Comparative effects of mammography-based screening strategies: summary and minor adaptations of 2023 Draft USPSTF review findings

Jennifer Pillay, Samantha Guitard, Lisa Hartling
Alberta Research Centre for Health Evidence, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada

Note: This summary was based on the draft USPSTF evidence review posted in 2023. The final version of this review was published in April 2024. One new non-randomized study (n=504,863; Sprague et al. doi: 10.1148/radiol.223142) from the US was included that compared digital breast tomosynthesis (DBT) alone with mammography alone and reported on screen-detected invasive cancers, advanced stage cancers, interval cancers and false positive (FP) recalls and biopsies. The review conclusions for FP recalls changed from no difference between modalities to lower with DBT versus mammography, with the certainty remaining at low. This study reported significantly lower FP rates for DBT at round 1 (absolute reduction of 3.4% [2.2 to 4.7%]) and round 2 (1.8% [0.7 to 3.0%) [both $p<0.05$], but not at round 3 or greater (1.5% [0.2 to 2.3%). Other data did not change the conclusions or direction of effect for other outcomes.

Introduction:

The Canadian Task Force on Preventive Health Care (task force) is updating their recommendations on screening for breast cancer, which will be informed by updates of the evidence on the benefits and harms of screening (versus no screening) and on the values and preferences about screening (i.e., the relative importance placed by patients on the outcomes). Findings from the draft evidence review informing the recent recommendations by the United States Preventive Services Task Force (USPSTF) on the comparative effects of different mammography-based screening strategies will also be considered. For more details on this process, please see the Research Plan and review protocols available on Open Science Framework: https://osf.io/xngsu/. This report describes the work and results of the summary and adaptation of the USPSTF review (Key Question 2 in the Research Plan).

Methods:

The Edmonton ERSC reviewed the draft (May 2023) evidence review[1] conducted for the USPSTF by the Kaiser Permanente Evidence-based Practice Centre (EPC) (led by Dr. Jillian Henderson) on the comparative effects of mammography-based screening strategies, and modified their Summary of

Evidence tables to i) add details about the studies (e.g. study time periods, specifics on comparisons) and findings (adding quantitative results, subgroup analysis findings), ii) add ratings of indirectness of the evidence to the strength of evidence (SOE) assessments to better align with the GRADE approach, and iii) provide narrative statements about the effects and their certainty for each outcome. Overall, there was reliance on the conduct and reporting of the EPC review authors for all data extraction, analysis, and assessment of study quality. The ERSC only used the data and summaries for outcomes of interest to the task force, but after working group input they added data for screen-detection of invasive cancers; this data could possibly be used as a surrogate together with stage distribution of invasive cancers (reduced advanced stage) and interval cancers to provide indirect evidence for breast-cancer mortality reduction. The ERSC communicated with the EPC review lead several times to get a thorough understanding of the rationale for the SOE assessments. The review lead confirmed that, at the time of their communications, there was no new RCT evidence or major changes to the conclusions between the draft and final (not publicly available as of Dec 4, 2023) versions of the report.

In term of indirectness, a major consideration as per discussion with the task force's Breast Cancer working group was the timing of the data collection for mortality and cancer treatment-related outcomes (e.g., receipt of chemotherapy as a surrogate for treatment morbidity from more advanced stage disease); data collected before 2014 was considered indirect to capture effects that may have changed due to better curative treatment outcomes since that time. Findings on stage distribution or treatment-related morbidity that relied on case-only analysis (only using data for those with cancer) was also rated down because these do not account for any differences in cancer incidence between groups. The timing of data collection for outcomes related to the accuracy, such as advanced cancer diagnosis and interval cancers, of digital versus film (assumed to be used pre-2014) mammography was not considered a serious consideration based on evidence showing fairly similar accuracy between these two methods.[2] Though there were expectations that higher rates of FPs may be reported by studies in the US versus Canada, if focusing on the *relative* effects between different screening strategies, the evidence on FPs was not considered to have serious indirectness. Otherwise, the ERSC examined indirectness in comparisons (e.g., when the comparison of interest was not used at each round of screening). In cases where the evidence was limited/applicable to a particular duration (e.g., 2 screening rounds) or population (e.g., age, dense breasts), the evidence was not rated down for indirectness but this applicability was noted in the narrative statements. The evidence for screen-detection of invasive cancers was not rated down for indirectness (in relation to the outcomes of interest by the task force), and should be interpreted with this in mind.

The USPSTF review authors did not apply thresholds of effect to determine whether any differences between the comparators across outcomes would be considered clinically important. Because

of this, the task force may later revise the conclusions when considering their thresholds (see KQ1 review).

Results:

Overview of USPSTF review methods

The focus of the USPSTF review was on the comparative effects between different mammography-based screening strategies. All eligible studies included a comparator group that received screening with mammography only in one of the study arms (digital [**DM**] and film mammography eligible); comparisons of interest were different screening ages (to start or stop screening), intervals, or screening eligibility criteria (i.e., based on personal risk for cancer including dense breasts or other means to predict risk), as well as alternative screening methods (e.g. digital breast tomosynthesis [**DBT**]), or additions to mammography (e.g., tomosynthesis, ultrasound or MRI).

Eligible study designs included RCTs, prospective and retrospective cohorts with a concurrent control and (for harms) nested case-control and cross-sectional studies from included trials or large population-based studies; they did include case-only analyses (all participants having cancer) for the stage distribution outcome. They excluded observational studies using paired designs (i.e., within-person comparisons) or from large registry or surveillance but using selected nonrepresentative subset of data, and were very strict for ecological/geographical exposure data (needing to have people with very similar characteristics enrolled in a very narrow [1 year] and concurrent time period). Nonrandomized studies were excluded if the comparison groups were highly selected based on factors that could influence breast cancer risk or health outcomes (e.g., family history or health status, breast density, access to health care, care seeking behaviors). For detection of cancer, stage distribution, FPs and FP biopsies, studies had to have at least two rounds of screening. Where multiple publications on similar analyses from the same registry (e.g., BCSC data) or observational cohort studies were available, the most recently analyzed data available were selected for inclusion in the review.

Poor quality studies were excluded; these were considered to have serious important limitations or a critical flaw that would likely affect the validity of study findings. For nonrandomized studies, a rating of poor-quality often resulted from: confounding (i.e., imbalances in baseline characteristics without proper statistical adjustment); no reporting of population characteristics by study arm; concerns about the classification of the intervention (e.g., self-reported screening interval, determination of diagnostic versus screening mammography); differences in follow-up procedures based on intervention arm; high or differential rates of attrition between groups; or evidence of possible selective reporting.

Searches were conducted in MEDLINE, Cochrane Central Register of Controlled Clinical Trials, and the Cochrane Database of Systematic Reviews, supplemented with reference lists of previous systematic reviews of breast cancer screening, through August 22, 2022. Ongoing surveillance methods

are used to find major new evidence and these new studies may change the results in the final review once available. The review team performed screening and assessments of risk of bias in duplicate and used verification for all data extraction. They assessed the SOE using EPC methods which considers study quality, precision, consistency and reporting bias.[3] Instead of rating the directness of the evidence and including this in the overall assessment of SOE, EPC authors comment on its applicability. The guidance is adapted from that used by GRADE, though does not at this time apply thresholds for decision-making; findings that are not statistically significant are interpreted as having no effect/difference regardless of the estimate of effect. Further, though they do not rate down evidence from a single study for "inconsistency" they typically rate down this evidence unless the single study is very well conducted and large. RCTs start at high SOE and nonrandomized studies (**NRSIs**) start at low SOE for benefits and moderate SOE for harms (which differs from GRADE). Noteworthy, the EPC took the perspective of adherence to/attenders of screening for the certainty assessments, whereas the task force typically prefers to assess the certainty of a "real world" perspective of screening programs where invitations/offers of screening are the main intervention of interest. This was evident for the one comparison in this review [supplemental MRI] where the effects were assessed using an offer/intention to screen population where adherence was suboptimal and findings were rated down; all other studies relied on data from attenders of screening. For the lowest ratings, their methods use the term "insufficient" quality though these have been converted in this report to "very low certainty" to reflect GRADE terminology. For outcomes were there was missing data there is notation of no data whereas this could also be interpreted as very low certainty evidence.

Overview of findings

The draft review included 19 studies[4-22], two of which were not used for this summary because they did not report outcomes of interest to the task force: one RCT[21] on DBT versus DM that to date has only reported on radiation exposure and rare harms, and one NRSI[22] on supplemental MRI that only reported on incidental findings. No RCTs but 13 NRSI were excluded due to poor quality (see **Appendix**), primarily due to confounding based on imbalances in baseline characteristics (without proper statistical adjustment), biased selection into study groups, and the absence of information on participant characteristics by study arm.

No RCTs reported on morbidity, mortality, or quality of life outcomes. No relevant benefit outcomes are yet reported for supplementation with MRI or US; findings relating to screen-detection of cancer or stage distribution were either not reported (MRI) or not eligible (US) due to the reporting after only one round of screening. Demographic characteristics of study participants were minimally described. Most studies included participants in their 40s to 60s, with two studies focusing on 40-49 years (triennial screening and supplemental ultrasound) and one on after age 70 years (age to stop). Only 6 of the 19 studies reported racial and/or ethnic characteristics (in five studies, White ethnicity was predominant

(73% to 92% of participants); one study focused on Hispanic/Latina (76%) participants (annual vs. biennial screening). Few subgroup findings were reported, though these have been extracted into the tables, and none were used for changes to the overall findings. The modified tables (following the textual summaries below) provide reference to tables or figures in the draft EPC evidence review that were used for the results. In the table, the narrative summaries use "may" when referring to low certainty evidence and "probably" when referring to moderate certainty evidence. In summary, adding concerns about indirectness only changed one certainty assessment (for screen-detection of invasive cancer from DBT vs. DM) but several of the summary statements include notation about the applicability of the findings.

The **Appendix** describes relevant studies/reviews conducted in Canada that were not included in this review or examined by the task force but were submitted for consideration by the task force by Canadian stakeholders.

Age to start screening

No evidence was included.

Age to stop screening

One fair-quality US NRSI[4] (n=1,058,013; data 1999-2008) examined continuation of screening versus stopping beyond 70 and 74 years among those with a high probability of living ≥10 more years. BC mortality, treatment-related morbidity (via need for aggressive treatments), and overdiagnosis were reported. The data used for treatment-related morbidity was used by the review authors to help capture overdiagnosis (and overtreatment), whereby if more ductal carcinoma in situ (DCIS) was found in the group continuing to screen this may reflect an increased receipt of lumpectomy with/without radiation. Despite study estimates of fewer deaths in those continuing to screen at age 70-74 years (8-year risk: 1 fewer death per 1,000 women [95% CI -2.3 to 0.1]; 2.7 vs. 3.7 per 1000), less aggressive treatments received (for 70-74 and 75-85 year age groups continuing screening), and overdiagnosis (using 8-year cumulative incidence) with more cancers diagnosed in those continuing to screen (14-18 per 1000 more [95% CI NR]; both groups), the evidence for all outcomes was rated as very low certainty due to risk of bias and imprecision. The ERSC added indirectness for the mortality and treatment outcomes due to the age of the study (and due to the case-only analysis for treatment morbidity) but this did not change the certainty of evidence. In summary, there is very low certainty evidence (for several outcomes due to no data) across all outcomes of interest to the task force for determining age to stop screening.

Annual versus biennial screening

Three fair-quality US NRSIs[5-7] were included and one or more studies reported on stage distribution of invasive cancers, FPs, FP biopsies, and interval cancers among those screening annually versus biennially. Evidence on stage distribution (n=15,440; data 1996-2012)[5] showed no significant differences in risk of stage IIB+ (range of adjusted RRs across age groups 0.98 to 1.17) or less favorable prognosis

(range of RRs 1.03 to 10.7) cancers for any age group and was rated as very low certainty for risk of bias and imprecision. The ERSC added indirectness from the use of case-only analysis but this did not change conclusions. Findings were of moderate certainty for increases in both FPs (n=903,495 [data 2005-2018][6]: about 50% vs. 35% people with FPs over 10 years [excluding prevalence screens; using discrete-time survival modeling based on round specific probabilities and accounting for censoring]; n=2,019 [data 2014-2015][7]: OR 2.2 [95% CI 1.7 to 2.8] over 9 years) and FP biopsies (n=903,495: approximately 115 vs. 66 per 1000 over 10 years)[6] with annual versus biennial screening. The large NRSI[6] reported rates for DM and DBT separately and found no differences between those screening methods. For interval cancers (n=15,440: 22.2% with annual vs. 27.2% with biennial DM over at least 2 screens)[5] there was very low certainty from risk of bias and imprecision. The addition of indirectness from using a case-only analysis did not change the certainty or conclusions. In summary, there is moderate certainty that FPs and FP biopsies probably increase by at least 1.5-fold with annual screening and very low certainty (mainly from missing data) for all other outcomes.

Annual versus triennial

One fair-quality RCT (n=76,022; data 1989-1996)[8] and one fair-quality NRSI (n=14,765; data 1985-1995)[9] reported various outcomes from examining annual versus triennial screening with DM. No significant differences were found by the NRSI for BC or all-cause mortality among 40 to 49-year-olds; findings were of very low certainty due to risk of bias, imprecision and indirectness (pre-2014 data for mortality outcomes). The RCT never reported on mortality outcomes as planned. This trial, studying 50 to 62-year-olds, found an increase in screen-detected invasive cancers (4.42 per 1000 versus 2.70 per 1000; RR: 1.64 [95% CI, 1.28 to 2.09]), no difference in incidence of invasive cancers over 3 years (6.26 per 1000 versus 5.4 per 1000; RR: 1.16, 95% CI 0.96 to 1.40), and no evidence of stage shift via similar rates and no statistical differences in tumor size, nodal status, grade, or prognostic index for all invasive cancers; findings for these three outcomes were rated as low certainty from risk of bias and evidence from a single study. Annual versus triennial screening was rated as having low certainty evidence for slightly fewer interval cancers over 3 years, relying mostly on the RCT findings (1.8 vs. 2.7 per 1,000 screened; RR: 0.68 [95% CI 0.50 to 0.92]) but noting some inconsistency with the findings of no difference in the NRSI. For overdiagnosis, the NRSI found no significant difference in *invasive* breast cancer incidence over mean 9.8 years (141.1 vs. 144.0 per 100,000 person-years; RR: 0.98 [95% CI 0.75 to 1.29]), but with very low certainty. Calculations of overdiagnosis often compare rates of all diagnoses, including DCIS. In summary, there was low certainty of some increase in screen-detection of invasive cancers and slightly fewer interval cancers, but no difference in stage shift for 50 to 69-year-olds screening annually versus triennially. No data on FPs or FP biopsies were reported and evidence on mortality and overdiagnosis was of very low certainty.

Digital breast tomosynthesis versus digital mammography

Three RCTs[10-12] and 6 NRSIs[6,13-17] compared DBT with DM. One good-quality RCT (RETomo; Italy [2014-2017]; n=26,877; 45-69 years),[12] one fair-quality RCT (Proteus Donna; Italy [2004-2017]; n=73,866; 46-68 years),[10] and one fair-quality NRSI (Norway [2014-2017]; n=98,927; 50-69 years)[13] compared DBT and DM with DM in the first round and used DM for both groups at round two. Another good-quality RCT (To-BE; Norway [2016-2020]; n=28,749; 50-69 years)[11] compared DBT (or synthetic DM [sDM]) with DM at round one and used DBT/sDM at round two. These four studies only reported on screen-detection of invasive cancer, stage shift, FPs, and interval cancers. One other NRSI (US; n=903,495)[6] reported on FPs and FP biopsies over 10 rounds (excluding prevalence screens; using discrete-time survival modeling based on round specific probabilities and accounting for censoring), and four other NRSI (3 US and 1 Sweden)[14-17] contributed to findings on interval cancers.

At round one, the three RCTs found that DBT detected more invasive cancers than DM (RR 1.41, 95% CI 1.20 to 1.64, n = 129,492) with absolute differences ranging from 0.6 to 2.4 more per 1000 screened; similar results were seen in the NRSI (2.3 more per 1000 screened; RR 1.43, 95% CI 1.22 to 1.67; unadjusted). At round two, the trials found no significant difference (RR 0.87, 95% CI 0.73 to 1.05), whereas the NRSI found lower detection for the DBT group (1.3 fewer per 1000 screened; RR 0.71, 95% CI 0.55, 0.92); unadjusted). Subgroup findings in two of the RCTs suggested that detection among those with dense breasts may have been lower at round two in the group receiving DBT at round one. The EPC rated this evidence as moderate SOE (due to inconsistency) for more detection with DBT, but the ERSC rated the evidence down further to low certainty due to the use of the same screening modalities in both groups at the second round (i.e., indirect comparison). For stage shift, the three RCTs found no differences in stage II or higher cancers between modalities at either round, one RCT[12] found no differences in stage III or higher at either round, and the RCTs and NRSI found no differences in tumor characteristics that inform staging such as tumor diameter, histologic grade, and node status. The EPC authors noted low power with the studies to detect these differences and the evidence was rated as low certainty due to risk of bias and imprecision. The ERSC replaced risk of bias concerns for indirectness from use of the same device at round two in both groups.

For FPs, the RCTs had mixed findings in round one; in round two, findings were consistent for no difference as may be expected from using the same device for both groups. The NRSI[6] using incident screens suggested slightly lower FP recall over 10 years with DBT versus DM with an annual interval (50% versus 56%) and similar rates with biennial screening (36% versus 38%). Risk of bias and inconsistency led to conclusions of low certainty for little to no difference in FPs. One RCT[11] and one NRSI[6] (estimating events over 10 cumulative rounds) found no difference in FP biopsies, but findings were of low certainty due to risk of bias and imprecision. Three RCTs[10-12] and 5 NRSI[13-17] reported on

intervals cancers. No difference was found in pooled analysis of the RCTs (RR = 0.87, 95% CI 0.64 to 1.17), whereas the NRSIs had mixed findings thought in part due to differing definitions and timepoints for outcome measurement. No subgroup effects were found based on age or breast density. The evidence for little-to-no difference in interval cancers was rated to have moderate certainty from imprecision (i.e., wide 95% CI).

In summary, DBT versus DM may lead to more screen-detected invasive cancer over two rounds of screening, but probably does not lead to fewer interval cancers and may not reduce advanced stage cancers. DBT may make little-to-no difference in FPs or FP biopsies. There were no data on mortality outcomes or overdiagnosis.

Supplemental MRI

One good-quality RCT (DENSE; The Netherlands [2011-2016]; n=40,373; 50-75 years; 100% extremely dense breasts)[18] compared an invitation to MRI versus no invitation after a negative screening mammogram result among those with extremely dense breasts. The trial is examining three rounds of screening but has only reported comparative effects on interval cancers after round one. Results found reduction in invasive interval cancers (2.2 vs. 4.7 per 1,000 invited to screening, RR 0.47, 95% CI 0.29 to 0.77) and any interval cancer (2.5 vs. 5.0 per 1000; RD -2.5, 95% CI, 1.0 to 3.7) after round one; follow-up was 2 years to identify the interval cancers. The EPC rated the evidence as low certainty, due to reliance on a single study and because of study limitations from the 59% adherence in the supplemental MRI group which was judged to possibly lead to underestimated effects. The specificity to one round and a population with extremely dense breasts was added to the summary statement rather than rating down more for indirectness due to timing or the population. No other outcomes have been reported for this trial. In summary, supplementing mammography with MRI for those with extremely dense breasts may reduce interval cancers. There were no data to examine on other potential harms or benefits.

Supplemental ultrasound

One fair-quality RCT (J-START; Japan [2007-2011]; n=72,717; 40-49 years; 58% with dense breasts among a sub-sample n=19,213)[19] and one fair-quality NRSI (US [2000-2013]; n=18,562; 30-80+ years; 65% BI-RADS 3/4 and 35% at “intermediate risk”)[20] were included. Authors of the RCT have only reported on one round of screening, therefore data related to screen-detected cancers were not eligible for the review. Neither study found a significant difference in interval cancers (RCT (invasive): 0.4 versus 0.8 per 1,000 screened; RR 0.58, 95% CI 0.31 to 1.08 and NRSI (invasive and DCIS): 1.5 vs. 1.9 per 1,000 screened; aRR 0.67, 95% CI 0.33 to 1.37) and this evidence was rated at low certainty due to risk of bias and imprecision. The trial provided data by breast density among a sub-set (n=19,213) of participants, and there were similar relative risks between intervention groups for each category (BI-RADs A/B and C/D). The ERSC added to the narrative statement that this finding was at one round. The

NSRI also reported on FP biopsies over about two screening rounds, finding higher rates in the group receiving supplemental US (52.0 vs 22.2 per 1000 screens; RR 2.23, 95% CI, 1.93 to 2.58); evidence was of low certainty due to risk of bias. In summary, supplementing DM with US may increase FP biopsies (possibly 2-fold over two rounds) and may not reduce interval cancers at the first round. No other outcomes were reported.

Canadian Task Force on Preventive Health Care

KQ2 Summary of Findings tables

Comments: i) red font are additions/revisions by ERSC to assessments or modifications to the outcome labelling; ii) the narrative statements in the What happens? column were developed by the ERSC based on the EPC wording and any notations added about the applicability; iii) cited Tables/Figures refer to those within the USPSTF draft report. When the outcomes for which there were no data are listed, the bold font refers to the outcomes considered critical or important by the Canadian task force.

Table 1. Summary of Findings, age to stop screening

Outcome	No. and design (study period and size) of included studies Study quality	Findings	GRADE certainty	What happens?
AGE TO STOP SCREENING				
Benefits				
Breast-cancer mortality	1 NRSI[4] (US Medicare; 1999-2008; n=1,058,013), Fair quality	Screening from age 70 to 74: 8-year risk of breast cancer mortality was 1 fewer death per 1,000 women who continued annual screening (2.7 per 1000 versus 3.7 per 1000; aRD: -1.0 [95% CI -2.3 to 0.1]; aHR: 0.78 [95% CI 0.63 to 0.95]) Screening beyond age 74 (75-85 yrs): No difference in 8-year estimated risk in breast cancer mortality (3.8 per 1000 versus 3.7 per 1000; RD: 0.07 [95% CI -0.93 to 1.3]; aHR 1.00 [95% CI 0.83 to 1.19])	**Very low** ⨁⊖⊖⊖ due to ROB, indirectness, and imprecision Indirectness: pre-2014 for mortality outcome; all had high probability of living ≥10 more years	We are very uncertain about the effects on BC mortality from continuing screening beyond 70 years.
Treatment-related morbidity (*USPSTF describe under overtreatment*)	1 NRSI[4] (US Medicare; 1999-2008; n= 1,058,013), Fair quality	Cancers diagnosed in those continuing versus stopping: Less likely to receive aggressive treatments: Radical mastectomy [70-74: 13.9% (13.4–14.5) versus 18.2% (17.0–19.4); 75-84: 14.2% (13.7–14.6) versus 17.0% (16.0–17.9)] Chemotherapy [70-74: 15.2% (14.7–15.8) versus 21.1% (20.0–22.1) versus; 75-84: 8.6% (8.3–9.1)]) versus 11.5% (10.6–12.3) versus More likely to receive: Lumpectomy [70-74: 52.6% (51.8–53.4) versus 36.5% (35.2–38.0); 75-84: 48.8% (47.9–49.5) versus 32.6% (31.5–33.8)	**Very low** ⨁⊖⊖⊖ due to ROB, indirectness, imprecision Indirectness: pre-2014 for treatment outcome; not estimated in screened versus unscreened population (i.e. not accounting for less cancers in stop strategy); high probability of living 10+ more years	We are very uncertain about the effects on treatment-related morbidity from continuing screening beyond 70 years.

		Radiotherapy [70-74: 51.0% (50.3–51.8) versus 39.9% (38.6–41.3) **(Table 10)**		
No data: All-cause mortality, Breast cancer morbidity, Health-related quality of life, Screen-detection of invasive breast cancer**,** Detection of invasive breast cancer over follow-up, **Stage distribution of screen-detected breast cancer, Stage distribution of any invasive breast cancer during follow-up**				
Harms				
Overdiagnosis	1 NRSI[4] (US Medicare; 1999-2008; n = 1,058,013), Fair quality	More cancers diagnosed in continue screening strategy: adjusted 8-year cumulative risk of breast cancer diagnosis 70-74: 5.3% versus 3.9% (95% CI NR) 75-85: 5.8% versus 3.9% (95% CI NR) (**Table 10**).	**Very low** ⊕⊖⊖⊖ due to ROB and imprecision Some but not serious indirectness: high probability of living ≥10 more years; only 8-years follow-up	We are very uncertain about the effects on overdiagnosis from continuing screening beyond 70 years.
No data: False-positive rate requiring imaging only or imaging plus biopsy (cumulative over multiple rounds), False-positive rate requiring imaging plus biopsy (cumulative over multiple rounds), Interval cancers				

Table 2. Summary of Findings, biennial versus annual screening

Outcome	No. and design (study period and size) of included studies Quality	Findings	GRADE	What happens?
Screening interval (annual versus biennial)				
Benefits				
Stage distribution of any invasive breast cancer	1 NRSI[5] (BCSC data US: 1996 to 2012; n = 15,440) Fair quality	40-79 years (data stratified by decade and menopausal status; case-only analysis): No difference in risk of stage IIB+ (range of aRRs 0.98 to 1.17) or less favorable prognosis (range of RRs 1.03 to 10.7) cancers diagnosed after a biennial compared with annual interval (≥2 rounds in group) for any age group (**Table 7**).	**Very low** ⨁⊖⊖⊖ due to ROB, **indirectness**, and imprecision Indirectness: comparison (case-only analysis i.e. lack of data on rates of cancer to interpret)	We are very uncertain about the effects on advanced stage cancers from screening annually versus biennially.
No data: Breast-cancer mortality, All-cause mortality, Treatment-related morbidity, Breast cancer morbidity, Health-related quality of life, Screen-detection of invasive breast cancer**,** Detection of invasive breast cancer over follow-up, **Stage distribution of screen-detected breast cancer**				
Harms				
False-positive rate requiring imaging only or imaging plus biopsy (cumulative over multiple rounds)	2 NRSI N=905,514 (US BCSC[6]; n = 903,495; 2005-2018 and US academic centre[7] n = 2,019; 2014-2015) Fair quality	BCSC: calculated estimated cumulative 10-years for DBT/sDM or DM screening approximately 50% of those undergoing annual screening had at least one false positive recall, compared with approximately 35% of those undergoing biennial screening (*not including prevalence screens; similar rates for DBT and DM)*. ~140-180 more per 1000 Subgroups: Age: Annual screening was associated with higher cumulative FPs for all age groups (i.e., DM: 40-49 19.4%, 50-59 20.0%, 60-69 18.6%, 70-79 17.3% more; DBT: 40-49 14.6%, 50-59 16.3%, 60-69 14.7%, 70-79 11.2% more) (**Appendix F Table 6**) Density: Annual screening was associated with higher cumulative FP recalls across density groups (less so with BI-RADS A) (**Figure 8**)	**Moderate** ⨁⨁⨁⊖ Some but not serious indirectness: US data but relative effects should be similar in Canada; no prevalent screen data included (so would underestimate this); data for DM still applies	Annual versus biennial screening with DM or DBT probably leads to more (possibly 1.5-fold) false positives.

		One NRSI from a US academic centre reported higher odds (OR 2.2, 95% CI 1.7 to 2.8) of a false positive result over a median of 8.9 years.		
False-positive findings at biopsy	1 NRSI (BCSC US[6]; 2005-2018; n=903,495) Fair quality	BCSC data calculated estimated cumulative 10-years for DBT/sDM or DM screening annual screening resulted in ~50 additional FP biopsies per 1,000 screened over 10 years (annual ~115 per 1,000 versus biennial ~66 per 1,000). (*not including prevalence screens; similar rates for DBT and DM*) (Appendix F Table 4) Subgroups: Age: Annual screening was associated with higher cumulative FP biopsies for all age groups (i.e., DM 40-49 5.2%, 50-59 5.6%, 60-69 5.2%, 70-79 4.1% more; DBT: 40-49 4.8%, 50-59 5.0%, 60-69 4.7%, 70-79 4.0% more)(**Appendix F Table 6**) Density: Annual screening was associated with higher cumulative FP biopsies across density groups (less so with BI-RADS A) (**Figure 9**)	**Moderate** ⊕⊕⊕⊖ Some indirectness: US data but relative effects should be similar in Canada; no prevalent screen data included (so would underestimate this); data for DM still applies	Annual versus biennial screening with DM or DBT probably leads to more (possibly 1.5 to 2.0-fold) false positive biopsies.
Interval cancers	1 NRSI (BCSC US[5]; 1996-2012; n=15,440) Fair quality	Unadjusted percent with interval cancer for people screened negative after an annual (22.2%; followed for 12 mos) or biennial screening (27.2%; followed for 24 mos) interval.	**Very low** ⊕⊖⊖⊖ due to ROB, indirectness, imprecision Indirectness: comparison (case-only analysis)	We are very uncertain about the effects on interval cancers from annual versus biennial screening.
No data: Overdiagnosis				

Table 3. Summary of Findings, triennial versus annual screening

Outcome	No. and design (study period and size) of included studies Quality	Findings	GRADE	What happens?
Screening interval (annual versus triennial)				
Benefits				
Breast-cancer mortality	1 NRSI (Finland[9]; 1985-1995; n=14,765) Fair quality 1 RCT (UK[5] 1989-1996) N=76,022 Fair quality	40-49 years: No difference in breast cancer mortality from annual versus triennial film mammography (20.3 versus 17.9 per 100,000 PY; RR 1.14, 95% CI 0.59 to 1.27) at 13 years. Intention-to-screen analysis. RCT never reported mortality outcome as planned.	**Very low** ⊕⊖⊖⊖ due to ROB, indirectness, and imprecision Indirectness: pre- 2014 for treatment outcome	We are very uncertain about the effects of annual versus triennial screening for breast-cancer mortality in 40 to 49-year-olds. No data was examined for older ages.
All-cause mortality	1 NRSI (Finland[9]; 1985-1995; n=14,765) Fair quality 1 RCT (UK[5] 1989-1996) N=76,022 Fair quality	40-49 years: No difference in all-cause mortality from annual versus triennial film mammography (230.9 versus 192.6 per 100,000 PY; RR 1.20, 95% CI 0.99 to 1.46) at 13 years. Intention-to-screen analysis. RCT never reported mortality outcome as planned.	**Very low** ⊕⊖⊖⊖ due to ROB, indirectness, and imprecision Indirectness: pre- 2014 for treatment outcome	We are very uncertain about the effects of annual versus triennial screening for all-cause mortality in 40 to 49-year-olds. No data was examined for older ages.
Screen-detection of invasive breast cancer	1 RCT (UK[8] 1989-1996) N=76,022 Fair quality	50-62 years: More invasive cancers screen-detected over 3 years with annual screening screen (4.42 per 1000 versus 2.70 per 1000; RR: 1.64 [95% CI, 1.28 to 2.09])	**Low** ⊕⊕⊖⊖ due to ROB (and single study) **Indirect outcome but did not rate down**	Annual versus triennial screening may lead to more screen-detected invasive cancers for 50 to 69-year-olds over 3 years.
Detection of all invasive breast cancers over follow-up	1 RCT (UK[8] 1989-1996) N=76,022 Fair quality	50-62 years: Total number of invasive cancers similar between groups over 3 years (6.26 per 1000 versus 5.4 per 1000; RR: 1.16, 95% CI 0.96 to 1.40)	**Low** ⊕⊕⊖⊖ due to ROB and imprecision **Indirect outcome but did not rate down**	Annual versus triennial screening may not lead to detection of more invasive cancers for 50 to 69-year-olds over 3 years.
Stage distribution of any invasive breast cancer	1 RCT (UK[8] 1989-1996; n=76,022 Fair quality	50-62 years: similar rates and no statistical differences by screening interval in tumor size, nodal status, grade, or prognostic index for all invasive cancers diagnosed over 3 years (**Table 6**).	**Low** ⊕⊕⊖⊖ due to ROB and imprecision	Annual versus triennial screening may make little-to-no difference for advanced stage cancers

		Stage II+ or III+ NR.	Some but not serious indirectness; no data specific to stage II+ or III+ reported; only 3 years of screening and limited ages but added applicability into conclusions	for 50 to 69-year-olds over 3 years. No data was examined for other ages.
No data: Treatment-related morbidity, Breast cancer morbidity, Health-related quality of life, Stage distribution of screen-detected breast cancer				
Harms				
Interval cancers	1 RCT (UK[8]: 1989-1996; n=76,022) Fair quality 1 NRSI (Finland[9]; 1985-1995; n=14,765) Fair quality	RCT (50-62 years) estimated 1 fewer invasive interval cancers in the annual screening arm (1.8 versus 2.7 per 1,000 screened; RR: 0.68 [95% CI 0.50 to 0.92]). NRSI (40-49 years) found no difference in interval cancer incidence (p = 0.22).	**Low** ⊕⊕⊖⊖ due to ROB and inconsistency Some but not serious indirectness; 3 years in RCT; added applicability into conclusions	Annual versus triennial screening may slightly reduce the number of invasive interval cancers for 50 to 69-year-olds over 3 years.
Overdiagnosis (data not used in review)	1 NRSI (Finland[9]; 1985-1995; n=14,765) Fair quality	NRSI (40-49 years): *invasive* breast cancer incidence over mean 9.8 years was similar for those invited to annual screening (141.1 per 100,000 person-years) and those invited to triennial screening (144.0 per 100,000 person-years) (RR: 0.98 [95% CI 0.75 to 1.29])	**Very low** ⊕⊖⊖⊖ due to ROB and indirectness Serious indirectness: no DCIS; applies to 40-49 years	We are very uncertain about the effects of annual versus triennial screening for overdiagnosis in 40 to 49-year-olds. No data was examined for older ages.
No data: False-positive rate requiring imaging only or imaging plus biopsy (cumulative over multiple rounds), False-positive rate requiring imaging plus biopsy (cumulative over multiple rounds)				

Table 4. Summary of Findings, digital breast tomosynthesis versus digital mammography

Outcome	No. and design (study period and size) of included studies Quality	Findings	GRADE	What happens?
Digital breast tomosynthesis versus digital mammography				
Benefits				
Screen-detection of invasive cancer	**3 RCTs with 2 rounds** N=129,492 2 Good quality (i) RETomo[12] Italy [2014-2017]; n=26,877; 45-69 years [9% BI-RADs 4]; DBT/DM versus DM but DM at 2nd round both groups 1 [45-49 years; 38%] or 2 [50-69] years later) (ii) To-BE[11] Norway [2016-2020]; n=28,749; 50-69 years [7% BI-RADs 4]; DBT/sDM versus DM but DBT/sDM at 2nd round for both groups 2 years later or next screening round) 1 Fair quality (Proteus Donna[10]; Italy [2004-2017]; n=73,866; 46-68 years [density NR]; DBT/DM versus DM but DM in 2nd round both groups [1 year 46-49 or 2 years 50-68]) **1 NRSI with 2 rounds** Norway[13] [2014-2017]; n=98,927; 50-69 years [density NR]; DBT/sDM	Round 1: DBT higher invasive cancer detection (3 RCTs pooled RR 1.41, 95% CI 1.20 to 1.64, I^2 8%, n = 129,492) with absolute differences ranging from 0.6 to 2.4 more per 1000 screened. Similar results were seen in the NRSI (2.3 more per 1000 screened; RR 1.43, 95% CI 1.22 to 1.67; unadjusted). Round 2: No significant difference was found (3 RCTs pooled RR 0.87, 95% CI 0.73 to 1.05, I^2 0%, n = 105,064). The NRSI found lower detection at round 2 for the study group screened with DBT/sDM at round one (1.3 fewer per 1000 screened; RR 0.71, 95% CI 0.55, 0.92); unadjusted). (**Figure 3**) Subgroups (**Table 9**): **RETomo RCT**: DBT resulted in a higher invasive cancer detection at the first round of screening for women ages 50 to 69 (RR: 1.60, 95% CI 1.10 to 2.30) and for women with nondense breasts [BI-RADS A/B] (RR: 1.80, 95% CI 1.10 to 3.00), but at the next round of screening when all were screened with DM, there was not a statistically significant difference in invasive cancer detection. (RRs 1.0 and 0.97). For women aged 45-49 and women with dense breasts (BI-RADS C/D) there was no statistical difference in the detection of invasive cancers at either round of screening (round 1: RR=1.9 (95% CI, 0.89 to 4.1) and RR=1.5 (95% CI, 0.94 to 2.5) (but still same direction as overall findings) and round 2: RR=0.50 (95% CI, 0.20 to 1.2) and RR=0.64 (95% CI, 0.34 to 1.2). **To-BE** (using Volpara Density Grade): Round 1: density grades 1-3 RRs 1.07 to 1.16 versus density grade 4 RR=1.97 (95% CI, 0.47 to 8.21)	**Low** ⨁⨁⊖⊖ due to inconsistency and indirectness Indirectness: serious concerns about use of same device at round 2 **Indirect outcome but did not rate down for this**	DBT versus DM may detect more invasive cancers over two rounds of screening.

	versus DM but 2nd round with DM 2 years later Fair quality	Round 2: density grades 1-3 (0.82 to 1.04) versus grade 4 RR=0.66 (95% CI, 0.26 to 1.70)		
Stage distribution of screen-detected breast cancer	Same as above (Author contact for To-Be data)	Stage II+: No significant differences within any of the 3 RCTs in the detection of Stage II+ at either round. Rates at round 1 were 1.2 per 1000 (Proteus Donna) or 1.3-1.6 per 1000 (RETomo & To-Be) in both groups. Results were inconsistent at round two with one trial nearing statistical significance for more stage II+ cancers from DBT/sDM (RETomo 1.2 versus 0.5 per 1000; RR 2.53 [95% CI 0.98 to 6.53]) and the other two trials in the direction of reduced stage II+ cancer in the DBT arm (Proteus Donna 0.7 versus 1.1 per 1000 and To-Be 1.4 versus 2.2 per 1000). Stage III+: RETomo (round 1: 0.2 versus 0.1 per 1000; round 2: 0.2 versus 0.3 per 1000). No clear evidence of stage shift. Stage not reported by NRSI. (**Table 8, Figure 4**) The three trials and NRSI reported tumor characteristics that inform staging such as tumor diameter, histologic grade, and node status. No statistically significant differences in these or other individual tumor prognostic characteristics were reported at the first or second round of screening for any of the included studies, but statistical power was limited for comparisons of less common tumor types. (**Table 8, Figures 5-7**) Subgroups: No information on the characteristics of the screen-detected tumors was provided by different population characteristics and risk markers (e.g., age, breast density, race/ethnicity, family history)	**Low** ⊕⊕⊖⊖ due to indirectness and imprecision Indirectness: serious concerns about use of same device at round 2	DBT versus digital mammography may make little-to-no difference for advanced stage cancers over two rounds.
No data: Breast-cancer mortality, All-cause mortality, Treatment-related morbidity, Breast cancer morbidity, Health-related quality of life, Detection of all invasive breast cancers over follow-up**, Stage distribution of any invasive cancer**				
Harms				
False-positive findings at screening	3 RCTs (see above) Good quality: RETomo[12], To-Be[11] Fair quality: Proteus Donna[10]	Three RCTs and one NRSI reported false positive recall rates at two rounds of screening, and results were mixed. In round 1 the RCTs had mixed findings (rates approx. 3-5%; Proteus Donna RR 1.22 versus To-Be RR 0.72) and in	**Low** ⊕⊕⊖⊖ due to ROB and inconsistency	DBT versus digital mammography may make little-to-no difference for false positives.

	2 NRSIs Norway[13]; n=98,927; see above) BCSC US[6] [2005-2018]; n=903,495 Fair quality	round 2 were consistent for no difference (but using same device) (**Figure 12)**. One NRSI calculated (using probabilities from mean 3.3 rounds) the estimated (via discrete-time survival modeling to account for censoring) cumulative probability of at least one false-positive recall over 10 years of screening and suggested slightly lower FP recall with DBT with annual interval (50% versus 56%) and similar rates with biennial screening (36% versus 38%). Subgroups: Age: RETomo, stratified by ages 45-49 and 50-69 with no significant differences at either round for either group Density: To-Be stratified by density suggested lower FPs at round 1 for 1/2 (RR: 0.58 [0.43 to 0.80] and 0.66 [0.54 to 0.81]) but not for 3/4; at round 2 no significant difference for any group (**Table 18**). BCSC data, in stratified analyses there was not a statistical difference in cumulative FPs among those with extremely dense breasts in any age group	Not serious indirectness: only 2 rounds in RCTs, round 2 in RCTs used similar device between groups (used for ROB); US data for multiple rounds but relative effects should be similar in Canada	(**Note**: **as per comments at beginning of this document, this was changed in the final version of the review to low certainty for a reduction with DBT**)
False-positive findings at biopsy	1 RCT (see above) To-Be[11] Good quality 1 NRSI BCSC US[6] (903,495) Fair quality	One trial reported no significant difference in false positive biopsy (round 1: RR: 0.85 (95% CI, 0.69 to 1.05); round 2: RR: 0.99 (95% CI: 0.80 to 1.24). One NRSI calculated (using probabilities from mean 3.3 rounds) the estimated cumulative probability of at least one false-positive biopsy over 10 years of screening and suggested no difference in cumulative FP biopsy for DBT versus DM regardless of screening interval (11-12% annual, 7% biennial). Subgroups: Density: To-Be stratified analysis by density, at round 1 significantly fewer biopsies with DBT in groups 1 (RR 0.57 [0.33 to 1.00] and 2 (RR 0.64 [0.46 to 0.89]), with higher from DBT for groups 3 RR 1.79 [1.23 to 2.61] and 4 RR 1.12 ($p<<0.05$). No significant differences at round 2 (using DM) (**Table 17**).	**Low** ⨁⨁⊖⊖ due to ROB and imprecision Not serious indirectness: only 2 rounds in RCTs, round 2 in RCTs used similar device between groups (used for ROB); US data for multiple rounds but relative effects should be similar in Canada	DBT versus digital mammography may make little-to-no difference for false positive biopsies.

		BCSC data, in stratified analyses there was not a statistical difference in cumulative FP biopsy among those with extremely dense breasts in any age group		
Interval cancers	3 RCTs (see above) Good quality: RETomo[12], To-Be[11] Fair quality: Proteus Donna[10] (12-month follow-up for those ages 45 to 49 years and 24-month follow-up for those ages 50 to 69 years) 5 NRSIs Fair quality 1 DBT versus DM (BCSC US[16] [2011-2018]; n=504,427; 40-79 years) 4 DBT/DM versus DM (2 US[15,17] [2015-2017 & 2011-2015], Norway[13] [2014-2017], Sweden[14] [2010-2015]) N=4,816,610	Three RCTs did not find difference in interval (invasive) cancer rates (pooled RR = 0.87, 95% CI 0.64 to 1.17, k = 3 RCT, n = 130,196, I^2 = 0%). (**Figure 10**) Five NRSI had inconsistent results - three did not find differences, one commercial claims registry study (US; n=4,580,698) reported more interval (invasive) cancers with DBT (adj difference: 0.07 per 1000 screens, 99% CI 0.01 to 0.12), and one (Sweden; n=40,107) comparing trial participants to an age-matched population reported fewer interval (invasive) cancers with DBT (1.4 versus 2.7 per 1,000, RR 0.53, 95% CI 0.32 to 0.87).(all differences small). (**Table 12**) There were no significant differences when studies (3 RCTs and 2 NRSIs) examined only DCIS (RCT RRs ~1.0). (**Table 12**) Subgroups: Age: RETomo and two NRSI reported no significant findings related to the relationship between age and interval cancer outcomes. (**Table 19**) Density: RETomo and To-Be, and one analysis of BCSC data, found no statistically significant differences in the incidence of interval cancer for the breast density stratified comparisons. (**Table 19**)	**Moderate** ⊕⊕⊕⊖ due to imprecision Some but not serious indirectness: studies differed in the timeline of follow up and method of identifying interval cancers; in RCTs data from round 1 only but NRSI had multiple rounds	DBT versus digital mammography probably makes little-to-no difference for interval cancers.
No data: Overdiagnosis (only reported on DCIS at round 1 in 3 RCTs (did not find differences in DCIS, screen-detected lesions that could contribute to over- detection, at round 1 (pooled RR 1.33, 95% CI 0.92 to 1.93, k = 3 RCT, n = 130,196, I^2 = 0%) or round 2 (pooled RR 0.75, 95% CI 0.49 to 1.14, k = 3 RCT, n = 130,196, I^2 =0%). (Figure 14)				

Table 5. Summary of Findings, supplemental MRI

Outcome	No. and design (study period and size) of included studies Quality	Findings	GRADE	What happens?
Mammography supplemented with MRI				
Benefits				
No data: Breast-cancer mortality, All-cause mortality, Treatment-related morbidity, Breast cancer morbidity, Health-related quality of life, Screen-detection of invasive breast cancer**,** Detection of invasive breast cancer over follow-up, **Stage distribution of screen-detected breast cancer, Stage distribution of any invasive cancer**				
Harms				
False-positive findings at screening	1 RCT (DENSE[17] The Netherlands [2011-2016]; n=40,373; 50-75; 100% extremely dense breasts (Volpara category D); invitation versus not to MRI after a negative screening mammogram result among biennial screening program) (2 of 3 rounds reported but only 1st reports comparative data) Good quality	NR (only for MRI group; 79.8 per 1,000 screened)	NA	NA
False-positive findings at biopsy		NR (only for MRI group; 62.7 per 1000 screened)	NA	NA
Interval cancers		Reduced invasive interval cancer (follow-up 2 years) with invitation to screening for those with extremely dense breasts and negative mammogram (2.2 versus 4.7 per 1,000 invited to screening, RR 0.47, 95% CI 0.29 to 0.77). Any interval cancer 2.5 versus 5.0 per 1000; RD -2.5 (95% CI, 1.0 to 3.7) Among the 20 interval cancers in MRI group, 4 were among those who had received MRI (59%). No subgroup analyses.	**Low** ⨁⨁⊖⊖ (single study and limitations from poor adherence 59%) Some indirectness: specific population and 1 round only but added to conclusions	Supplementing digital mammography with MRI may reduce interval cancers over the first round for individuals with extremely dense breasts
No data: Overdiagnosis				

Table 6. Summary of Findings, supplementation with ultrasound

Outcome	No. and design (study period and size) of included studies Quality	Findings	GRADE	What happens?
Mammography supplemented with US				
Benefits				
No data: Breast-cancer mortality, All-cause mortality, Treatment-related morbidity, Breast cancer morbidity, Health-related quality of life, Screen-detection of invasive breast cancer**,** Detection of invasive breast cancer over follow-up, **Stage distribution of screen-detected breast cancer, Stage distribution of any invasive cancer** (data in the RCT on detection of cancers and stage distribution was after only 1 round thus were ineligible for review)				
Harms				
False-positive findings at screening	1 RCT (J-START[18] Japan [2007-2011]; n=72,717; **40-49 years**; 58% dense breasts (among subset n=19,213); DM/US versus DM for 2 rounds [only 1 round reported]) Fair quality	NR (RCT: 48.0 per 1000 extra FPs from adding US; not reported in DM group)	NA	NA
False-positive findings at biopsy	1 NRSI (BSSC US[19] [2000-2013]; n=18,562; 30-80+ years; 65% BI-RADS 3/4 NR; 35% "intermediate risk") Fair quality	NRSI: 52.0 vs 22.2 per 1000 screens (RR 2.23 [95% CI, 1.93 to 2.58])	**Low** ⨁⨁⨀⨀ for ROB	Supplementing digital mammography with ultrasound may increase (possibly 2-fold over two rounds) false positive biopsies
Interval cancers	1 RCT and 1 NRSI See above	RCT (invasive): 0.4 (DM/US) versus 0.8 (DM) per 1,000 screened; RR 0.58, 95% CI 0.31 to 1.08 NRSI (invasive and DCIS): 1.5 (DM/US) versus 1.9 (DM) per 1,000 screened; aRR 0.67, 95% CI 0.33 to 1.37 Subgroups: Density: J-START stratified analysis (n=19,213), similar RRs and no statistically significant difference for either group (A/B versus C/D) (**Table 19**)	**Low** ⨁⨁⨀⨀ due to ROB and imprecision Indirectness: One round only but noted in conclusions	Supplementing digital mammography with ultrasound may not reduce interval cancers at the first round
No data: Overdiagnosis				

References

1. Henderson J, Webber E, Weyrich M, Miller M, Melnikow J. Screening for Breast Cancer: A Comparative Effectiveness Review for the U.S. Preventive Services Task Force. (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 75Q80120D00004). AHRQ Publication No. 23-05303-EF-1. May 2023. Rockville, MD: Agency for Healthcare Research and Quality; 2023. Available at: https://www.uspreventiveservicestaskforce.org/uspstf/document/draft-evidence-. review/breast-cancer-screening-adults
2. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et la. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology. 2008; 246(2): 376-83.
3. Berkman ND, Lohr, K.N., Ansari, M. et al. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2013.
4. Garcia-Albeniz X, Hernan MA, Logan RW, et al. Continuation of annual screening mammography and breast cancer mortality in women older than 70 years. Ann Intern Med. 2020;172(6):381-9. https://dx.doi.org/10.7326/M18-1199
5. Miglioretti DL, Zhu W, Kerlikowske K, et al. Breast tumor prognostic characteristics and biennial vs annual mammography, age, and menopausal status. JAMA Oncol. 2015;1(8):1069-77. https://dx.doi.org/10.1001/jamaoncol.2015.3084
6. Ho TH, Bissell MCS, Kerlikowske K, et al. Cumulative probability of false-positive results after 10 years of screening with digital breast tomosynthesis vs digital mammography. JAMA Netw Open. 2022;5(3):e222440. https://doi.org/10.1001/jamanetworkopen.2022.2440
7. McGuinness JE, Ueng W, Trivedi MS, et al. Factors associated with false positive results on screening mammography in a population of predominantly Hispanic women. Cancer Epidemiol Biomarkers Prev. 2018;27(4):446-53. https://dx.doi.org/10.1158/1055-9965.EPI-17-0009
8. Breast Screening Frequency Trial Group. The frequency of breast cancer screening: results from the UKCCCR Randomised Trial. United Kingdom Co-ordinating Committee on Cancer Research. Eur J Cancer. 2002;38(11):1458-64. https://www.doi.org/10.1016/s0959-8049(01)00397-5
9. Parvinen I, Chiu S, Pylkkänen L, et al. Effects of annual vs triennial mammography interval on breast cancer incidence and mortality in ages 40-49 in Finland. Br J Cancer. 2011;105(9):1388-91. https://doi.org/10.1038/bjc.2011.372
10. Armaroli P, Frigerio A, Correale L, et al. A randomised controlled trial of digital breast tomosynthesis versus digital mammography as primary screening tests: screening results over subsequent episodes of the Proteus Donna study. Int J Cancer. 2022. https://www.doi.org/10.1002/ijc.34161
11. Hofvind S, Moshina N, Holen AS, et al. Interval and subsequent round breast cancer in a randomized controlled trial comparing digital breast tomosynthesis and digital mammography screening. Radiology. 2021;300(1):66-76. https://dx.doi.org/10.1148/radiol.2021203936

12. Pattacini P, Nitrosi A, Giorgi Rossi P, et al. A randomized trial comparing breast cancer incidence and interval cancers after tomosynthesis plus mammography versus mammography alone. Radiology. 2022:211132. https://doi.org/10.1148/radiol.211132
13. Hovda T, Holen AS, Lang K, et al. Interval and consecutive round breast cancer after digital breast tomosynthesis and synthetic 2D mammography versus standard 2D digital mammography in BreastScreen Norway. Radiology. 2020;294(2):256-64. https://dx.doi.org/10.1148/radiol.2019191337
14. Johnson K, Lang K, Ikeda DM, et al. Interval Breast Cancer Rates and Tumor Characteristics in the Prospective Population-based Malmo Breast Tomosynthesis Screening Trial. Radiology. 2021;299(3):559-67. https://dx.doi.org/10.1148/radiol.2021204106
15. Conant EF, Beaber EF, Sprague BL, 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):109-16. https://dx.doi.org/10.1007/s10549-016-3695-1
16. Kerlikowske K, Su Y-R, Sprague BL, et al. Association of Screening With Digital Breast Tomosynthesis vs Digital Mammography With Risk of Interval Invasive and Advanced Breast Cancer. JAMA. 2022;327(22):2220-30. https://doi.org/10.1001/jama.2022.7672
17. Richman IB, Long JB, Hoag JR, et al. Comparative Effectiveness of Digital Breast Tomosynthesis for Breast Cancer Screening among Women 40-64 Years Old. J Natl Cancer Cent. 2021;03:03. https://dx.doi.org/10.1093/jnci/djab063
18. Veenhuizen SGA, de Lange SV, Bakker MF, et al. Supplemental Breast MRI for Women with Extremely Dense Breasts: Results of the Second Screening Round of the DENSE Trial. Radiology. 2021;299(2):278-86. https://dx.doi.org/10.1148/radiol.2021203633
19. Ohuchi N, Suzuki A, Sobue T, 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. Lancet. 2016;387(10016):341-8. https://dx.doi.org/10.1016/S0140-6736(15)00774-6
20. Lee JM, Arao RF, Sprague BL, et al. Performance of Screening Ultrasonography as an Adjunct to Screening Mammography in Women Across the Spectrum of Breast Cancer Risk. JAMA internal medicine. 2019;179(5):658-67. https://dx.doi.org/10.1001/jamainternmed.2018.8372
21. Heindel W, Weigel S, Gerß J, et al. Digital breast tomosynthesis plus synthesised mammography versus digital screening mammography for the detection of invasive breast cancer (TOSYMA): a multicentre, open-label, randomised, controlled, superiority trial. Lancet Oncol. 2022. https://doi.org/10.1016/s1470-2045(22)00194-2
22. Ganguli I, Keating NL, Thakore N, et al. Downstream Mammary and Extramammary Cascade Services and Spending Following Screening Breast Magnetic Resonance Imaging vs Mammography Among Commercially Insured Women. JAMA Netw Open. 2022;5(4):e227234. https://doi.org/10.1001/jamanetworkopen.2022.7234

Appendix

Studies excluded based on quality (for the Key question they were otherwise eligible for)

(Codes: E11 means poor quality; E4 means no relevant outcome reported for the Key Question [KQ1 breast cancer morbidity and breast cancer–specific or all-cause mortality; KQ2 incidence/detection of cancer and progression to advanced breast cancer; KQ3 harms false positives, interval cancers, overdetection; E7 means study design)

Note: the report states that 13 studies were exclude due to poor quality so two of the below citations (n=14) may report on the same study

1. Alsheik, NH, Dabbous, F, et al. Comparison of Resource Utilization and Clinical Outcomes Following Screening with Digital Breast Tomosynthesis Versus Digital Mammography: Findings From a Learning Health System. *Acad Radiol*. 26(5): 597-605. 2019. https://dx.doi.org/https://dx.doi.org/10.1016/j.acra.2018.05.026. **KQ1E4, KQ2E7, KQ3E11**

2. Chae, EY, Kim, HH, et al. Evaluation of screening whole-breast sonography as a supplemental tool in conjunction with mammography in women with dense breasts. *J Ultrasound Med*. 32(9): 1573-8. 2013. https://dx.doi.org/10.7863/ultra.32.9.1573. **KQ1E5, KQ2E5, KQ3E11**

3. Caumo, F, Montemezzi, S, et al. Repeat Screening Outcomes with Digital Breast Tomosynthesis Plus Synthetic Mammography for Breast Cancer Detection: Results from the Prospective Verona Pilot Study. *Radiology*. 298(1): 49-57. 2021. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2020201246. **KQ1E4, KQ2E11, KQ3E11**

4. Destounis, S, Arieno, A, et al. Initial experience with combination digital breast tomosynthesis plus full field digital mammography or full field digital mammography alone in the screening environment. *J Clin Imaging Sci*. 4: 9. 2014. https://dx.doi.org/https://dx.doi.org/10.4103/2156-7514.127838. **KQ1E4, KQ2E11, KQ3E11**

5. Field, LR, Wilson, TE, et al. Mammographic screening in women more than 64 years old: a comparison of 1- and 2-year intervals. *AJR Am J Roentgenol*. 170(4): 961-5. 1998. https://doi.org/10.2214/ajr.170.4.9530044. **KQ1E4, KQ2E11, KQ3E4**

6. Houssami, N, Bernardi, D, et al. Interval breast cancers in the 'screening with tomosynthesis or standard mammography' (STORM) population-based trial. *Breast*. 38: 150-153. 2018. https://dx.doi.org/https://dx.doi.org/10.1016/j.breast.2018.01.002. **KQ1E4, KQ2E11, KQ3E4**

7. Hunt, KA, Rosen, EL, et al. Outcome analysis for women undergoing annual versus biennial screening mammography: a review of 24,211 examinations. *AJR Am J Roentgenol*. 173(2): 285-9. 1999. https://doi.org/10.2214/ajr.173.2.10430120. **KQ1E4, KQ2E11, KQ3E11**

8. Kemp Jacobsen, K, Abraham, L, et al. Comparison of cumulative false-positive risk of screening mammography in the United States and Denmark. *Cancer Epidemiol*. 39(4): 656-63. 2015. https://dx.doi.org/10.1016/j.canep.2015.05.004. **KQ1E4, KQ2E4, KQ3E11**

9. Moorman, SEH, Pujara, AC, et al. Annual Screening Mammography Associated With Lower Stage Breast Cancer Compared With Biennial Screening. *AJR Am J Roentgenol*. 217(1): 40-47. 2021. https://dx.doi.org/10.2214/AJR.20.23467. **KQ1E4, KQ2E11, KQ3E4**

10. O'Meara, ES, Zhu, W, et al. Mammographic screening interval in relation to tumor characteristics and false-positive risk by race/ethnicity and age. *Cancer*. 119(22): 3959-67. 2013. https://doi.org/10.1002/cncr.28310. **KQ1E4, KQ2E11, KQ3E11**

11. Park, HL, Chang, J, et al. Mammography screening and mortality by risk status in the California teachers study. *BMC Cancer*. 21(1): 1341. 2021. https://doi.org/10.1186/s12885-021-09071-1. **KQ1E11, KQ2E4, KQ3E4**

12. Randall, D, Morrell, S, et al. Annual or biennial mammography screening for women at a higher risk with a family history of breast cancer: prognostic indicators of screen-detected cancers in South Wales, Australia. *Cancer Causes Control*. 20(5): 559-66. 2009. https://doi.org/10.1007/s10552-008-9264-0. **KQ1E4, KQ2E11, KQ3E4**

13. Simon, MS, Wassertheil-Smoller, S, et al. Mammography interval and breast cancer mortality in women over the age of 75. *Breast Cancer Res Treat*. 148(1): 187-95. 2014. https://doi.org/10.1007/s10549-014-3114-4. **KQ1E11, KQ2E11, KQ3E4**

14. Winch, CJ, Sherman, KA, et al. Toward the breast screening balance sheet: cumulative risk of false positives for annual versus biennial mammograms commencing at age 40 or 50. *Breast Cancer Res Treat*. 149(1): 211-21. 2015. https://dx.doi.org/10.1007/s10549-014-3226-x. **KQ1E4, KQ2E11, KQ3E11**

Description of relevant Canadian primary research studies not in USPSTF review (not eligible)

1. **Seely et al. Breast density and risk of interval cancers: the effect of annual versus biennial screening mammography policies in Canada. Canadian Association of Radiologists' Journal. 2022; 73(1); 90–100.**
   Methods: In Canada where most screening is biennial, the targets for interval cancer rate (ICR) are < 0.6 per 1000 for annual screening and < 1.2 per 1000 for 12 to 24 months after screening. Seely at al. evaluated the impact on ICR rate of population-based annual versus biennial screening policies for women with dense breasts, using information available from the Canadian Breast Cancer Screening Database (extremely dense breasts [BI-RADS 5th edition]; data from Manitoba [MB], Quebec, Ontario [ON] and Newfoundland [NL], Northwest Territories [NT]) or (for New Brunswick) the NB Cancer Network (dense breasts) for 2008-2010. Four jurisdictions (MB, ON, NL and NT) also included the radiologists' screening interval recommendation; radiologists may recommend an earlier return visit than the general policy if they deem it indicated via clinical impression, previous recall, breast density, and/or personal/family risk factors. Only NB and NT submitted data for women younger than 50 years old. Data were

included on women with no first-degree family history, known recommended screening interval, and not referred for follow-up testing. Annual screening mammography policies for dense breasts existed in NT, NL, and ON. In the six jurisdictions studied, 57% of mammograms were film, 22% digital radiography and 21% digital computed radiology. Interval cancers were defined as cancers occurring between regular screening visits, after a negative or benign mammographic assessment, i.e. within 1-2 years of the last screen. The ICR for the biennial (24 month) screening interval was presented as averaged annual rates (annualized) to facilitate comparison with the annual (12 month) screening.
Results: 288 of 148,575 women with (mostly extremely) dense breasts (representing 17.5% of women screened in these jurisdictions) had an interval cancer. In the jurisdictions routinely providing annual screening (NT, NL, ON; n=70,814), the rate of interval cancers was 0.89 per 1000 women screened/year (95% CI: 0.67-1.11) and for those providing biennial screening (n=77,761), the annualized ICR was 1.45 per 1000 women (95% CI: 1.19-1.72) screened/year, or 63% greater (0.56 per 1000 more; p = 0.0016). When screening interval policy and radiologists' recommendations were combined, there were 76,103 screened women eligible for the analysis, 87 of whom had invasive interval cancers: 65 of the 69,650 screened with annual recommendations had ICR 0.93/1000 (95% CI: 0.71-1.16) and 22 of 6453 screened with biennial recommendations had a higher annualized ICR 1.70/1000 (95% CI:0.70-2.71) (p = 0.0605).
Limitations: There is no comparison with women without dense breasts to know if the difference in ICRs for annual versus biennial screening differs between groups. [The authors report that in Canada between 2003 and 2010, the ICR among women with incident (subsequent) screens was between 1.25 and 1.28 per 1000 women within 12-24 months of their screen.] The actual timing of the screening interval for each woman was not evaluated. Findings were not presented by age and are most applicable to women aged 50 years and older, for use of film mammography, and for those accessing primary care to learn about annual screening (screening participants were likely not notified about breast density by the programs during these years). The study was not able to explore whether the increased ICR translated into worse outcomes for women with dense breasts as stage of cancers and survival could not be assessed.

2. **Wu and Warren. The added value of supplemental breast ultrasound screening for women with dense breasts: a single center Canadian experience. Canadian Association of Radiologists' Journal. 2022; 73(1); 101-106.**
Methods: The authors conducted a retrospective review of all handheld screening US exams performed on women (mean age 55 ± 10 years) with BI-RADS C (n=466) or D (n=229) dense breasts (based on previous full field digital mammography) at one breast imaging center from January 1 2019 to December 31 2019. This imaging was publicly insured in 2019 for those with a requisition from a healthcare provider, such that these results are for a first/prevalent screen with US. The sample included women without a family history of a first degree relative with breast cancer who were on a 2-year call back screening mammogram schedule, and those who are on an annual call back schedule due to a family history of breast cancer in a first degree relative. For women who requested a screening US and whose mammogram was due, the mammogram was arranged prior to the US. For those on a 2-year call back schedule for screening mammograms, the mammogram schedule was not altered. Therefore, the previous negative mammogram results were from between a few days up to 2 years prior to the US exam. In cases where women had preexisting abnormal mammographic or sonographic findings, only new US findings without previous mammographic correlates were included in the analysis. Rates of US-detected cancer and core biopsy were reported.

Results: 17 (2.4%) women had new abnormal (BI-RADS 0, 4, and 5) findings on breast screening US, of which 9 (1.3%) underwent core biopsies and 5 (0.7%; 7 per 1000) had proven invasive cancer. Three of the 5 patients had either personal or family history of breast cancer and the range of time between the screening mammogram and US for these 5 patients was from 10 to 182 days.
Limitations: It is evident from the results that the sample included some women with high risk for breast cancer (e.g., personal history). There is no comparison with women without dense breasts and the duration between the negative mammogram and US imaging was up to 6 months such that in some cases the "additional" cancers detected by US could have also been found by a mammogram 6 months later. The authors compare their findings with a previous review that found an incremental cancer detection rate (over mammography) of 1.8-4.1 per 1000 (combining incidence and prevalent screening), though some of the studies reviewed likely had the same limitations. The sample size is small and the screen-detection of cancers is considered an indirect outcome.

3. **A. Wilkinson et al. Impact of breast cancer screening on 10-year net survival in Canadian women age 40-49 years. J Clin Oncol 2023. 41(29):4669-77.**
   B. Wilkinson et al. The impact of organised screening programs on breast cancer stage at diagnosis for Canadian women aged 40–49 and 50–59. Curr. Oncol. 2022. 29: 5627–43.
   Methods: These studies relied on data from the Canadian Cancer Registry (linked to death information for survival and mortality outcomes) to compare for invasive cancers diagnosed 2002-2007 i) 10-year net survival rates, based on an algorithm controlling for background differences in mortality and using annual provincial population life tables; ii) incidence-based mortality rates over 10 years among the female populations in the jurisdictions; iii) age-specific incident rates; and for invasive cancers diagnosed 2010-2017 iv) stage at diagnosis, in the jurisdictions with organized screening programs that included (via annual recall, usually after self-referral) women age 40-49 years, designated as screeners (Northwest Territories, British Columbia, Alberta, Nova Scotia, and Prince Edward Island), with comparator programs that either did not offer or required referrals for screening and (except for Manitoba with biennial recalls) did not recall women (Yukon, Manitoba, Saskatchewan, Ontario, Quebec [except for data on stage at diagnosis], New Brunswick, and Newfoundland and Labrador). Stage migration in women aged 50–59 was investigated by comparing proportions of stage at diagnosis for women aged 50–59 in screener and comparator provinces. Data for the 40-49 year-olds is considered relevant to the Task Force's review on the effects of screening versus not (not the focus in this report) and comparisons between the effects of screening for those 40-49 versus 50-59 (i.e., use in determining which age to start screening) are limited by the lack of a no screening policy comparator in the 50-59 year age category. Any differences in the findings among those 50-59 years (all screened) based on the screening practices for women in their 40s may be somewhat informative to determine what happens for those aged 50-59 based on their screening exposure the decade prior.
   Results: Results for women 40-49 years indicated that i) 10-year net-survival of cancer was 1.9 percentage points higher ($P < .001$) in the screening jurisdictions (84.8% [95% CI 83.8 to 85.8]) compared with the comparators (82.9% [95% CI 82.3 to 83.5]) (stratified by age, findings were larger and only significant for those aged 45-49 [2.6 percentage points; p=0.001]), ii) the incidence-based breast-cancer mortality rate was significantly lower in screener versus comparator jurisdictions among women age 40-49 years at diagnosis (rate ratio, 0.92 [95% CI,

0.85 to 0.99]) and among women age 45-49 years (0.89; 95% CI, 0.81 to 0.98) but not among women age 40-44 years (0.95; 95% CI, 0.85 to 1.07) (absolute rates per group not provided), iii) there was no significant difference in age-specific incident rates (133.1 cases per 100,000 women in both cases; P = 0.976), and iv) women in comparator jurisdictions had lower proportions of stage I (33.3% vs. 39.9%, p < 0.001), and higher proportions of stages II (43.7% vs. 40.7%, p < 0.001), III (18.3% vs. 15.6%, p < 0.001) and IV (4.6% vs. 3.9%, p = 0.001) compared to their peers in screener jurisdictions (crude data/absolute rates by stage and comparator group were not reported, but 2010-2017 incident rates across all stages for 40-49 and 50-59 year-olds were approximately [from figure] 100-200 and 200-300 per 100,000, respectively).

Among women 50-59 years, 10-year net cancer-survival rates were not significantly different (-0.03 percentage points; p = 0.602) nor were the incidence-based mortality rates (rate ratio 0.94 [95% CI 0.85 to 1.04]; from figure) between jurisdictions screening versus not 40-49 year-olds. Age-specific breast-cancer incident rates were higher in comparator jurisdictions (238.4 vs 217.6 cases per 100,000 women; P < .001). The authors suggest that the lower incidence of breast cancer observed in women in their 50s in screener jurisdictions may signal the benefit of early detection in women age 40-49 years because of lead time and the treatment of less advanced cancers such as DCIS (rates not reported) which may be associated with a reduction in the incidence of invasive cancer in the next decade of life. Screening programs that included women in their 40s were associated with earlier stage migration in women in their 50s: based on screening practices for women aged 40–49, women aged 50–59 had lower proportions of stage I (44.5 vs. 46.8%, p <0.001), and higher proportions of stages II (37.2% vs. 36.0%, p = 0.003) and III (13.6% vs. 12.3%, p < 0.001) in the comparator versus screener groups. Among stage III cancers, there was a significant average decline of 2.5% per year (p < 0.001) which was mainly influenced by an annual reduction of 5.8% per year in the screener jurisdictions (p = 0.001). The overall trend for stage IV was not significant, but there was a significant average annual increase of 1.7% in metastatic disease among women in the comparator jurisdictions (p = 0.025) (a 10.3% increase over the 6 years).

Limitations: There was no ability to determine which breast cancers were screen-detected and which were symptomatically detected; the populations included all women regardless of risk for breast cancer or other screening eligibility factors, and the estimates for survival and mortality outcomes (from diagnoses 2002-2007) would be considered indirect due to advances in treatment that occurred during (e.g., availability of aromatase inhibitors and trastuzumab) and after this timeframe. The estimates of net-survival are prone to both lead and length time biases, though the 10-year follow-up would have mitigated these effects to some extent. The breast cancer incidence rates did not include non-invasive cancers (e.g., DCIS), which would be of interest for determining the possibility and extent of overdiagnosis. The importance of findings for stage at diagnosis, based on proportions of cancers, are hard to interpret without absolute rates especially since overall incident rates are relatively low in women 40-49 years. The analyses for all outcomes were based on screening policies, yet self-reported participation rates (Canadian Community Health Survey) in the screening jurisdictions between 2003-2017 ranged from 31.2% to 52.9% and in comparator jurisdictions from 18.1% to 51.8%. Further, there may have been difference between screener and non-screener jurisdictions in risk for breast cancer, access to care, available treatments, ethnic composition (e.g., via founder mutations in comparator jurisdiction Newfoundland and Labrador having lower survival statistics) that may have affected all but in particular mortality outcomes.

Description of other relevant Canadian research

Stakeholders submitted a recent recommendation by Ontario Health, and its related Health Technology Assessment (HTA) conducted by Ontario Health Technology Advisory Committee. https://www.hqontario.ca/evidence-to-improve-care/health-technology-assessment/reviews-and-recommendations/supplemental-screening-as-an-adjunct-to-mammography-for-breast-cancer-screening-in-people-with-dense-breasts

The recommendation is for "publicly funding supplemental screening as an adjunct to mammography for people with extremely dense breasts." Rationale for the recommendation focused on evidence showing that supplemental screening for people with dense breasts detects more cases of breast cancer and leads to fewer interval cancers despite no evidence describing the impact of supplemental screening on mortality. The recommendation was specific to people with extremely dense breasts, because the highest quality of clinical evidence was in this population and because the cost-effectiveness and budget impact of supplemental screening with ultrasound, digital breast tomosynthesis (DBT), or magnetic resonance imaging (MRI) were more favorable for this population. They recognized that the main harms of supplemental screening for dense breasts are false-positives and overdiagnosis, both of which may lead to unnecessary and burdensome health care treatments, with implications for both individuals and health systems.

The HTA evaluated evidence on accuracy, safety, effectiveness, and cost-effectiveness of supplemental screening (contrast-enhanced DM, DBT, MRI, US) for people with dense breasts, as well as the budget impact, experiences, preferences, understandings, and values of people with dense breasts and their health care providers; and ethical issues.

There are several differences between the eligibility criteria and methods for this HTA and the USPSTF review that likely led to differences in conclusions, such as: i) the HTA focused on studies only enrolling people with dense breasts (i.e., excluding studies with only sub-populations having dense breast), ii) the HTA included comparisons between different supplemental modalities (e.g. US vs. DBT), iii) the HTA included studies using within-participant comparisons (e.g., detection of cancer using MRI after a previous negative mammography [possibly up to two or more years earlier]), iv) the HTA included data on detection rates, stage of cancer, and FPs after only one round of screening (vs. USPSTF requiring two or more rounds), and v) differing methods for the certainty assessments (e.g., same data across reviews for fewer interval cancers from supplemental MRI but HTA gave high and USPSF low certainty). The HTA did not include any studies that evaluated supplemental DBT.

The task force's main decision to rely on the USPSTF review was due to its comprehensiveness across all comparisons of interest (e.g., age to start and stop, intervals, modalities). Further, the eligibility criteria of the USPSTF review (e.g., excluding within-person comparisons) aligned better with the task force's typical methodology.