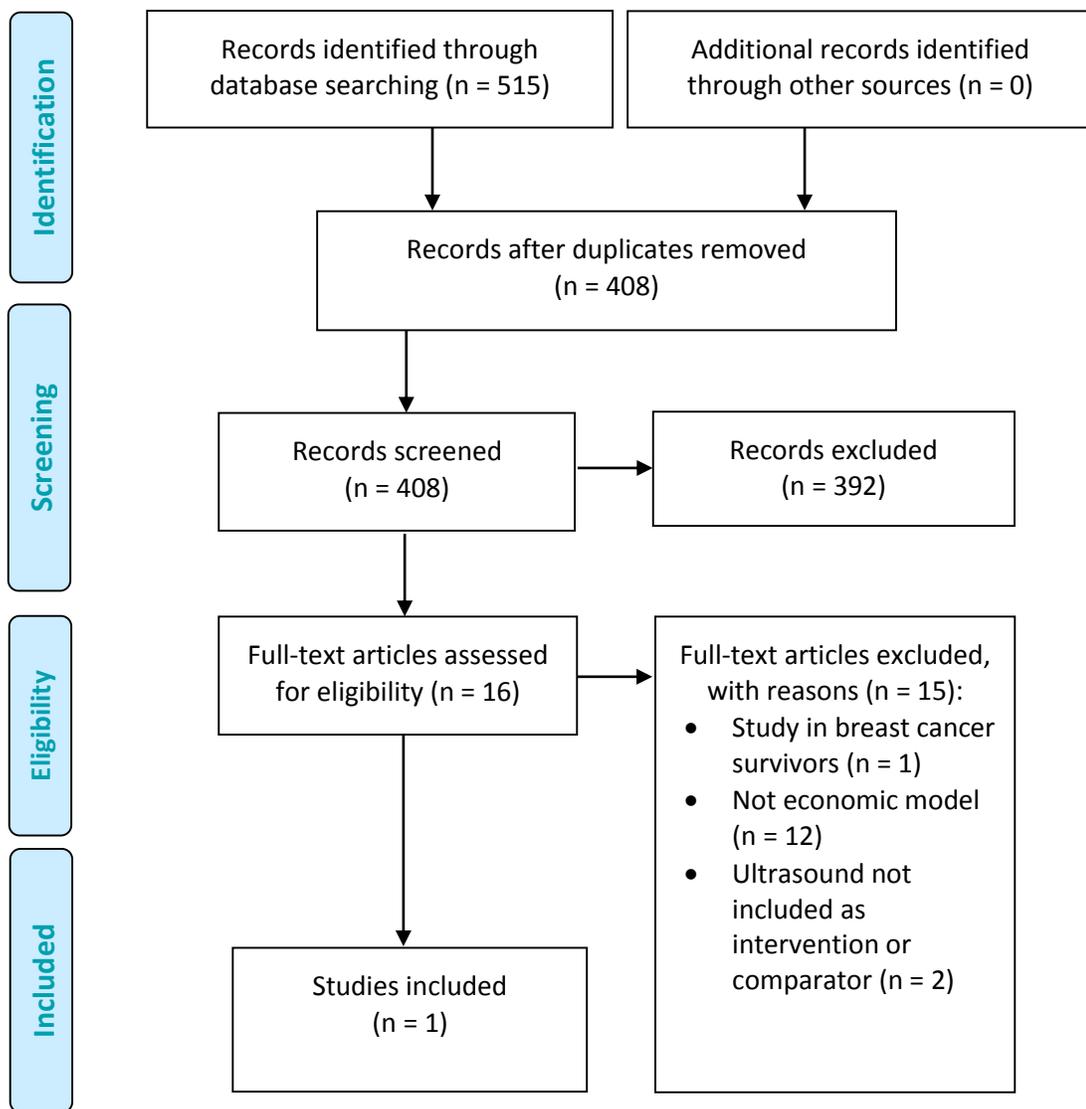


Figure 4: PRISMA Flow Diagram for Economic Evidence Review

Source: Adapted from Moher et al.³³

Critical Review

The only relevant study we identified was an American cost-utility analysis. Sprague et al⁵¹ compared screening with mammography alone and supplemental screening with ultrasound after a negative mammography result for women with dense breasts. Two populations and screening strategies were examined: biennial screening (every two years) in women aged 50 to 74 years and annual screening in women aged 40 to 74 years. The study used three validated microsimulation models to compare breast cancer outcomes, quality-adjusted life-years (QALYs) gained, and costs for mammography alone versus supplemental ultrasound after a negative mammography result.

Table 16 summarizes the study design and results. For biennial screening in women aged 50 to 74 years and with heterogeneously or extremely dense breasts, the authors estimated that adjunct ultrasound screening would result in an additional 2.1 life-years gained and 1.7 QALYs gained per 1,000 women screened, compared with mammography alone. With additional costs of \$560,000 per 1,000 women screened, the authors estimated incremental cost-effectiveness ratios of \$325,000 per life-year gained and \$338,000 per QALY gained for supplemental ultrasound versus mammography alone in this population.

For annual screening in women aged 40 to 74 years with heterogeneously or extremely dense breasts, 3.7 life-years and 3.0 QALYs were gained by using supplemental ultrasound versus mammography alone, at an additional cost of \$2.2 million per 1,000 women screened. This translates into incremental cost-effectiveness ratios of \$728,000 per life-year gained and \$776,000 per QALY gained.

Conclusions

The one relevant study we identified concluded that supplemental ultrasound screening of women with dense breasts might produce small incremental benefits but with substantially increased costs, resulting in unfavourable cost-effectiveness ratios.

Table 16: Results of Economic Literature Review—Summary

Author, Year, Location	Study Design, Perspective	Population, Comparator	Interventions	Results		
				Health Outcomes (per 1,000 Women)	Screening Costs, \$ ^a (per 1,000 Women)	Cost-Effectiveness
Sprague et al, 2015, United States ⁵¹	Cost-utility analysis, payer's perspective	Women aged 50–74 or 40–74 years with dense breasts after a negative mammography Comparator: mammography alone	Mammography + ultrasound after a negative mammography result for women with extremely dense breasts Mammography + ultrasound after a negative mammography result for women with heterogeneously or extremely dense breasts	<p>Biennial screening in women 50–74 years old</p> <p><i>Extremely dense breast:</i></p> <ul style="list-style-type: none"> Deaths averted: 0.30 Life-years gained: 1.2 years QALYs gained: 1.1 <p><i>Heterogeneously or extremely dense breast:</i></p> <ul style="list-style-type: none"> Deaths averted: 0.36 Life-years gained: 2.1 years QALYs gained: 1.7 <p>Annual screening in women 40–74 years old</p> <p><i>Extremely dense breast:</i></p> <ul style="list-style-type: none"> Deaths averted: 0.35 Life-years gained: 3.6 years QALYs gained: 3.1 <p><i>Heterogeneously or extremely dense breast:</i></p> <ul style="list-style-type: none"> Deaths averted: 0.43 Life-years gained: 3.7 years QALYs gained: 3.0 	<p>Biennial screening in women 50–74 years old</p> <p><i>Mammography alone:</i> 3.02 million</p> <p><i>Mammography + ultrasound, extremely dense breast:</i> 3.08 million</p> <p><i>Mammography + ultrasound, heterogeneously or extremely dense breast:</i> 3.39 million</p> <p>Annual screening in women 40–74 years old</p> <p><i>Mammography alone:</i> 5.15 million</p> <p><i>Mammography + ultrasound, extremely dense breast:</i> 5.42 million</p> <p><i>Mammography + ultrasound, heterogeneously or extremely dense breast:</i> 6.58 million</p>	<p>Biennial screening in women 50–74 years old</p> <p><i>For extremely dense breast:</i> \$246,000/QALY</p> <p><i>For heterogeneously or extremely dense breast:</i> \$325,000/QALY</p> <p>Annual screening in women 40–74 years old</p> <p><i>For extremely dense breast:</i> \$553,000/QALY</p> <p><i>For heterogeneously or extremely dense breast:</i> \$728,000/QALY</p>

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life-year.

^a 2014 US dollars.

BUDGET IMPACT ANALYSIS

We conducted a cost analysis from the perspective of the Ontario Ministry of Health and Long-Term Care to determine the estimated cost burden of providing adjunct ultrasound in breast cancer screening over the next 5 years. All costs are reported in 2015 Canadian dollars.

Objectives

The objective of this analysis was to evaluate the cost consequences of funding, under the Ontario Breast Screening Program, ultrasound as an adjunct to mammography for breast cancer screening in high-risk women contraindicated for magnetic resonance imaging (MRI).

This analysis was limited to the costs of screening and follow-up procedures; we did not evaluate the long-term costs of treatment for cancers found or potential savings from early detection.

Methods

Target Population

The target population was women with high risk for breast cancer who are screened with mammography but contraindicated for adjunct screening with MRI. Current practice in Ontario is for all high-risk women who are contraindicated for MRI to receive ultrasound as an adjunct screening test, regardless of the results of mammography (Personal communication, Dr. Anna Chiarelli, October 30, 2015, and Dr. Derek Muradali, November 3, 2015).

Resource

We used data from Cancer Care Ontario for women screened under the Ontario Breast Screening Program for the period July 2011 to March 2015. The dataset provided the number of high-risk women aged 30 to 69 years who were contraindicated for MRI and were screened with both mammography and ultrasound, the number of women who had an abnormal result from ultrasound after being screened with mammography, and the number of follow-up assessment procedures that those women received after their abnormal ultrasound. These follow-up assessment procedures included only cases of BIRADS 4 and BIRADS 5. Based on these data, we forecasted the corresponding numbers for the next 5 years.

Assumptions

- The number of additional high-risk women who are contraindicated for MRI each year would remain constant over the next 5 years (i.e., the number would grow by the same amount each year).
- All high-risk women contraindicated for MRI would receive bilateral screening (both breasts) by both mammography and ultrasound.

Number of High-Risk Women Screened With Both Tests

Table 17 shows the actual number of high-risk women who received both mammography and ultrasound screening from July 2011 to March 2015 and our estimates for April 2015 to March 2020. The number grew by 60, 49, and 50 each year, respectively, from 2011/12 to 2014/15.

This translates into an average increase of 53 women per year, and we conservatively assumed that the same trend would continue between April 2015 and March 2020.

Table 17: Number of High-Risk Women Screened by Mammography Plus Ultrasound, Actual and Forecast, 2011/12 to 2019/20

Number of High-Risk Women Screened by Mammography Plus Adjunct Ultrasound, by Year								
Actual ^a				Forecast				
2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19	2019/20
10	70	119	169	222	275	328	381	434

^aSource: Cancer Care Ontario, Ontario Breast Screening Program.

Number of Abnormal Ultrasound Results

The data from Cancer Care Ontario showed that between July 2011 and March 2015 a total of 368 women who were contraindicated for MRI were screened with both mammography and ultrasound under the Ontario Breast Screening Program. Of these, 22 women received an abnormal result on their ultrasound. This translates to a rate of 6%: for every 100 women screened with both mammography and ultrasound, 6 women would have an abnormal ultrasound. Based on this rate, we estimated the number of abnormal ultrasounds that would be expected among high-risk women screened under the Ontario Breast Screening Program from 2015/16 to 2019/20 (Table 18).

Table 18: Expected Number of Abnormal Ultrasound Results After Screening by Mammography Plus Ultrasound, Ontario Breast Screening Program, 2015/16 to 2019/20

Estimated Number of High-Risk Women Expected to Receive Abnormal Ultrasound Breast Cancer Screening Result, by Year				
2015/16	2016/17	2017/18	2018/19	2019/20
13	17	20	23	26

Number of Follow-Up Assessment Procedures

Among the 22 women who received an abnormal ultrasound after being screened with mammography and ultrasound under the Ontario Breast Screening Program from 2011/12 to 2014/15, 47 follow-up assessment procedures were done: 19 biopsies, 15 ultrasounds, 8 diagnostic mammography tests, and 5 surgical consultations. Based on this information, we estimated the number of follow-up assessment procedures following abnormal ultrasounds for the period 2015/16 to 2019/20 (Table 19).

Table 19: Estimated Number of Follow-Up Procedures for High-Risk Women Screened by Mammography Plus Ultrasound, 2015/16 to 2019/20

Procedures	Estimated Number of Follow-Up Assessments for High-Risk Women, by Year				
	2015/16	2016/17	2017/18	2018/19	2019/20
Biopsies	12	15	18	21	24
Ultrasound	10	12	14	16	19
Diagnostic mammography	5	6	7	9	10
Total	27	33	39	46	52

Canadian Costs

Tables 20 and 21 show the unit costs in 2015 for screening tests (screening mammography, screening ultrasound) and for follow-up assessment and diagnostic procedures (biopsies, surgical consultations, diagnostic mammography, physician consultations) from the Ontario Schedule and Benefits for Physician Services.²³

Table 20: Breast Screening Costs

Code	Screening Types	Unit Cost, \$		
		Technical	Professional	Total, \$
Mammography (screening)				
X172	Unilateral	28.05	16.90	44.95
X178	Bilateral	37.15	27.00	64.15
Mammography: signs or symptoms (assumed diagnostic)				
X184	Unilateral	28.05	16.90	44.95
X185	Bilateral	37.15	27.00	64.15
Ultrasound				
J127	Scan B-mode (per breast)	23.70	16.40	40.10

Source: Ontario Health Insurance Plan.²³

Table 21: Costs of Follow-Up Procedures After Abnormal Results of Screening Breast Ultrasound

Code	Description	Procedure	Unit Cost, \$		
			Technical	Professional	Total, \$
J149	Ultrasound guidance	Biopsy	47.30	36.85	84.15
A035		Surgical consultation		90.30	90.30
J127	Ultrasound	Scan B-mode (per breast)	23.7	16.40	40.10
X184	Diagnostic mammogram	Unilateral	28.05	16.90	44.95
X185		Bilateral	37.15	27.00	64.15
A005		Physician consultation		77.20	77.20

Source: Ontario Health Insurance Plan.²³

Scenario Analyses

The data from Cancer Care Ontario did not indicate how many of the 22 high-risk women who received an abnormal ultrasound result had also tested positive on their mammograms. However, as noted above, the practice in Ontario is for all high-risk women who are contraindicated for MRI to be screened by both ultrasound and mammography, regardless of the mammography results. We therefore proposed two scenarios for analysis to calculate the highest and the lowest costs associated with screening ultrasound:

Scenario 1: We assumed that all women received a negative result from screening mammography. In this scenario, all abnormal screening results were caught only by screening ultrasound.

Scenario 2: We assumed that all women received a positive result from screening by mammography and the results were confirmed by ultrasound. In this scenario, all abnormal screening results were caught by both mammography and ultrasound.

Cost Analysis

We calculated two cost components: costs of screening by mammography and ultrasound and costs of follow-up assessment procedures if an abnormal ultrasound was detected. For the first component, the total cost was calculated by multiplying the number of high-risk women contraindicated for MRI by the unit costs of mammography and ultrasound screening tests. For the second component, the total cost was calculated by multiplying the total number of follow-up procedures by the respective unit costs. The grand total was the sum of the total screening cost and the total follow-up cost.

Results

Base Case

Table 22 summarizes the estimated costs in Ontario over 5 years (2015/16 to 2019/20) of funding screening breast ultrasound as an adjunct to screening mammography in high-risk women who are contraindicated for MRI.

Table 22: Estimated Breast Cancer Screening Costs for High-Risk Women Contraindicated for MRI, Ontario, 2015 to 2019/20

Cost Components	Estimated Costs, \$, by Year				
	2015/16	2016/17	2017/18	2018/19	2019/20
Screening					
Mammography	14,241	17,641	21,041	24,441	27,841
Ultrasound	15,473	19,168	22,862	26,556	30,250
Total screening cost	29,715	36,809	43,903	50,997	58,091
Follow-up procedures					
Biopsies	2,102	2,604	3,106	3,608	4,110
Ultrasound	663	821	980	1,138	1,296
Diagnostic mammography	717	888	1,060	1,231	1,402
Total follow-up assessment cost	3,483	4,314	5,146	5,977	6,809
Grand total cost	33,197	41,123	49,049	56,974	64,900

Abbreviation: MRI, magnetic resonance imaging.

Scenarios Results

Scenario 1: Assuming that all abnormal results would be detected by ultrasound only (i.e., all women tested negative with screening mammography and positive with ultrasound), the cost of follow-up assessment procedures would be attributed to the costs associated with adding screening breast ultrasound as an adjunct to screening mammography for high-risk women who are contraindicated for MRI. In this case, the additional budget required to fund ultrasound would consist of the cost of adding ultrasound as the adjunct screening method and the cost of follow-up assessment procedures (Table 23).

Table 23: Additional Budget Required to Fund Breast Cancer Screening for High-Risk Women Contraindicated for MRI, Assuming Abnormal Results Are Detected by Ultrasound Only

Cost Components	Estimated Costs, \$, by Year				
	2015/16	2016/17	2017/18	2018/19	2019/20
Screening ultrasound	15,473	19,168	22,862	26,556	30,250
Follow-up assessment procedures	3,483	4,314	5,146	5,977	6,809
Total cost	18,956	23,482	28,007	32,533	37,058

Scenario 2: Assuming that all abnormal results would be detected by both screening mammography and ultrasound (i.e., all women with a positive screening mammography and positive screening ultrasound), the cost of follow-up assessment procedures would be attributed to mammography itself. In this case, the additional budget required would only be the cost of ultrasound as the adjunct screening method (Table 24).

Table 24: Additional Budget Required to Fund Breast Cancer Screening for High-Risk Women Contraindicated for MRI, Assuming Abnormal Results Are Detected by Both Mammography and Ultrasound

Cost Components	Estimated Costs, \$, by Year				
	2015/16	2016/17	2017/18	2018/19	2019/20
Screening ultrasound	15,473	19,168	22,862	26,556	30,250

Limitations

Our analysis was limited to the costs of breast cancer screening and follow-up assessment procedures for the small group of women who have high risk for breast cancer and are contraindicated for MRI. Given the nature of this analysis, we did not capture the costs of treating breast cancer, nor could we project the potential cost savings that might result if cases of breast cancer are detected by adjunct ultrasound screening. Understanding those costs and savings would require a full cost-effectiveness analysis including the development of a decision analytic model that considers various screening strategies, treatment pathways, and potential outcomes.

Additionally, the data we used did not include information needed to calculate the false-positive rate for screening mammography or screening ultrasound in Ontario. Therefore, we could not calculate the budget strictly required for adding screening breast ultrasound as an adjunct to screening mammography. For example, due to this data limitation, we could not confidently attribute the costs of follow-up assessment to either mammography or ultrasound. Our scenario analyses presented the least likely and most likely scenarios, given the unknown false-positive rates for the two screening methods. Furthermore, in our budget impact assessment, we considered costs only of follow-up cases of BIRADS 4 and BIRADS 5 since the costs for follow-up cases for BIRADS 3 were not available. The estimated total cost may, therefore, be an underestimation.

We did not compare the cost of screening mammography and ultrasound because all high-risk women in Ontario who are contraindicated with MRI receive both mammography and ultrasound screening regardless of the results of mammography. Therefore, the incremental effect of screening breast ultrasound based on local data is unclear.

Discussion and Conclusions

This analysis explored the additional costs of providing screening breast ultrasound as an adjunct to screening mammography among high-risk women who are contraindicated for MRI.

Based on data provided by Cancer Care Ontario, we estimated that each year between 2015/16 and 2019/20 about 53 additional high-risk women who are contraindicated for MRI would be screened with both mammography and ultrasound under the Ontario Breast Screening Program. This translates to a total of between 222 and 434 women who would be screened by

both methods each year over the next 5 years (Details are in Table 17). The additional annual cost of screening ultrasound would range from \$15,500 in 2015/16 to \$37,000 in 2019/20. If abnormal results are detected with screening ultrasound, the annual cost of follow-up assessment procedures would range from \$3,500 to \$6,800 in the next 5 years.

The number and costs of this additional screening and follow-up for women at high risk for breast cancer are very small and would likely have a small budget impact if ultrasound were publicly funded as an adjunct to screening mammography in Ontario.

ABBREVIATIONS

AMSTAR	Assessment of Multiple Systematic Reviews
BI-RADS	Breast Imaging Reporting and Data System
DCIS	Ductal carcinoma in-situ
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
MRI	Magnetic resonance imaging
NICE	(UK) National Institute for Health and Care Excellence
PPV	Positive predictive value
QALY	Quality-adjusted life-year

GLOSSARY

BRCA1 and BRCA2	BRCA1 and BRCA2 are genes that help control cell growth. People with abnormalities in these genes are at higher risk of certain cancers, including breast, prostate and ovarian cancer.
Diagnostic yield	The diagnostic yield is the proportion of persons screened in a study who are correctly diagnosed as having the condition tested.
False-negative	A test result that indicates a person does not have the disease or condition tested for when they actually do have it.
False-positive	A test result that indicates a person has the disease or condition tested for when they actually do not have it.
Incremental cost-effectiveness ratio (ICER)	Determines “a unit of benefit” for an intervention by dividing the incremental cost by the incremental effectiveness. The incremental cost is the difference between the cost of the treatment under study and an alternative treatment. The incremental effectiveness is usually measured as additional years of life or as “quality-adjusted life years.”
Positive predictive value	The proportion of people with a positive test result who actually have the disease or characteristic.
Quality-adjusted life-year (QALY)	A measurement that takes into account both the number of years gained by a patient from a procedure and the improved quality of those extra years (ability to function, freedom from pain, etc.). One QALY is expressed as a number between zero (no benefit) and one (perfect health). The QALY is commonly used as an outcome measure in cost–utility analyses.
Receiver operating characteristic (ROC) curve	A technique to visually represent the effectiveness of a diagnostic test by weighing the rate of true positive results (where the patient has the condition tested) against the rate of false positive results (positive findings among patients who do not have the condition tested).
Sensitivity	The ability of a test to accurately identify persons with the condition tested for (how well it returns positive results in persons who have the condition).
Specificity	The ability of a test to accurately identify persons who do not have the condition tested for (how well it returns negative results in persons who do not have the condition).
Statistical significance	The result of an analysis is considered to be statistically significant if the assumption tested (the “null hypothesis”) is sufficiently unlikely to be true. Typically, the result is considered statistically significant if there is less than a 5% chance that the result would have occurred if the null hypothesis were true.
True-negative	A test result where a person who does not have the disease or condition tested for is correctly identified as not having it.
True-positive	A test result where a person who has the disease or condition tested for is correctly identified as having it.

APPENDICES

Appendix 1: Literature Search Strategies

Clinical Evidence Review

We selected 1998 as the start date for the literature search because the first research evaluating ultrasound as an adjunct method of breast cancer screening was published that year. This decision was supported by experts in the field.

All Ovid MEDLINE(R) <1946 to Present>, Embase <1980 to 2015 Week 24>, EBM Reviews - Cochrane Central Register of Controlled Trials <May 2015>, EBM Reviews - Cochrane Database of Systematic Reviews <2005 to May 2015>, EBM Reviews - Database of Abstracts of Reviews of Effects <2nd Quarter 2015>, EBM Reviews - Health Technology Assessment <2nd Quarter 2015>, EBM Reviews - NHS Economic Evaluation Database <2nd Quarter 2015>.

Search Strategy:

-
- 1 exp Breast Neoplasms/ (613873)
 - 2 Carcinoma, Lobular/ (29066)
 - 3 ((breast* or mammar*) adj2 (cancer* or neoplas* or tumo?* or carcinoma* or adenoma or pre-cancer or precancer or dysplasia)).tw. (549709)
 - 4 exp *BRCA1* Protein/ or exp Genes, *BRCA1*/ (67196)
 - 5 exp *BRCA2* Protein/ or exp Genes, *BRCA2*/ (61641)
 - 6 (*BRCA** or ((high risk or increase* risk or hereditary or family histor* or genetic*) and breast*)).tw. (98890)
 - 7 exp Mammography/ (68175)
 - 8 mammogra*.mp. (79419)
 - 9 or/1-8 (833699)
 - 10 exp Ultrasonography/ (823558)
 - 11 (ultrasound* or ultrasonogra* or sonogra* or echotomograph* or ultrasonic or echogra* or echoscop* or ABUS or HHUS).mp. (931796)
 - 12 or/10-11 (1234145)
 - 13 exp Mass Screening/ (276638)
 - 14 "Early Detection of Cancer"/ (87110)
 - 15 (screen* or (earl* adj (diagnos* or detect*))).tw. (1398239)
 - 16 or/13-15 (1522578)
 - 17 9 and 12 and 16 (5918)
 - 18 limit 17 to english language [Limit not valid in CDSR,DARE; records were retained] (4969)
 - 19 18 use pmoz,cctr,coch,dare,clhta,cleed (1641)
 - 20 exp Breast Cancer/ (546215)
 - 21 ((breast* or mammar*) adj2 (cancer* or neoplas* or tumo?* or carcinoma* or adenoma or pre-cancer or precancer or dysplasia)).tw. (549709)
 - 22 *BRCA1* protein/ (14404)
 - 23 *BRCA2* protein/ (10214)
 - 24 (*BRCA** or ((high risk or increase* risk or hereditary or family histor* or genetic*) and breast*)).tw. (98890)
 - 25 exp Mammography/ (68175)
 - 26 mammogra*.mp. (79419)

- 27 or/20-26 (749506)
- 28 Ultrasound/ (115001)
- 29 Echography/ (302108)
- 30 (ultrasound* or ultrasonogra* or sonogra* or echotomograph* or ultrasonic or echogra* or echoscop* or ABUS or HHUS).mp. (931796)
- 31 or/28-30 (931796)
- 32 Mass Screening/ (136224)
- 33 Cancer Screening/ (63618)
- 34 Early Diagnosis/ (91643)
- 35 (screen* or (earl* adj (diagnos* or detect*))).tw. (1398239)
- 36 or/32-35 (1490525)
- 37 27 and 31 and 36 (5126)
- 38 limit 37 to english language [Limit not valid in CDSR,DARE; records were retained] (4265)
- 39 38 use emez (2651)
- 40 19 or 39 (4292)
- 41 limit 40 to yr="1998 -Current" [Limit not valid in DARE; records were retained] (3809)
- 42 remove duplicates from 41 (2748)

Economic Evidence Review

EBM Reviews - Cochrane Central Register of Controlled Trials <May 2015>, EBM Reviews - Cochrane Database of Systematic Reviews <2005 to May 2015>, EBM Reviews - Database of Abstracts of Reviews of Effects <2nd Quarter 2015>, EBM Reviews - Health Technology Assessment <2nd Quarter 2015>, EBM Reviews - NHS Economic Evaluation Database <2nd Quarter 2015>, Embase <1980 to 2015 Week 24>, All Ovid MEDLINE(R) <1946 to Present>

Search Strategy:

-
- 1 exp Breast Neoplasms/ (613976)
 - 2 Carcinoma, Lobular/ (29076)
 - 3 ((breast* or mammar*) adj2 (cancer* or neoplas* or tumo?r* or carcinoma* or adenoma or pre-cancer or precancer or dysplasia)).tw. (549600)
 - 4 exp *BRCA1* Protein/ or exp Genes, *BRCA1*/ (67204)
 - 5 exp *BRCA2* Protein/ or exp Genes, *BRCA2*/ (61646)
 - 6 (*BRCA** or ((high risk or increase* risk or hereditary or family histor* or genetic*) and breast*)).tw. (98866)
 - 7 exp Mammography/ (68182)
 - 8 mammogra*.mp. (79399)
 - 9 or/1-8 (833583)
 - 10 exp Ultrasonography/ (823617)
 - 11 (ultrasound* or ultrasonogra* or sonogra* or echotomograph* or ultrasonic or echogra* or echoscop* or ABUS or HHUS).mp. (931603)
 - 12 or/10-11 (1233981)
 - 13 exp Mass Screening/ (276652)
 - 14 "Early Detection of Cancer"/ (87122)
 - 15 (screen* or (earl* adj (diagnos* or detect*))).tw. (1397847)
 - 16 or/13-15 (1522193)
 - 17 economics/ (245733)

- 18 economics, medical/ or economics, pharmaceutical/ or exp economics, hospital/ or economics, nursing/ or economics, dental/ (691529)
- 19 economics.fs. (362604)
- 20 (econom* or price or prices or pricing or priced or discount* or expenditure* or budget* or pharmaco-economic* or pharmaco-economic*).tw. (627600)
- 21 exp "costs and cost analysis"/ (480120)
- 22 cost*.ti. (216411)
- 23 cost effective*.tw. (224905)
- 24 (cost* adj2 (util* or efficacy* or benefit* or minimi* or analy* or saving* or estimate* or allocation or control or sharing or instrument* or technolog*).ab. (140893)
- 25 models, economic/ (125304)
- 26 markov chains/ or monte carlo method/ (114508)
- 27 (decision adj1 (tree* or analy* or model*).tw. (30600)
- 28 (markov or markow or monte carlo).tw. (90649)
- 29 quality-adjusted life years/ (25520)
- 30 (QOLY or QOLYs or HRQOL or HRQOLs or QALY or QALYs or QALE or QALEs).tw. (43651)
- 31 ((adjusted adj (quality or life)) or (willing* adj2 pay) or sensitivity analys*s).tw. (86261)
- 32 or/17-31 (2122483)
- 33 9 and 12 and 16 and 32 (781)
- 34 limit 33 to english language [Limit not valid in CDSR,DARE; records were retained] (721)
- 35 34 use pmoz,cctr,coch,dare,clhta (246)
- 36 9 and 12 and 16 (5915)
- 37 36 use cleed (29)
- 38 35 or 37 (275)
- 39 exp Breast Cancer/ (546318)
- 40 ((breast* or mammar*) adj2 (cancer* or neoplas* or tumo?r* or carcinoma* or adenoma or pre-cancer or precancer or dysplasia)).tw. (549600)
- 41 BRCA1 protein/ (14407)
- 42 BRCA2 protein/ (10216)
- 43 (BRCA* or ((high risk or increase* risk or hereditary or family histor* or genetic*) and breast*)).tw. (98866)
- 44 exp Mammography/ (68182)
- 45 mammogra*.mp. (79399)
- 46 or/39-45 (749390)
- 47 Ultrasound/ (115001)
- 48 Echography/ (302109)
- 49 (ultrasound* or ultrasonogra* or sonogra* or echotomograph* or ultrasonic or echogra* or echoscop* or ABUS or HHUS).mp. (931603)
- 50 or/47-49 (931603)
- 51 Mass Screening/ (136234)
- 52 Cancer Screening/ (63630)
- 53 Early Diagnosis/ (91649)
- 54 (screen* or (earl* adj (diagnos* or detect*))).tw. (1397847)
- 55 or/51-54 (1490140)
- 56 Economics/ (245733)
- 57 Health Economics/ or exp Pharmacoeconomics/ (207908)
- 58 Economic Aspect/ or exp Economic Evaluation/ (371416)

- 59 (econom* or price or prices or pricing or priced or discount* or expenditure* or budget* or pharmaco-economic* or pharmaco-economic*).tw. (627600)
- 60 exp "Cost"/ (480120)
- 61 cost*.ti. (216411)
- 62 cost effective*.tw. (224905)
- 63 (cost* adj2 (util* or efficacy* or benefit* or minimi* or analy* or saving* or estimate* or allocation or control or sharing or instrument* or technolog*).ab. (140893)
- 64 Monte Carlo Method/ (46451)
- 65 (decision adj1 (tree* or analy* or model*)).tw. (30600)
- 66 (markov or markow or monte carlo).tw. (90649)
- 67 Quality-Adjusted Life Years/ (25520)
- 68 (QOLY or QOLYs or HRQOL or HRQOLs or QALY or QALYs or QALE or QALEs).tw. (43651)
- 69 ((adjusted adj (quality or life)) or (willing* adj2 pay) or sensitivity analys*s).tw. (86261)
- 70 or/56-69 (1737857)
- 71 46 and 50 and 55 and 70 (621)
- 72 limit 71 to english language [Limit not valid in CDSR,DARE; records were retained] (571)
- 73 72 use emez (314)
- 74 38 or 73 (589)
- 75 limit 74 to yr="1998 -Current" [Limit not valid in DARE; records were retained] (515)
- 76 remove duplicates from 75 (413)

Appendix 2: Diagnostic Accuracy Definitions and Calculations

Table A1: Summary of Calculations Used to Assess Diagnostic Performance

Measure	Formula
Sensitivity (true-positive rate)	$TP / (TP + FN)$
Specificity	$TN / (FP + TN)$
False-positive rate	$1 - (TN / (FP + TN))$
Positive predictive value (PPV)	$TP / (TP + FP)$
Diagnostic yield	Number of cancers / Number of screens
Number needed to screen (NNS)	$1 / [(Sensitivity^{Mammography+Ultrasound} \times Prevalence) - (Sensitivity^{Mammography} \times Prevalence)]$
Additional false-positives	$NNS \times [(Specificity^{Mammography} \times (1 - Prevalence)) - (Specificity^{Mammography+Ultrasound} \times (1 - Prevalence))]$

Abbreviations: FN, false-negative; FP, false-positive; TN, true-negative; TP, true-positive.

Appendix 3: Excluded Systematic Reviews of Studies of Women With Dense Breasts and no Risk Specified

Table A2: Summary of Systematic Reviews of Screening Breast Ultrasound as an Adjunct to Mammography in Women with Dense Breasts

Author, Year	Search Dates	Key Selection Criteria	Outcomes	Studies Included	Key Results	Review Conclusions
Tice et al, 2013 ⁵²	1945 to June 2013	<ul style="list-style-type: none"> Women with negative mammography and high breast density Imaging with ABUS or HHUS Exclusion: studies focused on high-risk women 	<ul style="list-style-type: none"> Mortality Diagnostic test characteristics 	14 HHUS, 3 ABUS	Adjunct ultrasound identified 2 to 3 more cancers than mammography alone; PPV of 7%, 98 additional recalls, 49 biopsies	<ul style="list-style-type: none"> Significant limitations and high uncertainty in results
Blue Cross Blue Shield et al, 2014 ⁵³	2000 to 2013	<ul style="list-style-type: none"> Asymptomatic women with dense breasts Normal screening mammogram Imaged with HHUS or ABUS 	<ul style="list-style-type: none"> Survival Quality of life Negative biopsy, unnecessary testing 	7 studies	<ul style="list-style-type: none"> 0.4–19 cancers per 1,000 screens 11–132 biopsies per 1,000 screens 	<ul style="list-style-type: none"> Existing evidence on ultrasound suggests that adding it to mammography will increase cancer yield while also increasing the number of recalls and unnecessary biopsies
Nothacker et al, 2009 ⁵⁴	2000 to 2008	<ul style="list-style-type: none"> Asymptomatic, primarily healthy women Negative mammogram and reporting on breast density (BI-RADS 2 to 4) Supplemental breast ultrasound 	<ul style="list-style-type: none"> Cancer detection rates, tumour characteristics Sensitivity, specificity, PPV, number of biopsies 	6 cohort studies, 2 with adequate follow-up	<ul style="list-style-type: none"> Diagnostic yield of 0.32% (0.23%–0.41%) Mean tumour size 9.9 mm, 90% with negative lymph node status Most cancers detected in women with dense breasts Biopsy rates: 2.3%–4.7% Mean PPV 15% (2%–28%) 	<ul style="list-style-type: none"> Supplemental ultrasound in women with mammographically dense breast tissue permits detection of small, otherwise occult, breast cancers; potential adverse impacts are associated with an increased biopsy rate Significant population heterogeneity and study design limitations
Scheel et al, 2015	2000 to 2013	<ul style="list-style-type: none"> Asymptomatic women with normal screening mammograms and dense breasts Screening with ABUS or HHUS 	<ul style="list-style-type: none"> Survival, quality of life, negative biopsy, unnecessary testing 	10 HHUS, 2 ABUS	<ul style="list-style-type: none"> Detected 0.3–7.7 cancers per 1,000 screens Additional 11.7–106.6 biopsies per 1,000 screens 	<ul style="list-style-type: none"> Existing evidence on HHUS ultrasound suggests that adding it to mammography will increase the cancer yield while also increasing number of recalls and unnecessary biopsies; insufficient evidence is available for ABUS

Abbreviations: ABUS, automated breast ultrasound; BI-RADS, Breast Imaging Reporting and Data System; HHUS, hand-held ultrasound; PPV, positive predictive value.

Appendix 4: Evidence Quality Assessment

Table A3: AMSTAR Scores of Included Systematic Reviews

Author, Year	AMSTAR Score	(1) Provided Study Design	(2) Duplicate Study Selection	(3) Broad Literature Search	(4) Considered Status of Publication	(5) Listed Excluded Studies	(6) Provided Characteristics of Studies	(7) Assessed Scientific Quality	(8) Considered Quality in Report	(9) Methods to Combine Appropriate	(10) Assessed Publication Bias	(11) Stated Conflict of Interest
Gartlehner et al 2013 ⁴⁸	10	✓	✓	✓	✓	✓	✓ ^b	✓ ^b	✓ ^b	✓ ^b	X	✓
NICE, 2013 ⁴⁰	6	✓	X	✓	X	X	✓	✓	✓	Not stated	X	✓

Abbreviations: AMSTAR, Assessment of Multiple Systematic Reviews.

^aMaximum possible score is 11. Details of AMSTAR score are described in Shea et al.²⁸

^bNo studies were identified in the review; however, descriptions of planned data extraction, synthesis and quality assessment were provided in the methods.

Table A4: Risk of Bias for Studies of Screening Breast Ultrasound (QUADAS-2)

Author, Year	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Riedl et al, 2012 ⁴¹	High ^a	Low	High ^b	High ^{c,d}	No	No	Low
Kuhl et al, 2010 ³⁸	Unclear ^e	Low	High ^b	High ^{c,d}	No	No	Low
Kuhl et al, 2005 ³⁹	Unclear ^e	Low	High ^b	High ^{c,d}	No	No	Low
Sardanelli et al, 2011 ⁴²	Unclear ^e	High ^f	High ^b	High ^{c,d}	No	No	Low
Berg et al, 2012 ³⁶	Unclear ^e	Low	High ^b	High ^{c,d}	Yes ^g	No	Low

Abbreviations: QUADAS, Quality Assessment of Diagnostic Accuracy Studies.

^aFirst 3 years of screening included only *BRCA* mutation carriers, but the study was expanded to high-risk women in year 3.

^bThe accuracy of the reference standard for positive tests was likely to correctly classify the condition; however, for negative tests the reference standard of clinical or phone follow-up is imperfect. The reference standard was interpreted with knowledge of the index tests. Despite these limitations, the reference standards used are considered the gold standard for screening studies.

^cNot all women eligible or enrolled were included in the analysis due to loss to follow-up or incomplete screening rounds.

^dOnly women who tested positive received biopsy, resulting in a potential for differential verification bias.

^eRisk of bias is unclear as studies did not report if consecutive or selected patients were recruited. Other measures contributing to patient selection were of low risk of bias through the use of prospective, paired study design and by avoiding inappropriate exclusion criteria.

^fResults of mammography, ultrasound, and magnetic resonance imaging were blinded to one another; however, results of the clinical breast exam were not blinded at time of ultrasound and mammography.

^gThe definition of high risk in this study did not meet the criteria for high risk in Ontario. Using Ontario guidelines, both intermediate and high-risk women were included.

Table A5: GRADE Evidence Profile for Comparative Diagnostic Accuracy of Mammography and Mammography With Adjunct Breast Ultrasound

Number of Studies (Design)	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Quality
Incremental Diagnostic Accuracy of Mammography + Adjunct Ultrasound Compared With Mammography Alone: BI-RADS Threshold ≥ 4						
4 (observational)	Very serious limitations (-2) ^a	No serious limitations	No serious limitations	Serious limitations ^b	Undetected	⊕⊕ Low
Incremental Diagnostic Accuracy of Mammography + Adjunct Ultrasound Compared With Mammography Alone: BI-RADS Threshold ≥ 3						
1 (observational)	Serious limitations (-1) ^a	No serious limitations ^e	Serious limitations (-1) ^d	Serious limitations (-1) ^c	Undetected	⊕ Very low

^aUnclear or biased patient selection and potential for differential verification bias. The reference standard was imperfect; however, biopsy and clinical follow-up are considered the gold standard for breast cancer screening studies. See QUADAS-2 Risk of Bias Assessment (Table A2).

^bThe confidence intervals around sensitivity were wide; however, results were not downgraded for imprecision as the overall effect was significant irrespective of a small number of cancers identified. Furthermore, meta-analytic methods used lead to conservative effect estimates.

^cThe confidence intervals around the individual summary estimates for sensitivity and specificity as well as the absolute difference in estimates were substantially wide.

^dThis study did not define high risk as it is defined in Ontario; women with a personal history of breast cancer or high-risk lesions were classified as high risk.

^eInconsistency could not be assessed as only one study was included in the evaluation.

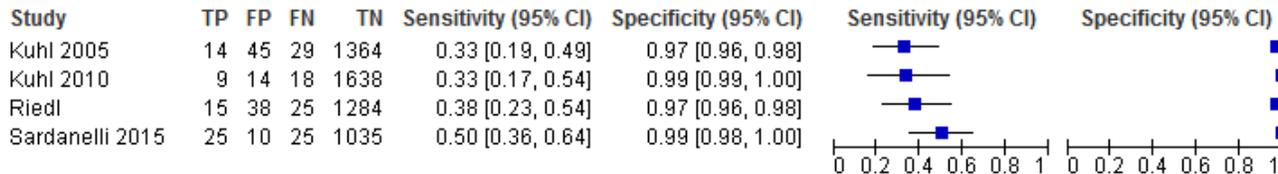
Appendix 5: Supplemental Results

Table A6: Summary of Screening Values in Primary Studies

Author, Year	Screens, N	TP		FP		TN		FN	
		M	M+US	M	M+US	M	M+US	M	M+US
All Women									
Riedl et al, 2015 ⁴¹	1,365	15	20	38	57	1,287	1,268	25	20
Sardanelli et al, 2011 ⁴²	1,095	25	30	10	24	1,035	975	25	18
Kuhl et al, 2010 ³⁸	1,452	14	21	45	155	1,364	1,254	29	22
Kuhl et al, 2005 ³⁹	1,679	9	13	14	27	1,638	1,625	18	14
Berg et al, 2012, Round 1 ³⁶	2,659	20	34	286	673	2,337	1,950	16	2
Berg et al, 2012, Round 2,3 ³⁶	4,814	39	57	414	752	4,325	3,987	36	18
No Personal History									
Berg et al, 2012 ³⁶	3,463	26	41	360	757	3,051	2,654	26	11
Kuhl et al, 2005 ³⁹	1,176	10	16	33	121	1,112	1,024	21	15

Abbreviations: FN, false-negative; FP, false-positive; M, mammography, N, number; TN, true-negative; TP, true-positive; US, ultrasound; R1, round 1; R2, round 2.

Mammography



Mammography + Ultrasound

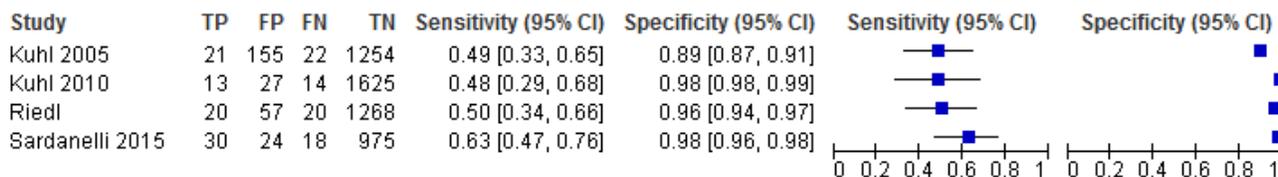


Figure A1: Forest Plots for Individual Study Sensitivity and Specificity

Abbreviations: CI, confidence interval; FN, false-negative; FP, false-positive; TN, true-negative; TP, true-positive.

Table A7: Diagnostic Outcomes of Adjunct Ultrasound Compared With Mammography Alone for High-Risk Women, Subgrouped by Personal History of Breast Cancer

Author, Year	Sub-group	Women, N (screens)	Diagnostic Yield, per 1,000 Screens			Sensitivity, %			Specificity, %			PPV, %		
			M	M+ US	Difference Between Groups (95% CI)	M	M+ US	Difference Between Groups (95% CI)	M	M+ US	Difference Between Groups (95% CI)	M	M+ US	Difference Between Groups (95% CI)
Positive Test: BI-RADS 4, 5														
Kuhl et al, 2005 ³⁹	Personal history	139 (276)	14.5	18.1	3.5 (NR)	33.3	41.7	8.4 (NR)	95.5	87.1	-8.4 (NR)	25.0	12.8	-12.2 (NR)
	No personal history	390 (1,176)	8.5	13.6	5.1 (NR)	32.3	51.6	19.3 (NR)	97.1	89.4	-7.7 (NR)	23.3	11.7	-11.6 (NR)
Positive Test: BI-RADS 3, 4, 5														
Berg et al, 2012 ³⁶	Personal history	1,426 (4,010)	8.2	12.5	4.2 (2.5, 6.2); <i>P</i> < .001	55.9	84.7	28.8 (18.6, 40.7); <i>P</i> < .001	91.4	83.1	-8.3 (-9.3, -7.4); <i>P</i> < .001	PPV1 ^a : 8.8 PPV2 ^b : 36.8	PPV1 ^a : 7.0 PPV2 ^b : 17.8	-1.9 (-3.5, -0.2); <i>P</i> = .02 -19 (-26.7, -11.7); <i>P</i> < .001
	No personal history	4,085 (3,463)	7.5	11.8	4.3 (2.3, 6.6); <i>P</i> < .001	50.0	78.8	28.8 (17.3, 40.4); <i>P</i> < .001	89.4	77.8	-11.6 (-12.8, -10.4); <i>P</i> < .001	PPV1 ^a : 6.7 PPV2 ^b : 32	PPV1 ^a : 5.1 PPV2 ^b : 11	-1.6 (-3.2, 0); <i>P</i> = .05 -21 (-29.1, -13.1); <i>P</i> < .001

Abbreviations: BI-RADS, Breast Imaging Reporting and Data System, CI, confidence interval; M, mammography; N, number; NR, not reported; NS, not significant; PPV, positive predictive value; US, ultrasound.

^a Calculated as the malignancy rate among cases that test positive (recommended for further testing, short-interval follow-up or biopsy) on screening.

^b Defined by authors as the malignancy rate among women with a positive screening test who underwent biopsy of the same lesion. These values could include biopsy resulting from a BI-RADS 3 diagnosis.

Table A8: Area Under the Receiver Operating Characteristic Curve for Mammography Plus Ultrasound Screening Compared With Mammography Alone for Women at High Risk for Breast Cancer

Author, Year	Screening Round	Area Under Receiver Operating Characteristic Curve (95% CI)	
		Mammography	Mammography + Ultrasound
Kuhl et al, 2010 ³⁸	All	0.66 (0.55, 0.77) ^a	0.77 (0.55, 0.88) ^a
Sardanelli et al, 2011 ⁴²	All	0.83 (0.76, 0.90)	0.87 (0.80, 0.93)
Berg et al, 2012 ³⁶	1	0.74 (0.63, 0.84)	0.94 (0.89, 0.99) ^b
	2	0.75 (0.65, 0.86)	0.89 (0.82, 0.97) ^b
	3	0.72 (0.64, 0.81)	0.82 (0.74, 0.89) ^b

Abbreviations: CI, confidence interval.

^a Calculated on a per-lesion basis.

^b Statistically significant difference from mammography screening alone.

REFERENCES

- (1) Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian cancer statistics 2015. [Internet]. Toronto (ON): Canadian Cancer Society; 2015 [cited 2015 Aug]. Available from: <http://www.cancer.ca/statistics>
- (2) Erbas B, Provenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in situ of the breast: a review. *Breast Cancer Res Treat.* 2006;97(2):135-44.
- (3) Narod SA, Iqbal J, Giannakeas V, Sopik V, Sun P. Breast cancer mortality after a diagnosis of ductal carcinoma in situ. *JAMA Oncol.* 2015;1(7):888-96.
- (4) Cancer Care Ontario. About the Ontario Breast Screening Program [Internet]. 2014 [cited 2015 May]. Available from: <https://www.cancercare.on.ca/pcs/screening/breastscreening/obsp/>
- (5) Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol.* 2007;25(11):1329-33.
- (6) Petrucelli N, Daly MB, Feldman GL. BRCA1 and BRCA2 hereditary breast and ovarian cancer. In: Pagon RA, Adam MP, Ardinger HH, et al., editors. *GeneReviews(R)*. [Internet]. Seattle (WA): University of Washington, Seattle; 1998 [updated 2013; cited 2015 Jul]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1247/>.
- (7) Leach MO, Boggis CR, Dixon AK, Easton DF, Eeles RA, Evans DG, et al. Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). *Lancet.* 2005;365(9473):1769-78.
- (8) Lakhani SR, Jacquemier J, Sloane JP, Gusterson BA, Anderson TJ, van de Vijver MJ, et al. Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. *J Natl Cancer Inst.* 1998;90(15):1138-45.
- (9) Lampic C, Thurfjell E, Bergh J, Sjoden PO. Short- and long-term anxiety and depression in women recalled after breast cancer screening. *Eur J Cancer.* 2001;37(4):463-9.
- (10) Steggle S, Lightfoot N, Sellick SM. Psychological distress associated with organized breast cancer screening. *Cancer Prev Control.* 1998;2(5):213-20.
- (11) Canadian Task Force on Preventive Health Care, Tonelli M, Connor Gorber S, Joffres M, Dickinson J, Singh H, et al. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. *CMAJ.* 2011;183(17):1991-2001.
- (12) Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev.* 2013;6:CD001877.
- (13) Health Quality Ontario. Screening mammography for women aged 40 to 49 years at average risk for breast cancer: an evidence-based analysis. *Ont Health Technol Assess Ser.* 2007;7(1).
- (14) Gradishar WJ, Anderson BO, Balassanian R, Blair SL, Burstein HJ, Cyr A, et al. Breast Cancer Version 2.2015: clinical practice guidelines in oncology. *J Natl Compr Canc Netw.* 2015;13(4):448-75.
- (15) Brem RF, Lenihan MJ, Lieberman J, Torrente J. Screening breast ultrasound: past, present, and future. *Am J Roentgen.* 2015;204(2):234-40.
- (16) Kerlikowske S, Zhu W, Tosteson A, Sprague B, Tice J. Identifying women with dense breasts at high risk for interval cancer: a cohort study. *Annals Int Med.* 2015;10(162):673-81.
- (17) Health Quality Ontario. Cancer screening with digital mammography for women at average risk for breast cancer, magnetic resonance imaging (MRI) for women at high risk: an evidence-based analysis. *Ont Health Technol Assess Ser.* 2010;10(3):1-55.
- (18) Warner E, Messersmith H, Causer P, Eisen A, Shumak R, Plewes D. Magnetic resonance imaging screening of women at high risk for breast cancer. Toronto (ON):

- Cancer Care Ontario; 2012 Aug 31. Program in Evidence-Based Care Evidence-Based Guideline No. 15-11 Version 2.
- (19) Marcus J, Watson P, Page D, Narod S, GM L, P T. Hereditary breast cancer: pathobiology, prognosis, and BRCA1 and BRCA2 gene linkage. *Cancer*. 1996;697:709.
 - (20) Lord SJ, Lei W, Craft P, Cawson JN, Morris I, Walleser S, et al. A systematic review of the effectiveness of magnetic resonance imaging (MRI) as an addition to mammography and ultrasound in screening young women at high risk of breast cancer. *Eur J Cancer*. 2007;43(13):1905-17.
 - (21) Warner E, Messersmith H, Causer P, Eisen A, Shumak R, Plewes D. Systematic review: using magnetic resonance imaging to screen women at high risk for breast cancer. *Ann Intern Med*. 2008;148(9):671-9.
 - (22) Cancer Quality Council of Ontario. Breast cancer screening participation [Internet]. 2013 [cited 2015 July]. Available from: <http://www.csqi.on.ca/cms/One.aspx?portalId=289784&pageId=296132>
 - (23) Ministry of Health and Long-Term Care. Schedule of benefits [Internet]. 2015 [cited 2015 May]. Available from: http://www.health.gov.on.ca/english/providers/program/ohip/sob/physserv/sob_master11_062015.pdf.
 - (24) Chiarelli AM, Prummel MV, Muradali D, Majpruz V, Horgan M, Carroll JC, et al. Effectiveness of screening with annual magnetic resonance imaging and mammography: results of the initial screen from the Ontario high risk breast screening program. *J Clin Oncol*. 2014;32(21):2224-30.
 - (25) Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane handbook for systematic reviews of diagnostic test accuracy*. Version 1.0. [Internet]. London: The Cochrane Collaborative; 2010 [cited 2015 Oct]. Available from: <http://srdta.cochrane.org>.
 - (26) Eberl MM, Fox CH, Edge SB, Carter CA, Mahoney MC. BI-RADS classification for management of abnormal mammograms. *J Am Board Fam Med*. 2006;19(2):161-4.
 - (27) Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-90.
 - (28) Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10.
 - (29) Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine*. 2011;155(8):529-36.
 - (30) Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the *Journal of Clinical Epidemiology*. *J Clin Epidemiol*. 2011;64(4):380-2.
 - (31) Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *ACP J Club*. 2008;149(6):2.
 - (32) Brozek JL, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. *Allergy*. 2009;64(8):1109-16.
 - (33) Moher D, Liberati A, Tetzlaff J, Altman DG, the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(6):e1000097.

- (34) Gartlehner G, Thaler K, Chapman A, Kaminski-Hartenthaler A, Berzaczy D, Van Noord MG, et al. Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk. The Cochrane database of systematic reviews. 2013;4:CD009632.
- (35) Gartlehner G, Thaler KJ, Chapman A, Kaminski A, Berzaczy D, Van Noord MG, et al. Adjunct ultrasonography for breast cancer screening in women at average risk: a systematic review. *Int J Evidence-Based Healthcare*. 2013;11(2):87-93.
- (36) Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA*. 2012;307(13):1394-404.
- (37) Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, Bohm-Velez M, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA*. 2008;299(18):2151-63.
- (38) Kuhl C, Weigel S, Schrading S, Arand B, Bieling H, Konig R, et al. Prospective multicenter cohort study to refine management recommendations for women at elevated familial risk of breast cancer: the EVA trial. *J Clin Oncol*. 2010;28(9):1450-7.
- (39) Kuhl CK, Schrading S, Leutner CC, Morakkabati-Spitz N, Wardelmann E, Fimmers R, et al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *J Clin Oncol*. 2005;23(33):8469-76.
- (40) National Institute for Health and Care Excellence. Classification and care of people at risk of familial breast cancer and management of breast cancer and related risks in people with a family history of breast cancer [Internet]. 2013 [cited 2015 Aug]. Available from: <http://www.nice.org.uk/guidance/cg164/resources/guidance-familial-breast-cancer-pdf>
- (41) Riedl CC, Luft N, Bernhart C, Weber M, Bernathova M, Tea MKM, et al. Triple-modality screening trial for familial breast cancer underlines the importance of magnetic resonance imaging and questions the role of mammography and ultrasound regardless of patient mutation status, age, and breast density. *J Clin Oncol*. 2015;33(10):1128-35.
- (42) Sardanelli F, Podo F, Santoro F, Manoukian S, Bergonzi S, Trecate G, et al. Multicenter surveillance of women at high genetic breast cancer risk using mammography, ultrasonography, and contrast-enhanced magnetic resonance imaging (the High Breast Cancer Risk Italian 1 Study): final results. *Invest Radiol*. 2011;46(2):94-105.
- (43) Riedl CC, Ponhold L, Flory D, Weber M, Kroiss R, Wagner T, et al. Magnetic resonance imaging of the breast improves detection of invasive cancer, preinvasive cancer, and premalignant lesions during surveillance of women at high risk for breast cancer. *Clin Cancer Res*. 2007;13(20):6144-52.
- (44) Sardanelli F, Podo F, D'Agnolo G, Verdecchia A, Santaquilani M, Musumeci R, et al. Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRIT Study): interim results. *Radiol*. 2007;242(3):698-715.
- (45) Chang JM, Koo HR, Moon WK. Radiologist-performed hand-held ultrasound screening at average risk of breast cancer: results from a single health screening center. *Acta Radiol*. 2015;56(6):652-8.
- (46) Giuliano V, Giuliano C. Improved breast cancer detection in asymptomatic women using 3D-automated breast ultrasound in mammographically dense breasts. *Clin Imaging*. 2013;37(3):480-6.
- (47) Shen S, Zhou Y, Xu Y, Zhang B, Duan X, Huang R, et al. A multi-centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women. *Br J Cancer*. 2015;112(6):998-1004.
- (48) Gartlehner G, Thaler K, Chapman A, Kaminski-Hartenthaler A, Berzaczy D, Van Noord MG, et al. Mammography in combination with breast ultrasonography versus

- mammography for breast cancer screening in women at average risk. The Cochrane database of systematic reviews. 2013;4:CD009632.
- (49) American College of Breast Radiology. ACR practice parameter for the performance of a breast ultrasound examination [Internet]. 2014 [cited 2015 May]. Available from: http://www.acr.org/~media/ACR/Documents/PGTS/guidelines/US_Breast.pdf.
- (50) Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, et al. Breast-cancer screening--viewpoint of the IARC Working Group. *N Engl J Med*. 2015;372(24):2353-8.
- (51) Sprague BL, Stout NK, Schechter C, Van Ravesteyn NT, Cevik M, Alagoz O, et al. Benefits, harms, and cost-effectiveness of supplemental ultrasonography screening for women with dense breasts. *Ann Intern Med*. 2015;162(3):157-66.
- (52) Tice J, Ollendorf D, Lee J, Pearson S. The comparative clinical effectiveness and value of supplemental screening tests following negative mammography in women with dense breast tissue: a technology assessment. [Internet]. 2013 [cited 2015 Jul] Available from: <http://ctaf.org/sites/default/files/assessments/ctaf-final-report-dense-breast-imaging-11.04.2013-b.pdf>.
- (53) Blue Cross Blue Shield A, Kaiser P. Special report: screening asymptomatic women with dense breasts and normal mammograms for breast cancer. *Technol Eval Cent Assess Program Exec Summ*. 2014;28(15):1-2.
- (54) Nothacker M, Duda V, Hahn M, Warm M, Degenhardt F, Madjar H, et al. Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. A systematic review. *BMC Cancer*. 2009;9:335.

About Health Quality Ontario

Health Quality Ontario is the provincial advisor on the quality of health care. We are motivated by a single-minded purpose: **better health for all Ontarians.**

Who We Are

We are a scientifically rigorous group with diverse areas of expertise. We strive for complete objectivity, and look at things from a vantage point that allows us to see the forest and the trees. We work in partnership with health care providers and organizations across the system, and engage with patients themselves, to help initiate substantial and sustainable change to the province's complex health system.

What We Do

We define the meaning of quality as it pertains to health care, and provide strategic advice so all the parts of the system can improve. We also analyze virtually all aspects of Ontario's health care. This includes looking at the overall health of Ontarians, how well different areas of the system are working together, and most importantly, patient experience. We then produce comprehensive, objective reports based on data, facts and the voice of patients, caregivers and those who work each day in the health system. As well, we make recommendations on how to improve care using the best evidence. Finally, we support large scale quality improvements by working with our partners to facilitate ways for health care providers to learn from each other and share innovative approaches.

Why It Matters

We recognize that, as a system, we have much to be proud of, but also that it often falls short of being the best it can be. Plus certain vulnerable segments of the population are not receiving acceptable levels of attention. Our intent at Health Quality Ontario is to continuously improve the quality of health care in this province regardless of who you are or where you live. We are driven by the desire to make the system better, and by the inarguable fact that better has no limit.

Permission Requests: All inquiries regarding permission to reproduce any content in Health Quality Ontario reports should be directed to EvidenceInfo@hqontario.ca.

About the Ontario Health Technology Advisory Committee (OHTAC)

How to Obtain OHTAC Recommendation Reports

Health Quality Ontario
130 Bloor Street West, 10th Floor
Toronto, Ontario
M5S 1N5
Tel: 416-323-6868
Toll Free: 1-866-623-6868
Fax: 416-323-9261
Email: EvidenceInfo@hqontario.ca
www.hqontario.ca
ISSN 1915-7398 (online)
ISBN 978-1-4606-8378-1 (PDF)

© Queen's Printer for Ontario, 2016