

# **Report on Breast Density**

**Andy Coldman**

**12<sup>th</sup> July 2018**

## Sections

- p. 4 1.0 Executive Summary**
- p. 8 2.0 Introduction**
- p. 10 3.0 Measurement of Breast Density**
- p. 14 3.1 Distribution of Breast Density**
- p. 15 3.1.1 Number of Women Affected in British Columbia**
- p. 17 3.2 Breast Density and Risk of Breast Cancer**
- p. 20 3.3 Breast Density and Risk of Interval Breast Cancer**
- p. 22 3.4 Performance Targets for Interval Cancer Rates**
- p. 24 3.5 Communication of Breast Density**
- p. 26 3.6 The Effect of Density on Breast Cancer Prognosis**
- p. 28 3.7 The Effect of Density on False Positive Mammography Examinations**
- p. 29 4.0 Framework for Evaluating Studies of Supplemental Screening in Women not at High Risk of Breast Cancer Receiving Primary Screening by Mammography**
- p. 36 5.0 Intervention Following a Dense Breast Classification**
- p. 39 5.1 Breast Ultrasound (BU)**
- p. 44 5.1.1 The J-START Trial**
- p. 50 5.2 Breast Tomography (BT)**
- p. 56 5.2.1 Analysis of Data from the Oslo Tomosynthesis Screening Trial**
- p. 58 5.3 Breast Magnetic Resonance Imaging (MRI)**
- p. 62 6.0 Recommendations**
- p. 64 7.0 References**

## Figures:

p. 30 - **Figure 1:** Schematic of Usual Model for Cancer Screening

p. 31 - **Figure 2:** Schematic for Overdiagnosis Model of Cancer Screening

## Tables:

p. 12 - **Table 3.1:** Comparison of Summary of Results on Stability of BIRADS Density from US Task Force and from BC Analysis for Consecutive Screens

p. 14 - **Table 3.2:** Distribution of BIRADS density (%) by Age for 1,178,262 Women Reported by the Breast Cancer Surveillance Consortium

p. 15 - **Table 3.3:** Distribution of BIRADS density (%) by Age for 484,375 Screening Rounds in the BC Cancer Breast Screening Program

p. 16 - **Table 3.4:** Number of Women Classified as Dense in One-Year of Screening Using the Reported Program Density Distribution and The Number of Women Screened in 2015

p. 17 - **Table 3.5:** Summary Relative Risks and Confidence Intervals for General Population Incidence Studies as Summarized by McCormack

p. 18 - **Table 3.6:** Rates of Invasive Breast Cancer by Screening Cycle Length, Annual or Biennial, and by Age for Participants in the BCCBSP Program

p. 19 - **Table 3.7:** Summary Relative Risks and Confidence Intervals by Percent Density and Mode of Detection for Women Participating in Screening in Canada

p. 20 - **Table 3.8:** 12-month Interval Cancer Rates by Age and Density in a Cohort of Women from The Breast Cancer Surveillance Consortium

p. 26 - **Table 3.9:** Proportion with Tumour Size  $\geq 15$ mm and Positive Lymph Nodes by BIRADS Breast Density Screened Biennially

p. 26 - **Table 3.10:** Proportion with Tumour Size  $\geq 15$ mm and Positive Lymph Nodes for Screen Detected and 24-Month Interval Cancers by BIRADS Breast Density Screened Biennially in the Analysis of BCCBSP Data

p. 41-43 - **Table 5.1:** Summary Outcomes for Studies Utilising Supplementary Screening with Ultrasound

p. 45 - **Table 5.2:** Screen Detected and Interval Cancers for the two arms of the J-START Trial

p. 46 - **Table 5.3:** Stage Distribution of Screen Detected and Interval Cancers for the two arms of the J-START Trial by Mode of Detection and Test

p. 47 - **Table 5.4:** Adjusted Counts by Stage of Screen Detected by Ultrasound in the Intervention Arm and Estimated Interval Cancers for the J-START Trial

p. 49 - **Table 5.5:** Assumed Sensitivity for Mammography and Ultrasound in Simulation Study and Period Sensitivity from BCCBSP Study

p. 52-55 - **Table 5.6:** Summary Outcomes by Study Utilising Supplementary Screening with Breast Tomography

p. 56 - **Table 5.7:** Projected Results Over a Two-Year Period of Adding Breast Tomography to Routine Biennial Mammography Screening in a Population of 100,000 Women Using Results from the Oslo Tomosynthesis Screening Trial and Norwegian Breast Screening Program

p. 60-61 - **Table 5.8:** Summary Outcomes by Study Utilising Supplementary Screening with Breast Magnetic Resonance Imaging

## 1.0 Executive Summary

The purpose of this review is to examine evidence, expert opinion and BC Cancer Breast Screening Program data regarding breast density, breast cancer risk and breast screening. Breast density is an area of interest for breast cancer screening programs since it influences their performance. The objective of breast screening is to reduce the risk of breast cancer death among participants. Screening reduces the risk of breast cancer death when it causes cancers in participants to be diagnosed at a sufficiently less advanced stage of disease than they otherwise would be if they were not screened. Screening can also have unintended harms and its benefits must clearly outweigh the harms.

This review was conducted using three primary sources of information, 1) Published studies in the scientific literature, 2) Interviews with key informants and 3) Analysis of data from the BC Cancer Breast Screening Program (BCCBSP).

### *Measurement and Distribution of Breast Density*

Breast density is characterised by areas of whiteness on mammograms due to higher absorption of x-rays by fibrous and glandular tissue and can obscure identification of breast cancer at screening; this is termed masking. Breast density is measured in British Columbia using the Breast Imaging-Reporting and Data System (BIRADS) which includes a 4-point density scale (A, B, C and D) completed by the radiologist. Practice in the United States designates BIRADS C or D as being “dense”. Commonly in Canada a category similar to BIRADS D is used to define dense. In an average year the BCCBSP would classify approximately 86,000 women as BIRADS C and 20,000 women as BIRADS D. In British Columbia, breast density declines by age and some ethnic groups, particularly East- and South East-Asians, have higher average density than others. In BCCBSP density measurement on consecutive mammograms is variable with 36% of those classified as BIRADS D not being classified as D on the next mammogram. During this period the program did not record which BIRADS version was used (4<sup>th</sup> or 5<sup>th</sup> edition) which may have contributed to some of the classification differences. The increasing availability of computer algorithms, which utilize digital mammography images, provide automated continuous measured densities and offer some opportunity for improved and more flexible density assignment. Computer algorithms continue to evolve and BIRADS remains the most commonly used system in clinical use. Monitoring of BIRADS performance and future opportunities offered by automated assessment is warranted.

### *Breast Density and Breast Cancer Risk*

Breast cancer risk has been shown to vary with breast density with greater densities being at higher risk. In an individual woman, the multiplicity of known risk factors implies that estimating risk based upon a single factor, such as breast density, will have limited predictive power. Interval breast cancers, those not detected by screening but diagnosed between screening visits, are more common among women with dense breasts as the joint effect of increased risk and masking interact synergistically. Estimates from the BCCBSP indicate that women aged 40-49 who are BIRADS D have approximately 1.8 times the age-specific average risk of interval cancer and 2.2 times the average risk for women 50-74.

## *Communication of Breast Density*

The increased rates of interval breast cancers among women with denser breasts has resulted in concern, with the suggestion that screening programs should inform women of their density status and initiate altered approaches to screening for women with higher density. Many US States require, by law, that women be informed of a finding of dense breasts after attending screening. In Canada one program informs women of a dense finding and most record it as part of a screening visit, with five Provinces recommending annual mammography screening for those they designate as dense. The BCCBSP currently provides density assessments to women upon request and makes no special recommendations for future screening. Organized screening programs in other countries typically do not inform women of their density or take special measures. Key informant interviews revealed a wide range of opinion on the advisability and approach to dissemination of density to women and their physicians with consensus only that if it is to be done it will require comprehensive education for women and their physicians. A randomized trial (RCT), conducted in British Columbia, found that it is possible to inform women of their breast density, in the absence of altered screening, without lasting negative psychological effects.

## *Supplemental Screening of Women with Dense Breasts*

The elevated rates of interval breast cancer are a concern in women with dense breasts. Reducing the rate of interval cancer, especially advanced cancer, is an accepted metric of improved screening performance. While no organisation issuing guidelines recommends supplemental screening of women on the basis of breast density alone, it has been extensively implemented in parts of the United States and Canada. The following four approaches have been suggested and were reviewed:

**Annual Mammography:** Studies in women over age 50 have found little difference in the prognosis of cancers in women receiving annual mammography from those screened biennially. Average risk women in BCCBSP varied in the frequency of screening and retrospective comparison of those returning annually versus those returning biennially found that there was no difference in the proportion of interval cancers among those with BIRADS D.

**Supplementary Breast Ultrasound (BU):** Many published studies have shown that cancers are identified by BU in women with dense breasts following a negative mammogram. A single RCT has been conducted in Japanese women aged 40-49 and found that supplementary BU increased cancer detection and reduced the interval cancer rate in the first round of biennial screening, with modest reduction in advanced cancers. Further follow-up, including the second round, of this study will provide more definitive results and give insight into the potential of supplemental BU.

**Supplementary Breast Tomography (BT):** Many published studies have shown that cancers are identified by BT in women with dense breasts following a negative mammogram. One advantage of BT is that, when combined with routine mammography, the overall abnormal call rate is lower than for mammography alone because of the ability of BT to resolve insignificant abnormalities. Studies of BT are ongoing but to date no RCT has been published and observational studies show no reduction in interval cancers.

**Supplementary Breast Magnetic Resonance Imaging (MRI):** MRI is currently used in women at very high risk (typically  $\geq 25\%$  lifetime risk) of breast cancer. No RCT has been

undertaken and only a single observational study in lower risk women has been published. This technology is resource intensive and availability within British Columbia is limited.

### *Recommendations*

The BC Cancer Breast Screening Program has indicated, by its decision to mandate BIRADS breast density assessment as part of the screening visit, that breast density is a significant factor in the provision of breast screening to the population of BC. Nevertheless, policies regarding breast density within the BCCBSP are limited and the following actions are recommended:

#### **Recommendation 1**

Develop a plan to communicate breast density results in British Columbia that involves:

- a. a process to understand the communication needs of British Columbia screening participants, and their physicians, in relation to breast density;
- b. the use of BC Breast Screening Program data to develop breast density risk information that is relevant to the BC population;
- c. a review of existing information materials, in British Columbia and elsewhere, to develop messaging for breast density as a risk factor in the context of other recognized factors that influence the likelihood of breast cancer.

#### Rationale:

##### 1.a

- Best practice in the communication of breast density information is unclear
- Practice within Canada, the United States and elsewhere is heterogenous
- Key informants expressed differing views on the desirability of various approaches to the communication of breast density
- User consultation and focus groups provide a mechanism for identifying the preferred community approach at this time

##### 1.b

- The BCCSP maintains an excellent longitudinal database capturing information on breast density and other breast cancer risk factors
- This database allows the calculation of specific screening outcomes (disease detection, interval cancers, staging, false positives, etc.) for British Columbia screening participants
- Quantitative information available from the scientific literature is based upon varied patient populations
- Estimates in the scientific literature used a variety of statistical analytic techniques which do not permit straightforward generalisation to the BC population

##### 1.c

- Existing breast density information is diverse and potentially confusing
- Research has indicated that women often over-estimate their breast cancer risks and the influence of various risk factors
- Breast density is one of many factors which influence breast cancer risk

- Breast density on its own is a poor discriminator of breast cancer risk
- Breast density influences the ability of mammography to identify breast cancer

## **Recommendation 2**

Continue to utilize the BIRADS density scoring within the BCCBSP, but continuously assess its performance and monitor the scientific literature for opportunities for improvement.

Rationale:

- BIRADS density is a commonly used clinical tool and is the most common scale used in North America
- BIRADS is subjectively assessed and inter-radiologist and consecutive measurements show variation
- BIRADS categories are clinically defined and may not be the most suitable for separating subjects based upon likelihood of developing breast cancer, or having breast cancer diagnosed before the next screening round
- Evolving mammography technology may result in changes in future performance of BIRADS
- Automated density assessment is an area of active research and future improvements seem likely

## **Recommendation 3**

Supplemental screening of women with dense breasts is not recommended at this time. The Breast Screening Program should monitor ongoing results of RCT's of supplemental screening in women with negative screening mammography.

Rationale:

- Breast Ultrasound (BU), Breast Tomography (BT) and Breast Magnetic Resonance Imaging (MRI) are all able to identify cancers in women following a normal mammography examination
- Increased detection is a necessary, but not a sufficient, requirement for benefit resulting from supplemental screening
- Use of supplemental BU and MRI will increase the number of women requiring further testing with an increase in the number of false-positive screens
- BU has the most supporting evidence for potential benefit having been demonstrated to reduce the rate of interval cancers
- In the single RCT of BU it had a modest effect on the likelihood of cancer being identified at an advanced stage, but this was based on results from the first round of supplementary screening using BU and further follow-up is required
- No guideline committee currently recommends supplemental screening with BU, BT or MRI on the basis of breast density alone
- No Canadian screening program currently recommends supplemental screening with BU, BT or MRI on the basis of breast density alone
- Further evidence is required to evaluate the benefits and harms of adding supplemental screening in selected mammographically negative women

## 2.0 Introduction

Breast density relates to the appearance of white areas on mammograms indicating increased absorption of radiation by fibro-glandular tissue compared to fat in the breast which appears dark (1). Breast density contributes to risk of breast cancer and the ability to identify early cancers on mammograms. The widespread adoption of mammographic screening has resulted in increased interest and study of the implications of breast density. The effect of breast density on risk of breast cancer has been extensively studied by epidemiologists (2,3) and it has been found to be a consistent risk factor which is correlated with some other well described breast cancer risk factors (4) and for different populations (5). However, unlike most recognized risk factors for breast cancer, breast density will not be known to the woman unless it has been provided following a mammogram. Many states in the United States now require that breast density be reported to women undergoing breast screening (6-8) with some supporting supplemental screening in women determined to have dense breasts.

The purpose of this report is review some of the evidence related to breast density and risk of breast cancer, both from the scientific literature and experience in British Columbia and provide an assessment of the current state of knowledge. The contents of this report were derived from three principal sources.

1. Interviews with key informants. Key informants (Appendix 1) were identified and interviewed, with most interviews being conducted by phone. Informants were of two key types: stakeholders within Canada with interests in breast screening and/or experts on breast screening or breast cancer researchers. Interviews were conducted using a standard set of questions (Appendix 1). Key informants were asked to identify influential publications containing information on breast density and screening.
2. Literature review. Relevant literature was sought using searches in PubMed and Google Scholar. A systematic review was not attempted but key words “breast density”, “breast density review”, “breast density meta-analysis”, “breast density and screening” were used to identify relevant publications in English. The following approach was used. Published peer reviewed meta-analyses conducted by recognized sources were assumed to be representative summaries of relevant literature published prior to their review cut-off date and used to identify relevant studies. Depending upon the particular issue potential relevant publications occurring after the systematic review cut-off date were reviewed online for relevant content. Particular emphasis was placed upon more recent publications (without formal cut-off date) both as a source of relevant results and for potential relevant citations which they contained. Studies based on data from Canada were given special attention when encountered. Publications provided by key informants were also reviewed and included in the same way.
3. Analysis of British Columbia Data. Data collected through the BC Cancer Breast Screening Program, (BCCBSP), formerly the Screening Mammography Program of BC, were analysed to determine the distribution of density in British Columbia breast screening participants and their rates of breast cancer. This analysis was intended to allow findings from the literature to be placed within the context of British Columbia.



Data from British Columbia is based upon a period when both analogue and digital mammography was in use and there was insufficient data to separate outcomes by technology.

### 3.0 Measurement of Breast Density

Mammographic pattern and breast cancer risk was extensively studied by Wolfe in the 1970's (9,10) who described a 4-category classification scale and demonstrated how breast cancer risk varied with category. The Wolfe classification is no longer used in either research or clinical practice having been replaced by other measures of density. The most common classification used in clinical practice in North America is the Breast Imaging-Reporting and Data System (BIRADS) 4-category density scale developed by the American College of Radiology. This system has changed over time and its current form (Fifth Edition) is intended to provide a subjective indication of the likelihood of lesion obscuration (11) using a four-category scale:

- A: Fatty
- B: Scattered Fibroglandular
- C: Heterogeneously dense
- D: Extremely Dense

Earlier versions of this scale also used four categories but were a subjective assessment of the dense tissue expressed as proportions of breast area (<25%, 25-50%, 50-75% and >75%). Legislation and clinical practice in the United States usually refers to women who are BIRADS C or D as having dense breasts. In Canada many screening programs have included density assessments by the interpreting radiologist recorded as a binary classification using either <50% vs ≥50% or <75% vs ≥75%.

Researchers in breast etiology and epidemiology have tended to use quantitative indices (12) which relate to the quantity of the breast deemed to be dense. The most common approach is to estimate, by varied means, the proportion of the breast classified as dense and group proportions into discrete intervals. The digitisation of mammography has resulted in the development of several algorithms which process digital images to provide estimates of breast density using diverse scales (13). This is an active area of research with new algorithms being developed (14) and existing algorithms undergoing evaluation in a number of environments (15-18). Automated measures of breast density performed on the captured mammography image offer the advantage of reductions in reclassification associated with inter-rater and intra-rater variation associated with radiologist-based measurement used in BIRADS. Also, computed continuous scales offer the opportunity for user specified categories for potential clinical use and would also allow the inclusion of other patient characteristics (e.g. age, etc.) to provide better estimates of breast cancer risk (14,17). Comparative analyses suggest that automated measures provide superior risk discrimination to radiologist recorded measures (18,19). A study comparing different methods of quantitative assessment of breast density found that automated techniques were valid alternatives to more labour-intensive approaches (20). A study of serially collected mammograms found that automated methods provided greater consistency than human readers for density assignment (21).

To examine the performance of the BIRADS density scale among BC radiologists a sample of pairs of consecutive screens, where density was reported, were examined and the consistency of reported density examined. Both mammograms in a pair were required to be either captured on

analogue or on digital equipment although a comparison in the United States found no difference radiologist reported in density assessment by equipment type (22). There were 48,254 pairs of mammograms which satisfied the eligibility criteria (for methodology and detailed results see Appendix 3). The US Preventive Services Task Force (23) reports that in 77% of consecutive examinations the BIRADS category was unchanged where the same radiologist reported both times whereas this fell to 68% when different radiologists reported. In British Columbia the comparable rates were 74% and 62% respectively. It is possible that a longer screening interval, 2-years, contributed to the lower level of agreement in British Columbia. Also, the British Columbia data did not record which version of BIRADS was used (4<sup>th</sup> or 5<sup>th</sup> Edition)<sup>1</sup> in each interpretation and this likely contributed to disagreement as the two editions lead to different classifications (24,25). In clinical use the usual outcome is to classify breasts as dense or not. As previously noted in the United States the designation of a woman having dense breasts is BIRADS C or D, whereas in some parts of Canada  $\geq 75\%$  is used, which approximately corresponds to BIRADS D. Using C+D as dense, 80.3% (11,403 of 14,202) of those initially dense were classified as dense on the second mammogram whereas using D as dense, 63.7% (1,804 of 2,833) of those initially dense were classified as dense on the second mammogram in the review of British Columbia data (Appendix Tables 7 and 8).

In the British Columbia data, it is clear that using C+D as the definition of dense results in a classification which is more stable than that based upon D alone since the proportion changed is lower between the two measurements. However, while the proportion changed is greater for D (as dense) the number changed is higher when using C+D as dense. For example, based on the measurement on the first mammogram of C or D, 2,799 (5.8% of the total sample) classified as dense would be classified as not dense on the second mammogram, compared to 1,029 (2.1% of the total sample) when D alone is used. While density is not anticipated to be constant, the comparative frequency of change would suggest that attention would need to be paid to its application for screening. The British Columbia data were also analysed by woman's age at the first mammogram in the pair of examinations (Appendix Table 9). This indicates that there is an influence of age with lower proportions changed for younger women (using either D or C+D as dense). However, the influence of age is fairly weak.

Variation in the stability of serial BIRADS breast density assessments for individual radiologists was not examined in the British Columbia analysis but large variations have been reported amongst radiologists in the United States (26-28) although studies from other countries have found good levels of agreement (29,30). Variability in density assignment can potentially result in reduced measured relationship between density and breast cancer risk and reduce the usefulness of BIRADS classification as a tool for use in a screening program.

It should be noted that the circumstances under which BIRADS density was captured in BCCBSP may be unrepresentative of performance for any clinical use in the future. Clinical management of women was not dependent on BIRADS density and reporting was voluntary. Data was drawn from a period during which the 4<sup>th</sup> Edition of BIRADS was replaced by the

---

<sup>1</sup> In the 4<sup>th</sup> Edition the density categories are D1, D2, D3 and D4. For simplicity we will use A, B, C and D as in the 5<sup>th</sup> edition.

current version (5<sup>th</sup> Edition, (31)). A United States study found that variability increased using the fifth compared to the fourth edition of BIRADS (24).

Melnikow et al (23) published a review of evidence for US Preventive Services Task Force with a view to key questions the first of which was:

1. What are the accuracy and reproducibility of BI-RADS determination of breast density?

In addressing this question, the committee noted that there was no gold standard for density measurement so that they used consistency in classification as a surrogate for accuracy. They primarily use published data collected from United States sources. They find that consistency of density classification in consecutive screens is higher when the same radiologists reports both compared to when different radiologists report. They note that variations in individual radiologist performance can greatly exceed that of group average. They do not include studies using other (not BIRADS) methods of classifying density. While of interest it would seem that, for the purposes of this report, that the results obtained from the British Columbia analysis are more directly relevant. Table 3.1 compares some results from the US Preventive Services Task Force evidence review and those found in the BC analysis.

**Table 3.1:** Comparison of Summary of Results on Stability of BIRADS Density from US Task Force and from BC Analysis for Consecutive Screens

	% Change	
	USPS TF Summary	BC Results*
% BIRADS in Different Category (A, B, C, D) on Subsequent Screen by Same Radiologist	23%	26%
% BIRADS in Different Category (A, B, C, D) on Subsequent Screen by Different Radiologist	32%	38%
Change in Dense Assignment (Dense=C+D) on Subsequent Screen by Same Radiologist	14%	11%
Change in Dense Assignment (Dense=C+D) on Subsequent Screen by Different Radiologist	19%	15%

\*source: Appendix Tables 7 and 8

The key informant interviews (Appendices I) included a question of the appropriateness of BIRADS as a scale for measurement density in a clinical setting. Among respondents knowledgeable about breast density different opinions were expressed about the suitability of BIRADS for measuring density by BCCBSP:

- some felt it was the most appropriate tool for clinical use and by BCCBSP while
- some felt that BIRADS was suboptimal and that continuous quantitative scales were superior.

Those preferring quantitative scales appeared to do so because of perceived superior consistency, possibly by automated measurement, and stronger discrimination of breast cancer risk. Respondents were not asked, and none expressed a preference for human versus automated measurement. An advantage of continuous scales is that also permit arbitrary dichotomization of the screening population for clinical use rather than relying on scales, such as BIRADS, which use qualitatively predefined thresholds.

### 3.1 Distribution of Breast Density

Several estimates of the distribution of breast density, using BIRADS, are available from locations within the United States. A commonly quoted guide (6,32) is A:10%, B:40%, C:40% and D:10%, so that ~50% of women would be considered to have dense breasts in the United States. However, density varies considerably with age so that averages depend on the age distribution of the cohort used. Sprague reports BIRADS density by age for a large cohort of American women (33) and an excerpt of their data is presented in Table 3.2. Overall the distribution is similar to the 10:40:40:10 commonly used with approximately 50% being considered dense. However, the data shows that for women aged 40-49 61.7% would be considered dense (C&D) while for women 70-79 31.9% would be classified as dense. Recent data from the BC Cancer Breast Screening Program (BCCBSP) reported a distribution of A:17%, B:43%, C:33%, D:8% among those reporting in 2017<sup>2</sup>. Consequently in 2017, assuming 260,000 screens are performed by BCCBSP there would be approximately 86,000 and 20,000 women classified as BIRADS C and D respectively.

**Table 3.2:** Distribution of BIRADS Density (%) by Age for 1,178,262 Women Reported by the Breast Cancer Surveillance Consortium (33)

Age	BIRADS Breast Density			
	A	B	C	D
40-49	6.2	32.1	46.9	14.8
50-59	10.8	41.6	39.2	8.4
60-69	14.9	47.9	32.7	4.6
70-79	16.3	51.7	28.8	3.1
40-79	11.2	41.7	38.5	8.7

An analysis undertaken to investigate the relationship between breast density and breast cancer risk in the BCCBSP database provided the distribution by age given in Table 3.3: full details of this analysis are given in Appendix 3. The distribution in Table 3.3 is different from that in Table 3.2 which is related to the use of Screening Rounds, which are weighted by screening frequency, as the unit of measurement in Table 3.3 and also is based upon different reporting periods with different BIRADS versions. It should also be noted that the BCCBSP populations being screened are different from many US jurisdictions since it excludes those with a history of breast cancer and proven genetic predisposition. Another publication (34) using data from the Breast Cancer Surveillance Consortium has almost identical distribution of extremely dense (D) women by age as shown in Table 3.3. Both Tables 3.2 and 3.3 show similar trends with categories BIRADS C & D becoming less frequent as age increases. Density was also examined by ethnic group in the BCCBSP data and women self-identifying as East or South East Asians tended to have denser breasts, and First Nations less dense breasts than other participants (Appendix Table 2).

<sup>2</sup> C McGahan, BCCBSP, personal communication

In conclusion data available suggests that BCCBSP participants generally have a lower density distribution than that reported for the United States and that similar relationships with age are present.

**Table 3.3:** Distribution of BIRADS Density (%) by Age for 484,375 Screening Rounds in the BC Cancer Breast Screening Program

Age	BIRADS Breast Density			
	A	B	C	D
40-49	18.4	31.5	36.2	13.9
50-59	28.4	37.6	26.8	7.2
60-69	37.0	39.9	19.4	3.9
70-74	40.3	40.7	16.4	2.7
40-74	28.9	36.7	26.5	7.8

### 3.1.1 Number of Women Affected in British Columbia

Sprague (33) attempted to estimate the number of women in the United States with dense breasts by weighting age- and body-mass-index (BMI) - specific rates of density in women undergoing screening, by known population age and BMI distributions. Breast density is known to be inversely related to BMI and directly related to age. While of some epidemiologic interest, any intervention in women with dense breasts in British Columbia would first require them to have known density which would effectively correspond to them being participants in the BCCBSP program. Consequently, the number of affected women would be the number of women with dense breasts in the program. Using the BCCBSP program reported distribution of density and the screening program volume in 2015, it is possible to estimate the number of women who would be affected in the program: this is given in Table 3.4. Four scenarios are considered for the designation “dense”:

- a single BIRADS C or D finding,
- two consecutive BIRADS C or D findings
- a single BIRADS D finding,
- two consecutive BIRADS D findings

Requiring two consecutive reading to be consistent in classification reduces the number of women designated as dense, and given interpretation variability, is likely to result in them being at higher risk than those classified on the basis of a single measurement. That is false-positive “dense” assignments are likely to be reduced but false-negatives will be increased. The estimates in Table 3.4 is a little different from that obtained using BCCBSP data for 2017 and may be a reflection of the increased reporting in 2017.

**Table 3.4:** Number of Women Classified as Dense in One-Year of Screening Using the Reported Program Density Distribution and The Number of Women Screened in 2015 (35)

	Number Screened in 2015 = 255,534	
Dense Classification	Proportion (%)	Estimated Number
Single BIRADS C or D Screen	39.0	99,700
Single BIRADS D Screen	7.0	17,900
Two Consecutive BIRADS C or D Screens	31.3	79,980
Two Consecutive BIRADS D Screens	4.5	11,500



### 3.2 Breast Density and Risk of Breast Cancer

In analysing the effect of breast density on breast cancer risk it is common to separate interval cancers for special consideration.

#### Overall Risk of Breast Cancer

A meta-analysis of density and breast cancer risk was published in 2006 (36) which reviewed English language publications prior to 2005. This review only included 3 publications which used the then existing BIRADS density classification. The studies reviewed (37-39) all found an increasing risk of breast cancer as BIRADS category increased. Relative risks (reference category=fatty) were: Scattered, 2.04 [95%CI:1.56-2.67]; Heterogeneously Dense, 2.81 [95%CI=2.13-3.71]; and Extremely Dense, 4.08 [95%CI=2.96-5.63]. In these studies, the most common category was Scattered with less than 10% of women falling in either the Fatty or Extremely Dense categories. The Meta-Analysis (36) also summarized studies where percent density was assessed and used to categorize women. In this analysis 5 categories were used (<5%, 5-24%, 25-49%, 50-74%,  $\geq$ 75%). The overall results are summarized in Table 3.5.

**Table 3.5:** Summary Relative Risks and Confidence Intervals for General Population Incidence Studies as Summarized by McCormack (36)

<b>Breast Density</b>	<b>Combined RR</b>	<b>95%CI</b>
<5%	1.0 (reference)	
5-24%	1.79	1.48-2.16
25-49%	2.11	1.70-2.63
50-74%	2.92	2.49-3.42
$\geq$ 75%	4.64	3.64-5.91

The meta-analysis (36) also presented results by age, time post study entry and time post mammography. Results did not demonstrate significant variations although RR's of cancer detection following mammography in women with  $\geq$ 75% breast density was the highest for any category.

Since BMI and breast density are negatively correlated, adjustment for BMI frequently increases the strength of the overall relationship between breast cancer risk and breast density. For example, adjustment for BMI increased the relative risk (for breast density >50% vs <10%) from 3.0 to 3.9 in a US study (40). Interactions with other breast cancer risk factors have been reported (41). In a Canadian study (for breast density >75% vs <10%) the adjusted RR is (all cases) 4.7 which declines to 3.3 with no adjustment (12).

Data was extracted from the BCCBSP database to examine the relationship between recorded breast density using BIRADS and risk of breast cancer over a 2-year screening cycle. Full details of the definitions, patient population, methods and results are given in Appendix 3. Some of the results are summarized Table 3.6. Excluding women classified as high risk, the annualized estimates for annual and biennial screening cycles are given by age in Table 3.6.

When examining either of the annualized estimates (annual or biennial) it can be seen that risk increases regularly with density for 40-49 age group but not for 50-74 age-group so that in this latter group density does not discriminate overall cancer risk well. This irregularity in the older group is likely because of the wide age-range included and the negative correlation between age and density and the positive correlation between age and cancer risk. A further contributor is the use of a screening round as a unit of analysis. A consequence of this is that the screen detected rate is the rate at the next screen following the negative screen which was the source of the density assignment. Expressed as relative annualized rates for biennial screening, compared to the average for women aged 40-49, the ratios were as follows: A, 0.69; B, 0.86; C, 1.19; and D, 1.63. For women aged 50-74 the same ratios with density were: A, 0.79; B, 1.05; C, 1.27; and D, 0.96 (Average Annualized Rate for 50-74 was 2.74 per 1,000).

**Table 3.6:** Rates of Invasive Breast Cancer by Screening Cycle Length, Annual or Biennial, and by Age for Participants in the BCCBSP Program (Source: Appendix Tables 3 and 4)

		Rates of Invasive Cancer per 1000					
		Annual Screening			Biennial Screening		
Age	Density	Screen Detected	Interval <12 months	Annualized (SD+PS)	Screen Detected	Interval <24 months	Annualized (SD+PS ÷2)
40-49	A	0.85	0.09	<b>0.94</b>	1.58	0.68	<b>1.13</b>
	B	0.98	0.49	<b>1.47</b>	1.82	0.97	<b>1.40</b>
	C	1.15	0.55	<b>1.70</b>	2.15	1.73	<b>1.94</b>
	D	1.50	1.46	<b>2.96</b>	2.79	2.53	<b>2.66</b>
	All	1.18	0.57	<b>1.75</b>	1.86	1.40	<b>1.63</b>
50-74	A	2.39	0.17	<b>2.56</b>	3.46	0.85	<b>2.16</b>
	B	3.23	0.41	<b>3.64</b>	4.69	1.08	<b>2.89</b>
	C	3.04	0.80	<b>3.84</b>	4.40	2.54	<b>3.47</b>
	D	1.45	1.84	<b>3.29</b>	2.11	3.12	<b>2.62</b>
50-59	All	1.77	0.50	<b>2.27</b>	2.78	1.31	<b>2.05</b>
60-74	All	3.33	0.48	<b>3.81</b>	5.23	1.51	<b>3.37</b>

Many recent studies of breast density and breast cancer risk have examined the relationship between breast density and other known breast cancer risk factors. A Canadian study (12) which drew subjects from the Breast Screening Programs in British Columbia and Ontario and the Canadian National Breast Screening Study classified subjects by percent density (<10%, 10-24%, 25-49%, 50-74%, ≥75%) and found similar results for all cancers to those presented in Table 3.5 although the reference level was not the same. Table 3.7 provides some results from the study by Boyd (3) where estimates are adjusted for BMI, reproductive factors, menopausal history, HT and family history of breast cancer. Boyd (3) reviews studies using percent density published prior to 2011.

A detailed discussion of the risk of breast cancer detection by mammography is not provided here. However, we note that since benefit from screening comes via the screen detection of cancer then the rate of screen detection should be approximately proportional to the benefit of screening. On this basis women with BIRADS A (see Table 3.6) aged 40-49 receive the least

benefit from biennial mammography screening since they have the lowest screen detection rate (1.58 per 1,000) whereas women BIRADS B aged 50-74 receive the most benefit with a rate of 4.69 per 1,000 for biennial screening.

**Table 3.7:** Summary Relative Risks and Confidence Intervals by Percent Density and Mode of Detection for Women Participating in Screening in Canada (12)

Breast Density	Method of Detection			Total
	Screen Detection	Interval <12 months	Interval ≥12 months	
<10%	1.0 (ref)	1.0 (ref)	1.0 (ref)	1.0 (ref)
10-24%	1.6 (1.2-2.2)	2.1 (0.9-5.2)	2.0 (1.2-3.4)	1.8 (1.4-2.2)
25-49%	1.8 (1.3-2.4)	3.6 (1.5-8.7)	2.6 (1.5-4.6)	2.1 (1.6-2.6)
50-74%	2.0 (1.3-2.9)	5.6 (2.1-15.3)	3.1 (1.6-6.2)	2.4 (1.8-3.3)
≥75%	3.5 (2.0-6.2)	17.8 (4.8-65.9)	5.8 (2.1-15.5)	4.7 (3.0-7.4)

### 3.3 Breast Density and Risk of Interval Breast Cancer

While breast density is related to overall breast cancer risk special interest focuses on its relationship to the risk of interval cancer following mammography screening. Data from a Canadian study (Table 3.7) shows that the relative risk of an interval cancer within 12 months was the most elevated (OR=17.8) for density  $\geq 75\%$  relative to the lowest category ( $<10\%$ ). The gradient of relative risk with density for the  $< 12$  months category is stronger (larger) than for screen detected cancers. This is reflected in the proportion of cases detected at screening compared to that detected at screening or occurring within 12 months of a negative screen (frequently referred to as sensitivity) by percent density (based upon (12) using raw data),  $<10\%$ : 0.94, 10-24%: 0.89, 25-49%: 0.87, 50-74%: 0.76 and  $\geq 75\%$ : 0.68. Examination of the variation, by density, in the overall and  $\geq 12$  months interval RR's in Table 3.7 shows them to be quite similar indicating that most of the gradient in  $\geq 12$  months interval cancers is associated with the effect of density on breast cancer risk rather than an effect of masking.

Table 3.6 provides similar results to Table 3.7 from the analysis of BCCBSP data although they are presented as absolute rather than relative rates: the risk of interval breast cancer increases progressively by density in both age groups in Table 3.6. Measured over a screening period of 24 months among women 40-49 the relative risk of interval cancers compared to the 40-49 average, increased with BIRADS density categories: A: 0.49, B: 0.69, C: 1.24 and D: 1.81 (Table 3.6). For women aged 50-74 over a 24-month screening period the relative risk of interval cancers, compared to the 50-74 average, increased with BIRADS density categories A: 0.60, B: 0.76, C: 1.79 and D: 2.20 (Average for 50-74 = 1.42 per 1,000). As an internal reference point the 24-month interval cancer rate (Appendix Table 4) in women 40-49 deemed at high risk, 3.13 per 1,000, exceeded the rates for any of the density categories, but for women 50-74 the rate, 2.30 per 1,000 (Appendix Table 4), was less than that for C and D densities in that age group (Table 3.6). These high-risk women are identified to be screened annually with mammography and their corresponding 1-year interval cancer rates are less than 1 per 1,000 (Appendix Table 4).

An analysis of Breast Cancer Surveillance Consortium Data provides observed 12-month interval cancer data in a cohort of 362,730 screened women contributing 825,586 screens (34). Some results from their analysis are summarized in Table 3.8. Although the age intervals in these results (Table 3.8) are not the same as those presented for BC (Table 3.6) they provide a fairly similar pattern with density providing a larger gradient in interval cancer risk than age.

**Table 3.8:** 12-month Interval Cancer Rates by Age and Density in a Cohort of Women from The Breast Cancer Surveillance Consortium (34)

Age	12-Month Interval Cancer Rates/1,000 by BIRADS Breast Density			
	A	B	C	D
40-49	0.19	0.26	0.76	0.98
50-59	0.14	0.33	0.80	1.11
60-69	0.23	0.49	0.96	1.13
70-74	0.35	0.55	1.15	3.45

### *Summary of Density and Breast Cancer Risk*

The preceding findings indicate that use of breast density as a single factor in a screening program population will not result in a risk separation as large as that in some published findings. This arises primarily because these published results use a more detailed density scale intended to identify risk and control for other breast cancer risk factors which are often positively correlated with risk but negatively correlated with density (e.g. age, body mass index). Using density as a single factor to divide a population into lower and higher risk would lead to considerable misclassification. A more personalized risk assessment would be required to achieve improved classification and methods are developed in several publications (42,43).

Generally, the results of the BCCBSP analysis replicated ones found in the literature for density with increases in risk associated with increasing density (23). This was not true for screen detected cancers in women 50-74 possibly because of confounding by age and other risk factors. It does indicate that selecting women aged 50-74 on the basis of density alone will not result in the identification of a very high-risk group for breast cancer. However, all ages showed a relationship between increased density and higher rates of interval cancer.

The strongest relationship between breast density and breast cancer risk in the BCCBSP data was seen for interval breast cancers. This agrees with results from the literature and aligns with the concept of masking. BIRADS density provides a consistent discrimination of the risk of interval cancer.

In the key informant interviews with those claiming a familiarity with breast density there was a consensus that increasing breast density is associated with increasing risk of breast cancer and decreasing detectability (masking) of breast cancer by mammography.

### 3.4 Performance Targets for Interval Cancer Rates

There is no generally accepted definition of what constitutes unacceptably high interval cancer rates. Interval cancers occur at all ages in all density categories. In 2013 BCCBSP changed screening policy so that women aged 40-74 with a first-degree family history of breast cancer are recalled annually. Making this change was based upon the ability to maintain a reasonably high-screen detection rate (so that a reasonable balance is maintained with the likelihood of false-positive screens) and reduce the elevated interval cancer detection rate over the pre-existing 24-month screening interval. Examination of Appendix Table 3 in Appendix 3 shows that the measured rate of interval cancers in women aged 40-74 with density D (2.84 per 1,000) exceeds that measured in similar high-risk women (2.46 per 1,000) but other density categories are less. On the basis of this comparison there would be a case for considering those with density D to have a sufficiently high interval cancer rate to make them candidates for altered approaches if effective measures are available.

In an analysis Kerlikowske (34) suggests that an interval cancer rate of 1 per 1,000 or greater in the first twelve months following negative mammography represents a high interval cancer rate (for primary mammography screening). This designation appears to be based on the reasoning that most women are screened annually in the US (thus 12 months) and that interval cancers should represent no more than 1/3 of the rate of screen detected cancers and that average overall incidence is ~4 cases per 1,000. However, breast cancer incidence varies strongly with age, so it is difficult to apply this rate across age groups. For example, a recent monitoring report of breast screening in Canada (44) reports a detection rate of 1.6, 2.8, 4.7 and 6.5 per 1,000 for ages 40-49, 50-59, 60-69 and 70-74 respectively in previously screened women returning biennially. Furthermore, in Canada, screening for average risk women is every two years, so it is necessary to consider interval cancers within 24 months of screening.

In the British Columbia analysis rates of interval cancer over 24 months for density categories B, C and D all exceed 1 per 1,000 interval cancers over two years (Table 3.6). However, published studies using data from British Columbia and elsewhere have found that there is little if any advantage in annual versus biennial screening at the population level (45,46) and consequently criteria based on 12 month rates cannot be directly applied to 24 month interval cancer rates. The Breast Screening in Canada Report (44) establishes program targets in of <0.6 per 1,000 and <1.2 per 1,000 for interval cancers within 0-12 and 12-24 months post-screen respectively and a screen detection rate of >3.0 per 1,000 for subsequent screens in the 50-69 age group. Table 3.6 shows that these interval targets are met for the three age ranges (40-49, 50-59 and 60-74) considered in the analysis of British Columbia data and the Canadian Report (44) finds that the screen detection target is met for the applicable 50-69 age group in British Columbia. The 0-12 months Canadian interval cancer rate target, 0.6 per 1,000, is below the threshold proposed by Kerlikowske (34), 1.0 per 1,000, as would be expected since the first is for average performance and the second an upper threshold. It is clear that the acceptable performance in Canada for a 24-month interval cancer rate, as high as 1.8 per 1,000, is different from the United States based upon a 12-month interval, <1.0 per 1,000. One possible approach to adapt the threshold provided by Kerlikowske (34) to a Canadian context is to apply the ratio of the two 12-month rates to the 12-24 months Canadian target to provide a high level at 12-24 months which yields a rate of 2.0 per 1,000 ( $1.2 \times 1.0/0.6$ ) so that a upper threshold for 0-24 months would be a rate of

>3.0 per 1,000. Examination of Table 3.6 shows that only BIRADS D for women 50-74 has a 24-month rate of interval cancer which exceed 3.0 per 1,000 and that only BIRADS D, for both age groups considered, has a 0-12 months interval cancer rate which exceeds 1 per 1,000.

Another way of viewing performance thresholds would be to express them in terms of the proportion of cancers identified by screening over the 24-month screening interval. Using the target national average screen detection rate of >3.0 per 1,000 and the inferred upper threshold for the 24-month interval cancer rate of 3.0 per 1,000, then the proportion screen detected, PSD is

$$PSD = \frac{> 3.0}{3.0 + > 3.0} \Rightarrow 0.5.$$

That is, we would aim that the proportion of cancers screen detected in participant subgroups should exceed 50%. Results from Table 3.6 shows that PSD (proportion screen detected) for women with BIRADS D screened biennially is 52% (2.79/5.32) for women 40-49 and 40% (2.22/5.23) for women 50-74 whereas for BIRADS C it is 55% (2.15/3.88) and 63% (4.4/6.94) respectively. For BIRADS D women the value of PSD for annual screening is little changed 51% (age 40-49) and 44% (age 50-74) from that for biennial screening indicating that annual mammography does not substantially improve this performance measure in this group.

In summary there are no accepted targets for rates of interval cancer in population subgroups of women undergoing mammography screening in Canada. Adaptation of targets proposed in the United States to Canada suggests that women with BIRADS D density would not meet the performance targets based upon analysis of BCCBSP data. Using British Columbia data approximately 8% of participants are BIRADS D but this does vary with age.

### 3.5 Communication of Breast Density

There does not seem to be any accepted best practice in the communication of breast density to women or their family physicians. Many states in the United States now require that breast density be reported to women undergoing breast screening (6-8), so that it is also reported to the physician. Some states support supplemental screening in women determined to have dense breasts. Reporting in the United States is based upon BIRADS and C & D are considered as “dense”. In Canada some provinces manage those they consider dense ( $\geq 75\%$ ) and allocate them to a higher risk group (along with women with a first-degree family history) and recall them for annual mammography screening. Five provinces notified the family physician but only Ontario routinely notifies the woman of a finding of dense breasts following screening. It appears that generally those considered not-dense are not notified of their status. Canadian screening programs do not recommend other modalities for screening women on the basis of breast density alone whereas it is common practice in the United States.

A randomized trial conducted using participants in the BCCBSP examined communication of breast density to women with reported density  $\geq 50\%$  (47). Women and their physicians were provided with information on breast cancer risk factors, breast density and risk reduction strategies by including supplemental material included with the communication of screening results. No specific supplemental imaging interventions were described. Subjects were randomized to receive or not receive the material and were tested at 4 weeks and 6 months, by telephone survey, on knowledge of breast density and a number of behavioural dimensions (e.g. anxiety, worry). Although some differences were noted between the two arms at 4 weeks these had declined by 6 months and were no longer significant. There was no longer term follow-up done to determine if subsequent screening behaviour was affected. Comparatively few subjects contacted their physician (21%) and physician-based outcomes were not collected.

The Canadian research study (47), although well designed is not an implementation study and several factors would need to be considered:

- participants were volunteers and needed to provide informed consent so they may be unrepresentative of the general population of screening participants
- the sample size was modest and rare events ( $< 1$  per 100) may not be reliably captured
- the study was performed in an era when on-line information on breast density was limited: the curation of information accessed by the woman would not be so great today
- family physician experiences were not measured

In the interviews of key respondents there was considerable heterogeneity in opinion about the communication of breast density:

- Several felt that it should be automatically provided to both screened women and their physicians
- Some felt it should be directly provided to the physician only, who could then use it within the context of clinical care, i.e. communicate it or not as deemed appropriate by the provider’s judgement and the patient’s needs



- Some felt that density should only be proactively reported if it would affect recommended care
- All felt that any communication of density should only be made within the context of a well-developed education strategy aimed at both physicians and women
- Several respondents expressed the opinion that by collecting density information the program would need to explicitly justify non-communication to women
- Several respondents felt that breast density should be reported alone but within the context of a general communication of breast cancer risk which could go as far as the production of individualized estimates of breast cancer risk for women in screening
- Some respondents felt that communication of density be limited to only high breast density (i.e. non-high density should not be reported)
- One respondent favoured reporting only following a request (current BCCBSP practice)

Respondents favouring direct communication of density frequently advocated tailored clinical management based upon density. Several indicated that it was potentially counterproductive, and possibly harmful, to communicate a dense categorisation but not indicate potential specific actions as a result of it, however not everyone was asked their opinion of the communication of density without associated clinical recommendations.

### 3.6 The Effect of Density on Breast Cancer Prognosis

Mammographic breast screening is known to reduce breast cancer mortality and the reduction in mortality is commensurate with reductions in tumor size at and likelihood of lymph node involvement at diagnosis (48). Increased breast density has been linked to larger cancers and positive lymph nodes at diagnosis (49). Higher breast density has also been linked to increased breast cancer mortality and this effect may be stronger than its effect on incidence alone (2). A study that controlled for stage at presentation found no effect of increased density (50) indicating that any effect of density on prognosis acts principally through the stage at presentation and risk of disease. It is to be anticipated that women at increased risk of breast cancer (for reasons of density, age etc.) are more likely to die from breast cancer than those at lower risk. Data from the Breast Cancer Surveillance Consortium (51) is summarized in Table 3.9 and includes both screen-detected and 24-month interval cancers in women screened biennially. They find that tumour size at diagnosis and the proportion with positive nodes increases with increasing density. Similar gradients between density and size and nodal involvement are seen in those screened annually (51).

**Table 3.9:** Proportion with Tumour Size  $\geq 15$ mm and Positive Lymph Nodes by BIRADS Breast Density Screened Biennially (51)

	Proportions (%) by Breast Density			
	A	B	C	D
Tumour Size $\geq 15$ mm	42.4	46.0	53.8	61.1
Positive Nodes	23.5	25.1	30.6	32.1

BCCBSP program data was also analysed to examine the effect of density on tumour size and lymph node involvement at diagnosis for those undergoing biennial screening. Full results are contained in Appendix Table 6 and are summarized in Table 3.10.

**Table 3.10:** Proportion with Tumour Size  $\geq 15$ mm and Positive Lymph Nodes for Screen Detected and 24-Month Interval Cancers by BIRADS Breast Density Screened Biennially in the Analysis of BCCBSP Data

	Mode of Detection	Proportions (%) by Breast Density			
		A	B	C	D
Tumour Size $\geq 15$ mm	Screen Detected	27	27	39	38
	Interval	58	61	67	64
	<b>Overall</b>	<b>33</b>	<b>34</b>	<b>50</b>	<b>52</b>
Positive Nodes	Screen Detected	11	19	27	24
	Interval	25	42	27	40
	<b>Overall</b>	<b>14</b>	<b>24</b>	<b>27</b>	<b>32</b>

Overall the patterns seen in Tables 3.9 and 3.10 are similar with increasing proportion of larger tumours and proportion of lymph node involvement as density increases, although the number of cancers is small in the British Columbia data. The separate breakdown for interval cancers and screen detected cancers in the British Columbia analysis finds that the difference by density seems to be stronger for the screen detected than for the interval cancers. This finding is not unexpected since the masking will tend to affect screen detection primarily. However, it can be seen that the difference in prognostic profile of the screen detected and interval cancers is large in each density category so that increased screen detection with reductions in interval cancers, in any density category, presents an opportunity for improvement in prognosis. A recent large Dutch study (52) found that while improved therapy had increased survival for all cancer stages, stage at presentation (tumour size and lymph node involvement) was still a major influence on long term outcomes.

### **3.7 The Effect of Density on False Positive Mammography Examinations**

Increased density is associated with an increased likelihood of a false positive screening mammography. Kerlikowski (51) provides estimates, by BIRADS density and by age and mammography screening frequency (annual, biennial and triennial), of 10-year risks of a false positive mammogram. Over a fixed time period (10 years) they find that false positives increase with screening frequency and decrease with age. False positives are highest for BIRADS density C, slightly lower for D with A having about half the rate of C and B intermediate between A and C. The relative proportions vary little with screening frequency and age although the absolute rates do. The analysis of BC data did not attempt to replicate the foregoing United States analysis, but false-positives were examined using the screening round data set. This data set does not include first screens. The general results were the same as reported by Kerlikowski (51): that BIRADS density C had the highest rate of false positives followed by D, B and A. Younger women had higher false-positive rates and the longest intervals had the highest rate of false-positives. Details of the analysis are contained in Appendix 3 with results in Appendix Table 5.

#### **4.0 Framework for Evaluating Studies of Supplemental Screening in Women not at High Risk of Breast Cancer Receiving Primary Screening by Mammography**

When considering routine screening with mammography, the immediate outcome is to diagnose some cancers earlier than they would otherwise be diagnosed. Using the conventional schematic for screening (Figure 1) some cancers are screen detected at an earlier stage and the risk of death from those cancers is reduced. However, there is another model for the effect of screening which sees cancers with little or no malignant potential being diagnosed because of screening (overdiagnosis) with minimal effect on cancers with true malignant potential (Figure 2). The models are not mutually exclusive and the phenomena they represent coexist in mammographic screening for breast cancer. For example, the Euroscreen study consortium estimated 6.5% overdiagnosis, expressed as a proportion of cancer risk beyond the onset of screening, (53), and breast cancer mortality reductions of 38% (54). These percentages are based upon different denominators but, overall, they translate into approximately 2 breast cancer deaths prevented for each breast cancer overdiagnosis (55). The Independent UK Panel commissioned to consider the harms and benefits of mammography screening concluded that 3 breast cancer deaths were prevented for each overdiagnosis (56). Much has been written on overdiagnosis and the mortality benefit of breast screening and different authors have developed wildly different estimates (57). It is probably reasonable to conclude that ratio of 2-3 deaths are prevented for each overdiagnosis indicating that overdiagnosis is a significant harm associated with mammography screening. When considering supplemental (adjunctive) screening, added to mammography, it is not clear whether further overdiagnosis can occur.

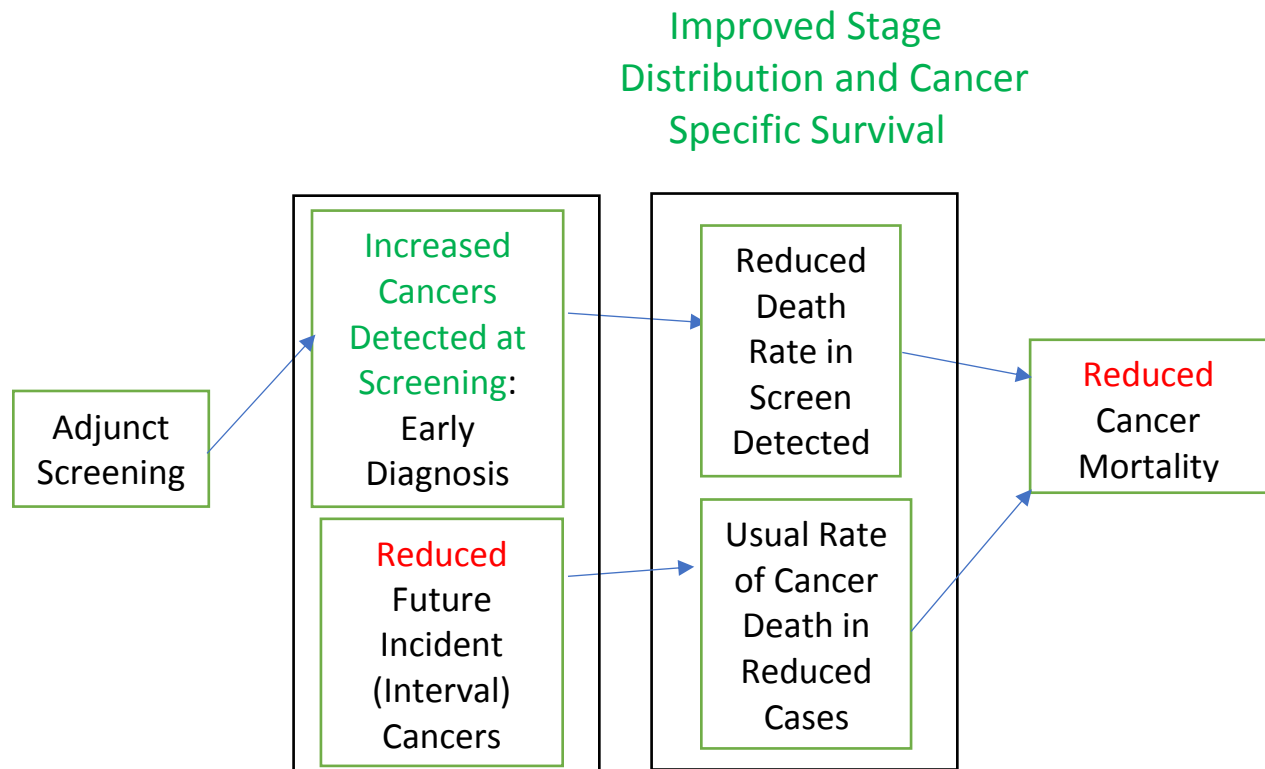
Identification of overdiagnosis, in any context, is challenging. Three main methods have been used:

1. Long-term follow-up of randomized trials where subjects were randomized to mammography screening for several rounds compared to a non-screened control group.
2. Analysis of populations before and after screening was introduced with statistical adjustment for breast cancer trends, prevalence effects etc.
3. Statistical simulation models of breast cancer comparing scenarios with and without screening.

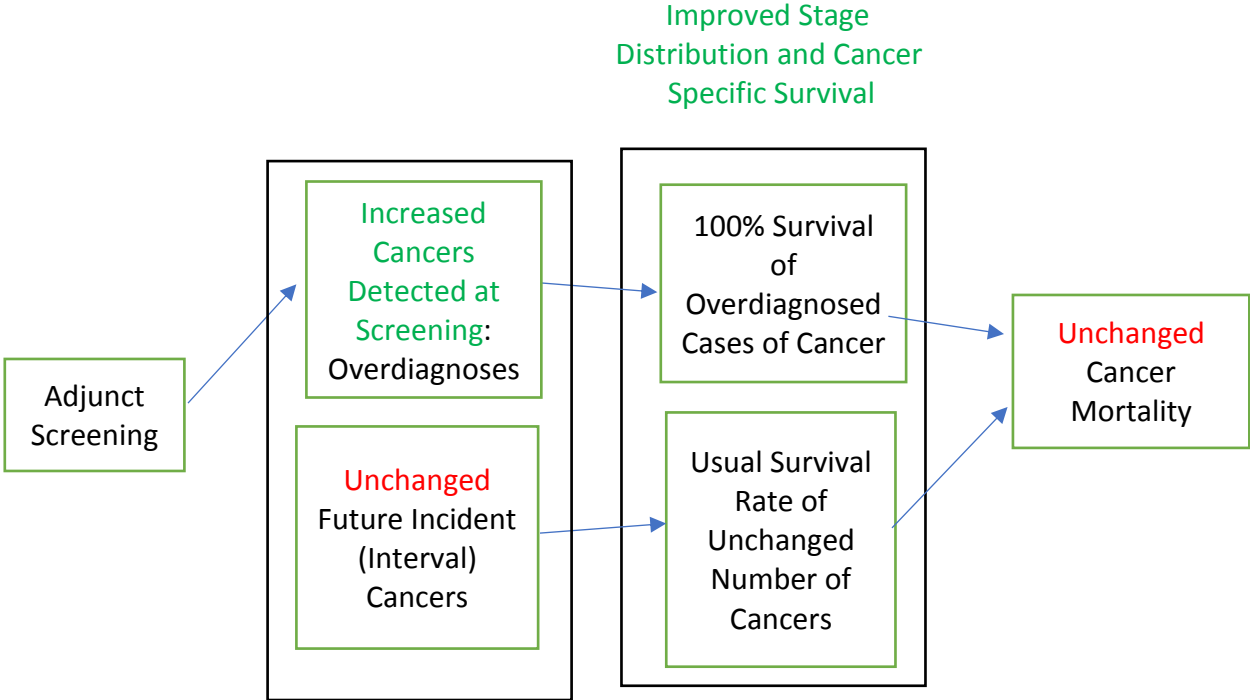
Each of these methods has its strengths and weaknesses although expert guideline committees have tended to favour estimates based upon the first of the above methods (56,58). So far there have been no attempts to estimate overdiagnosis associated with adjunct screening and it would seem that given current data availability only approaches using method 3 would be possible.

When considering primary screening with mammography the presence of early detection and overdiagnosis imply that any single screen detected cancer must be one of the two types, that is

**Figure 1:** Schematic of Usual Model for Cancer Screening



**Figure 2:** Schematic for Overdiagnosis Model of Cancer Screening



S1: Early detection of a cancer which would be symptomatically diagnosed at a later time in the absence of primary screening, or,  
S2: A cancer which would not be diagnosed at a later time (overdiagnosis) in the absence of screening.

When considering screening supplementary to mammography a cancer detected at a supplementary screen can be one of three types:

SS1: Earlier detection of a cancer which would be symptomatically diagnosed at a later time and not be identified by subsequent mammography screening

SS2: Detection of a cancer which would not be diagnosed at a later time (supplemental - overdiagnosis) in the absence of supplementary screening.

SS3: Earlier detection of a cancer which would be screen detected at a later time by subsequent mammography screening

Just as for the situation when considering overdiagnosis and mammography screening, it is not possible to know which category an individual cancer diagnosed at supplementary screening belongs too. When examining data on cancer detection by supplemental screening it is not possible to determine how the cancers are distributed across the three categories given above.

Given an absence of direct evidence of the effect of supplementary screening on overdiagnosis is there any indirect evidence? One area of indirect evidence is comparing groups which have received more mammography screening with those receiving less to see if “extra mammography screening” results in extra overdiagnosis. Although no trials have directly compared more versus less screening with mammography, several RCT’s involved eligibility for population screening, in both arms, at the conclusion of the trial screening in the experimental arm. In an analysis published in 2005 (59) Moss finds that overdiagnosis disappears in RCT’s eligible for follow-on population mammography screening in both arms. In the most recently conducted RCT of mammography screening (60) for which data is available, there is no evidence of overdiagnosis in the cohort screened between the age 40-49 in the trial (compared to none) where both arms were eligible for population screening at age 50 (61). In particular the cumulative incidence of in-situ breast cancers, a sub-type believed to be the most susceptible to overdiagnosis by mammography, is doubled in the intervention arm at the conclusion of trial screening but by year 20 (when both arms have been eligible for population screening for 10 years) the rates have equalized. It would appear that the majority of evidence would indicate that extra screening, using a test for which overdiagnosis is accepted to occur, results in no increase in overdiagnosis. This seems to align with most simulation models of breast cancer which assume that overdiagnosis accrues because of long-lead times or that there is a reservoir of low malignancy potential breast cancers which are a potential source of overdiagnosis. However, it must be recognized that the preceding comparison is based upon “extra” mammography screening only and supplementary screening by another technology may have different results.

RCT’s of mammography screening were able to correlate reduced breast cancer mortality with a more favourable prognostic profile for screen detected cancers versus symptomatically detected



cancers (48) which is compatible with the classical model of screening. The presence of overdiagnosis complicates this relationship but does not negate it. However, the presence of a third category when considering supplemental screening, cancers detected by supplemental screening which would otherwise be detected by subsequent mammography screening, does introduce an added level of complexity. Consequently, while supplementary screen detected cancers must have a good prognostic profile to have the potential to provide long-term benefits it will not be sufficient to guarantee they do.

Irwig et al (62) discuss what evidence is required for the introduction of new screening tests for breast cancer. Although not explicitly stated their presentation seems to focus on replacement tests, that is, what evidence is required for a test being considered to replace mammography. They indicate that for tests, such as mammography, where early detection has been shown to result in mortality reductions and that the new candidate test is similar in nature to the accepted one (they give the example of digital and analogue mammography) that it may be sufficient to use a dual test cross-sectional reading study design. Contemporary examples of similar tests may be a comparison of MRI versus abbreviated MRI or comparing breast tomography plus digital mammography to breast tomography with synthetic 2D. In such a design subjects receive both tests and detected cancers are compared and contrasted. However, for supplemental screening the added nature of the new test means that there is really no comparison between two tests being made but that an extra test is being evaluated in a sub-population: those negative upon mammography. Supplemental tests can only increase cancers detected, false-positive rates and possibly overdiagnoses. In such a situation their recommendation for cases where screening is known to be of benefit, but the new test is not similar to the existing test would seem to apply (62): that a short-term RCT comparing the alternatives (supplementary screening versus none) is required.

In their discussion of a short term RCT Irwig et al (62) stress that the primary outcome of interest is interval cancers where one is looking to see a reduction in interval cancers with supplementary screening, particularly those with a poor prognostic profile. For primary screening there are only two possibilities for a screen detected cancer, as discussed earlier (S1 or S2 above). But for supplemental screening there are three possibilities (SS1, SS2 or SS3 above). Thus, while an improvement associated with reduction in interval cancers is reasonable to require this may not measure the whole benefit, as earlier detection by supplementary screening of cancer destined to be screen detected by primary mammography screening at a later time may also offer some benefit.

A further issue not discussed by Irwig (62) is the prevalence effect of a first screen. Screening with mammography for the first time frequently shows a “prevalence” effect with increased cancer diagnoses and false-positives when compared to rates at subsequent screens. This is recognized and results in different performance targets for first and subsequent mammography screens (44). First use of supplementary screening in a population may also behave in a similar manner with increased cancer detection and false positives at its initial application. This possibility implies that an RCT of adjuvant screening should consist of at least two rounds with sufficient subsequent follow-up to resolve potential overdiagnoses from early detection.

We may create an equation as follows:

$$\# \text{ screen detected by SS} = \#SS1 + \#SS2 + \#SS3 \text{ [1]}$$

where, # indicates the number in each category defined for supplemental screening earlier. With no comparison group that were not exposed to supplemental screening it is impossible to determine how the screen detected break down into the component subgroups. However, we can deduce some things from Equation [1]. If we let R refer to rate so that  $RS1$  is the rate of interval cancers following primary screening with mammography (as defined earlier), etc. then if

$$R \text{ screen detected by SS} > RS1, \text{ [2]}$$

this implies that  $RSS2 + RSS3 > 0$ , that is, some of the cancers screen detected by supplemental screening are either overdiagnoses or earlier detection of cancers which, in the absence of supplemental screening, would be screen detected by subsequent mammography. Using the relationship [2] in practice is not clear because of the uncertain choice of the time period for interval cancer calculation. Also, as remarked above there may be a difference in the quantity  $R$  screen detected depending on whether it is based upon a first screen or subsequent screens. When based upon subsequent screens the screen detected rate at supplementary screening will not exceed the interval cancer rate from mammography unless SS2 or SS3 occurs. For example, the screen detection at annual (biennial) supplementary screening cannot exceed the annual (2-year) interval cancer rate following mammography screening. While there may be extra benefit from the earlier supplemental screening mediated detection of cancers detectable by future mammography (SS3) this does have implications for the effectiveness of mammography since it implies that future detection rates by mammography will decline.

When examining data on cancers detected as a result of supplemental screening one is presented with a mixture of the 3 SS types with no knowledge about the various proportions in the mix. Thus, data on screen detection by supplemental screening provides no clear picture of potential benefit. It must exist for supplemental screening to have benefit, but its existence does not guarantee benefit. Examining the prognostic profile (e.g. stage, size, etc.) of cancers detected by supplemental screening provides some indication of the potential for benefit. If the profile of such cancers is similar to that of interval cancers following negative mammography, then benefit is unlikely. If the prognostic profile of cancers detected by supplemental screening is good, then there is a potential for long term benefit. However, since the supplemental screen detected cancers are composed of the three types, in unknown proportions, the magnitude of any associated benefit will not be clear from examination of those cancers alone. Consequently, it is necessary to undertake careful “accounting” to compare cancers detected at supplemental screening with their counterfactual counterparts in the absence of supplemental screening. This can only be convincingly done within the context of a randomized control trial as described earlier.

In conclusion, to understand the effect of supplemental screening it is necessary to have randomized trial evidence which provides data on the effect of supplemental screening on interval cancers and mammographic cancer detection beyond the first supplemental screening

round. Data on cancer detection by supplemental screening especially at the first round provides evidence for potential effectiveness but not for actual effectiveness. The following quantities need to be determined from research on supplemental screening:

1. Rates of cancer detection by supplemental screening in initial and subsequent rounds of screening
2. Interval rates of cancer in those receiving and those not receiving supplemental screening
3. Rates of cancer detection by primary screening (mammography) by screening round
4. Prognostic profile of all cancers detected in those receiving and those not receiving supplemental screening

When the above information is available it will then be possible to predict long term benefits and costs using a validated simulation model.

## 5.0 Intervention Following a Dense Breast Classification

Assuming the communication of breast density to an individual woman, a reasonable question is to ask what, if anything should follow? Most published discussion centres around supplemental imaging of one form or another. However, several possible options exist:

1. The communication of appropriate age-specific information alone
2. Provide individualized breast cancer risk assessment
3. Provide advice/interventions for breast cancer risk/breast density reduction
4. Increase the frequency of mammography screening (done in some Canadian provinces), or double read original screening mammograms
5. Recommend supplemental screening following a negative mammogram

The object of Options 1 and 2 is to provide information to place a woman's density category in a wider context. More detailed assessment will indicate the magnitude of personal risk in relationship to the average woman in British Columbia. Option 3 has some appeal, especially among some women disposed towards preventive actions. For example, moderate physical exercise has been found to reduce the risk of both pre- and post-menopausal breast cancer by approximately 20% (63) and reducing alcohol consumption may lower breast density (64). However, primary preventive measures are limited. For example, reduction in fat intake in a randomized control trial conducted in British Columbia was not associated with declines in breast cancer risk (65) probably because carbohydrate intake increased. Chemoprevention is known to be effective to reduce the risk of breast cancer. Tamoxifen, in the preventive setting, has been found in randomized control trials to result in a significant 22% (66) to 29% (67) reduction in risk of subsequent breast cancer but no associated reduction in breast cancer deaths. Analysis of one of these trials showed that Tamoxifen reduced breast density and that reduction in density in the first 12 months following starting Tamoxifen was associated with reduction in breast cancer risk (68). Exemestane has been found to result in a significant 65% reduction in breast cancer incidence in one preventative randomized control trial but again no reduction in deaths (69). Increased frequency of mammography (Option 4) has the appeal of simplicity from a program perspective since the facilitating mechanism (letter recall) and approach (mammography) already exists within the program. Although this may be an intuitive approach it is not clear that this provides significant advantage since the underlying problem is unchanged: mammographic discovery of lesions in areas of breast density. Analysis of the BC data did not find that annual mammography was associated with a lower proportion of interval cancers in women with the densest breast (BIRADS D). Also, early recall for mammography does depend on the stability of classification of such women. If this classification is unstable then a proportion of women previously classified as dense will not be classified as dense at early recall and vice-versa. Second reading of mammograms has been found in a number of studies (70-74) to increase breast cancer detection, and is standard practice in most European breast screening programs, but its utility in women with dense breasts is unclear. Potentially double reading could be used to improve consistency of density assignment prior to any referral for supplementary screening. Supplemental screening (Option 5) following a negative mammogram classified as dense has the practical advantage that intervention is immediate following the time that the density classification is made. Superficially a recommendation for further imaging is not

dissimilar to that made in most cases when the mammogram leads to the identification of a suspicious area within the breast: refer for further imaging.

As for any screening, supplemental screening should only be undertaken if there is a demonstration that the benefits outweigh the risks. The principal targeted benefit is the reduction in mortality from breast cancer and the risks are false-positives and overdiagnosis. D’Orsi and Sickles (75) discuss the use of definitions commonly used in discussing mammography screening when applied to other screening modalities. While false-positives may be estimated from a variety of studies mortality reductions and overdiagnosis may only be reliably estimated from sufficiently large and lengthy randomized control trials (RCTs) as discussed earlier. While some RCTs of supplemental screening are ongoing, none are of sufficient power to measure anything but very large changes in breast cancer mortality or overdiagnosis and the previously discussed (Section 4.0) approaches are required.

Three major modalities are considered as potential supplementary (to digital mammography – DM) modalities for screening women with dense breasts: breast ultrasound (BU), breast magnetic resonance imaging (MRI) and breast tomography (BT). BU and MRI may both be considered as potential supplementary screening approaches whereas BT, and in some studies MRI, is viewed more as a replacement for DM when synthetic 2-D images are co-produced or perhaps without such images.

Several organisations have considered supplemental screening. The International Agency for Research in Cancer (76) in 2015 updated its handbook on breast cancer screening and concluded that evidence was lacking for the effect of supplemental BU in women with dense breasts and for both BT for all women and MRI for high risk women. In 2015 the American College of Obstetricians and Gynecologists’ Committee on Gynecologic Practice (77) published that “the College does not recommend routine use of alternate or adjunctive tests to screening mammography in women with dense breasts who are asymptomatic and have no additional risk factors”. Melnikow et al (23) published a review of evidence for the US Preventive Services Task Force (USPSTF) and one of the questions to be addressed was the following:

When performed after a negative mammogram in women found to have dense breasts, what is the effectiveness of supplemental screening with breast ultrasonography, MRI, or breast tomosynthesis on proximate clinical outcomes, including cancer detection rates, DCIS detection rates, stage at diagnosis, recall rates, biopsy rates and interval cancer rates?

The review committee noted that no studies have addressed the effect of supplemental screening on breast cancer morbidity and mortality. The committee identified studies meeting minimum quality criteria which provided information on the outcomes in key questions. They report substantial variability, within modalities, of cancer detection rates, recall rates etc. This is partially due to the heterogeneity of the patient populations included in the studies and makes the application of such rates to BCCBSP of unknown accuracy. The resulting USPSTF report (78) concludes “that the current evidence (mid 2015) is insufficient to assess the balance of benefits and harms of adjunctive for breast cancer using BU, MRI or BT, or other methods in women identified to have dense breasts on an otherwise negative screening mammogram”.

In the sections that follow we will review each modality separately: Breast Ultrasound, Breast Tomosynthesis and Breast Magnetic Resonance Imaging. Guidelines have been developed for the reporting of diagnostic accuracy tests (79), which include tests used for screening, however, these guidelines have not been widely adopted by authors or journals. Many of the studies which are summarized in the reviews provide statistics described as sensitivity and specificity of the respective tests without definition of the clinical reference standard used (79). Given an absence of a gold standard for the presence of breast cancer, the resulting statistics do not represent classical sensitivity or specificity. Furthermore, different formulations for calculating these statistics are used so that the resulting quantities are seldom comparable across studies. Some studies only report on women following negative mammography. Consequently, we present study detection specific rates etc. which, although no more comparable, are well defined and commonly understood. No distinction will be made on the BIRADS version used in classifying density.

## 5.1 Breast Ultrasound (BU)

Scheel (80) reviews hand held and automated breast ultrasound in women with dense breasts. In their review of published research, they note that patient populations are variable and results on detection rates, biopsy rates etc. are correspondingly varied. In their review they identify no studies which measured interval cancer rates. They also report none of the 8 national United States medical societies they identified recommended the routine use of ultrasound for screening in any population.

In reviewing studies of BU here, results for hand held (HHUS) and automated (ABUS) are not separated and no attempt was made to differentiate results between these two technologies. A single randomized trial was found where ultrasound was included with mammography in primary screening and compared to mammography alone, the J-START trial (81). This study (J-START) will be discussed separately below and we will discuss here the studies where there are no randomized controls that are summarized in Table 5.1. Studies differed in their inclusion of ultrasound with some using it in women following a negative mammography screen and some combining mammography and ultrasound so that all women received both. Where both technologies were used the patient population usually included women whose breasts were not considered dense: wherever possible results presented in Table 5.1 are restricted to those with dense breasts.

All the studies reported that the addition of ultrasound resulted in significant levels of ultrasound (only) detected cancer, where significant represents at least 50% of the rate detected by mammography (where reported) from which the study cohort was drawn (Table 5.1). Reported cancer detection rates by ultrasound in mammography negative episodes varied from a low of 0.4 per 1,000 examinations (82) to a high of 25.2 per 1,000 (83). Where information on mammography detection was available the detection rate by supplementary ultrasound was typically  $\geq 50\%$  of that reported for primary mammography screening in the same study population. Most studies reported results from a single screening episode with BU or where multiple episodes were included did not report them separately. Exceptions were the studies by Berg (84) and Weigert (85). In both these studies detection rates at the first screening episode seemed higher (potential prevalence effect) although the Weigert (85) study does not follow individual women but reports cross-sectional outcomes in successive years following the adoption of breast density legislation in Connecticut. In the trial reported by Berg, ACRIN 6666, (84) the rate of abnormal ultrasound examinations drops in the second and third rounds. In the ACRIN 6666 trial (84) the rate of cancer detection by mammography does not seem to change across screening rounds and ultrasound detects cancer at about half the rate of mammography in rounds 2 and 3. Across the studies recall rates vary from a low of 1.4% (86) to a high of 15.1% (84), with the latter being that in the first round of ACRIN 6666. Several authors report on interval cancer rates (83,84,86-89) following supplementary ultrasound however most of these studies provide no comparisons in the absence of BU. One study which does (88) provides interval cancer rates in women without BU who do not have dense breasts and finds that the rate is similar to those in women with dense breasts receiving BU supplemental screening. The second (89) compares interval cancer rates before and after the inclusion of ultrasound in routine screening of women with dense breasts and find that the rates decline after ultrasound was introduced. However, the extra detection associated with the use of ultrasound, 7.6 per 1,000

(12.3-4.7) is much greater than the decline in the interval cancer rate, 1.2 per 1,000 (1.5-0.3) suggesting that many of the extra cancers detected would not have been interval cancers in the absence of BU. Given the observational uncontrolled nature of the study by Giuliano (89) it would not be appropriate to attach much weight to the magnitude of the forgoing values. Many of the studies report on the prognostic profile of ultrasound detected cancers and generally they seem to be comparable to those detected by mammography (e.g. see (80)) although fewer in-situ cancers are detected. The lack of a comparison group makes it difficult to assess the significance of these observations.

Collectively these studies confirm that BU detects cancers in women with dense breasts who have negative mammography. The cancers detected have a prognostic profile which is comparable to those detected by mammography although in-situ cancers are less common amongst cases ultrasound detected only by ultrasound. The rate of cancer detection by ultrasound exceeds the reduction in future interval cancers so that the cancers detected include future screen detectable (by mammography) cases and overdiagnoses in unknown proportions.



**Table 5.1:** Summary Outcomes for Studies Utilising Supplementary Screening with Ultrasound\*

Study	Design	Population	Definition of Dense	Cancer Detection Rate/ 1,000	Recall Rate - %	Biopsy Rate - %	Interval Cancer Rate /1,000
Berg 2012 ACRIN 6666 (84)	Cohort series – 3 annual screens HHUS	High Risk Population age>25 2,662 women	BIRADS C or D in at least 1 quadrant	Round 1 US   M- = 5.9 M= 7.5  Round 2 US   M-=4.0 M= 6.4  Round 3 US   M-=4.2 M=9.9	Round 1 US   M- = 15.1% M=11.5%  Rnds 2-3 US   M- = 7.4% M = 9.4%	Round 1 US   M- = 7.8% M=2.4%  Rnds 2-3 US   M- = 5.0% M=2.0%	US- & M- Round 1 0.85  Round 2 1.8  Round 3 1.4
Brancato 2007 (82)	Series HHUS	All ages 5,227 women	BIRADS D3 or D4	Combined M- only US   M- = 0.4	US   M- = 2.1%	US   M- = 0.6%	NR
Brem 2014 SomoInsight (90)	Series ABUS	Self-referral >25 15,318 women	BIRADS D3 or D4	US   M- = 2.3 M= 5.4	US   M- = 15.8%  M =15.0%	US   M- = 4.4%  M =4.0%	NR
Corsetti 2011 (88)	Series HHUS	Presenting for screening – may include symptoms 7,224 screens	BIRADS D3 or D4	Rounds are combined US   M-=4.4 M= 2.8	US   M- =5.5%	US   M- =5.5%	US- & M- 1.1 Compare to 1.0 in women D1 or D2 w/o US
Girardi 2013 (91)	Series HHUS	Presenting for screening Asymptomatic 9,960 women	BIRADS C & D M-	US   M-=2.2	As biopsy	1.9% (all densities)	NR

**Table 5.1:** Summary Outcomes for Studies Utilising Supplementary Screening with Ultrasound\* (cont)

Study	Design	Population	Definition of Dense	Cancer Detection Rate/ 1,000	Recall Rate - %	Biopsy Rate - %	Interval Cancer Rate /1,000
Giuliano 2013 (89)	Before /after Retro ABUS	Asymptomatic – no high risk Control=4076 Interv=3,418	>50% with Wolf	US & M =12.3 M= 4.7	NR	NR	US & M = 0.3 M +=1.5
Grady 2017 (92)	Retro Series ABUS	Asymptomatic – high risk 5,638	BIRADS C & D + elevated risk	US   M=4.2 M= 7.4	US   M- = 3.2% M= 11.4%	NR	?
Hooley 2012 (93)	Retro Series HHUS	Diagn & Screening	Mixed risk C & D	Results Not Summarized			
Kelly 2010 (87)	Series	Asymptomatic 6,425 exams on 4,419 women	mixed risk Includes BIRADS D3 and D4	US   M=4.6 M= 4.4	US   M- =7.2% M= 4.2%	US   M- =1.5% M= 1.2%	USM = 2.3
Tagliafico <sup>3</sup> 2016 ASTOUND (94)	Double read with Tomo and HHUS	Mammo Neg screenees, Italy >38yrs 3,231	BIRADS C & D	US   M= 7.1	US   M- = 2.7%	US   M- = 1.5%	NR
Venturini 2013 (95)	HHUS + MRI	40-49 year olds – high risk 1,666 women	All densities	US   M= 2.4 M=7.2		US   M- =0.9% M=0.1%	NR

<sup>3</sup> Compares Ultrasound and Tomosynthesis

**Table 5.1:** Summary Outcomes for Studies Utilising Supplementary Screening with Ultrasound\* (cont)

Study	Design	Population	Definition of Dense	Cancer Detection Rate/ 1,000	Recall Rate - %	Biopsy Rate - %	Interval Cancer Rate /1,000
Weigert 2017 (85)	Series M-, Conn. HHUS	All comers 1 <sup>st</sup> year: 2706 2 <sup>nd</sup> year: 3351 3 <sup>rd</sup> year: 4128 4 <sup>th</sup> year: 3331	BIRADS C & D	1 <sup>st</sup> year =4.1 2 <sup>nd</sup> year = 2.7 3 <sup>rd</sup> year = 2.7 4 <sup>th</sup> year = 3.0	1 <sup>st</sup> year = 4.1% 2 <sup>nd</sup> year =5.4% 3 <sup>rd</sup> year =3.6% 4 <sup>th</sup> year =1.6%	Same as previous column? All 4 & 5's rec for biopsy	NR
Wilczek 2016 (86)	ABUS	40-74 screenees 1,668 women	BIRADS C & D	US   M-= 2.4 M=4.2	US   M- = 1.4% M=1.4%	US   M- = 0.7% M=0.7%	3.1
Youk 2011 (83)	HHUS	21-74 High risk Asian 1,507 Exams	BIRADS D3 and D4	US   M- = 25.2	US   M- = 7.8%		3.3

\* M refers to a result with mammography, M- refers to negative mammography, US | M- refers to a result with ultrasound in women with negative mammography, US & M refers to a result for US and M combined

### 5.1.1 The J-START Trial

The only randomized control trial of adding ultrasound to digital mammography screening is the J-START trial (81). The randomized trial format provides the opportunity to apply the framework described in the earlier section (Section 4.0). The J-START randomized trial compares mammography plus clinical breast examination (CBE), the control arm, to mammography plus ultrasound plus CBE, the intervention arm, in Japanese women aged 40-49 (96). Results from the first (of two) planned biennial rounds have been published with 24-month follow-up following the first round including identification of interval cancers in the more than 70,000 women randomized in this trial. Although Japanese women are known to have a lower risk of breast cancer than North American women this relates primarily to post-menopausal breast cancer so that this cohort should not be greatly dissimilar in terms of risk to British Columbia women of the same age (97). The major limitation of this study, for the purposes of this review, is that they do not limit entry to women with dense breasts and have not yet reported results by density. They do report (81) that 58% of women enrolled in the trial have dense breasts which would not be dissimilar to the proportion for British Columbia using BIRADS C or D as dense. Consequently, the results published to date would likely be more representative of all BC women aged 40-49 rather than those with high density. Also, the inclusion of CBE makes this unrepresentative of usual breast screening in British Columbia. Despite these limitations this study represents the best source of information on the effect of incorporating ultrasound into routine screening based on mammography.

Some results are summarized in Table 5.2. The cancer detection rate in the control arm (mammography plus CBE) of the J-START trial is 3.3 per 1,000 which lies between the British Columbia mammography detection rate on first screens of 3.8 per 1,000 and on subsequent screens of 2.4 per 1,000 (35). The addition of ultrasound results in a detection rate of 5.0 per 1,000 in the intervention arm of the trial, an increase of 1.8 per 1,000 compared to the usual screening arm (81). Based upon the intervention arm the addition of ultrasound finds 61 further cancers, which corresponds to a relative increase in detection of 50% (57% of cancers detected by CBE are excluded) which is similar to the 54% increase following ultrasound screening seen in the ACRIN 6666 trial (84). The false-positive rate in the intervention arm was 12.8% and was 8.6% in the control arm, so that the addition of ultrasound increased the false positive rate by 4.2%. The false positive rate in the control arm (mammography plus clinical breast examination), 8.6%, was lower than the abnormal call rate on first screens, 17.5%, or subsequent screens, 9.9%, among women 40-49 reported by BCCBSP for 2015 (35).

The 24-month interval rate of invasive cancer in the control arm of J-START was 0.75 per 1,000 and if cases detected by CBE only are counted as interval cancers the rate increases to 0.95 per 1,000. This is somewhat lower than the rate of invasive interval cancer found at 24 months in the analysis of British Columbia data (Appendix Table 3) which was 1.4 per 1,000 among women 40-49. The number of interval cancers in the intervention arm was 18 versus 35 (43 if those found by CBE only are considered interval cancers) in the control arm so that the number of interval cancers following ultrasound has declined to 51% (42%) of the value without ultrasound. Corsetti (88) in their observational study found that the rate of interval cancer in

women with dense breasts (C & D) following ultrasound was 10% higher than women with non-dense breasts who did not receive ultrasound. Using rates derived from the British Columbia analysis (Appendix Table 3) this would correspond to a decline to 52% of the interval cancer rate (without ultrasound) which is similar to that seen in the Japanese trial.

**Table 5.2:** Screen Detected and Interval Cancers for the two arms of the J-START Trial (81)

Modality*	Number Randomized	Screen Detected (Rate /1,000)	Interval within 24 months (Rate /1,000)	Total (Rate /1,000)
Intervention (M, US, CBE)	36,752	184 (5.0)	18 (0.49)	202 (5.5)
Control (M, CBE)	35,965	117 (3.3)	35 (0.97)	152 (4.2)
Difference	-	67 (1.8)	-17 (-0.48)	50

\*M=Digital Mammography, US=ultrasound (HHUS), CBE=Clinical Breast Examination

The J-START trial confirms the pattern in additional screen detected cancers which seemed to be evident or inferred from the studies discussed in Table 5.1, that is

- The addition of ultrasound increases cancer detection
- The interval cancer rate declines following ultrasound detection
- The magnitude of the decline in interval cancers is smaller than the increase in screen detected cancers in the first cycle

Following the earlier discussion (Section 4.0) the net difference in cancers must be a composite of future (mammography) interval cancers, future mammography detectable cancers and overdiagnoses.

Some further perspective can be identified by examining stage distribution of the cancers, see Table 5.3, which contains data extracted from the trial publication (81). Table 5.3 shows that the stage distribution of those detected by ultrasound only is favourable with a lower proportion of stage II+ than detected by mammography in either of the trial arms. However, the interval cancers show a different trend with those in the control arm having a better prognostic profile than in the intervention arm (Table 5.3). When examined overall the proportion of stage II+ cancers in the Intervention Arm, 23%, is lower than that in the Control Arm, 32%. Using proportions does not give the whole picture as the denominators and numbers of cases in each group is different. Using stage II or worse (II+) the framework (Section 4.0) presented earlier would indicate that a better picture can be obtained by examining rates, which given the randomization is equivalent to examining counts. For example, the number of stage II+ is 3 (48-45) less in the Intervention Arm than in the Control Arm with 34 more Stage I cancers and 16 more in-situ cancers (with 3 cases of unknown stage).

**Table 5.3:** Stage Distribution of Screen Detected and Interval Cancers for the two arms of the J-START Trial by Mode of Detection and Test (81)

Arm	Mode of Detection	TEST	Number of Cases	Stage		
				0	I	II+
Intervention	Screen Detected	M or CBE	122*	45 (37%)	52 (43%)	25 (20%)
	Screen Detected	US not M or CBE	59†	9 (15%)	39 (66%)	11 (19%)
	Interval	-	18	1 (6%)	8 (44%)	9 (50%)
	TOTAL		199	55 (28%)	99 (50%)	45 (23%)
Control	Screen Detected	M or CBE	117	31 (26%)	48 (41%)	38 (32%)
	Interval	-	35	8 (23%)	17 (49%)	10 (29%)
	TOTAL		152	39 (26%)	65 (43%)	48 (32%)

\*1 case had missing stage and was omitted from this table

†2 cases had missing stage and were omitted from this table

It is possible to undertake a more refined analysis of the first-round results from the J-START trial, while acknowledging that such an analysis is provisional depending, as it does, on the results of the first (prevalent) round incorporating ultrasound. First, we note that the contribution of mammography to screen detection alone should be the same in the two arms, but the actual contribution will be subject to stochastic variation. Consequently, removing mammography detected cancers from both arms (in the first round) would remove one source of random variation between the arms. It can be seen in Table 5.3 that in both arms the number detected by mammography or CBE is similar: 117 v 123, although their stage distribution is not as similar. Secondly, in British Columbia, CBE is not part of programmatic breast screening so that cases detected by CBE only, in either arm, would not be detected under normal program screening in British Columbia. We will assume that all cases detected by CBE only would occur as interval cancers before the next screen: there are 8 such cases in the Control Arm and 0 in the intervention arm (6 cases had positive CBE and negative mammography but all 6 had positive ultrasound). Consequently, this corresponds to assuming that there are 43 interval cancers in the Control Arm with 18 remaining in the Intervention Arm. Of course, the stage at diagnosis for the 8 CBE detected cancers in the control group will likely be different than observed if they had not been screen detected. We will assume that they have the same stage distribution as the 35 true interval cancers in the control group as given in Table 5.3.

**Table 5.4:** Adjusted Counts by Stage of Screen Detected by Ultrasound in the Intervention Arm and Estimated Interval Cancers for the J-START Trial (81)

		Stage			Total
		0	I	II+	
Intervention Arm	US Detected Cancers in M-Interval Cancers	13* (19%)	42 (63%)	12 (18%)	67
	Interval Cancers	1 (6%)	8 (44%)	9 (50%)	18
	Total	14 (16%)	50 (59%)	21 (25%)	85
Control Arm	Interval Cancers	10 (23%)	21 (49%)	12 (28%)	43†

\*2 cases had missing staging and were assigned to be stage 0

†includes 8 detected by CBE only, stage distribution is estimated from the 35 true interval cases

Table 5.4 shows the results of the calculation outlined in the previous paragraph: mammography detected cancers are removed, CBE only detected cancers are assumed to be interval cancers and have the same stage distribution as other interval cancers. This estimation indicates that the effect of adding ultrasound to mammography in the first round of screening, including follow-up, would be to reduce the interval cancer rate by 58% with a net increase in the number of DCIS, stage I and of stage II+ cancers diagnosed. Given the number of ultrasound detected cancers, 67, versus the estimated reduction in the number of interval cancers, 25, it seems unlikely that all the extra cancers detected by ultrasound would have been future interval cancers since this would imply a lead-time for ultrasound exceeding that for mammography. In the absence of data from future rounds, we can estimate how the addition of ultrasound would translate into added detection by making the simplifying assumption that it would be equal to the reduction in interval cancer for subsequent rounds of screening. From the analysis of BC data (Appendix Table 4) for age 40-49 screened biennially the interval cancer rates (per 1,000) are currently by density, C=1.73, D=2.53 and for age 50-74 C=2.54, D=3.12. Multiplying the preceding values by 58% indicates that the projected rate of added screen detection (prevented interval cancers) would be (per 1,000) for age 40-49 C=1.00, D=1.46 and for age 50-74 C=1.47, D=1.81. Consequently, if the primary purpose of adding ultrasound it to reduce interval cancers, then the yield per ultrasound performed will be generally below that of primary mammography whose lowest rate of screen detection is 1.58 for women 40-49 with density A (appendix Table 4).

Comparison of Tables 5.3 and 5.4 indicates a somewhat different effect of supplementary BU on the stage distribution of cancers in the two arms. Table 5.3 indicates increases in the frequency of stage 0 and I cancers with a small reduction in stage II+s in the BU arm. Table 5.4 indicates increases in the frequencies of cancers of all stages in the BU arm. This discrepancy arises from the dissimilar stage distribution of mammographically detected cancers in the two arms where such cases are included in Table 5.3 but not in Table 5.4.

In analysing the effect of supplementary ultrasound on interval cancers we have included all cancers regardless of stage. However, survival of ductal carcinoma-in-situ and stage I breast cancers are extremely high, both exceeding 95% at 5-years (52), and earlier supplementary screen detection of such future interval cancers will be of modest net benefit. Consequently, it is more important to concentrate on the reduction in more advanced cancers, e.g. stage II+. Examination of Table 5.3 shows that amongst the interval cancers the number of stage II+ are reduced from 10 to 9 by the addition of supplemental screening with the addition of 11 screen detected cases. The predicted distributions, after removal of the effect of CBE diagnoses, shows a similar picture with an estimated net reduction of 3 interval Stage II's with 12 screen-detected. It should be noted that the small number of stage II+s makes inferences unreliable.

In summary, the J-START trial has confirmed earlier observational studies that supplementary ultrasound increases cancer detection and reduces the 24-month rate of interval cancer with increased detection being much greater than the reduction in interval cancers. The trial also shows that despite the favourable profile of the ultrasound detected cancers that the number of advanced stage (II or greater) interval cancers are not greatly reduced in the first cycle of supplementary ultrasound.

### *Simulation Study of Sprague*

One study (98) attempted to simulate the effect of supplementary ultrasound on outcomes (benefits, harms and cost-effectiveness) using 3 models of breast cancer developed within the CISNET consortium (99) which incorporate United States data and costing. This study examined cohorts from age 40 to end-of-life where women were screened with mammography at different frequencies (annual or biennial), for different age ranges (50-74 or 40-74) and with options for added ultrasound (none, BIRADS D only and BIRADS C & D only). The study found that supplementary ultrasound had higher cost-effectiveness when used for D alone (US\$246,000 per QALY) than when used for C & D combined (US\$325,000 per QALY). The study undertook sensitivity analysis (varied assumed parameter values) for the ultrasound sensitivity, specificity and cost but found this had little effect on the conclusions. The authors (98) did not report "sensitivity analysis" for the performance parameters for mammography and the assumed sensitivities are given in Table 5.5 below. Several aspects of the assumed sensitivities in the simulation study should be noted. First, all values are the same for BIRADS C and D. Secondly the lowest sensitivity rates are all for BIRADS A. In contrast C and D have quite different period sensitivities in the BCCBSP analysis and BIRADS A has the highest sensitivity. In the BCCBSP analysis BIRADS C is more similar to BIRADS B than it is to BIRADS D. Results from the BCCBSP analysis are in accord with other published studies, e.g. (100). Also, in the simulation study the sensitivity of ultrasound is lower than that of mammography in every density category although these may not be strictly comparable. While it is difficult to decompose results from simulations using such complex models, the assumed cost-effectiveness of ultrasound must be dependent on the false negative rates assumed for mammography. The high assumed sensitivity for mammography in BIRADS D will tend to reduce the potential effectiveness of ultrasound in such individuals so that the simulated estimates of cost-effectiveness in BIRADS are likely to be high (cost-per-QALY).



**Table 5.5:** Assumed Sensitivity for Mammography and Ultrasound in Simulation Study (98) and Period Sensitivity from BCCBSP Study

Source	BIRADS Density	Mammography				Ultrasound
		Age 40-49		Age 50-74		All Ages
		Annual	Biennial	Annual	Biennial	All
Sprague 2015 (98)	A	69%	76%	76%	82%	-
	B	82%	87%	87%	90%	-
	C	74%	80%	80%	85%	55%
	D	74%	80%	80%	85%	55%
BCCBSP	A	90%	95%	93%	95%	-
	B	67%	79%	89%	92%	-
	C	68%	80%	79%	85%	-
	D	51%	66%	44%	53%	-

## 5.2 Breast Tomography (BT)

Cole and Pisano (101) reviewed published studies on BT and report published detection rates, recall rates and aspects of technical performance related to different device manufacturers. They report that detection rates are increased, and false positives decreased.

While it has been suggested that Breast Tomography (BT) could be used in women with dense breasts most research has focussed on it as a supplement or replacement for full field digital mammography (DM) in all women. Following approved practice for inclusion in screening in the United States most research has compared DM to DM plus BT so that its use is as a supplementary modality. However, recent research has focussed on the creation and use of a synthetic 2D image (by BT) to replace DM so that only a single radiation exposure is required to produce a screening examination including 2D and 3D imaging or, in some cases, to use BT alone. The incorporation of BT holds the promise of increased breast cancer detection, via increased sensitivity, and a reduction of abnormal calls because of the ability to resolve, without further work-up, insignificant abnormalities which are seen by the projection of compressed 3D breasts onto 2D DM images. One would anticipate that BT would have increased sensitivity for the detection of tumours located in a region of low density but where regions of high density overlap the 2D images produced in DM. As BT relies upon the same method of imaging (X-rays) as digital mammography it will likely be subject to many of the same limitations.

Because of the focus of BT as a potential enhancement, or replacement, for DM most research has included women of all densities. The studies reviewed are summarized in Table 5.6. Results are given in this table by density, if the authors provided density specific results, however, only one study was restricted to women with dense breasts.

As for studies in ultrasound provided earlier, rates (detection, recall and biopsy) when BT was incorporated showed a wide variation which was likely related to differences in the patient populations as was reflected in detection rates by DM alone. Studies were principally cross-sectional studies where the contribution of DM and BT were apportioned in women receiving DM with BT, or before/after series where experiences were reported for periods prior to and after a switch to include BT in screening. No randomized trials have been reported with women randomized to include BT or not in DM screening although such a trial is currently recruiting in North America with centres in Canada, including British Columbia. All studies show an increase in screen detected cancers, which is not surprising since BT is added to DM, with a relative increase of ~50% and a range of 10-70%. For those studies where DM and BT interpretation are integrated, so that not all abnormal DM examinations are recalled, a reduction in the recall rate with a relative reduction that averages ~20% (of the rate for DM) and ranges from 10% to 57%. Studies reporting biopsy rates do not indicate much change by the inclusion of BT in screening compared to DM alone.

Individual studies provide unique pieces of information. McDonald (102) reported the experience following several cycles of screening with BT and found that there appeared to be a “prevalence” effect with first time addition of BT with cancer detection and recall rates declining

after first use. Tagliafico (94) reports interim results from the ASTOUND trial where women with dense breasts and negative mammography received both supplementary BU and BT. They report that approximately twice as many cancers were found by ultrasound: 7.1/1,000 versus BT 3.7/1,000 with a higher recall rate 2.7 versus 2.0% and a higher biopsy rate 1.5% versus 1.1%. Rafferty (103) reports detection by density with the addition of BT and finds that increased detection is highest for BIRADS C with a smaller increase in BIRADS D.

Three studies (102,104,105) report interval cancer rates for periods when screening used only DM and following the addition of BT to screening. In each of these studies the interval cancer rate was not reduced, despite substantial increases in the rate of screen detection with increased screen detection exceeding the mammography only interval cancer rate in two of the studies (102,104) and being 75% of the rate in the third (105).

**Table 5.6:** Summary Outcomes by Study Utilising Supplementary Screening with Breast Tomography

Study	Design	Population	Dense	Cancer detection Rate/1,000 screens	Recall Rate %	Biopsy Rate %	Interval Cancer Rate /1,000
Bernardi 2016 STORM 2	Reading M v M+BT v SynM +BT	>49 attending screening Italy C&D:2,592	C & D	M=7.7 M+BT=13.1 SynM+BT=13.9	M= 4.0% M+BT=5.0% SynM+BT=6.1%	NR	NR
Ciatto 2013 STORM	Reading M v M+BT (integrated)	Asymptomatic >48 C&D:1,215	C & D	M=4.1 M+BT=6.6 BT only=6.6	M=7.3% M or T=6.6%- conditional	NR	NR
Conant 2016 PROSPR	Before/After with multiple centres	40-74 Screenees – no prev. BC M:142,883 M+BT:55,998	All M:C+D =34% M+BT:C+D =40.2%	All M=4.4 M+BT=5.9	All M=10.4% M+BT=8.7%	All M=1.8% M+BT=2.0%	M=0.46 M+BT = 0.60
Friedewald 2014	Before/After Adding Tomo – 13 centres	Screening Pop? Average age 56-57 M:281,187 M+BT:173,663	All	Invasive M=2.9 M+BT=4.1	M=10.6 M+BT=8.9	M=1.8 M+BT=1.9	NR
Haas 2013	Centres with and w/o Tomo	Presenting for screening – high risk, C+D: M=2,158 M+BT=2,639	All	All: M=5.2 M+BT=5.8	C: M:16.7 M+BT=10.2 D: M:15.6 M+BT=6.7		

**Table 5.6:** Summary Outcomes by Study Utilising Supplementary Screening with Breast Tomography (Cont.)

Study	Design	Population	Dense	Cancer detection Rate/1,000 screens	Recall Rate %	Biopsy Rate %	Interval Cancer Rate /1,000
Houssami 2014 STORM again <sup>4</sup>	Cohort M v M+BT (integrated)	Asymptomatic >48 7,292	All	Integrated Double Read, All M=5.3 M+BT=8.1	Double, All M=4.5% M+BT =3.5%		M+BT=0.8
Lang 2016 Malmo Trial MBTST	Series One view BT v 2view DM	Screening Participants 40-74	All	C&D alone M=9.8 BT=13.7	All cases M=2.6% BT=3.8%	NR	NR
Maxwell 2017	2 rounds cross over RCT	High risk 40-49 screening – subsequent 1,227 608 & 619	All		M=2.4% M+BT2=2.2%	NR	NR
McCarthy 2014	Before/After Tomo	Screening – All comers C&D: M =3,489 MT =5,056	All	Inv Ca: C&D M= 3.4 M+BT=4.7	M=12.8 M+BT=10.8	NR	NR
McDonald 2016	Before/After Tomo with 3 years of data	Asymptomatic screening/no history of cancer M=10,728 M+BT1=11,007 M+BT2=11,157 M+BT3=11,576	All C=30% D=2%	C&D only M=5.2 M+BT1=6.6 M+BT2=7.4 M+BT3=8.6	All Densities M=10.4 M+BT1=8.8 M+BT2=9.0 M+BT3=9.2	All Densities M=1.8 M+BT1=2.0 M+BT2=1.9 M+BT3=1.9	All Densities M=0.5 M+BT1=0.7

<sup>4</sup> Found that double reading of M increased detection rate by 25%.

**Table 5.6:** Summary Outcomes by Study Utilising Supplementary Screening with Breast Tomography (Cont.)

Study	Design	Population	Dense	Cancer detection Rate/1,000 screens	Recall Rate %	Biopsy Rate %	Interval Cancer Rate /1,000
McDonald supplement	As above but restricted	Only include subjects with previous screen	All	Rates/1000 M+BT1 = 11.2 M+BT1   M=13.0 M+BT2   (M+BT1,M) =6.2 M+BT3   (M+BT1,M+BT2,M) =7.3	M+BT1 = 13.0 M+BT1   M=11.1 M+BT2   (M+BT1,M) =7.8 M+BT3   (M+BT1,M+BT2,M) =5.9		
Rafferty 2016	As Friedewald but by density			Invasive C: M=3.0 M+BT = 4.5 D: M=1.9 M+BT = 2.6	C: M=12.8 M+BT = 11.0 D: M=11.4 M+BT = 9.8		NR
Rose 2013	Pre/Post Tomo Start	Screening subjects! M=13,856 MT=9,499	All	Invasive M = 2.8 M+BT = 4.3	All M = 8.7 M+BT = 5.5 (C+D: M=10.6 M+BT=6.9)	All M = 1.5% M+BT = 1.4%	NR
Skaane et al 2018 OTST	Pre (2 rounds) v Post (1 round) – Post was voluntary	Norwegian Screening Subjects 50-69 Pre-59,877 Post-24,301	All	M = 6.3 M+BT = 9.3	M = 4.2 M+BT = 3.4	NR	2-year M=2.0 M+BT=2.1

**Table 5.6:** Summary Outcomes by Study Utilising Supplementary Screening with Breast Tomography (Cont.)

Study	Design	Population	Dense	Cancer detection Rate/1,000 screens	Recall Rate %	Biopsy Rate %	Interval Cancer Rate /1,000
Tagliafico <sup>5</sup> 2016 ASTOUND	Double read with Tomo and HHUS	Mammo Neg screenees >38yrs	C & D	BT   M- = 3.7	BT   M- = 2%	BT   M- = 1.1%	NR

---

<sup>5</sup> Compares Ultrasound and Tomo

### 5.2.1 Analysis of Data from the Oslo Tomosynthesis Screening Trial

The most recent publication using data from the Oslo Tomosynthesis Screening Trial (105) provides information on cancers and their prognostic profile identified during the trial as well as comparison data from the Norwegian Breast Cancer Screening Program which contributed participants to the trial. In order to provide an easy comparison between DM and DM+BT the results from the trial have been extended to a hypothetical cohort of 100,000 over a period of single screening round of 2-years (the extent of results available) and are summarized in Table 5.7. The intent of this table is to be able to compare appropriate results as laid out in the framework (Section 4.0). Since the interval cancer rate does not fall following the inclusion of BT, but extra cancers are screen detected, the overall number of cancers diagnosed increases, by an anticipated 3.2 per 1,000. Despite a lower rate of nodal involvement among screen cancers detected by DM+BT, compared to those detected by DM alone, the extra cancers detected results in an increase of the total number with positive nodes, with an estimated increase rate of 0.2 per 1,000.

**Table 5.7:** Projected Results Over a Two-Year Period of Adding Breast Tomography to Routine Biennial Mammography Screening in a Population of 100,000 Women Using Results from the Oslo Tomosynthesis Screening Trial and Norwegian Breast Screening Program (105)

	Screening Modality			
	Digital Mammography		Digital Mammography Plus Breast Tomography	
	Screen Detected	Interval	Screen Detected	Interval
Rate per 1,000	6.31	1.97	9.34	2.10
Expected Number (100,000 women)	631	197	934	210
Proportion with positive nodes	16.9%	36.4%	12.8%	37.3%
Expected Number with positive Nodes	106.6	71.7	119.6	78.3
Total Expected Cases	828		1144	
Total Expected Cases with +ve Nodes	178.3		197.9	

In conclusion, the inclusion of breast tomography with digital mammography screening has been repeatedly shown to reduce the rate of false positive screens, by about 20%, through the resolution of insignificant abnormalities. Similarly, the inclusion of BT has been shown to increase the detection rate of screening, by approximately 50%, of that detected by digital mammography alone. No reduction in interval cancers rates have been reported following the inclusion of BT in routine screening. Most studies were conducted in populations without regard



to breast density. Reported results in women with dense breasts do not appear to be different than that for all women. The literature on BT in screening is rapidly expanding, although it will undoubtedly be several years before any results from a randomized control trial are available.

### 5.3 Breast Magnetic Resonance Imaging (MRI)

Breast Magnetic Resonance Imaging has been primarily tested and used as a supplemental screening modality in those judged to be at sufficiently high risk of breast cancer. This is most frequently expressed as either having a proven genetic predisposition (typically BRCA1 or 2), a history of chest wall irradiation or an estimated lifetime risk of breast cancer in excess of 20-25% based upon a predictive risk model (15). The average woman with dense breasts (either C or D) in British Columbia will not have a lifetime risk greater than 25% so that MRI screening is not currently recommended in any jurisdiction for such women. Applying results from very high-risk populations to general screening populations with dense breasts is unclear. Table 5.8 provides some research findings in women at very high risk where MRI is used. Study participants are often young, subject to screening with MRI plus mammography plus or minus ultrasound on a 1-year screening interval, with potential interleaving of screening modalities (106). In these participants, it is clear that MRI is more sensitive than digital mammography, with or without ultrasound, with few cancers identified by these other modalities which are not identified by MRI. In these high risk women interleaved MRI and mammography was predicted to be the most cost-effective use of the two modalities (107) with a predicted cost per QALY of US\$74,200 however it is unclear whether the model predicted gains are accurate. Studies of screened cohorts provide varying observed survival benefits (108,109). Given that any gains in lower risk women (e.g. those with increased density) would almost certainly be lower than for these high-risk women, whose lifetime risks can be in excess of 50% (110), the frequency of MRI screening would need to be correspondingly lower without loss of relative effectiveness, to provide comparable cost-effectiveness.

Some research has investigated the potential for MRI screening in women at lower risk. ACRIN 6666 (84) included a sub-study of exiting women after the third round of combined screening of DM and BU (see Table 5.8 - Berg). That study identified a large detection rate of breast cancer by MRI, 17.5/1,000, in women negative by DM and BU. Contrasted with the interval cancer rate of 1.3/1,000 and mammography detection rate of 8.2/1,000 in earlier rounds of this study such a detection rate is equivalent to 2 years-worth of diagnoses. The rate of recall was very high at 26%. Despite having consented to participate in the MRI sub-study 42% of eligible women declined MRI when offered and those accepting were at elevated risk compared to those that declined (111). A second study of interest (112) screened a cohort of average risk women by adding MRI to DM plus or minus BU (see Kuhl in Table 5.8). In women DM ± BU negative, MRI had a yield of 14.2 invasive breast cancers per 1,000 in the first round and 5.7 per 1,000 in the second round. No interval cancers were observed in an estimated almost 5,000 years of follow-up. The abnormal call rate was much lower in the second round 1.6% vs 4.2% and is lower than in studies based on high risk women. It is clear from these studies that MRI is a very sensitive technology for identifying breast cancer and that few cancers identifiable by mammography are MRI negative at simultaneous testing. Current research using abbreviated protocols for breast MRI is examining whether MRI screening can be performed faster which, if successful, will reduce one barrier to its use in more general screening situations (113).

Should the feasibility of MRI screening be demonstrated in lower risk populations, the low additional yield by mammography seen to date, would suggest its likely use would be to replace DM rather than be a component of multimodality screening. The low uptake, even among motivated women, would seem to make it challenging for general use. Research into this technology for general screening applications is very early and longer-term follow-up of screened subjects is necessary before any assessment of efficacy could be made. In their assessment IARC felt that there was inadequate evidence for the use of MRI in high-risk women (76) and there is much less evidence for average risk women.

**Table 5.8:** Summary Outcomes by Study Utilising Supplementary Screening with Breast Magnetic Resonance Imaging

Study	Design	Population	Dense	Abnormal	Cancer detection Rate/ 1000 screen	Recall Rate %	Biopsy Rate	Interval Cancer
Berg 2012 ACRIN 6666 (84)	Cohort - MRI sub-study at BU screen 3	High Risk Population age>25 -612 women	C D in at least 1 quadrant	≥3 Integrated assessment	M=8.2 M&BU=11.3 MRI   M=19.6 MRI   (M&BU)- =17.5	M=8.5% M&BU=16.3% MRI =26.0%	NR	Afetr M&BU is 1.3/1,000
Kriege 2004 (114)	Cohort Annual Mammography/MRI	Very High Risk -1909 women mean age 40	All	≥3	M=4.3 MRI=7.7 MRI   M= 5.6 M   MRI- = 2.2	M=5.4% MRI=10.8%		5536 yrs FU 5 cancers observed
Kuhl 2017 (112)	Cohort – MRI+M±BU	Average Risk - 2120	All A (12%), B (27%) C (38%) D (22%)	≥4	1 <sup>st</sup> screen Inv MRI   M±BU- = 14.2 Subsequent Screens MRI & M ±BU MRI=5.7 Only 1 cancer seen on M		MRI   M±BU- = 4.2% Subsequent Screens MRI & M ±BU MRI=1.6%	Est. 4887 years FU 0 cancers observed

**Table 5.8:** Summary Outcomes by Study Utilising Supplementary Screening with Breast Magnetic Resonance Imaging (cont)

Reidl 2015 (115)	Cohort MUSMRI	Very High Risk median age 44 559 women for 1365 rounds	C & D's		M=9.3 BU=7.7 MRI=18.5 All Densities: First M&BU&MRI =30.5 Subsequent M&BU&MRI =23.5	M=4.0% BU=3.9% MRI=12.8% All Densities: MRI +ve round 1 v subsequent: 15.4% v 8.2%		1191 rounds: ~1 year per round: 1 interval cancer
------------------------	---------------	--	---------	--	--	---	--	---

## 6.0 Recommendations

The BC Cancer Breast Screening Program has indicated, by its decision to mandate BIRADS breast density assessment as part of the screening visit, that breast density is a significant factor in the provision of breast screening to the population of BC. Nevertheless, policies relating to breast density within the BCCBSP remain to be further developed. The following recommendations relate to the communication (Recommendation 1), measurement (Recommendation 2) and supplementary screening (Recommendation 3).

### Recommendation 1

Develop a plan to communicate breast density results in British Columbia that involves:

- d. a process to understand the communication needs of British Columbia screening participants, and their physicians, in relation to breast density;
- e. the use of BC Breast Screening Program data to develop breast density risk information that is relevant to the BC population;
- f. a review of existing information materials, in British Columbia and elsewhere, to develop messaging for the breast density risk factor in the context of other recognized factors that influence the likelihood of breast cancer.

Rationale:

1.a

- Best practice in the communication of breast density information is unclear
- Practice within Canada, the United States and elsewhere is heterogeneous
- Key informants expressed differing views on the desirability of various approaches to the communication of breast density
- User consultation and focus groups provide a mechanism for identifying the preferred community approach at this time

1.b

- The BCCSP maintains an excellent longitudinal database capturing information on breast density and other breast cancer risk factors
- This database allows the calculation of specific screening outcomes (disease detection, interval cancers, staging, false positives, etc.) for British Columbia screening participants
- Quantitative information available from the scientific literature is based upon varied patient populations
- Estimates in the scientific literature used a variety of statistical analytic techniques which do not permit straightforward generalisation to the BC population

1.c

- Existing breast density information is diverse and potentially confusing
- Research has indicated that women often over-estimate their breast cancer risks and the influence of various risk factors
- Breast density is one of many factors which influence breast cancer risk
- Breast density on its own is a poor discriminator of breast cancer risk

- Breast density influences the ability of mammography to identify breast cancer

## **Recommendation 2**

Continue to utilize the BIRADS density scoring within the BCCBSP, but continuously assess its performance and monitor the scientific literature for opportunities for improvement.

Rationale:

- BIRADS density is a commonly used clinical tool and is the most common scale used in North America
- BIRADS is subjectively assessed and inter-radiologist and consecutive measurements show variation
- BIRADS categories are clinically defined and may not be the most suitable for separating subjects based upon likelihood of developing breast cancer, or having breast cancer diagnosed before the next screening round
- Evolving mammography technology may result in changes in future performance of BIRADS
- Automated density assessment is an area of active research and future improvements seem likely

## **Recommendation 3**

Supplemental screening of women with dense breasts is not recommended at this time. The Breast Screening Program should monitor ongoing results of RCT's of supplemental screening in women with negative screening mammography.

Rationale:

- Breast Ultrasound (BU), Breast Tomography (BT) and Breast Magnetic Resonance Imaging (MRI) are all able to identify cancers in women following a normal mammography examination
- Increased detection is a necessary, but not a sufficient, requirement for benefit resulting from supplemental screening
- Use of supplemental BU and MRI will increase the number of women requiring further testing with an increase in the number of false-positive screens
- BU has the most supporting evidence for potential benefit having been demonstrated to reduce the rate of interval cancers
- In the single RCT of BU it had a modest effect on the likelihood of cancer being identified at an advanced stage, but this was based on results from the first round of supplementary screening using BU and further follow-up is required
- No guideline committee currently recommends supplemental screening with BU, BT or MRI on the basis of breast density alone
- No Canadian screening program currently recommends supplemental screening with BU, BT or MRI on the basis of breast density alone
- Further evidence is required to evaluate the benefits and harms of adding supplemental screening in selected mammographically negative women

## 7.0 References

- (1) Boyd NF, Martin LJ, Bronskill M, Yaffe MJ, Duric N, Minkin S. Breast tissue composition and susceptibility to breast cancer. *J Natl Cancer Inst* 2010 Aug 18;102(16):1224-1237.
- (2) Chiu SY, Duffy S, Yen AM, Tabar L, Smith RA, Chen HH. Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. *Cancer Epidemiol Biomarkers Prev* 2010 May;19(5):1219-1228.
- (3) Boyd NF, Martin LJ, Yaffe MJ, Minkin S. Mammographic density and breast cancer risk: current understanding and future prospects. *Breast Cancer Research* 2011;13(6):223.
- (4) Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* 2000;92(13):1081-1087.
- (5) Masala G, Ambrogetti D, Assedi M, Bendinelli B, Caini S, Palli D. Mammographic breast density and breast cancer risk in a Mediterranean population: a nested case-control study in the EPIC Florence cohort. *Breast Cancer Res Treat* 2017:1-7.
- (6) Price ER, Hargreaves J, Lipson JA, Sickles EA, Brenner RJ, Lindfors KK, et al. The California breast density information group: a collaborative response to the issues of breast density, breast cancer risk, and breast density notification legislation. *Radiology* 2013;269(3):887-892.
- (7) Nayak L, Miyake KK, Leung JW, Price ER, Liu YI, Joe BN, et al. Impact of Breast Density Legislation on Breast Cancer Risk Assessment and Supplemental Screening: A Survey of 110 Radiology Facilities. *Breast J* 2016;22(5):493-500.
- (8) Kressin NR, Gunn CM, Battaglia TA. Content, readability, and understandability of dense breast notifications by state. *JAMA* 2016;315(16):1786-1788.
- (9) Wolfe JN. Breast patterns as an index of risk for developing breast cancer. *AJR Am J Roentgenol* 1976 Jun;126(6):1130-1137.
- (10) Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer* 1976;37(5):2486-2492.
- (11) Winkler NS, Raza S, Mackesy M, Birdwell RL. Breast density: clinical implications and assessment methods. *Radiographics* 2015;35(2):316-324.
- (12) Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356(3):227-236.



- (13) Brandt KR, Scott CG, Ma L, Mahmoudzadeh AP, Jensen MR, Whaley DH, et al. Comparison of clinical and automated breast density measurements: implications for risk prediction and supplemental screening. *Radiology* 2015;279(3):710-719.
- (14) Abdollell M, Tsuruda KM, Lightfoot CB, Payne JI, Caines JS, Iles SE. Utility of relative and absolute measures of mammographic density vs clinical risk factors in evaluating breast cancer risk at time of screening mammography. *Br J Radiol* 2016;89(1059):20150522.
- (15) Warwick J, Birke H, Stone J, Warren RM, Pinney E, Brentnall AR, et al. Mammographic breast density refines Tyrer-Cuzick estimates of breast cancer risk in high-risk women: findings from the placebo arm of the International Breast Cancer Intervention Study I. *Breast Cancer Research* 2014;16(5):451.
- (16) van der Waal D, den Heeten GJ, Pijnappel RM, Schuur KH, Timmers JM, Verbeek AL, et al. Comparing visually assessed BI-RADS breast density and automated volumetric breast density software: a cross-sectional study in a breast cancer screening setting. *PLoS One* 2015;10(9):e0136667.
- (17) Duffy SW, Morrish OWE, Allgood PC, Black R, Gillan MGC, Willsher P, et al. Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer. *Eur J Cancer* 2018 Jan;88:48-56.
- (18) Destounis S, Johnston L, Highnam R, Arieno A, Morgan R, Chan A. Using volumetric breast density to quantify the potential masking risk of mammographic density. *Am J Roentgenol* 2017;208(1):222-227.
- (19) Gilbert FJ, Tucker L, Gillan MG, Willsher P, Cooke J, Duncan KA, et al. Breast density assessment. 2015.
- (20) Eng A, Gallant Z, Shepherd J, McCormack V, Li J, Dowsett M, et al. Digital mammographic density and breast cancer risk: a case-control study of six alternative density assessment methods. *Breast cancer research* 2014;16(5):439.
- (21) Holland K, van Zelst J, den Heeten GJ, Imhof-Tas M, Mann RM, van Gils CH, et al. Consistency of breast density categories in serial screening mammograms: A comparison between automated and human assessment. *The Breast* 2016;29:49-54.
- (22) Harvey JA, Gard CC, Miglioretti DL, Yankaskas BC, Kerlikowske K, Buist DS, et al. Reported mammographic density: film-screen versus digital acquisition. *Radiology* 2013;266(3):752-758.
- (23) Melnikow J, Fenton JJ, Whitlock EP, Miglioretti DL, Weyrich MS, Thompson JH, et al. Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the US Preventive Services Task Force Supplemental Breast Cancer Screening in Women With Dense Breasts. *Ann Intern Med* 2016;164(4):268-278.

- (24) Irshad A, Leddy R, Ackerman S, Cluver A, Pavic D, Abid A, et al. Effects of Changes in BI-RADS Density Assessment Guidelines (Fourth Versus Fifth Edition) on Breast Density Assessment: Intra- and Interreader Agreements and Density Distribution. *Am J Roentgenol* 2016;207(6):1366-1371.
- (25) Youk JH, Kim SJ, Son EJ, Gweon HM, Kim J. Comparison of visual assessment of breast density in BI-RADS 4th and 5th editions with automated volumetric measurement. *Am J Roentgenol* 2017;209(3):703-708.
- (26) Spayne MC, Gard CC, Skelly J, Miglioretti DL, Vacek PM, Geller BM. Reproducibility of BI-RADS Breast Density Measures Among Community Radiologists: A Prospective Cohort Study. *Breast J* 2012;18(4):326-333.
- (27) Gard CC, Aiello Bowles EJ, Miglioretti DL, Taplin SH, Rutter CM. Misclassification of Breast Imaging Reporting and Data System (BI-RADS) Mammographic Density and Implications for Breast Density Reporting Legislation. *Breast J* 2015;21(5):481-489.
- (28) Sprague BL, Conant EF, Onega T, Garcia MP, Beaber EF, Herschorn SD, et al. Variation in Mammographic Breast Density Assessments Among Radiologists in Clinical Practice A Multicenter Observational Study Variation in Mammographic Breast Density Assessments Among Radiologists. *Ann Intern Med* 2016;165(7):457-464.
- (29) Redondo A, Comas M, Macia F, Ferrer F, Murta-Nascimento C, Maristany M, et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *Br J Radiol* 2012;85(1019):1465-1470.
- (30) Ekpo EU, Ujong UP, Mello-Thoms C, McEntee MF. Assessment of interradiologist agreement regarding mammographic breast density classification using the fifth edition of the BI-RADS Atlas. *Am J Roentgenol* 2016;206(5):1119-1123.
- (31) D'Orsi CJ. *ACR BI-RADS atlas: breast imaging reporting and data system.* : American College of Radiology; 2013.
- (32) Breast density and breast cancer risk: a practical review. *Mayo Clinic Proceedings: Elsevier*; 2014.
- (33) Sprague BL, Gangnon RE, Burt V, Trentham-Dietz A, Hampton JM, Wellman RD, et al. Prevalence of mammographically dense breasts in the United States. *JNCI: Journal of the National Cancer Institute* 2014;106(10).
- (34) Kerlikowske K, Zhu W, Tosteson AN, Sprague BL, Tice JA, Lehman CD, et al. Identifying Women With Dense Breasts at High Risk for Interval Cancer A Cohort Study Identifying Women With Dense Breasts at High Risk for Interval Cancer. *Ann Intern Med* 2015;162(10):673-681.
- (35) BC Cancer Agency. *Screening Mammography Program 2016 Annual Report.* 2017;2016.

- (36) McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2006 Jun;15(6):1159-1169.
- (37) Vacek PM, Geller BM. A prospective study of breast cancer risk using routine mammographic breast density measurements. *Cancer Epidemiol Biomarkers Prev* 2004 May;13(5):715-722.
- (38) Lam PB, Vacek PM, Geller BM, Muss HB. The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in Vermont. *Cancer* 2000;89(2):369-375.
- (39) Ziv E, Tice J, Smith-Bindman R, Shepherd J, Cummings S, Kerlikowske K. Mammographic density and estrogen receptor status of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2004 Dec;13(12):2090-2095.
- (40) Yaghjian L, Colditz GA, Collins LC, Schnitt SJ, Rosner B, Vachon C, et al. Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to tumor characteristics. *J Natl Cancer Inst* 2011;103(15):1179-1189.
- (41) Wong CS, Lim GH, Gao F, Jakes RW, Offman J, Chia KS, et al. Mammographic density and its interaction with other breast cancer risk factors in an Asian population. *Br J Cancer* 2011 Mar 1;104(5):871-874.
- (42) Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Ann Intern Med* 2011;155(1):10-20.
- (43) Darabi H, Czene K, Zhao W, Liu J, Hall P, Humphreys K. Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. *Breast Cancer Research* 2012;14(1):R25.
- (44) Canadian Partnership Against Cancer. Breast Cancer Screening in Canada: Monitoring and Evaluation of Quality Indicators - Results Report, January 2011 to December 2012. 2017(Toronto: Canadian Partnership Against Cancer).
- (45) White E, Miglioretti DL, Yankaskas BC, Geller BM, Rosenberg RD, Kerlikowske K, et al. Biennial versus annual mammography and the risk of late-stage breast cancer. *J Natl Cancer Inst* 2004;96(24):1832-1839.
- (46) Wai ES, D'yachkova Y, Olivotto IA, Tyldesley S, Phillips N, Warren LJ, et al. Comparison of 1- and 2-year screening intervals for women undergoing screening mammography. *Br J Cancer* 2005 Mar 14;92(5):961-966.

- (47) Bottorff JL, Ratner PA, Johnson JL, Hislop TG, Buxton JA, Zeisser C, et al. Women's responses to information on mammographic breast density. *CJNR (Canadian Journal of Nursing Research)* 2007;39(1):38-57.
- (48) Chen T, Jonsson SH, Lenner P. Effect of mammographic service screening on stage at presentation of breast cancers in Sweden. *Cancer* 2007;109(11):2205-2212.
- (49) Bertrand KA, Tamimi RM, Scott CG, Jensen MR, Pankratz VS, Visscher D, et al. Mammographic density and risk of breast cancer by age and tumor characteristics. *Breast Cancer Research* 2013;15(6):R104.
- (50) Gierach GL, Ichikawa L, Kerlikowske K, Brinton LA, Farhat GN, Vacek PM, et al. Relationship between mammographic density and breast cancer death in the Breast Cancer Surveillance Consortium. *J Natl Cancer Inst* 2012;104(16):1218-1227.
- (51) Kerlikowske K, Zhu W, Hubbard RA, Geller B, Dittus K, Braithwaite D, et al. Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. *JAMA internal medicine* 2013;173(9):807-816.
- (52) Saadatmand S, Bretveld R, Siesling S, Tilanus-Linthorst MM. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients. *BMJ* 2015 Oct 6;351:h4901.
- (53) Puliti D, Duffy SW, Miccinesi G, De Koning H, Lynge E, Zappa M, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen* 2012;19(1\_suppl):42-56.
- (54) Broeders M, Moss S, Nyström L, Njor S, Jonsson H, Paap E, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen* 2012;19(1\_suppl):14-25.
- (55) Paci E. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen* 2012;19(1\_suppl):5-13.
- (56) Independent UK Panel on Breast cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;380(9855):1778-86.
- (57) Bleyer A, QWelch H. Effect of Three Decades of Screening Mammography on Breast-Cancer Incidence. *New England Journal of Medicine* 2012;367:1998.
- (58) The Canadian Task Force on Preventive Health Care. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. *CMAJ* 2011;183(17):1991-2001.
- (59) Moss S. Overdiagnosis and overtreatment of breast cancer: overdiagnosis in randomised controlled trials of breast cancer screening. *Breast Cancer Research* 2005;7(5):230-234.

- (60) Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet* 2006 Dec 9;368(9552):2053-2060.
- (61) Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. *The Lancet Oncology* 2015;16(9):1123-1132.
- (62) Irwig L, Houssami N, Armstrong B, Glasziou P. Evaluating new screening tests for breast cancer. *BMJ* 2006 Mar 25;332(7543):678-679.
- (63) Neilson HK, Farris MS, Stone CR, Vaska MM, Brenner DR, Friedenreich CM. Moderate-vigorous recreational physical activity and breast cancer risk, stratified by menopause status: a systematic review and meta-analysis. *Menopause* 2017 Mar;24(3):322-344.
- (64) Ziembicki S, Zhu J, Tse E, Martin LJ, Minkin S, Boyd NF. The Association between Alcohol Consumption and Breast Density: A Systematic Review and Meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2017 Feb;26(2):170-178.
- (65) Martin LJ, Li Q, Melnichouk O, Greenberg C, Minkin S, Hislop G, et al. A randomized trial of dietary intervention for breast cancer prevention. *Cancer Res* 2011 Jan 1;71(1):123-133.
- (66) Powles TJ, Ashley S, Tidy A, Smith IE, Dowsett M. Twenty-year follow-up of the Royal Marsden randomized, double-blinded tamoxifen breast cancer prevention trial. *J Natl Cancer Inst* 2007;99(4):283-290.
- (67) Cuzick J, Sestak I, Cawthorn S, Hamed H, Holli K, Howell A, et al. Tamoxifen for prevention of breast cancer: extended long-term follow-up of the IBIS-I breast cancer prevention trial. *The lancet oncology* 2015;16(1):67-75.
- (68) Cuzick J, Warwick J, Pinney E, Duffy SW, Cawthorn S, Howell A, et al. Tamoxifen-induced reduction in mammographic density and breast cancer risk reduction: a nested case-control study. *J Natl Cancer Inst* 2011;103(9):744-752.
- (69) Goss PE, Ingle JN, Alés-Martínez JE, Cheung AM, Chlebowski RT, Wactawski-Wende J, et al. Exemestane for breast-cancer prevention in postmenopausal women. *N Engl J Med* 2011;364(25):2381-2391.
- (70) Georgian-Smith D, Moore RH, Halpern E, Yeh ED, Rafferty EA, D'Alessandro HA, et al. Blinded comparison of computer-aided detection with human second reading in screening mammography. *AJR Am J Roentgenol* 2007 Nov;189(5):1135-1141.
- (71) Brem RF. Blinded comparison of computer-aided detection with human second reading in screening mammography: the importance of the question and the critical numbers game. *AJR Am J Roentgenol* 2007 Nov;189(5):1142-1144.

- (72) Klompenhouwer EG, Voogd AC, Den Heeten GJ, Strobbe LJ, Tjan-Heijnen VC, Broeders MJ, et al. Discrepant screening mammography assessments at blinded and non-blinded double reading: impact of arbitration by a third reader on screening outcome. *Eur Radiol* 2015;25(10):2821-2829.
- (73) Klompenhouwer EG, Voogd AC, den Heeten GJ, Strobbe LJ, de Haan AF, Wauters CA, et al. Blinded double reading yields a higher programme sensitivity than non-blinded double reading at digital screening mammography: a prospected population based study in the south of The Netherlands. *Eur J Cancer* 2015;51(3):391-399.
- (74) Taylor-Phillips S, Wallis MG, Jenkinson D, Adekanmbi V, Parsons H, Dunn J, et al. Effect of using the same vs different order for second readings of screening mammograms on rates of breast cancer detection: a randomized clinical trial. *JAMA* 2016;315(18):1956-1965.
- (75) D’Orsi CJ, Sickles EA. To Seek Pereflection or Not? That is the Question. *Radiology* 2012;265:9-11.
- (76) 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):2353-2358.
- (77) Committee on Gynecologic Practice. Committee Opinion No. 625: Management of women with dense breasts diagnosed by mammography. *American College of Obstetricians and Gynecologists. Obstet Gynecol* 2015;125(3):750-751.
- (78) Siu AL. Screening for breast cancer: US preventive services task force recommendation Statement Screening for breast cancer. *Ann Intern Med* 2016;164(4):279-296.
- (79) Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016 Nov 14;6(11):e012799-2016-012799.
- (80) Scheel JR, Lee JM, Sprague BL, Lee CI, Lehman CD. Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts. *Am J Obstet Gynecol* 2015 Jan;212(1):9-17.
- (81) Ohuchi N, Suzuki A, Sobue T, Kawai M, Yamamoto S, Zheng Y, et al. Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomised controlled trial. *The Lancet* 2016;387(10016):341-348.
- (82) Brancato B, Bonardi R, Catarzi S, Iacconi C, Risso G, Taschini R, et al. Negligible advantages and excess costs of routine addition of breast ultrasonography to mammography in dense breasts. *Tumori* 2007;93(6):562.

- (83) Youk JH, Kim E, Kim MJ, Kwak JY, Son EJ. Performance of hand-held whole-breast ultrasound based on BI-RADS in women with mammographically negative dense breast. *Eur Radiol* 2011;21(4):667-675.
- (84) 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):1394-1404.
- (85) Weigert JM. The Connecticut experiment; The third installment: 4 Years of screening women with dense breasts with bilateral ultrasound. *Breast J* 2017;23(1):34-39.
- (86) Wilczek B, Wilczek HE, Rasouliyan L, Leifland K. Adding 3D automated breast ultrasound to mammography screening in women with heterogeneously and extremely dense breasts: Report from a hospital-based, high-volume, single-center breast cancer screening program. *Eur J Radiol* 2016;85(9):1554-1563.
- (87) Kelly KM, Dean J, Comulada WS, Lee S. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *Eur Radiol* 2010;20(3):734-742.
- (88) Corsetti V, Houssami N, Ghirardi M, Ferrari A, Speziani M, Bellarosa S, et al. Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. *Eur J Cancer* 2011;47(7):1021-1026.
- (89) Giuliano V, Giuliano C. Improved breast cancer detection in asymptomatic women using 3D-automated breast ultrasound in mammographically dense breasts. *Clin Imaging* 2013 May-Jun;37(3):480-486.
- (90) Brem RF, Tabár L, Duffy SW, Inciardi MF, Guingrich JA, Hashimoto BE, et al. Assessing improvement in detection of breast cancer with three-dimensional automated breast US in women with dense breast tissue: the SonoInsight Study. *Radiology* 2014;274(3):663-673.
- (91) Girardi V, Tonegutti M, Ciatto S, Bonetti F. Breast ultrasound in 22,131 asymptomatic women with negative mammography. *Breast* 2013 Oct;22(5):806-809.
- (92) Grady I, Chanisheva N, Vasquez T. The Addition of Automated Breast Ultrasound to Mammography in Breast Cancer Screening Decreases Stage at Diagnosis. *Acad Radiol* 2017.
- (93) Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. *Radiology* 2012;265(1):59-69.
- (94) Tagliafico AS, Calabrese M, Mariscotti G, Durando M, Tosto S, Monetti F, et al. Adjunct screening with tomosynthesis or ultrasound in women with mammography-negative dense breasts: interim report of a prospective comparative trial. *Journal of Clinical Oncology* 2016;34(16):1882-1888.

- (95) Venturini E, Losio C, Panizza P, Rodighiero MG, Fedele I, Tacchini S, et al. Tailored breast cancer screening program with microdose mammography, US, and MR Imaging: short-term results of a pilot study in 40–49-year-old women. *Radiology* 2013;268(2):347-355.
- (96) Ohuchi N, Ishida T, Kawai M, Narikawa Y, Yamamoto S, Sobue T. Randomized controlled trial on effectiveness of ultrasonography screening for breast cancer in women aged 40–49 (J-START): research design. *Jpn J Clin Oncol* 2010;41(2):275-277.
- (97) Youlten DR, Cramb SM, Yip CH, Baade PD. Incidence and mortality of female breast cancer in the Asia-Pacific region. *Cancer Biol Med* 2014 Jun;11(2):101-115.
- (98) 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):157-166.
- (99) Alagoz O, Berry DA, de Koning HJ, Feuer EJ, Lee SJ, Plevritis SK, et al. Introduction to the Cancer Intervention and Surveillance Modeling Network (CISNET) breast cancer models. *Medical Decision Making* 2018;38(1\_suppl):3S-8S.
- (100) Weigel S, Heindel W, Heidrich J, Hense H, Heidinger O. Digital mammography screening: sensitivity of the programme dependent on breast density. *Eur Radiol* 2017;27(7):2744-2751.
- (101) Cole EB, Pisano ED. Tomosynthesis for breast cancer screening. *Clin Imaging* 2016 Mar-Apr;40(2):283-287.
- (102) McDonald ES, Oustimov A, Weinstein SP, Synnestvedt MB, Schnall M, Conant EF. Effectiveness of digital breast tomosynthesis compared with digital mammography: outcomes analysis from 3 years of breast cancer screening. *JAMA oncology* 2016;2(6):737-743.
- (103) Rafferty EA, Durand MA, Conant EF, Copit DS, Friedewald SM, Plecha DM, et al. Breast cancer screening using tomosynthesis and digital mammography in dense and nondense breasts. *JAMA* 2016;315(16):1784-1786.
- (104) Conant EF, Beaber EF, Sprague BL, Herschorn SD, Weaver DL, Onega T, et al. Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: a cohort study within the PROSPR consortium. *Breast Cancer Res Treat* 2016;156(1):109-116.
- (105) Skaane P, Sebuødegård S, Bandos AI, Gur D, Østerås BH, Gullien R, et al. Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo Tomosynthesis Screening Trial. *Breast Cancer Res Treat* 2018:1-8.
- (106) BC Cancer. Screening/Early Detection: Breast. 2015; Available at: <http://www.bccancer.bc.ca/health-professionals/clinical-resources/cancer-management-guidelines/breast/breast#Screening-Early-Detection>. Accessed 03/11, 2018.



- (107) Cott Chubiz JE, Lee JM, Gilmore ME, Kong CY, Lowry KP, Halpern EF, et al. Cost-effectiveness of alternating magnetic resonance imaging and digital mammography screening in BRCA1 and BRCA2 gene mutation carriers. *Cancer* 2013;119(6):1266-1276.
- (108) Møller P, Stormorken A, Jonsrud C, Holmen MM, Hagen AI, Clark N, et al. Survival of patients with BRCA1-associated breast cancer diagnosed in an MRI-based surveillance program. *Breast Cancer Res Treat* 2013;139(1):155-161.
- (109) Gareth ED, Nisha K, Yit L, Soujanya G, Emma H, Massat NJ, et al. MRI breast screening in high-risk women: cancer detection and survival analysis. *Breast Cancer Res Treat* 2014;145(3):663-672.
- (110) Møller P, Maehle L, Vabø A, Clark N, Sun P, Narod S. Age-specific incidence rates for breast cancer in carriers of BRCA1 mutations from Norway. *Clin Genet* 2013;83(1):88-91.
- (111) Berg WA, Blume JD, Adams AM, Jong RA, Barr RG, Lehrer DE, et al. Reasons women at elevated risk of breast cancer refuse breast MR imaging screening: ACRIN 6666. *Radiology* 2009;254(1):79-87.
- (112) Kuhl CK, Strobel K, Bieling H, Leutner C, Schild HH, Schrading S. Supplemental breast MR imaging screening of women with average risk of breast cancer. *Radiology* 2017;283(2):361-370.
- (113) Kuhl CK, Schrading S, Strobel K, Schild HH, Hilgers R, Bieling HB. Abbreviated breast magnetic resonance imaging (MRI): first postcontrast subtracted images and maximum-intensity projection—a novel approach to breast cancer screening with MRI. *Journal of Clinical Oncology* 2014;32(22):2304-2310.
- (114) Kriege M, Brekelmans CT, Boetes C, Besnard PE, Zonderland HM, Obdeijn IM, et al. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *N Engl J Med* 2004;351(5):427-437.
- (115) Riedl CC, Luft N, Bernhart C, Weber M, Bernathova M, Tea MM, 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. *Journal of Clinical Oncology* 2015;33(10):1128-1135.

## **Appendix 1**

## **List of Key Informants**

Professor Norman Boyd, Medical Biophysics, University of Toronto

Professor Michael Burgess, W. Maurice Young Centre for Applied Ethics, University of British Columbia

Dr Stephen Chia, Chair, Breast Cancer Tumour Group, BC Cancer

Dr Cathy Clelland, Family Practice Oncology Network, BC Cancer

Ms Jennie Dale and Ms Michelle di Tomaso, Dense Breasts Canada

Mr Greg Doyle, Chair, Breast Cancer Screening Network, Canadian Partnership Against Cancer

Dr Paula Gordon, Medical Director, Breast Health Program, BC Women's Hospital and Medical Advisor, Dense Breasts Canada

Professor Joy Melnikow, Director, Centre for Health Care Policy and Research, UC Davis

Dr Sylvia Robinson, Director, Lifetime Prevention Schedule and Screening, BC Ministry of Health

Professor Edward Sickles, Radiology, School of Medicine, UC San Francisco

Professor Nancy Wadden, Faculty of Medicine, Memorial University and Chair, Mammography Accreditation Program, Canadian Association of Radiologists

Dr Charlotte Yong-Hing, BC Cancer and BC Radiological Society

## **Questionnaire for Key Informants**

Key informant:

Date/time of Interview:

Medium: in person, telephone, questionnaire

- 1) In your opinion is breast density an important question for the women of BC?
- 2) In your opinion is breast density reporting an important clinical question in the BC Medical community?
- 3) In your opinion is breast density an important predictor of breast cancer risk?
- 4) In your opinion is breast density an important prognostic factor for mammography detection?
- 5) What are key publications that support breast density as an important risk/prognostic factor for breast cancer?
- 6) BCCBS will be recording breast density using the BI-RADS scale. In your opinion is this the most best way to record breast density for BCCBS?

7) If breast density is to be reported outside of the screening program what approach would you favour?

- a) Do not report it at all
- b) The measured density reported to the family physician alone (complete or ?)
- c) The measured density reported to the family physician and subject (complete or ?)
- d)
- e) Should information on other breast risk factors be included?
- f)
- g) Should some estimate of risk be provided (based on density and/or other risk factors)?
- h)
- i) Should any density (or risk) specific management guidance be provided to the physician by BCCBS?
- j) Should any density (or risk) specific management be organized by BCCBS?

8) What, if any, specific recommendations or options (screening and/or non-screening) would you make for women based upon their measured breast density?

A: Fatty

B: Scattered fibroglandular

C: Heterogeneously dense

D: Dense

9) Any further factors that should be considered in relation to this issue?

## Appendix 2

## **Key Informant Interviews**

To provide perspective on the issues associated with breast density and screening a group of key informants were identified. Identification of key informants was done through consultation with BC Cancer Breast Screen (BCCBS). Individuals were chosen from segments and organisations associated with breast screening, advocacy organisations and scientists with a background in research into breast density and breast cancer risk. Individuals were not chosen, or asked to speak as, representatives of the organisations with which they were affiliated, but to provide their own perspective.

Each key informant was interviewed, either by phone or in person. Each interview was conducted using a common questionnaire (see Appendix 1). Questions were constructed to elicit information the respondents' views in the following domains:

- The importance of breast density on breast cancer risk
- The influence of breast density on breast cancer detection by mammography (masking)
- Appropriate scales used to record breast density for use by screening programs
- Who should receive information on breast density and how should it be provided
- Appropriate screening for women by BIRADS breast density category

Questions were primarily open ended were not intended to facilitate quantitative summary measures but to elicit the range of opinions and any issues the respondents felt to be key. Informants were of two key types: stakeholders within Canada with interests in breast screening or experts on breast screening or breast cancer researchers. Consequently, aggregate results from the interviews are presented as themes or issues. Respondents were not drawn from any defined population so that summary measures are not really appropriate.

## Results from Key Informant Interviews

Several of the respondents were selected for their expertise in areas allied to the use and reporting of breast density and did not claim expertise in the science or current state of evidence regarding breast density and screening. As a consequence not all respondents answered all the questions. In attempting to provide a qualitative summary of responses the summary will be based on only those professing knowledge of the question subject area.

Fourteen key respondents were identified and contacted. One was non-responsive and 13 interviews (see Appendix 1 for respondents) were conducted.

### Domain 1: Association of density and breast cancer risk

Among respondents who claimed some familiarity with breast density there was a consensus that increasing breast density is associated with increasing risk of breast cancer however measured.

### Domain 2: Association of density and masking in mammography

Among respondents who claimed some familiarity with breast density there was a consensus that increasing breast density is associated with decreasing detectability (masking) of breast cancer by mammography.

### Domain 3: Appropriateness of BIRADS as a scale for measurement density in a clinical setting

Among respondents knowledgeable about breast density different opinions were expressed about the suitability of BIRADS for measuring density by BCCBS:

- some felt it was the most appropriate tool for clinical use and by BCCBS while
- some felt that BIRADS was suboptimal and that continuous quantitative scales were superior.

Those preferring quantitative scales appeared to do so because of perceived superior consistency, possibly by automated measurement, and stronger discrimination of breast cancer risk. Respondents were not asked, and none expressed a preference for human versus automated measurement. An advantage of continuous scales is that also permit arbitrary dichotomization of the screening population for clinical use rather than relying on scales, such as BIRADS, which use qualitatively predefined thresholds.

### Domain 4: Communication of density to women and their health providers

All respondents addressed this domain. There was considerable heterogeneity in opinion about the communication of breast density:

- Several felt that it should be automatically provided to both screened women and their physicians



- Some felt it should be directly provided to the physician only who could then use it within the context of clinical care, i.e. communicate it or not as deemed appropriate by the provider's judgement
- Some felt that density should only be proactively reported if it would affect recommended care
- All felt that any communication of density should only be made within the context of a well-developed education strategy aimed at both physicians and women
- Several respondents expressed the opinion that by collecting density information the program would need to explicitly justify non-communication to women
- Several respondents felt that breast density should be reported within the context of a general communication of breast cancer risk which could go as far as the production of individualized estimates of breast cancer risk for women in screening
- Some respondents felt that communication of density be limited to only high breast density (i.e. non-high density should not be reported)
- One respondent favoured reporting only following a request (current BCCBS practice)

Respondents favouring direct communication of density also advocated tailored clinical management based upon density. Several indicated that it was counterproductive, and possibly harmful, to communicate density but not indicate potential specific actions as a result of density.

#### Domain 5: Clinical management based upon breast density

There was no consensus about how to use breast density information clinically:

- Some said that reductions in mammography frequency, in those with low breast density, should be considered whereas others indicated that this should not be considered
- A few respondents indicated that women with high density (BIRADS D) should receive annual mammography
- Several respondents indicated that women with dense breasts (BIRADS C or D) should receive breast ultrasound as a supplement to routine mammography
- Others provided recommendations which varied with BIRADS density and involved breast tomography, ultrasound and/or breast MRI
- Some respondents made recommendations that differ from current practice but these applied to all women irrespective of density

In answering this question some respondents proposed a new set of screening guidelines which included all women and some limited themselves to proposals for additional screening of those classified as dense: no change, more frequent mammography, *offer* supplemental tests (ultrasound or MRI) or *recommend* supplemental tests.

Respondents were also asked about potential contraindications to management which involved supplemental or altered screening. Several mentioned *cost* as an issue to be considered, two mentioned that the magnitude of any benefit with altered screening was not proven, one mentioned increasing false-positives associated with supplemental screening. No respondent indicated the possibility of increased over diagnosis of breast cancer as a potential concern.

## **Conclusions from Key Informant Interviews**

There was unanimity among those knowledgeable interviewees that increasing breast density was associated with increased risk of breast cancer and reduced sensitivity for screen detection by mammography. There was less consensus about how best to measure breast density. A continuous scale leads to greater discrimination and, if measured reproducibly, has clear advantages for population use. There was considerable diversity in opinion on the communication of breast density. A common theme, if there was any, was that any communication should only be done within the context of a comprehensive program strategy which would provide appropriate information to the respondent. There was a similar lack of consensus about approaches to the preventive clinical management of women with increased density. Responses generally favoured supplemental screening of one type or another: increased mammography, breast CT, ultrasound or MRI. No respondent advocated interventions aimed at reducing breast cancer risk.

## **Appendix 3**

### **Analysis of BCCBS Density Data**

**Authors: Andy Coldman, Colleen Mcgahan, Yvonne Zheng**

## Summary of Analysis of BC Data

Data sets were extracted from the BCCBS data base on women participating in screening in the period 2011-2014 inclusive. Eligible screens were required to have breast density reported using the BIRADS scale which some centres were reporting during this period. Data was also extracted on breast cancers reported to the BC Cancer Registry for the period 2011-2016 which were linked to screening records to identify cases of breast cancer occurring among the study participants. The object of analyses conducted were to examine the effect of density within screens performed by the BCCBS on the following factors:

1. The distribution of density in the screened population of BC
2. The effect of density on risk of breast cancer
3. The effect of density on the likelihood of detection by mammography screening
4. The effect of density on the distribution of prognostic factors among breast cancers
5. The effect of density on the likelihood of a false-positive mammogram
6. The stability of reported density in serial screens

It was recognized that this analysis would only be available some BCCBS centres and that the study population may not be representative of the overall screening population. The analysis was based upon screening rounds, a period following a screen up to and including the next screen or study endpoint. Consequently, screening rounds are proportional to screening frequency so that resulting distribution of patient specific factors may be affected.

## Summary of Study Findings

### Distribution of Density

Reported density was found to vary by age and ethnic group. Density declined by age and was higher in South-East Asians. Overall the density distribution was A: 29%, B: 37%, C: 27% and D: 8%<sup>6</sup>. This distribution was “less dense” than commonly reported and may have been related to the use of screening rounds as the unit of measurement rather than women and the transition from BIRADS Edition 4 to Edition 5.

### Risk of Breast Cancer

*Risk of Screen Detected Breast Cancer:* Risk of invasive screen detected breast cancer was measured using the rate at two years. Among women 40-49 the relative risk of screen detection increased with BIRADS density categories A: 0.79, B: 0.91, C: 1.07 and D: 1.39. For women aged 50-74 the relative risk of screen detected cancers varied with BIRADS density categories A: 0.86, B: 1.16, C: 1.09 and D: 0.52. For women 50-74 relative risk of invasive cancers did not increase monotonically with density which was likely the result of confounding with other factors, particularly age, in this group.

---

<sup>6</sup> Recent data following more widespread utilisation of BIRADS within BCCBS indicates a distribution of A:17%, B:42%, C:32%, D:7% and not reported 2% - J Sam, personal communication.

*Interval Cancers Following Mammography Screening:* Risk of invasive interval breast cancer was measured over a screening period of two years. Among women 40-49 the relative risk of interval cancers increased with BIRADS density categories A: 0.49, B: 0.69, C: 1.24 and D: 1.81. For women aged 50-74 the relative risk of interval cancers increased with BIRADS density categories A: 0.54, B: 0.70, C: 1.63 and D: 2.0. Rates of interval cancer exceeded 1 per 1,000 in the year following a negative screen for women of any age (40-74) who were BIRADS D: other density categories had rates less than 1 per 1000.

*Overall Risk of Breast Cancer:* Risk of invasive breast cancer was measured over a screening period of two years by summing screened detected and interval cancer rates. Among women 40-49 the relative risk of invasive cancer, with respect to the average, increased with BIRADS density categories A through D as follows: A: 0.68, B: 0.82, C: 1.15 and D: 1.56 respectively. For women 50-74 relative risk of invasive cancers with density was: A: 0.75, B: 1.08, C: 1.25 and D: 0.95 respectively.

Generally, the results replicated ones found in the literature for density with increases in risk associated with increasing density (1). This was not true for screen detected cancers in women 50-74 which was likely due to the confounding by age and other risk factors. It does indicate that selecting women aged 50-74 on the basis of density alone will not result in the identification of a group at higher risk than average for the age group. All ages showed a relationship between increased density and higher rates of interval cancer.

### **Distribution of Breast Cancer Prognostic Factors**

Two factors were examined proportion with size >15mm and proportion with positive nodes. Screen detection was associated with better prognostic profiles regardless of density. The prognostic profile of interval cancers did not vary with density. Density did influence the profile of screen detected cancers with dense (either D v A+B+C or C+D v A+B) associated with cancers with worse prognostic profile. This difference was smaller than the difference between screen detected and interval cancers suggesting that increased mammographic screen detection would still improve prognosis.

### **Likelihood of a False-Positive Mammogram**

Density was associated with the likelihood of a false-positive screening mammogram. The principal relationship was that BIRADS density A was lower than the other categories which were similar to one another. Age and time since preceding screen also influenced the false-positive rate.

### **Stability of Reported Density in Serial Screens**

Pairs of consecutive screens where density was reported were examined and the consistency of reported density examined. Using C+D (D) as the clinical definition of dense, then 80% (64%) initially classed as dense were dense on the second mammogram. Stability increased when the

same radiologist reported both screens but only 1/3 of pairs had the same radiologist. Woman's age did influence the stability of the reported density category but this effect was not strong.

### **Implications of the Analysis**

Relationships seen in other studies (breast cancer risk, masking) were evident in the British Columbia data. BIRADS density classification in the data did separate women into groups with varied risk, however, overall differences in risk were not large and the main discriminating power was for interval cancer risk. In women 50-74 BIRADS density alone was a poor predictor of overall risk. The qualitative nature and operator variability in BIRADS classification make it less than optimal for management of screening subjects. The prognostic profiles of cancers varied with density with the principal driver of this effect being the reduced proportion of screen detected cancers.

Benefit from screening mammography is approximately proportional to the rate of screen detection although other factors such as age, etc influence true benefit. Although this was only seen explicitly in women aged 40-49, women with denser breasts have higher rates of screen detection and thus are likely to benefit more from screening mammography than others. Nevertheless, women with dense breasts, particularly BIRADS D, have comparatively high rates of interval cancer and reductions in the rate of interval cancer are likely to lead to benefits if such reductions are associated with improved prognosis.

## **Introduction**

It is generally recognized that breast density affects the outcomes for mammography screening for breast cancer. It does so in two principle ways at the individual level:

1. Breast density is related to the risk of breast cancer
2. Breast density influences the ability to detect breast cancer with mammography (masking)

Consequently, at the provincial level the distribution of breast density in the screening population influences the performance of screening program.

While there are many peer reviewed articles which provide information on the relationship between breast density and breast cancer screening, the application of their results to British Columbia is subject to some uncertainty:

1. Various measures of density are used in the literature
2. The distribution of breast density is known to vary across populations
3. Results are frequently reported in the form of odds ratios or relative risks and consequently calculation of rates for British Columbia requires further information
4. Results are frequently presented with control of other known breast cancer risk factors and the effect of such control at the aggregate level can be unclear

In light of the above issues it was decided to undertake an analysis of data collected by the BC Cancer Breast Screening (BCCBS) as part of providing screening to BC women.

## **Objectives of the Analysis**

1. To examine the relationship between recorded breast density and screening outcomes in BCCBS screening participants.
2. To determine the stability of breast density classification in consecutive BCCBS screening mammograms

## **Data Used**

BIRADS density classification has been reported by several radiologists and centres as part of screening provision within BCCBS. Data was extracted from the BCCBS database for screens performed between 1<sup>st</sup> January 2011 and 31<sup>st</sup> December 2014 where BIRADS density was reported. This period was chosen so that notification of any cancer cases was complete and the data period was reasonably current. The eligibility criteria were:

- Women had to have a BCCBS screen reported between January 1<sup>st</sup> 2011 and December 31<sup>st</sup> 2014.
- Women were between 40 and 74 years of age at the time of a screen performed in the study period
- Women had no personal history of breast cancer

## Methods

As outcomes ductal carcinoma-in-situ (DCIS) and invasive cancer were included as breast cancers. Lobular carcinoma-in-situ was not included as breast cancer. A breast cancer was classified as screen-detected if it was diagnosed within the 12 months following an abnormal BCCBS screening mammogram<sup>7</sup>. All other breast cancers were classified as post-screen or incident cancers. Interval cancers were defined as post-screen cancers within the recommended screening interval (usually 24 months).

The screens performed in individuals within the study period were decomposed into screening rounds where a screening round begins immediately following a screen, and ends with the first of

1. *The next screen*
2. *A post-screen cancer*
3. *31<sup>st</sup> December 2014*

Screening rounds have 5 possible outcomes, O:

1. *if round ends at 31/12/2014*
2. *if round ends with a normal screen*
3. *if round ends with an abnormal screen but no breast cancer detected*
4. *if round ends with a screen detected cancer*
5. *if round ends with a post-screen cancer*

In addition to the length of the screening round the following information was extracted for the following factors for the screen at the start of the screening round

1. an indicator of whether this was the first ever screen
2. high risk flag (1<sup>st</sup> degree family history)
3. the age (categorized as [40,45), [45,50), ..., [75,80), [80+))
4. the result of the screen (normal, abnormal)
5. image type (analogue, digital)
6. ethnic group (East Asian, First Nations, other)
7. the BIRADS density (A, B, C, D)

Rates of screen detected cancer will depend on time since preceding screen (screening round length). There are two potential approaches to include the effect round length into an analysis: 1) group data based on having similar interval lengths (e.g. 10-20 months) or 2) model the effect of interval length as a continuous variable. Interval lengths are clustered in the data because of the effect of population recommendations (biennial screening etc.) and the system of reminder letters built around those recommendations. For this reason, the first approach was used and screening rounds which ended in a screen (O = 2, 3 or 4) were grouped into annual (<18

---

<sup>7</sup> This definition is commonly used in the published literature but is not identical to that used by BCCBS in reporting statistics.



months), biennial (18 - <30 months) and triennial (30 - <42 months). Screening recommendations were changed in early 2014, so that most women 40-49 were no longer recalled for annual screening, to one where all women 40-74 designated at high risk (first degree family history) were recalled for annual screening and non-high risk recalled biennially. Women who had not attended following earlier reminders received further reminders at 3 years. Consequently, screen detected cancers cluster at annual anniversaries and will be analysed using binomial models. Screen detected cases occurring <18 months will be considered to be the result of annual screens, 18-30 months biennial and 30-42 months triennial. In contrast to screen detected cancers, post-screen cancers occur continuously in time and are less directly influenced by screening policy and do not cluster at specific time-points. Consequently, post-screen cancers will be analysed using time-to-event models. Estimated cumulative rates at 12, 24 and 36 months will be assumed to correspond to annual, biennial and triennial screening respectively.

Over-diagnosis is known to be an unwanted consequence of mammography screening. DCIS identification by mammographic screening is believed to contribute disproportionately to over-diagnosis and DCIS identification at older ages has less potential to provide short-term benefit through the future reduction of breast cancer mortality. Consequently, the subsequent analysis of the effect of breast density will be limited to the effect on invasive cancers whilst recognizing that the diagnosis of DCIS leads to that woman exiting the study cohort.

## Results

There were 485,375 screening rounds contributed by 238,132 women in the study sample. As a consequence of the study design first screens could not be included in screening rounds, so that only subsequent screens could contribute to outcomes, but could be the start of a screening round. There were 292,521 screens included (60.3% of rounds ended in a screen). The use of screening rounds resulted in the study sample having a “younger” age distribution than that of the population screened in a year because of the contribution of more rounds by women under 50 as screening was annual for most of the period included in the study. Within the study period there were 2,225 cancers of which 386 were DCIS. Of the cancers, 1,313 were classified as screen detected (SD) so that the overall rate of screen detection (invasive +DCIS) was 4.5/1,000<sup>8</sup>. Thirty seven out of 41 (90%) of centres contributed some screening rounds with BIRADS density recorded. The distribution of the included factors over the screening rounds is given in Appendix Table 1.

### Distribution of BIRADS Density Classification

Density was significantly related to several of the other factors considered but most relationships were not strong. In particular, high risk status (a factor used in mammography screening recommendations) appeared to be unrelated to reported breast density (Appendix Table 2). Two factors which were related to BIRADS density were age and ethnic group (Appendix Table 2). As has been reported, density declined with age with over 50% screening rounds being reported as C or D at age 40-44 and less than 20% at age 70-74. Ethnic affiliation is self-reported by women at their first participation in BCCBS. East and south-east Asian (E/SE) women had high reported density with the reported proportion C or D exceeding that in the lowest age group (40-44). Women reporting First Nations (FN) heritage had lower reported breast density and contributed a small proportion (2.2%) of the screening rounds. The relationship with ethnic status was somewhat confounded with age, with 19% of rounds in E/SE women performed in 40-44 age group compared to 11% for Other (not E/SE or FN) women with this difference reversed at the oldest age-group (70-74). However, the age distribution for FN women was more similar to that of E/SE and their reported proportion BIRADS C or D was less than that of the Other group.

### Risk of Screen Detected and Interval Breast Cancers

Estimating the influence of density on true risk of breast cancer in a screened population is complex since screen detection is subject to the masking effect of high density as well as the effect of screening on advancing the time at diagnosis. As an approximation to the risk of breast cancer we use the sum of the rates of screen-detected and interval cancers. This measure has the advantage that it relates to screening rounds and permits the comparison within rounds of similar length and is appropriate for its intended use in the accompanying report. The quantities will be calculated separately for rounds of one, two and three years in length. For comparison purposes

---

<sup>8</sup> Based upon the age distribution of the study sample we would have a predicted rate of 4.6/1,000 using rates for subsequent screens performed in 2015 (SMP Annual Report 2016 – Table 7).

an “annualized” rate is also calculated so that the presented results are per-year in each case. For the estimation of the post-screen cancer rates rounds following an abnormal mammogram (with no cancer detected) have been removed. The rationale for this is that the focus of this analysis is the influence of density following a normal screening mammogram: women with an abnormal mammogram receive further diagnostic testing and may receive enhanced follow-up. Such women would represent a specific sub-group and examining outcomes in such women would require supplementary data on any extra testing and follow-up that they received.

Appendix Table 3 shows the calculated SD and post screen (PS) invasive breast cancer rates by screening round length (annual, biennial and triennial) with the average annualized cancer risk. These rates have not been smoothed using any multivariate model and are subject to stochastic variation. In particular the SD rates derived from triennial screening rounds are less precise because of the smaller number of such intervals. Similarly, the estimation of the rate of PS cancers in the third year is also based on limited data. Ages were grouped into 40-49, 50-59 and 60-74 to provide stability of estimates by age. The rate of PS cancer does not vary much by age although the SD rate shows a strong gradient with age. For annual and biennial screening rounds the risk of breast cancer shows the expected pattern for high-risk status with higher screen detection and post screen cancer rates. The annualized estimates by density are less consistent but still show the expected pattern with overall risk increasing with measured density. The estimate for PS cancers shows a strong positive relationship with increasing density. The relationship for screen detected cancers shows a U-shaped relationship with rates highest for B density pattern, presumably because of opposing trends from risk and masking by increased density.

Through most of the study period women aged 40-49 were recalled annually while women 50-74 were recalled biennially. Current recommendations would recall all women aged 40-74 biennially with those 40-49 being advised to weigh the benefits and risks of mammography screening before being screened. Given this dichotomy between the recommendations for those 40-49 and 50-74 it is appropriate to consider them separately. Results for the two age groups are presented in Appendix Table 4. For screened women aged 40-49 the annualized estimate of risk increases regularly with density whereas for women 50-74 BIRADS D has lower risk than BIRADS C. This is likely due to residual confounding with age since the 50-74 age-group is broad and density declines with age. However, the incidence of PS cancer increases regularly with density in both age groups and it can be seen that it is the decline in screen detected cancer at higher densities which leads to the relationship seen in the annualized rate. Consequently, any information on density and breast cancer risk provided to women or practitioners would probably wish to narrow the age bands used.

Benefit associated with mammography screening accrues from the screen detection of existing clinically significant cancer. Whilst biennial mammography may identify a smaller proportion of cancers in women with dense breasts, it does identify more cancers by density among women 40-49 and this relationship would likely hold in women 50-74 if broken down into narrower age-groupings (Appendix Table 4). Consequently, while the relative effectiveness of mammography may decline with density, its absolute effectiveness is likely to increase<sup>9</sup>.

---

<sup>9</sup> Assuming other factors effecting mortality to be equal across density categories.

To illustrate the results we use the findings for a biennial screening cycle and examine the sum of screen detected and interval cancers. For a two-year period in women aged 40-49 we found that there would be an average rate of 3.4 invasive cancers per 1,000 with the corresponding rates by density of A: 2.3, B: 2.8, C: 3.9 and D: 5.3 for relative rates (compared to the average) of A: 0.68, B: 0.82, C: 1.15 and D: 1.56 respectively. For a two-year period in women aged 50-74 we found that there would be an average rate of 5.5 invasive cancers per 1,000 with the corresponding rates by density of A: 4.1, B: 5.8, C: 6.9 and D: 5.2 for relative rates (compared to the average) of A: 0.75, B: 1.08, C: 1.25 and D: 0.95 respectively. The non-regularity of the risk with increasing density in the 50-74 age group likely results from confounding with other breast cancer risk factors, particularly age.

For a two-year period in women aged 40-49 we found an average rate of 1.4 interval cancers per 1,000 with the corresponding rates by density of A: 0.68, B: 0.97, C: 1.73 and D: 2.53 per 1,000 giving relative rates (compared to the average) of A: 0.49, B: 0.69, C: 1.24 and D: 1.81. For a two-year period in women aged 50-74 we found an average rate of 1.56 interval cancers per 1,000 with the corresponding rates by density of A: 0.85, B: 1.08, C: 2.54 and D: 3.12 per 1,000 for relative (compared to the average) rates of A: 0.54, B: 0.70, C: 1.63 and D: 2.0.

As a point of comparison, among women aged 40-49, the rate of overall invasive cancer rates over 2-years in the *high risk* group was 7.25 per 1,000 for a ratio of 2.13 (7.25/3.4) and 3.13 per 1,000 for interval cancers for a ratio of 2.24 (3.13/1.4). For high-risk aged 50-74 the ratio of overall invasive cancer in the high risk was 8.08 per 1,000 for a ratio of 1.47 (8.08/5.5) and 2.30 per 1,000 for interval cancers for a ratio of 1.47 (2.30/1.56).

## **False Positives**

Density has the potential to influence the likelihood of a false-positive screening mammogram. This effect was examined in the data (Appendix Table 5). The false-positive rate decreased with age, was higher in those whose previous mammogram was false-positive, was higher for intervals over 30 months in length and did not vary by risk status. The rate of false-positives did vary by density and tended to increase with density. Most notably for those classified as BIRADS A the rate was the only category of all the factors considered for which the false-positive rate was less than 5%. While the rate of false-positives does vary with density the variance is comparable to the influence of other factors and does not suggest that the effect of density on false-positives would influence the use of mammography. This analysis is based upon screening rounds so that no first screens are included.

## Prognosis

The effect of screening on breast cancer is to “convert” incident cases into screen detected cancers so that the benefit from screening accrues due to the prognosis of screen detected cancers (following treatment) being better than if they were diagnosed as incident cancers. Any influence of density on the prognosis of cancers can therefore alter the benefit of screening in addition to density’s effect upon the likelihood that a cancer can be screen detected. To examine any influence two prognostic factors were selected:

1. Tumour size (invasive only) categorized as  $\leq 15\text{mm}$  versus  $>15\text{mm}$
2. Axillary Nodal Status: Positive versus negative

Tumour size is expected to be strongly related to screen detection and will be strongly influenced by lead time bias. Whilst nodal status will also be affected by lead time bias it would be anticipated that this effect would be less direct.

Invasive cases were selected from the first data set and attention restricted to SD cancers diagnosed in a biennial screening round (18-30 months) and interval cancers diagnosed prior to 24 months. No attempt was made to standardize the distribution of interval cancers to the risk of interval cancer<sup>10</sup>. The distribution of tumour size and nodal status was examined by mode of detection (SD and interval) and by BIRADS density. The results are given in Appendix Table 6. A major difference is seen between SD and interval cancers. For SD versus interval cancers: proportion  $>15\text{mm}$ , 65% versus 31% and for positive nodes, 34% versus 19%. Density did not influence the prognostic profile of interval cancers but did influence the prognostic profile of screen detected cancers with cancers screen detected in dense breasts being larger and with higher rates of nodal involvement than those detected in non-dense breasts (Appendix Table 6) using A+B v C+D as groupings, for size  $>15\text{ mm}$ , 27% versus 39% and for positive nodes 16% versus 27%, respectively. However, the principal difference in prognostic profile is between the screen detected and the interval cancers with interval cancers being larger and having higher rates of nodal involvement.

---

<sup>10</sup> Because of the frequency of annual screening cancers occurring in the first 12 months will be overrepresented in the interval cancers occurring before 24 months.

## Stability of BIRADS Density Measurement

When using BIRADS classified breast density the stability of the measurement is an important consideration since serial changes in individual classification present challenges in the delivery of appropriate services. Several studies have examined reproducibility of classification which provides an indication of consistency. In practice, consecutive mammograms individually report density so that density estimates are subject not just to intra- and inter-reader variability, but also variability associated with mammography equipment and changes in density which may be associated with participants hormonal influences, aging etc. It is thus likely that density measurements will be less stable than indicated by reproducibility studies alone.

To study stability of a sample of consecutive mammogram density classification data was extracted from the BCCBS database. The data was as follows:

- The first two mammograms performed on an individual performed between 1<sup>st</sup> January 2011 and 31<sup>st</sup> December 2013
- Both mammograms had to have density recorded without any intervening mammograms
- Both performed between 40 and 74 years of age
- Both performed on either analogue or digital equipment

There were 48,254 pairs of mammograms which satisfied the eligibility criteria. As anticipated there was a high level of stability with both measurements providing the same density classification with 65.7% (31,692 of 48,254) being unchanged and 98.1% (47,339 of 48,254) being within  $\pm 1$  category (Table 7).

In clinical use the important outcome is to classify breasts as dense or not. As previously noted the designation of density may correspond to C+D (US practice) or to D alone. For C+D 80.3% (11,403 of 14,202) of those initially dense were classified as dense on the second mammogram. Defining D as dense, then 63.7% (1,804 of 2,833) of those initially dense were classified as dense on the second mammogram (Appendix Table 7). Since density generally declines with age it is anticipated that the average density would be lower on the second mammogram than on the first although the reverse trend was observed. This was likely due to some early adoption of the 5<sup>th</sup> edition of BIRADS which is associated with increased density compared to the 4<sup>th</sup> edition. Analysis by age (data not shown) found that this trend was consistent at all ages. The distribution of density in this sample (stability sample) was similar to that sample used to examine risk.

From these results it is clear, that, in this data, using C+D as the definition of dense results in a classification which is more stable (than that based upon D alone) since the proportion changed is lower between the two measurements. However, while the proportion changed is greater for D (as dense) the number changed is higher using C+D as dense. For example, based on the measurement on the first mammogram of C or D, 2,799 (5.8% of the total sample) classified as dense would be classified as not dense on the second mammogram, compared to 1,029 (2.1% of the total sample) when D alone is used. As density seems to increase when using the 5<sup>th</sup> Edition of BIRADS (versus 4<sup>th</sup>) it may be anticipated that the absolute variability may change when used

in the population. While density is not anticipated to be constant, the comparative frequency of change would suggest that attention would need to be paid to its application for screening.

One factor which may influence the classification of density of individual mammograms is the reporting radiologist. It can be anticipated that inter-radiologist variability will exceed intra-radiologist variability. To examine this information on reporting radiologist was extracted from the data (contained in Table 7) and is reported in Table 8 where the reporting radiologist was the same for both mammograms in the pair. There were 15,303 pairs of mammograms which satisfied the eligibility criteria so that 31.5% of the stability sample had both mammograms reported by the same radiologist.

As anticipated there was a higher level of consistency when the same radiologist reported with both measurements providing the same density classification with 74.0% (compared to 65.7% for unmatched radiologists) being unchanged and 98.8% (versus 98.1%) being within  $\pm 1$  category (Appendix Table 8). This indicates that consistency is increased for the same reporting radiologist. When different radiologists reported the two mammograms the results for density classification were 61.8% for unchanged and 97.8% being within  $\pm 1$  category.

Results were also analysed by woman's age at the first mammogram in the pair and are presented in Appendix Table 9. This indicates that there is an influence of age with lower proportions changed for younger women (using either D or C+D as dense). However, the influence of age is fairly weak. The results also indicate that in current screening participation and practice in BC the majority of mammograms classified as dense (either BIRADS D or C+D) occur in screens performed in women 40-49 and for BIRADS D 65% of cases are in women 40-49. This pattern is also seen in Appendix Table 2.

It must be noted that the circumstances under which BIRADS density was captured in BCCBS may be unrepresentative of performance for any clinical use in the future. Clinical management of women was not dependent on BIRADS density and reporting was voluntary. Data was selected to be drawn from a period for which the 4<sup>th</sup> Edition of BIRADS was most likely to be used and performance with the current version (5<sup>th</sup> 2013) may be different.



## Discussion

The analysis presented was on screening rounds so that the population distribution by density presented would not correspond to the population average but is weighted by relative attendance rates during the study period and the screening centres which recorded BIRADS. BIRADS density assessment was not used for clinical management within the BCCBS so it is always possible that distributions could change if it were to be used clinically. Anticipated relationships were seen between age, ethnic status and reported density.

Reported relationships between breast cancer risk and risk of interval breast cancer were generally seen in the BCCBS data. Literature reports often provide larger estimated values for relative rates, but this is likely due to variable definition of the baseline and “dense” categories and control for other factors known to influence breast cancer risk, some of which are negatively correlated with density (e.g. age). By using the average for screening participants as the referent value the resulting value can be directly interpreted using the reported BCCBS rates. Interestingly, in the population interval cancer rates differ little by age with a steep gradient with age for screen detected cancer. Whereas for density the steep gradient with age is for interval cancer. This does provide indirect support for similar screening frequencies across different age ranges but that screening approaches may have to be tailored by density.

One approach used within some Canadian Provinces, which recommend biennial mammography screening, is to screen women with >75% density (similar to BIRADS D) on annual basis. One would anticipate that, for this to be useful, that the proportion of post-screen cancers would be reduced compared to that for biennial screening. For women with BIRADS D aged 40-49 the ratio (SD/SD+PS) for biennial versus that for annual is 0.52 versus 0.51, and for 50-74 is 0.40 versus 0.44, so that there is not much evidence of considerable change. In contrast, for high risk women, in whom annual mammography screening is recommended in British Columbia and elsewhere in Canada, for those aged 40-49 the ratio (SD/SD+PS) for biennial versus that for annual is 0.57 versus 0.72 and for 50-74 is 0.72 versus 0.83. While for women aged 40-49 the rate of interval cancer was higher for those at high risk than any of the density categories (Table 4) this was not true in the older age group where the interval cancer rate was higher in both C and D categories than in the high-risk group.

Density was found to influence the frequency of false positive screens (Appendix Table 5), but the effect was not of sufficient magnitude to influence screening approaches. The proportion of false positives, for any density category, did not exceed those seen on first screens (BCCBS Annual Report 2016).

A major difference is seen in the prognostic profile between SD and interval cancers (Table 6). For SD versus Interval on proportion >15mm, 65% versus 31% and for positive nodes, 34% versus 19%. The results appear to indicate that density has little if any effect on the prognostic profile of interval cancers. Since interval cancers are detected by other means than mammography it is understandable how their profile would not be strongly affected by a factor that influences mammography sensitivity. On the other hand, density does have some influence on the prognostic profile of screen detected cancers with cancers screen detected in dense breasts being larger and with higher rates of nodal involvement than those detected in non-dense breasts

(Appendix Table 6). This suggests that screen detection in women with denser breast may offer less benefit than screen detection in women with less dense breasts.

Using C+D to constitute “dense breasts” 80.3% (11,403 of 14,202) of those initially dense were classified as dense on the second mammogram whereas using D to represent dense breasts then 63.7% were classified as dense on the second screen (Appendix Table 7). Although C+D is more reproducible proportionately it leads to more changes in designation: using C or D, 2,799 (5.8% of the total sample) classified as dense would be classified as not dense on the second mammogram, compared to 1,029 (2.1% of the total sample) when D alone is used. Consistency was improved when the same radiologist reported both mammograms but this occurred in less 1/3 of cases. Inconsistency will tend to weaken the measured relationship between breast density and breast cancer risk.

**Appendix Table 1: Distribution of Factors by Screening Rounds**

Factor	Level	Count	%
First Screen Prior to Round	No	435,389	89.7
	Yes	49,986	10.3
High Risk	No	403,830	83.2
	Yes	81,545	16.8
Age at Beginning of Screening Round	40-44	57,190	11.8
	45-49	86,107	17.7
	50-54	80,631	16.6
	55-59	81,425	16.8
	60-64	76,359	15.7
	65-69	62,354	12.9
	70-74	41,309	8.5
Screen Result at Prior Screen	Normal	446,005	91.9
	Abnormal	39,370	8.1
Image Type of Prior Screen	Analogue	248,794	51.3
	Digital	236,581	48.7
Ethnic Group	East Asian	63,620	13.1
	First Nations	10,646	2.2
	Other	411,109	84.7
BIRADS Density at Prior Screen	A	140,370	28.9
	B	178,180	36.7
	C	128,835	26.5
	D	37,990	7.8
Cancer Detected in Round	Screen Detected - DCIS	291	13.1
	Screen Detected - Invasive	1,022	45.9
	Post-Screen - DCIS	95	4.3
	Post-Screen - Invasive	817	36.7

**Appendix Table 2:** Distribution of Screening Rounds by Density Cross-tabulated with Age, Ethnic Group and High-Risk Status

Factor	Level	Total Rounds (Column %)	Density (Row %)			
			A	B	C	D
Age	40-44	57,190 (11.8%)	9,806 (17.2)	17,533 (30.7)	21,255 (37.2)	8,596 (15.0)
	45-49	86,107 (17.7%)	16,568 (19.2)	27,544 (32.0)	30,609 (35.6)	11,386 (13.2)
	50-54	80,631 (16.6%)	20,121 (25.0)	29,249 (36.3)	24,238 (30.1)	7,023 (8.7)
	55-59	81,425 (16.8%)	25,953 (31.9)	31,747 (39.0)	19,121 (23.5)	4,604 (5.7)
	60-64	76,359 (15.7%)	27,267 (35.7)	30,299 (39.7)	15,642 (20.5)	3,151 (4.1)
	65-69	62,354 (12.8%)	24,023 (38.5)	25,014 (40.1)	11,216 (18.0)	2,101 (3.4)
	70-74	41,309 (8.5%)	16,632 (40.3)	16,794 (40.7)	6,754 (16.4)	1,129 (2.7)
Ethnic Group	East Asian	63,620 (13.1%)	8,606 (13.5)	17,275 (27.2)	25,450 (40.0)	12,289 (19.3)
	First Nations	10,646 (2.2%)	4,455 (41.9)	3,940 (37.0)	1,873 (17.6)	378 (3.6)
	Other	411,109 (84.7%)	127,309 (31.0)	156,965 (38.2)	101,512 (24.7)	25,323 (6.2)
High Risk	No	403,830 (83.2%)	117,204 (29.0)	147,234 (36.5)	107,199 (26.6)	32,193 (8.0)
	Yes	81,545 (16.8%)	23,166 (28.4)	30,946 (38.0)	21,636 (26.5)	5,797 (7.1)

**Appendix Table 3:** Risk of Screen Detected (SD) and Post-Screen (PS) Invasive Breast Cancer by Age, High-Risk Status and Density

Factor	Level	Rates of Cancer per 1000								
		Annual Screening			Biennial Screening			Triennial Screening		
		SD	PS	Annualized (SD+PS)	SD	PS	Annualized (SD+PS ÷2)	SD	PS	Annualized (SD+PS ÷3)
Age*	40-49	1.18	0.57	<b>1.75</b>	1.86	1.40	<b>1.63</b>	2.27	2.49	<b>1.59</b>
	50-59	1.77	0.50	<b>2.27</b>	2.78	1.31	<b>2.05</b>	3.39	2.30	<b>1.90</b>
	60-74	3.33	0.48	<b>3.81</b>	5.23	1.51	<b>3.37</b>	6.38	2.50	<b>2.96</b>
Density*	A	1.41	0.15	<b>1.56</b>	3.29	0.82	<b>2.06</b>	3.67	1.48	<b>1.72</b>
	B	1.81	0.43	<b>2.24</b>	4.21	1.06	<b>2.64</b>	4.70	1.70	<b>2.13</b>
	C	1.61	0.69	<b>2.30</b>	3.75	2.26	<b>3.01</b>	4.19	4.34	<b>2.84</b>
	D	1.13	1.64	<b>2.77</b>	2.64	2.84	<b>2.74</b>	2.95	4.19	<b>2.38</b>
High Risk	No	1.83	0.51	<b>2.34</b>	3.56	1.42	<b>2.49</b>	4.24	2.45	<b>2.23</b>
	Yes	3.27	0.84	<b>4.11</b>	6.35	2.46	<b>4.41</b>	7.56	3.77	<b>3.78</b>

\*The high-risk screening rounds are removed in the analysis of these factors

**Appendix Table 4:** Risk of Screen Detected and Post-Screen Invasive Breast Cancer by Age, High-Risk Status and Density for age 40-49 and 50-74.

		Rates of Cancer per 1000								
		Annual Screening			Biennial Screening			Triennial Screening		
Factor	Level	SD	PS	Annualized (SD+PS)	SD	PS	Annualized (SD+PS ÷2)	SD	PS	Annualized (SD+PS ÷3)
Age 40-50 at Screening Round Entry										
Density*	A	0.85	0.09	<b>0.94</b>	1.58	0.68	<b>1.13</b>	2.00	1.09	<b>1.03</b>
	B	0.98	0.49	<b>1.47</b>	1.82	0.97	<b>1.40</b>	2.30	1.56	<b>1.29</b>
	C	1.15	0.55	<b>1.70</b>	2.15	1.73	<b>1.94</b>	2.72	3.87	<b>2.20</b>
	D	1.50	1.46	<b>2.96</b>	2.79	2.53	<b>2.66</b>	3.52	3.38	<b>2.30</b>
High Risk	No	1.13	0.57	<b>1.70</b>	1.96	1.40	<b>1.68</b>	2.33	2.49	<b>1.61</b>
	Yes	2.37	0.93	<b>3.30</b>	4.12	3.13	<b>3.63</b>	4.90	3.80	<b>2.90</b>
Age 50-74 at Screening Round Entry										
Density*	A	2.39	0.17	<b>2.56</b>	3.46	0.85	<b>2.16</b>	3.96	1.57	<b>1.84</b>
	B	3.23	0.41	<b>3.64</b>	4.69	1.08	<b>2.89</b>	5.36	1.72	<b>2.36</b>
	C	3.04	0.80	<b>3.84</b>	4.40	2.54	<b>3.47</b>	5.04	4.51	<b>3.18</b>
	D	1.45	1.84	<b>3.29</b>	2.11	3.12	<b>2.62</b>	2.41	5.17	<b>2.53</b>
High Risk	No	2.78	0.49	<b>3.27</b>	4.01	1.42	<b>2.72</b>	4.96	2.40	<b>2.45</b>
	Yes	4.00	0.82	<b>4.82</b>	5.78	2.30	<b>4.04</b>	7.14	3.78	<b>3.64</b>

\*The high-risk screening rounds are removed in the analysis of these factors

**Appendix Table 5:** False Positive outcomes for Screening Rounds by Age, Density, High Risk, Previous Abnormal and Length of Interval

Factor	Level	Results of Screening Rounds		
		Number	% Normal	% False Positive
Age	40-44	35,564	92.0	<b>8.0</b>
	45-49	56,396	92.2	<b>7.8</b>
	50-54	46,324	92.9	<b>7.1</b>
	55-59	47,658	94.2	<b>5.8</b>
	60-64	46,301	94.2	<b>5.8</b>
	65-69	37,397	94.3	<b>5.7</b>
	70-74	21,589	94.7	<b>5.3</b>
Density	A	88,526	95.1	<b>4.9</b>
	B	105,985	93.2	<b>6.8</b>
	C	74,595	92.0	<b>8.0</b>
	D	22,123	92.4	<b>7.6</b>
High Risk	No	227,138	93.4	<b>6.6</b>
	Yes	64,091	93.3	<b>6.7</b>
Previous Abnormal	No	270,310	93.6	<b>6.4</b>
	Yes	20,919	90.7	<b>9.3</b>
Length of Screening Round (months)	<18	111,967	93.4	<b>6.6</b>
	18≤ & <30	158,761	93.6	<b>6.4</b>
	30≤ & <42	20,501	92.1	<b>7.9</b>

**Appendix Table 6:** Distribution of Tumour Size ( $\leq 15$  mm v  $>15$  mm) and Nodal Involvement (No/Yes) by Mode of Detection (SD at 18-30 months v Interval  $\leq 24$  months) and BIRADS Breast Density for Invasive Cancers

	Mode of Detection						Overall*	
	Screen Detected 18-30 months			Interval Diagnosed $<24$ months				
Density	Number	% $>15$ mm	% Node +	Number	% $>15$ mm	% Node +	% $>15$ mm	% Node +
A	143	27	11	71	58	25	33	14
B	213	27	19	116	61	42	34	24
C	117	39	27	163	67	27	50	27
D	21	38	24	70	64	40	52	32
A+B+C	473	30	19	350	65	33	39	22
A+B	356	27	16	187	64	36	34	20
C+D	138	39	27	233	66	33	50	28
ALL	494	31	19	420	65	34	40	25

\*weighted of screen detected and interval cancers using rates from Table 3



**Appendix Table 7:** Results of paired BIRADS measurements of Breast Density

Density on First Mammogram	Density on Second Mammogram				Total
	A	B	C	D	
A	11,850	3,888	367	42	16,147 (33.5%)
B	3,558	10,954	3,184	209	17,905 (37.1%)
C	208	2,502	7,084	1,575	11,369 (23.6%)
D	29	60	940	1,804	2,833 (5.9%)
Total	15,645 (32.4%)	17,404 (36.1%)	11,575 (24.0%)	3,630 (7.5%)	48,254 (100%)

**Appendix Table 8:** Results of paired BIRADS measurements of Breast Density where the Reporting Radiologist is the same.

Density on First Mammogram	Density on Second Mammogram				Total
	A	B	C	D	
A	4,473	807	59	18	5,357 (35.0%)
B	789	3,857	877	70	5,593 (36.6%)
C	31	587	2,278	533	3,429 (22.4%)
D	5	4	201	714	924 (6.0%)
Total	5,298 (34.6%)	5,255 (34.3%)	3,415 (22.3%)	1,335 (8.7%)	15,303 (100%)

**Appendix Table 9.** For women with BIRADS C or D at Earlier Screen: Proportion Classified as Dense at Second Screen by Age Group at Earlier Screen.

Age at First Mammogram	Result on First Mammogram		Second Mammogram Classified as Dense			
	Number C	Number D	Dense = D		Dense = C+D	
			Number	%	Number	%
40-44	2,313	756	497	65.7	3,069	84.4
45-49	3,267	1,082	706	65.2	4,349	82.4
50-54	1,753	412	250	60.7	2,165	79.1
55-59	1,362	234	145	62.0	1,596	75.6
60-64	1,175	169	100	59.2	1,344	76.3
65-69	799	99	56	56.6	898	76.9
70-74	487	57	36	63.2	419	77.0

## References

(1) Melnikow J, Fenton JJ, Whitlock EP, Miglioretti DL, Weyrich MS, Thompson JH, et al. Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the US Preventive Services Task Force. *Supplemental Breast Cancer Screening in Women With Dense Breasts*. *Ann Intern Med* 2016;164(4):268-278.