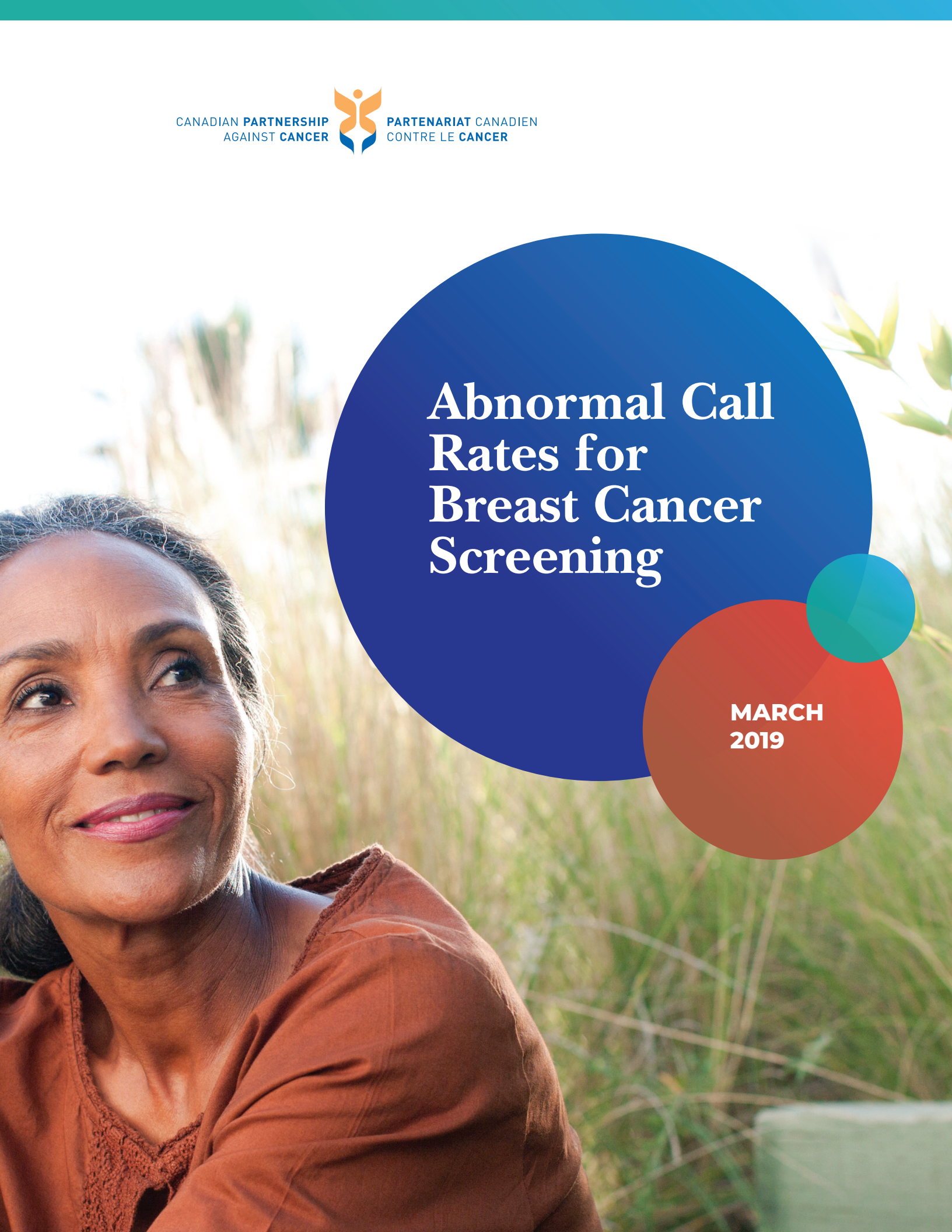
CANADIAN PARTNERSHIP AGAINST CANCER
PARTENARIAT CANADIEN CONTRE LE CANCER

Abnormal Call Rates for Breast Cancer Screening

MARCH 2019



This work was undertaken to directly inform the development of the ***Pan-Canadian Framework for Action to Address Abnormal Call Rates in Breast Cancer Screening*** led by Partnership.

Production of this report has been made possible through a financial contribution from Health Canada, through the Canadian Partnership Against Cancer.

If you wish to use the information therein please ensure appropriate attribution by referencing: *Centre for Effective Practice, Canadian Partnership Against Cancer. Abnormal Call Rates for Breast Cancer Screening. Prepared for the Canadian Partnership Against Cancer*. 2019 Mar 31.

Canadian Partnership Against Cancer
145 King Street, Suite 900, Toronto ON M5H 1J8
Phone: 416-915-9222, Fax: 416-915-9224

For more information on this publication, please email:
screening@partnershipagainstcancer.ca

About the CEP

The Centre for Effective Practice is one of the largest independent knowledge translation organizations for primary care in Canada. We bring together knowledge, evidence, expertise and resources to ensure frontline providers have the information they need to deliver high quality care and improve patients' health outcomes. In Ontario alone, our work reaches over 16,000 healthcare providers each year through established relationships with key organizations, medical schools, colleges and associations.

Our mission is to close the gap between evidence and practice in health care.

Table of Contents

1. Executive Summary

Background

This evidence review explores the upper and lower range for the abnormal call rate (ACR) target in Canada and internationally, to inform the development of an action plan that will include an identified approach to setting an appropriate national ACR target. The purpose of this report is to present the varying international targets for abnormal recall, identify different rationales for target-setting, and examine performance to identify key insights to inform appropriate target-setting in Canada.

Methodology

A comprehensive, multi-faceted search strategy was designed to identify key resources in both the indexed and grey literature. An extensive grey literature search was required for this topic because organizations that report on performance, quality improvement initiatives, and groups that develop or promote quality measures are typically not focused on publishing in indexed journals.

Key Findings

The Canadian national target for ACR in women aged 50-69 is <10% of initial screens and <5% of subsequent screens, and between 2004 and 2014 the abnormal call rates measured for initial and subsequent screens were higher than these targets.

ACR Rates vary significantly across countries, and even across jurisdictions. In countries with screening periods similar to Canada, between 2012 and 2013 the abnormal call rate ranged from 13.4% in France (PPV = 5.4%) to a low of 4.8 in Finland (PPV = 10.2%) for initial screening, and from a high of 8.4% in France (PPV = 7.9%) to a low of 2.0% in the Netherlands (PPV=29.5%) for subsequent screening.

Recent national, provincial, and international quality measures for abnormal call rates, cancer detection rates, and positive predictive values for both initial and subsequent screens are presented in Tables 1-3 for ease of reference.

The European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis (4th ed) have set an acceptable recall rate for initial screening at <7%, with a desirable level of <5%, and an acceptable rate of recall for subsequent screens at <5%, with a desirable level of <3%. This is similar to the recommendations of the NHS Breast Screening Program which states that <10% is acceptable for prevalent screening, and <7% is achievable, with <7% being acceptable for incident screens, and <5% achievable, although it should be noted that the UK invites women for screening every 3 years, and the European Guidelines recommend screening every 2 years.

A number of studies have recently been published that examine and discuss appropriate ACR ranges, and address the trade-offs between ACR and other important performance metrics associated with mammography, including PPV, CDR, sensitivity and specificity.

Overall, there is significant heterogeneity in optimal ACR targets suggested by the current literature. Qualitatively, in order to balance the trade-offs between ACR and other important performance metrics, evidence suggests a range of 5%-14%. It has been found that ACR below the 5% has the potential to compromise sensitivity and CDR, while ACR above 14% has the potential to compromise PPV and specificity and can affect patient anxiety and system costs.

It is important to note that many organizations provide disparate guidance regarding breast cancer screening, and well as distinct targets for recommended performance measures, often based on the same pool of research. Differences in recommendations between groups and across international organizations are often a result of the relative value placed on various outcomes, as well as differing priorities of the women being screened. A detailed examination of the significance various groups place on outcomes, benefits, and costs is beyond the scope of this report, but Canadian priorities must be taken into consideration when establishing ACR and related targets.

2. Introduction

The Centre for Effective Practice was engaged by the Canadian Partnership Against Cancer to develop a brief report on abnormal call rates (ACR) for breast cancer screening, with a focus on providing the most recent data available regarding:

- Performance measures in international, national, and provincial mammographic breast cancer screening programs
- International and national benchmarks and targets set for ACR and related measures such as cancer detection rate and positive predictive value of screening
- Current research regarding appropriate targets for these measures

Breast cancer screening is designed to identify potential breast cancers, increasing the likelihood of early detection. Abnormal call rate is an important quality indicator of mammogram imaging and interpretation, and is most informative when considered in the context of cancer detection rate, breast cancer incidence rate, post-screen cancer rate, and positive predictive value (PPV) (1).

Though screening has been shown to reduce morbidity, and is recommended by many international breast radiology bodies (2), there is significant international variation in performance measures, and a need to set appropriate targets for ACR, PPV and CDR. Overly high ACR could be indicative of increased incidence of false-positives and unnecessary follow-up tests. The experience of being recalled for further investigation can cause significant stress and anxiety for women – lasting beyond the screening and investigation process, and regardless of outcome (3,4). Though it has been found that most women aged 50–69 appear to identify reduction in breast cancer mortality as outweighing concerns about potential harms (5), high recall rates merit consideration.

Conversely, while false positives have been reported to cause women “significant psychological distress”(4), research shows that reducing ACRs can negatively impact detection, leading to missed cancers (false negatives); and that furthermore, little is known about which mammographic features or breast cancer appearances are affected by reduced ACR (6). International target-setting for ACR attempts to strike a balance between high detection sensitivity, low false-positive incidence, system costs, and clinical workload.

International target recall rates range widely. Although many countries use a standardized method of reporting mammograms (BI-RADS), a variety of factors impact ACR, including technology, single or double reading structures, reader experience, mammographic features, screening interval, and patient age (1).

3. Methodology

A comprehensive search was conducted to identify the most current, high-quality information using a rigorous and methodologically sound approach. Although a full systematic review was beyond the scope of this project, a multi-faceted search strategy was designed to identify key resources in both the grey and indexed literature. The search was limited to English language resources published between 2000 and 2019. Considering results across this time frame ensured that older guidelines and benchmarks that are still being used by screening programs were not unduly excluded, but due to changes in technology and mammography programs, it was important to use the most up-to-date version of a publication where possible. Additional details about the search strategy are provided in Appendix A.

Indexed Literature Search

The indexed literature was searched using Ovid MEDLINE in order to capture relevant resources published in traditional academic and peer reviewed literature. A combination of controlled vocabulary (MeSH terms), and free text words were used in combination with wildcard symbols to identify variations in spelling and variant word endings. The search was designed to be sensitive, with titles and abstracts manually reviewed for applicability and inclusion.

Grey Literature Search

Organizations that report on performance or quality improvement initiatives, and groups that develop or promote quality measures are typically not focused on publishing in indexed journals, so an extensive grey literature search was conducted.

To ensure the grey literature search was comprehensive, multiple complementary approaches were used. These include:

1. Searching key indicator repositories and the websites of developers of quality measures across broad health care topics (e.g. NICE Standards and Indicators).
2. Identifying and searching within resources and initiatives that provide quality measures related to cancer screening or to the cancer care continuum (e.g. Cancer Australia's National Cancer Control Indicators).
3. Searching the websites and resources of international, national, and provincial health agencies and quality improvement programs (e.g. Canadian Institute for Health Information).
4. Examining websites for provincial, national, and international breast cancer screening programs (e.g. CPAC's list of cancer screening programs across Canada).
5. Searching the websites and resources of national and international organizations related to breast cancer screening (e.g. Canadian Association of Radiologists).

6. Executing a targeted search for high quality clinical practice guidelines or systematic reviews related to breast cancer screening to identify any relevant evidence or quality measures (e.g. the Cancer Guidelines Database, Canadian Task Force for Preventive Health Care.)
7. Conducting an extensive supplementary grey literature search using Google to identify additional resources not captured above.

Key Definitions

Throughout this report, abnormal call rate (ACR) is used to refer to the percentage of women who were referred for diagnostic investigation because of an abnormal screening mammogram result. Cancer detection rate (CDR) is defined as the number of women with screen detected cancer per 1,000 women who had a screening mammogram, and is typically presented as overall detection rate (i.e. CIS and invasive cancer) unless otherwise stated. Positive Predictive Value (PPV) refers to the percentage of breast cancers detected among women recalled due to abnormal mammographic findings (i.e. PPV-1). PPV is particularly important when considering ACR; high PPV indicates that many women recalled for further assessment have breast cancer.

Note Regarding Variations in Terminology

It is important to emphasize the variations in terminology used internationally to describe elements related to abnormal call rate. These include terms such as call back rate, recalled for assessment, call rate, abnormal interpretation rate, proportion of persons attending further assessment after a positive screen test result, etc. Similarly, initial and subsequent screens are also referred to by a variety of terms in the international literature, such as prevalent and incident screenings in the UK.

Because of the significant variation in nomenclature and indexing in this area, a wide variety of search terms was used to identify potential resources, which were then examined individually for relevance.

Future research in this area will need to recognize and incorporate this breadth of terms, to ensure that all relevant information is captured and significant research is not overlooked.

4. Abnormal Call Rates in Canada

In Canada, population-based breast cancer screening programs exist across all provinces and territories except Nunavut (1,7). The current national target for ACR in women 50-69 is <10% of initial screens and <5% of subsequent screens, but on the whole, these targets are not being met (1). In fact, the national ACR has risen steadily over the last two decades, while the PPV has gone down. Table 1 provides the rates of ACR, CDR and PPV in Canada from 2004-2014.

The abnormal call rate target published in the third edition of the Guidelines for Monitoring Breast Cancer Screening Program Performance (8) provides references for the target and evidence supporting the measure, which consists of research and international targets published between 2002 and 2010.

Updated international targets and relevant research are provided in the following sections of this report.

Table 1: Abnormal Call Rate, Cancer Detection, and Positive Predictive Value in Canada

Indicator	Target	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
abnormal call rate (%) initial screen	<10	12.4	12.3	12.3	10.3	11.5	12.7	13.7	14.8	15.8	16.7	16.6
abnormal call rate (%) subsequent screen	<5	6.4	6.1	6	6	6.1	6.2	6.4	7.1	7.4	7.6	7.6
invasive cancer detection rate (per 1,000 screens) initial screen	>5	4.5	4.3	4.7	3.3	4.5	4.4	4.7	4.9	4.8	-	-
invasive cancer detection rate (per 1,000 screens) subsequent screen	>3	3.5	3.7	3.6	3.8	3.5	3.6	3.7	3.7	3.7	-	-
positive predictive value (%) initial screen	≥5	4.7	4.5	4.8	5.1	5	4.4	4.3	4.2	4	-	-
positive predictive value (%) subsequent screen	≥6	7	7.6	7.5	7.8	7.1	7.3	7	6.6	6.3	-	-

Adapted from Breast Cancer Screening in Canada: Monitoring and Evaluation of Quality Indicators (1)

The current target PPV for initial screening mammography in Canada is ≥5, and the target for subsequent screens is ≥6 (9). While PPV is recognized internationally as a valuable indicator for screening program performance, recent studies have acknowledged the limitations of a relying on a simple numeric value for PPV to determine program effectiveness, and must be considered in context with ACR and CDR (10).

Many factors may be contributing to the wide variation in abnormal call rates observed in programs across Canada (1). Quality indictors vary widely by age, suggesting the need for new age-specific targets (1). Choice of screening technology has been noted to impact recall rates in Quebec (9), and even in programs with standards in place for quality imaging, it can be difficult to obtain a high-quality image. A Quebec study found that 49.7% of a randomly selected sample of screens at an accredited centre did not meet the Canadian Association of Radiologists (CAR) criteria for image quality (10). Cancer detection and recall rates are also impacted by radiologist training and years of experience (3) and length of time between a patient's screening intervals (1). Table 2 provides a comparison of provincial indicators for the periods 2005-2006 and 2011-2012.

Table 2: Provincial Breast Cancer Screening Quality Indicators

Province	Abnormal Call Rate				Positive Predictive Value			
	2005-2006		2011-2012		2005-2006		2011-2012	
	Initial	Subsequent	Initial	Subsequent	Initial	Subsequent	Initial	Subsequent
BC	15.9	5.7	17.5	6.4	4.5	7.8	5.4	7.7
AB	6.9	3.1	13.9	6.1	5.2	15.5	5	6.7
SK	13.7	5.3	11.4	4	3.9	9.7	4.3	11.5
MB	9.5	4.7	9.1	4.2	6.9	10.8	6.3	12.1
ON	11.9	6.8	13.5	7.1	4.4	6.3	4.2	6
QC	15	7.3	19.3	9.2	4.3	7.3	3.3	5.9
NB	15.3	6.9	17.7	8.9	3.6	5.3	3.6	4.3
NS	8.4	4.5	14.2	5.1	7.3	10.4	7	8.8
PE	-	-	20.8	11.9	-	-	3	4.9
NL	11.1	8.3	14	6.3	4.2	5.3	6	6.5
NT	8.1	9.6	11.3	4.9	-	-	-	-

Province	Detection Rate: In situ (per 1000 screens)				Detection Rate: Invasive			
	2005-2006		2011-2012		2005-2006		2011-2012	
	Initial	Subsequent	Initial	Subsequent	Initial	Subsequent	Initial	Subsequent
BC	1.4	1.1	1.6	1	5.8	3.4	7.9	3.9
AB			1.3	0.7			5	2.9
SK	1.1	0.9	0.9	0.7	4.2	4.2	4	3.9
MB	1.1	0.9	1.8	0.9	5.5	4.2	3.9	4.2
ON	1	0.7	1.1	0.7	4.3	3.6	4.4	3.5
QC	1.5	1.1	1.2	1	4.9	4.2	5.2	4.4
NB	1	0.7	1.2	0.8	4.2	2.8	4.9	2.9
NS	1	0.8	1.1	0.9	5.2	3.9	8.6	3.5
PE	-	-	-	1.2	-	-	-	4.6
NL	0.5	0.6	1.2	0.7	4	3.5	7.1	3.4
NT	-	-	-	-	-	-	-	-

Adapted from Breast Cancer Screening in Canada: Monitoring and Evaluation of Quality Indicators and Organized Breast Cancer Screening Programs in Canada - Report on Program Performance in 2005 and 2006. (1,11)

5. International Targets and Measures

Current International Performance Measures

To support the development of Canadian benchmarks the most recently published international metrics are presented in Table 3. Changes in mammographic technology, program coverage, and specific breast cancer screening guidelines make it important to capture the most recently available data for ACR, PPV and/or CDR across countries in order to provide comparable benchmarks.

The primary age group presented in Table 3 (age 50-69/70 years) aligns with the most common international guidelines for quality assurance in breast cancer screening, and is the age group with performance measures most frequently reported by programs. Most centres use two-view digital mammography with double-reading, and follow a recommended screening period of every two years in this age group, with the exception of the UK which recommends a three year interval, and the American College of Radiology guidelines which recommend annual screening (12).

A single metric is not sufficient for evaluating performance, and ACR must be considered in the context of other relevant measures such as CDR or PPV. For example, a consistently low abnormal call rate combined with a low CDR could be driven by sub-optimal performance, or a high PPV without additional context could suggest that only high-probability lesions were investigated further, which would allow more subtle lesions to slip through screening (17). To ensure the usefulness of this data, the most recent ACR measures for 16 countries are presented in Table 3 in conjunction with published CDR and/or PPV values. Quality measures which were published only as a combined overall rate, rather than differentiating between initial and subsequent screens have not been included.

Table 3: International Measures of Abnormal Call Rate for Breast Cancer Screening

Country	Screen Interval (years)	Double Reading (Y/N)	Age Range	Initial or Subsequent Screening	Period Measured	ACR (%)	CDR / 1000 screens	PPV (%)	Year / Reference
Australia	2	Yes	50-74	Initial	2016	11.3	-	9.4	2018 (13)
Australia	2	Yes	50-74	Subsequent	2016	3.7	-	18.5	2018 (13)
Denmark	2	Yes	50-69	Initial	2013	5.0	7.01	14.9	2017 (14,15)
Denmark	2	Yes	50-69	Subsequent	2013	2.2	6.70	32.6	2017 (14)
Finland	2	Yes	50-69	Initial	2012	4.8	4.91	10.2	2017 (14)
Finland	2	Yes	50-69	Subsequent	2012	2.3	5.78	25.2	2017 (14)
France	2	Yes	50-69	Initial	2012	13.4	8.24	5.4	2017 (14)
France	2	Yes	50-69	Subsequent	2012	8.4	6.35	7.9	2017 (14)
Germany	2	Yes	50-69	Initial	2012	9.3	7.72	-	2017 (14)
Germany	2	Yes	50-69	Subsequent	2012	3.1	5.59	-	2017 (14)
Ireland	2	Yes	50-64	Initial	2013	8.0	8.82	11.2	2017 (14)
Ireland	2	Yes	50-64	Subsequent	2013	2.8	5.64	20.4	2017 (14)
Italy	2	Yes	50-69	Initial	2013	10.0	4.61	4.8	2017 (14)
Italy	2	Yes	50-69	Subsequent	2013	4.8	4.24	9	2017 (14)
Netherlands	2	Yes	50-69	Initial	2013	6.4	7.77	12.2	2017 (14,15)
Netherlands	2	Yes	50-69	Subsequent	2013	2.0	5.89	29.5	2017 (14,15)
Norway	2	Yes	50-69	Initial	2016	7.6	5.90		2017 (16)
Norway	2	Yes	50-69	Subsequent	2016	2.6	5.50		2017 (16)
Spain	2	Yes	50-69	Initial	2013	10.4	5.25	6	2017 (14,15)
Spain	2	Yes	50-69	Subsequent	2013	3.2	4.26	13.4	2017 (14,15)
Sweden [a]	2	Yes	50-69	Initial	2013	2.9	5.81	20	2017 (14,15)
Sweden [a]	2	Yes	50-69	Subsequent	2013	2.4	6.13	26.2	2017 (14,15)
UK: England	3	Yes	50-70	Initial	2016-2017	7.8	8.30		2017 (17)
UK: England	3	Yes	50-70	Subsequent	2016-2017	3.0	7.60		2017 (17)
UK: Northern Ireland	3	Yes	50-70	Initial	2013	7.0	6.16	8.8	2017 (14)
UK: Northern Ireland	3	Yes	50-70	Subsequent	2013	2.6	6.38	24.6	2017 (14)
UK: Scotland	3	Yes	50-70	Initial	2013-2014	10.1	7.17	7.2	2017 (14)
UK: Scotland	3	Yes	50-70	Subsequent	2013-2014	3.7	7.77	21.2	2017 (14)
UK: Wales	3	Yes	50-70	Initial	2013	9.2	10.38	11.5	2017 (14)
UK: Wales	3	Yes	50-70	Subsequent	2013	3.9	10.10	26.1	2017 (14)
United States	1 [b]	No [c]	50-69	Initial	1996-2009	17.6	9.70		2015 (18)
United States	1 [b]	No [c]	50-69	Subsequent	1996-2009	8.8	4.30		2015 (18)

[a] Stockholm-Gotland [b] Some facilities perform screening every 2 years [c] Some facilities use double reading

United States

The United States does not have a single national standards-setting body providing guidance for mammography screening. Groups such as the American College of Radiology (ACR), the American Cancer Society (ACS), the U.S. Preventive Services Task Force (USPSTF), and the American College of Obstetricians and Gynecologists (ACOG) all provide different recommendations for breast cancer screening. This variation in guidance results in screening in the US being performed annually or biennially, beginning at different ages, typically between the ages of 40 or 50 (19). This presents significant challenges for benchmarking and measuring quality cross-country. Differences in recommendations between these groups and across international organizations are frequently due to the relative value placed on various outcomes.

The more opportunistic screening in the US results in lower positive predictive value, and screening settings more common in the US translate into higher sensitivity and lower specificity. The more frequent screening (i.e., annual for many US cases) makes it difficult to use US data for international benchmarking, as the screening frequency is expected to lead to inherent differences in outcomes (20). The threat of malpractice litigation and a lack of a single payor also biases US performance toward accepting more false-positive results (20).

Considering US data in the international context is further challenged by differences in benchmarking methods. The fragmented US system leads to mammography facilities reporting outcomes separate from treatment facilities, so outcomes for positive screening examinations are not universally tracked.

The Breast Cancer Surveillance Consortium (BCSC) and the National Mammography Database (NMD) are two large national databases of breast cancer screening performance in the United States. BCSC and NMD both report on data at low levels (physician and facility, respectively) rather than by population (21). Neither BCSC or NMD provide current data regarding separate benchmarks for initial and subsequent screens, and instead combine the data into one measure of ACR (22). The most recent measures identified for separate initial and subsequent screening was for the period 1996-2009, and is presented in Table 3.

United Kingdom

Under the NHS Breast Screening Programme, women aged 50-70 are invited for screening every three years. Consolidated standards for the NHS Breast Screening Programme (NHSBSP) were revised in 2017, and due to the recency of the updated standards, many programs still reference the previous 2005 version in their reporting. Both the 2005 and 2017 consolidated standards recommend the performance threshold for referral to assessment rates as follows:

Acceptable (“minimum standard” in 2005): < 10% (prevalent screen), < 7% (incident screen)
Achievable (“target” in 2005): <7% (prevalent screen), <5% (incident screen) (23,24)

In 2017, cancer detection rates were withdrawn as a performance measure, and replaced with age standardised detection ratios for invasive cancers, with an acceptable target of 1.00, and an achievable target of 1.40 (23).

In England, in 2016-2017, 7.8% of women aged 50-70 attending screening for the first time (i.e. prevalent screening) were referred for assessment. Of women who had previously been screened in the past five years (i.e. incident screening), 3.0% were referred for assessment. The cancer detection rate (per 1,000 women screened) for this same time period was 8.3 for prevalent screens, and 7.6 for incident screens when the last screen was conducted within 5 years (12).

Norway

The Norwegian Breast Cancer Screening Program targets women aged 50 to 69 for mammographic screening biennially. The program started as a pilot in 1995 and became nationwide in 2005, with the transition from screen-film to digital mammography occurring between 2000 and 2011 (16).

Between 1996 and 2016, abnormal recall rates averaged 5.5% for prevalent screening examinations and 2.6% for subsequent examinations. Recall rates varied across screening areas during this time, between 3.5%-11.4% for prevalent screens, and from 1.6% to 5.3% for subsequent screens. The overall recall rate decreased from a high of 6.1% in 1996 to 3.6% in 2016 (16).

Between 1996 and 2016, 18% of women recalled due to abnormal mammographic findings were diagnosed with breast cancer. During this time period, PPV was lower for prevalent screening examinations (11%) than for subsequent screens (22%), and detection rates for prevalent screening examinations were 6.4/1000, and for subsequent screening examinations were 5.3/1000. For subsequent screening examinations, PPV increased with age from 13% for women under 55, to 29% for women older than 64 (16).

Between 1996 and 2016, 5.6 screen-detected breast cancers were detected per 1000 screening examinations, with a relatively stable rate of 5.3-5.9/1000 screening examinations since the beginning of the program. The rates were relatively stable within each age group, and as expected, increased with increasing age among subsequently screened women (16).

Netherlands

From 1989 to 1996, women in the Netherlands aged 50-69 were invited to attend biennial mammography screening. In 1997, the invitation was extended to women aged 70-75. After 2006, digital mammography replaced film based mammography (12). The National Evaluation Team for Breast Cancer Screening in the Netherlands (NETB) monitors and evaluates the effectiveness of the National Breast Cancer Screening Programme, and in 2013 reported an initial ACR of 6.4%, and a subsequent ACR of 2.0%, with PPV of 7.77% and 5.89%, respectively (14).

Australia

BreastScreen Australia previously focused on biennial screening of women aged 50–69 years, and beginning July 2013 expanded to include women aged 70–74 years. In 2016, 11.3% of women aged 50–74 years were recalled for further assessment after their first breast screen (age-standardised data), and 3.7% of women attending a subsequent screen were recalled for assessment (13). The increase in ACR for women attending screening for the first time over the past few years corresponds with an increase in the detection of invasive breast cancers. The balance of these two measures suggests that the increase in the recall to assessment rate from 11% to 12% for the first screening round may be considered acceptable (13).

Europe

The European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis (4th ed) have set an acceptable recall rate for initial screening at < 7%, and a desirable level of < 5%, with an acceptable rate of recall for subsequent screens of <5%, and a desirable level of <3% (25). As of 2016, population-based breast cancer screening programmes had been implemented or were being established in 25 of the 28 EU Member States (14).

Hofvind et al analysed performance measures in European biennial mammographic population-based screening programs for women aged 50–51 to ages 68–69, captured in the European Network for Information on Cancer project (EUNICE) (3). Data was included from 20 programs in 17 European countries, based on 7,658,586 screening examinations performed between 2005 and 2007 (26). The rate of further assessment ranged from 2.2%-15.6% for initial screening tests, and between 1.2%-10.5% for subsequent tests in women aged 50–69. The PPV also differed substantially, between 4.9%–24.2% in initial screening tests, and 6.8%–49.5% in subsequent screening tests (3).

Additional Considerations: France and Switzerland

Recent reviews of breast cancer screening programs conducted in Switzerland and France recommended against continuing the screening programs as they were being offered (27,28). In 2016, the French Minister of Health released an independent report regarding mammography screening that proposed two options: end the current national breast screening program, or end the program and replace it with a radically reformed program (27). In 2014, the Swiss Medical Board released a report that stated that new systematic mammography screening programs should not be introduced, that time limits should be placed on existing programs, and that screening quality be evaluated (28).

Examining these issues is beyond the scope of this report, but it is important to highlight the larger debate regarding the very nature and benefits of breast cancer screening when considering how international trends and targets may contribute to setting Canadian performance measures.

6. Research Examining Appropriate Targets

Breast cancer screening quality monitoring programs incorporate multiple metrics for performance not only because of these trade-offs and differing priorities, but because single metrics can be misleading. For example, a low ACR may be driven by good performance (high specificity) or poor performance (low sensitivity); high PPV may be driven by poor performance if lesions with lower probabilities of being malignant are not being acted on. Different jurisdictions have approached the components and trade-offs of screening programs in different ways, and this is reflected in their selection of indicators and benchmarks.

Despite the existence of specific Canadian and international targets for optimal ACR ranges, actual rates continue to vary (6). Lehman and colleagues reviewed the digital mammography interpretations of 359 radiologists across 95 U.S. centres between 2007 and 2013, and determined that only 59% of radiologists achieved recall rates within the acceptable range of 5%–12%, and only 63% achieved the recommended rates for specificity (29). Conversely, 92.1% of the same group of radiologists achieved recommended cancer detection rates (CDR), and 97.1% achieved sensitivity rates within the recommended range (29). These results suggest a potential disconnect between appropriate targets for ACR and their relationship with other performance metrics.

Setting target ACR thresholds is arbitrary without balancing the trade-offs between ACR and other important performance metrics associated with mammography, such as positive predictive value, sensitivity, specificity, and cancer detection rates. Optimizing the balance between these metrics results in the setting of both minimum and maximum ACR targets, thus resulting in a target ACR range that can be converted to a point estimate that represents the ideal ACR.

Rationale for Setting a Maximum ACR Threshold

As ACR increases and more patients are recalled, there is a risk of increasing the number of false positives (FP), which are also more common in younger women, and women with dense breasts (30,31). The rationale for minimizing false positives is both patient-centred and economical. A systematic review conducted by the US Preventative Services Task Force examining the harms of over-diagnosis concluded that women who receive false-positive results were more likely to report higher levels of anxiety, distress, and breast cancer specific worry (30). While eventual negative results alleviated anxiety in some women, others reported persistent anxiety despite receiving negative results from subsequent tests (30). Harm from radiation exposure also must be considered, and the same review concluded that radiation-induced cancers as a result of radiation exposure during screening leads to an estimated 2-11 deaths per 100,000 women who receive digital mammography (30). There is also a significant economic cost associated with high false-positive rates, as the average weighted cost per recall in the United States is estimated to be over $1000 (32), thus presenting a significant source of potential economic waste.

Blanks and colleagues analysed the screening program data from the English National Health Service Breast Cancer Screening Program during the years of 2009-2016, in tandem with data from a similar Dutch screening program order to model the relationship between ACR and CDR (33). A total of 11.3 million screening tests were included. Their models suggested that diminishing increases in CDR occur when ACR is set above 7%, and that increases beyond this upper ACR limit are associated with higher FP recalls without significant increases in CDR (33). The authors concluded that, while the optimal ACR may differ slightly in different populations, these results justify the setting of a maximum ACR in cancer screening programs in order to minimize FP rates (33).

Another frequent way to measure screening performance is by the number of additional work-ups per additional cancer detected (AW/ACD) (34). Schell and colleagues analyzed first (n=171,104) and subsequent (n=1,872,687) mammogram results from Breast Cancer Surveillance Consortium in the United States from 1996-2001 in order to model sensitivity as a function of ACR. When ACR was above 10% for first mammograms and 6.7% for subsequent mammograms, AW/ACD increased drastically, which suggests that these represent appropriate maximum ACR thresholds (34).

Positive predictive value is also important to consider when evaluating the success of a screening program. An earlier study, in which mammography data (n= 215,665) was obtained from the Carolina Mammography Registry between 1994-1998, demonstrated that as ACR increased beyond 4.8%, sensitivity increased very little, while PPV began decreasing significantly when ACR was 5.9% or higher (35). These results suggest that ACR between 4.9%-5.5% achieve the best trade-off of sensitivity and PPV (35). It is important to note that this study employed the use of film-screening mammography, therefore the suggested range for optimal ACR may differ from that of programs utilizing digital mammography (DM). However, more recent evidence suggests that the general negative correlation between ACR and PPV is maintained when programs transition to using DM (36).

Rationale for Setting a Minimum ACR Threshold

It has been argued that, while metrics such as PPV and specificity remain important in the context of mammography, the primary goal of most screening programs is to detect as many breast cancers as possible, as early as possible, in order to maximize the opportunity for successful treatment (37). Setting ACR to optimize PPV may lead to cancers being missed, therefore, the meaningful goal of setting a lower threshold for ACR is to optimize the trade-offs between ACR and sensitivity (37). Multiple studies have examined the association between ACR and sensitivity, with some early results suggesting that ACR above 4.9% optimizes sensitivity (35).

Further support for setting a minimum threshold for ACR comes from a study by Burnside and colleagues, in which data collected from the UK National Health Service Breast Screening Programme from 2005-2008 (n=5,126,689) was retrospectively analyzed to determine whether or not lower ACR lead to an increase in interval cancers (38). A significant negative correlation between ACR and interval cancer rate was observed, which demonstrates the need to set a minimum threshold for ACR (38).

Cancer detection rate is also an important performance metric linked to ACR. In a recent study, the results of mammograms (n=188,959) performed between 2007-2012 at an American center were analyzed in order to identify the optimal ACR based on CDR (39). Physicians' recall rates were stratified into categories (less than 10%, 10% to less than 12%, 12% to less than 14%, and 14% or higher). A significantly higher CDR was seen for ACR that were 12% or higher, and it was determined that target ACR thresholds of 10% or lower may be too low due their reduced CDR (39). Moreover, the results from this study also suggest a maximum ACR threshold due to minimal increases in CDR when ACR is increased above 14%, and therefore concludes that the optimal ACR range is 12%-14% (39).

Qualitative Considerations

In addition to quantitative metrics such as CDR, PPV, sensitivity, and specificity, the qualitative performance of mammography should also be considered when determining appropriate ACR targets. Specifically, an understanding of the types of cancer that are missed by lowering ACR may be important. A recent study sought to test whether or not certain mammographic appearances of breast cancer are missed when radiologists read at lower target ACR (6). In an experimental (as opposed to clinical) setting, an enriched set of mammogram results (n=200, comprising 180 normal cases and 20 abnormal cases) from the BreastScreen New South Wales digital imaging library was created. Radiologists reviewed the cases in three separate reading sessions using different target ACR: "free recall" (i.e. radiologists could recall as many cases as they thought necessary), 15%, and 10%. Regardless of target ACR, cancers with stellate mass features were most likely to be recalled by all participants. In the 15% recall condition, cancers with mixed features of calcification plus architectural distortion were most likely to be missed, and when ACR was further reduced to 10%, non-specific density was the most common feature that was missed (6). These results suggest that there can be specific mammographic features of breast cancer that are missed when ACR is lowered.

Overall, there is significant heterogeneity in optimal ACR targets suggested by the current literature. Qualitatively, in order to balance the trade-offs between ACR and other important performance metrics, evidence suggests a range of 5%-14%. ACR below the 5% has the potential to compromise sensitivity and CDR, while ACR above 14% has the potential to compromise PPV and specificity.

Factors Affecting ACR

There are a multitude of factors that affect ACR and other mammography screening performance metrics. An in-depth analysis of these factors (e.g. different types of screening, screening intervals, radiologist training etc.) was conducted by Risk Science International (RSI) in a report prepared for the Partnership in 2018 *titled Evidence Synthesis on Factors Associated with Abnormal Call Rate in Breast Cancer Screening*. A brief summary of the factors influencing ACR is provided here, with a focus on research published since the date of the search conducted for the RSI report.

One factor that has had the potential to affect ACR on an international scale is the shift from film-screen mammography to digital mammography (DM) techniques, which can facilitate easier acquisition,

storage, and transfer of images (39,40). Blanks and colleagues retrospectively analyzed annual screening data from the English National Health Service Breast Cancer Screening Program from 2009-2016 (a total of 11.3 million screening episodes) in order to determine CDR and ACR. This data, in combination with estimates of DM usage derived from official records that were kept as individual screening units within the program converted from using film-screen mammography to DM suggested that the overall CDR was 14% greater with the introduction of DM, while ACR remained relatively unchanged (40). In a similar study, data from the Reggio Emilia Breast Cancer Screening Program were analyzed from 2011 (n=42,240), when all mammography in the program was done by film-screen, and 2012 (n=45,196), the year that the program converted completely to DM (36). While no significant effects on CDR were observed as a result of the switch to DM, there was an increase in ACR from 3.3%-4.4% (36). Interestingly, although the average ACR increased year to year, by December 2012 (i.e. 12 months after the switch to DM) ACR had decreased back to a rate similar to that seen in 2011 when film-screening mammography was used. This pattern suggests a potential learning curve and may indicate that the introduction of DM has no significant impact on ACR once the initial learning curve is complete (36).

Different digital techniques may also have different effects on ACR. A recent systematic review and random effects meta-analysis of 17 studies that compared 2-dimensional DM to digital breast tomosynthesis (DBT) concluded that the use of DBT reduced ACR while improving CDR (41). One study included in the review used mammogram results (n=12,781) from an American medical centre, and retrospectively compared full-field digital mammography (FFDM) with DBT (42). The use of DBT significantly reduced ACR from 16%–14% while increasing CDR by 33% (which did not reach statistical significance) (42). In 2019, another comparison of FFDM and DBT in which 22,055 mammogram results from three American outpatient sites were retrospectively analysed showed similar results (43). Both false positives (FP) and ACR were significantly less likely with DBT (FP=91%, ACR=8%) as compared to FFDM (FP=93%, ACR=10.6%), thus providing further, recent evidence for DBT's lowering effect on ACR, which supports its continued use in screening programs (43).

Procedural changes have been shown to impact ACR. Mullen and colleagues explored two methods for achieving lower ACR. The first intervention involved raising awareness, by having radiologists personally review their recalls as well as increasing group awareness of team ACR. The second intervention used double readings of potential recalls. Overall, both methods led to decreases in ACR, with the first intervention (personal review of recalls and group awareness of team ACR) showing a larger effect (4).

7. References

1. Canadian Partnership Against Cancer. Breast Cancer Screening in Canada: Monitoring and Evaluation of Quality Indicators - Results Report January 2011 to December 2012 [Internet]. 2017. Available from: https://content.cancerview.ca/download/cv/prevention_and_screening/screening_and_early_diagnosis/documents/breast_cancer_screening_canada_monitoring_evaluating_report_2011_12p?attachment=0

2. Sardanelli F, Aase HS, Álvarez M, Azavedo E, Baarslag HJ, Balleyguier C, et al. Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Israel, Lithuania, Moldova, The Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and Turkey. Eur Radiol [Internet]. 2017;27(7):2737–43. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5486792/

3. Hofvind S, Ponti A, Patnick J, Ascunce N, Njor S, Broeders M, et al. False-Positive Results in Mammographic Screening for Breast Cancer in Europe: A Literature Review and Survey of Service Screening Programmes. J Med Screen [Internet]. 2012 Sep 1;19(1_suppl):57–66. Available from: https://doi.org/10.1258/jms.2012.012083

4. Mullen LA, Panigrahi B, Hollada J, Panigrahi B, Falomo ET, Harvey SC. Strategies for Decreasing Screening Mammography Recall Rates While Maintaining Performance Metrics. Acad Radiol. 2017 Dec;24(12):1556–60.

5. Klarenbach S, Sims-Jones N, Lewin G, Singh H, Thériault G, Tonelli M, et al. Recommendations on screening for breast cancer in women aged 40–74 years who are not at increased risk for breast cancer. CMAJ [Internet]. 2018 Dec 10;190(49):E1441–51. Available from: http://www.cmaj.ca/content/190/49/E1441

6. Mohd Norsuddin N, Mello-Thoms C, Reed W, Rickard M, Lewis S. An investigation into the mammographic appearances of missed breast cancers when recall rates are reduced. Br J Radiol [Internet]. 2017 Jun 16;90(1076):20170048. Available from: https://www.birpublications.org/doi/10.1259/bjr.20170048

7. Canadian Partnership Against Cancer. 2018 Cancer System Performance Report [Internet]. Canadian Partnership Against Cancer; 2018. Available from: https://www.systemperformance.ca/report/2018-cancer-system-performance-report/

8. Canadian Partnership Against Cancer. Report from the Evaluation Indicators Working Group. Guidelines for Monitoring Breast Cancer Screening Program Performance (3rd Edition). [Internet]. 2013. Available from: http://www.cancerview.ca/idc/groups/public/documents/webcontent/guideline_monitoring_breast.pdf

9. Théberge I, Vandal N, Langlois A, Pelletier É, Brisson J. Detection Rate, Recall Rate, and Positive Predictive Value of Digital Compared to Screen-Film Mammography in the Quebec Population-Based Breast Cancer Screening Program. Can Assoc Radiol J. 2016 Nov;67(4):330–8.

10. Guertin M-H, Théberge I, Dufresne M-P, Zomahoun HTV, Major D, Tremblay R, et al. Clinical Image Quality in Daily Practice of Breast Cancer Mammography Screening. Can Assoc Radiol J. 2014 Aug;65(3):199–206.

11. Public Health Agency of Canada. Organized Breast Cancer Screening Programs in Canada - Report on Program Performance in 2005 and 2006 [Internet]. 2011. Available from: http://www.phac-aspc.gc.ca/cd-mc/publications/cancer/obcsp-podcs05/pdf/breast-cancer-report-eng.pdf

12. Ebell MH, Thai TN, Royalty KJ. Cancer screening recommendations: an international comparison of high income countries. Public Health Rev [Internet]. 2018 Mar 2;39(1):7. Available from: https://doi.org/10.1186/s40985-018-0080-0

13. Australian Institute of Health and Welfare. BreastScreen Australia Monitoring Report 2018. Cancer series no. 112. 2018. Cat. no. CAN 116. [Internet]. 2018 Oct. Available from: https://www.aihw.gov.au/reports/cancer/breastscreen-australia-monitoring-report-2018/contents/table-of-contents

14. Ponti, A., Anttila, A., Ronco, G., Senore, C., Basu, P., Segnan, N. Cancer screening in the European Union: Report on the implementation of the Council recommendation on cancer screening [Internet]. Brussels: European Commission. Prepared by the International Agency for Research on Cancer in collaboration with CPO Piemonte and University Hospital "Città della Salute e della Scienza" (Turin, Italy) and the Mass Screening Registry/Finnish Cancer Registry (Helsinki, Finland).; 2017. Available from: https://ec.europa.eu/health/sites/health/files/major_chronic_diseases/docs/2017_cancerscreening_2ndreportimplementation_en.pdf

15. Altobelli E, Rapacchietta L, Angeletti PM, Barbante L, Profeta FV, Fagnano R. Breast Cancer Screening Programmes across the WHO European Region: Differences among Countries Based on National Income Level. Int J Environ Res Public Health [Internet]. 2017 Apr;14(4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5409652/

16. Hofvind S, Tsuruda K, Mangerud G, Ertzaas A. The Norwegian Breast Cancer Screening Program 1996-2016 Celebrating 20 years of organised mammographic screening. In: Cancer Registry of Norway. Cancer in Norway 2016 - Cancer incidence, mortality, survival and prevalence in Norway. [Internet]. Oslo: Cancer Registry of Norway; 2017. Available from: https://www.kreftregisteret.no/en/General/Publications/Cancer-in-Norway/cancer-in-norway-2016/

17. NHS Digital. Breast Screening Programme, England - 2016-17 [PAS]: National Statistics [Internet]. NHS Digital. 2018. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/breast-screening-programme/breast-screening-programme-england---2016-17

18. Jacobsen KK, O'Meara ES, Key D, Buist DS, Kerlikowske K, Vejborg I, et al. Comparing sensitivity and specificity of screening mammography in the United States and Denmark. Int J Cancer J Int Cancer [Internet]. 2015 Nov 1;137(9):2198–207. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4537675/

19. Domingo L, Hofvind S, Hubbard RA, Román M, Benkeser D, Sala M, et al. Cross-national comparison of screening mammography accuracy measures in U.S., Norway, and Spain. Eur Radiol [Internet]. 2016 Aug 1;26(8):2520–8. Available from: https://doi.org/10.1007/s00330-015-4074-8

20. D'Orsi CJ, Sickles EA. 2017 Breast Cancer Surveillance Consortium Reports on Interpretive Performance at Screening and Diagnostic Mammography: Welcome New Data, But Not as Benchmarks for Practice. Radiology [Internet]. 2017 Feb 28;283(1):7–9. Available from: https://pubs.rsna.org/doi/10.1148/radiol.2017170181

21. Lee CS, Moy L, Friedewald SM, Sickles EA, Monticciolo DL. Harmonizing Breast Cancer Screening Recommendations: Metrics and Accountability. Am J Roentgenol [Internet]. 2018 Feb;210(2):241–5. Available from: https://www.ajronline.org/doi/full/10.2214/AJR.17.18704

22. Lee CS, Bhargavan-Chatfield M, Burnside ES, Nagy P, Sickles EA. The National Mammography Database: Preliminary Data. Am J Roentgenol [Internet]. 2016 Feb 11;206(4):883–90. Available from: https://www.ajronline.org/doi/10.2214/AJR.15.14312

23. Public Health England. NHS Breast Screening Programme: consolidated standards [Internet]. 2017 p. 28. Available from: https://www.gov.uk/government/publications/breast-screening-consolidated-programme-standards

24. NHS Digital. Consolidated guidance on standards for the NHS Breast Screening Programme. NHSBSP 60 [Archived Content] [Internet]. 2005. Available from: https://webarchive.nationalarchives.gov.uk/20150506220951/http://www.cancerscreening.nhs.uk//breastscreen/publications/nhsbsp60.html

25. Perry N, Perry-Broeders-de Wolf-Törnberg-Holland-von Karsa, Europäische Kommission, editors. European guidelines for quality assurance in breast cancer screening and diagnosis [Internet]. 4th ed. Luxembourg: Off. for Official Publ. of the Europ. Communities; 2006. 416 p. Available from: http://www.euref.org/european-Guidelines

26. Giordano L, Von Karsa L, Tomatis M, Majek O, De Wolf C, Lancucki L, et al. Mammographic Screening Programmes in Europe: Organization, Coverage and Participation. J Med Screen [Internet]. 2012 Sep 1;19(1_suppl):72–82. Available from: https://doi.org/10.1258/jms.2012.012085

27. Barratt A, Jørgensen KJ, Autier P. Reform of the National Screening Mammography Program in France. JAMA Intern Med. 2018 Feb 1;178(2):177.

28. Chiolero A, Rodondi N. Lessons From the Swiss Medical Board Recommendation Against Mammography Screening Programs. JAMA Intern Med. 2014 Oct 1;174(10):1541.

29. Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DSM, Kerlikowske K, et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. Radiology [Internet]. 2017 Apr;283(1):49–58. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5375631/

30. Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L. Harms of Breast Cancer Screening: Systematic Review to Update the 2009 US Preventive Services Task Force Recommendation [Internet]. 2016. Available from: https://www.uspreventiveservicestaskforce.org/Page/Document/evidence-summary-harms-of-screening-for-breast-cancer/breast-cancer-screening1

31. Posso M, Louro J, Sánchez M, Román M, Vidal C, Sala M, et al. Mammographic breast density: How it affects performance indicators in screening programmes? Eur J Radiol [Internet]. 2019 Jan 1;110:81–7. Available from: https://www.ejradiology.com/article/S0720-048X(18)30407-8/abstract

32. Bonafede MM, Miller JD, Lenhart GM, Nelson J, Fajardo LL. Health Insurer Burden of Patient Recall Following Breast Cancer Screening Mammography. Value Health [Internet]. 2014 May 1;17(3):A82. Available from: https://www.valueinhealthjournal.com/article/S1098-3015(14)00530-0/abstract

33. Blanks RG, Given-Wilson RM, Cohen SL, Patnick J, Alison RJ, Wallis MG. An analysis of 11.3 million screening tests examining the association between recall and cancer detection rates in the English NHS breast cancer screening programme. Eur Radiol. 2019 Feb 4;

34. Schell MJ, Yankaskas BC, Ballard-Barbash R, Qaqish BF, Barlow WE, Rosenberg RD, et al. Evidence-based Target Recall Rates for Screening Mammography. Radiology. 2007 Jun;243(3):681–9.

35. Yankaskas BC, Cleveland RJ, Schell MJ, Kozar R. Association of Recall Rates with Sensitivity and Positive Predictive Values of Screening Mammography. Am J Roentgenol [Internet]. 2001 Sep 1 [cited 2019 Mar 29];177(3):543–9. Available from: https://www.ajronline.org/doi/full/10.2214/ajr.177.3.1770543

36. Campari C, Rossi PG, Mori CA, Ravaioli S, Nitrosi A, Vacondio R, et al. Impact of the Introduction of Digital Mammography in an Organized Screening Program on the Recall and Detection Rate. J Digit Imaging. 2016 Apr;29(2):235–42.

37. Hardesty LA, Klym AH, Shindel BE, Chough DM, Sumkin JH, Gur D. Is Maximum Positive Predictive Value a Good Indicator of an Optimal Screening Mammography Practice? Am J Roentgenol [Internet]. 2005 May 1;184(5):1505–7. Available from: https://www.ajronline.org/doi/10.2214/ajr.184.5.01841505

38. Burnside ES, Vulkan D, Blanks RG, Duffy SW. Association between Screening Mammography Recall Rate and Interval Cancers in the UK Breast Cancer Service Screening Program: A Cohort Study. Radiology. 2018 Jul;288(1):47–54.

39. Grabler P, Sighoko D, Wang L, Allgood K, Ansell D. Recall and Cancer Detection Rates for Screening Mammography: Finding the Sweet Spot. Am J Roentgenol [Internet]. 2017 Jan;208(1):208–13. Available from: https://www.ajronline.org/doi/full/10.2214/AJR.15.15987

40. Blanks RG, Wallis MG, Alison R, Kearins O, Jenkins J, Patnick J, et al. Impact of Digital Mammography on Cancer Detection and Recall Rates: 11.3 Million Screening Episodes in the English National Health Service Breast Cancer Screening Program. Radiology. 2019 Mar;290(3):629–37.

41. Marinovich ML, Hunter KE, Macaskill P, Houssami N. Breast Cancer Screening Using Tomosynthesis or Mammography: A Meta-analysis of Cancer Detection and Recall. JNCI J Natl Cancer Inst. 2018 Sep 1;110(9):942–9.

42. Powell JL, Hawley JR, Lipari AM, Yildiz VO, Erdal BS, Carkaci S. Impact of the Addition of Digital Breast Tomosynthesis (DBT) to Standard 2D Digital Screening Mammography on the Rates of Patient Recall, Cancer Detection, and Recommendations for Short-term Follow-up. Acad Radiol. 2017 Mar;24(3):302–7.

43. Honig EL, Mullen LA, Amir T, Alvin MD, Jones MK, Ambinder EB, et al. Factors Impacting False Positive Recall in Screening Mammography. Acad Radiol. 2019 Feb;[Epub ahead of print].

8. Appendix A: Literature Search Strategy

Although a full systematic review was beyond the scope of this project, a comprehensive, multi-faceted search strategy was designed to identify key resources in both the grey and indexed literature. The search was conducted between February and March 2019, and was limited to English language resources published between 2000-2019.

There is significant variation in nomenclature and indexing in this area, such as:

- Abnormal call rate
- Abnormal recall rate
- Recall rate
- Call back rate
- Recall to assessment rate
- Abnormal interpretation rate
- Proportion of women recalled for assessment

To capture each of the above concepts, the general terms "call rate" "recall" "call back" and "abnormal" were used to identify results, and then the full text of these resources were individually reviewed to determine relevancy.

Search Strategy for Indexed Literature

The indexed literature was searched using Ovid MEDLINE in order to capture relevant resources published in traditional academic literature.

A combination of controlled vocabulary (MeSH terms), and free text words were used in combination with wildcard symbols to identify variations in spelling and variant word endings. The search was designed to be sensitive, with titles and abstracts manually reviewed for applicability and inclusion.

Database(s): Ovid MEDLINE: Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE® 1946-Present

#	Searches	Results
1	exp *Mammography/	16905
2	exp Breast Neoplasms/	274099
3	exp "Early Detection of Cancer"/	20669
4	exp Mass Screening/	120056
5	screen$.mp.	746613
6	3 or 4 or 5	761327
7	1 and 2 and 6	7532
8	*Quality Assurance, Health Care/	31481
9	*Total Quality Management/	7306
10	*Health Status Indicators/	10654
11	*Quality Indicators, Health Care/	8116
12	*"Outcome and Process Assessment (Health Care)"/	9430

13	8 or 9 or 10 or 11 or 12	64310
14	(indicator$ or (quality adj assess$) or (quality adj care) or (quality adj metric$) or (quality adj measur$) or (quality adj improvem$) or (performance adj measure$) or (system$ adj performance) or program performance or benchmark$ or program results).m_titl.	61967
15	7 and (13 or 14)	167
16	(abnormal call rate$ or abnormal recall rate$ or recall rate$ or call back rate$ or abnormal call or recall to assessment or recalled for assessment or abnormal interpretation or abnormal result or recalled for additional or recall to assessment or further test$).mp	15201
17	7 and 16	368
18	15 or 17	505
19	limit 18 to (english language and yr="2000 - 2019")	384
Terms below were suggested by the ACR Advisory Team		
20	"1-specificity".mp.	447
21	7 and 20	2
22	exp *False Positive Reactions/	636
23	exp *"Predictive Value of Tests"/	1618
24	7 and (22 or 23)	30
25	24 not 18	27
26	21 not 18	1
27	25 or 26	28
28	limit 27 to (english language and humans and yr="2000 - 2019")	25
29	false positive*.mp.	69999
30	exp "Predictive Value of Tests"/	189302
31	exp False Positive Reactions/	26945
32	(7 and (29 or 30 or 31)) not 18	886
33	limit 32 to (english language and humans and yr="2000 - 2019")	668
34	33 not 28	644

Notes regarding search terminology:

- Capitalized terms followed by a / indicate controlled vocabulary MeSH terms
- .m_titl indicates that the terms are being searched in the title field
- .mp indicates that the terms are being searched in multiple fields, including the title, abstract or keyword fields
- $ is a wildcard that allows for truncation (e.g. assess$ will return results for assessment, assessing, etc.)
- adj indicates that two terms need to be adjacent to each another, in either direction. (e.g. quality adj assess$ will return results for ‘quality assessment’ and ‘assessing quality’, etc.)

Search Strategy for Grey Literature

Organizations that report on performance, quality improvement initiatives, and groups that develop or promote quality measures are typically not focused on publishing in indexed journals, so an extensive grey literature search was conducted for this topic.

The CEP has an established resource list of almost one hundred indicator repositories, indicator developers, and Canadian and international health agencies and quality improvement programs which were reviewed as part of the search strategy. This list was complemented by a list of topic-specific organizations suggested by the ACR Advisory Team, including The American College of Radiology, NHS Breast Screening Programme (NHSBSP), the European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis, and Cancer Australia's National Cancer Control Indicators.

To ensure a comprehensive the grey literature search, multiple complementary approaches were used to identify relevant material. These include:

1. Searching key indicator repositories and the websites of developers of quality measures across broad health care topics (e.g. NICE Standards and Indicators).
2. Identifying and searching within resources and initiatives that provide quality measures related to cancer screening or to the cancer care continuum (e.g. Cancer Australia's National Cancer Control Indicators).
3. Searching the websites and resources of international, national, and provincial health agencies and quality improvement programs. (e.g. Canadian Institute for Health Information).
4. Examining websites for provincial, national, and international breast cancer screening programs. (e.g. CPAC's list of cancer screening programs across Canada).
5. Searching the websites and resources of national and international organizations related to breast cancer screening (e.g. Canadian Association of Radiologists).
6. Executing a targeted search for high quality clinical practice guidelines or systematic reviews related to breast cancer screening to identify any relevant evidence or quality measures (e.g. the Cancer Guidelines Database, Canadian Task Force for Preventive Health Care.)
7. Conducting an extensive supplementary grey literature search using Google to identify additional resources not captured above. Multiple terms were used to capture literature related to each concept.