Evidence Synthesis

Number 231

Screening for Breast Cancer: A Comparative Effectiveness Review for the U.S. Preventive Services Task Force

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 75Q80120D00004, Task Order No. 75Q80121F32004

Prepared by:
Kaiser Permanente Evidence-based Practice Center
Kaiser Permanente Center for Health Research
Portland, OR

Investigators:
Jillian T. Henderson, PhD, MPH
Elizabeth M. Webber, MS
Meghan Weyrich, MPH
Marykate Miller, MS
Joy Melnikow, MD, MPH

AHRQ Publication No. 23-05303-EF-1
May 2023

This report is based on research conducted by the Kaiser Permanente Research Affiliates Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 75Q80120D00004, Task Order No. 75Q80121F32004). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help healthcare decision makers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of healthcare services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information (i.e., in the context of available resources and circumstances presented by individual patients).

The final report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: Tracy Wolff and Howard Tracer at AHRQ; current and former members of the U.S. Preventive Services Task Force who contributed to topic deliberations; Patricia Ganz, MD, Gerald Gartlehner, MD, Nehmat Houssami, MBBS, MPH, MEd, PhD, Karla Kerlikowske, MD, and Lisa Newman, MD, MPH, FACS, FASCO, who provided expert review of the draft report; Erin Abramsohn, DrPH, on behalf of the Centers for Disease Control and Prevention, Sarah Temkim, MD, with the Office of Research on Women's Health, Anna Maria Nápoles, PhD, MPH, and Arielle Gillman, PhD, MPH, on behalf of the National Institute on Minority Health and Health Disparities for providing federal partner review of the draft report; and Heidi Nelson, MD, MPH, for mentoring and expert input. Thanks to Eric (Rocky) Feuer from the National Cancer Institute and Steve Scoppa for technical assistance obtaining detailed SEER breast cancer surveillance data. We also thank Leslie Perdue, MPH, Jill Pope, Melinda Davies, MA, and Tracy Beil, MS, for technical and editorial assistance.

Structured Abstract

Objective: We conducted this systematic review to support the U.S. Preventive Services Task Force (USPSTF) in updating its recommendations on breast cancer screening. Our review addresses the comparative effectiveness of breast cancer screening for improving health outcomes. The review compares different strategies regarding when to screen (e.g., age to start/stop screening, screening interval), screening modalities (e.g., digital breast tomosynthesis [DBT] vs digital mammography [DM]), supplemental screening, or screening strategies defined by breast cancer risk-markers.

Data Sources: We searched MEDLINE, Cochrane Central Register of Controlled Clinical Trials, and the Cochrane Database of Systematic Reviews and the reference lists of previous systematic reviews of breast cancer screening for relevant studies published through August 22, 2022.

Study Selection: We reviewed 10,378 abstracts and assessed 419 full-text articles for inclusion against prespecified inclusion criteria. Eligible studies were conducted in asymptomatic adults eligible for breast cancer screening without clinically significant genetic markers or syndromes associated with high breast cancer risk. Randomized trials and nonrandomized studies of interventions (NRSI) with concurrent comparison groups that reported data over multiple rounds of screening were included to compare health outcomes (e.g., breast cancer mortality) and intermediate outcomes (e.g., risk of advanced cancer); study criteria were broader for identifying potential screening harms. The review was limited to studies conducted in countries with "very high" Human Development Index scores.

Data Analysis: We conducted dual independent critical appraisal of all included studies and extracted study details and outcomes from fair- or good-quality studies. We narratively synthesized results by key question and for each screening comparison. We used random-effects meta-analyses to estimate pooled effects when appropriate. We graded the overall strength of evidence as high, moderate, low, or insufficient based on criteria adapted from the EPC Program.

Results: Health outcomes (KQ1) associated with different screening programs were reported in only two fair-quality NRSIs that addressed the age to stop screening or screening interval. For invasive cancer detection (KQ2), two studies addressed the effect of screening frequency on the characteristics of detected cancers, including one fair-quality RCT of multiple rounds of screening and one fair quality cases-only analysis from the Breast Cancer Surveillance Consortium (BCSC). Four studies of DBT compared with DM, three RCTs [2 good- and 1 fair-quality] and one NRSI reported screening outcomes from more than one round of screening and were included for KQ2. These studies reported characteristics of cancers detected at each round, necessary to assess whether screening resulted in stage shift toward less advanced cases with better prognosis. All 19 studies were included to examine potential harms of different screening approaches (KQ3).

Ages to start or stop screening. One fair-quality NRSI reported an emulated trial analysis of Medicare data (N=264,274) comparing the age to stop screening with reported breast cancer mortality and all-cause mortality (KQ1). Continued screening between the ages of 70 and 74 was

associated with decreased 8-year breast cancer mortality compared with a cessation of screening after age 70 (1 fewer death per 1000 women screened), but no difference was found with continued versus discontinued screening from ages 75 to 84.

Harms (KQ3). Limited evidence on potential risks of overdiagnosis and overtreatment was reported, with more diagnosis and treatment occurring with continued screening, without a mortality benefit.

Interval of Screening. A study conducted in Finland during the years 1985 to 1995 assigned participants (N=14,765) to annual or triennial screening invitations and reported similar breast cancer mortality and all-cause mortality between the two study groups (KQ1). Intermediate cancer detection and progression outcomes (KQ2) were reported in one fair-quality RCT (n = 76,022) in the United Kingdom comparing annual or triennial screening and in one fair-quality registry study using Breast Cancer Screening Consortium (BCSC) data (N = 15,440) to compare annual with biennial screening intervals. The characteristics of tumors diagnosed among those screened with annual versus triennial intervals did not differ in the RCT, though more cancers diagnosed were screen-detected with annual screening (RR: 1.64, 95% CI, 1.28 to 2.09).

In the nonrandomized study, all reported results were stratified by age or hormonal status. Detection of stage IIB+ cancers and cancers with less favorable prognostic characteristics did not differ by screening interval for any reported age groups. Comparisons by menopausal status suggested that premenopausal women with a biennial interval directly preceding their breast cancer diagnosis were at increased risk of stage IIB or higher tumors (RR: 1.28 [95% CI, 1.01 to 1.63], p=.04) and tumors with less favorable prognostic characteristics (RR: 1.11 [95% CI, 1.00 to 1.22], p=.047). For post-menopausal individuals, there was no statistical difference in tumor characteristics by the screening interval preceding diagnosis. The study did not conduct formal tests for interaction in the subgroup comparisons. Neither study reported mortality outcomes, so it is unclear whether these findings would have clinically significant effects on health outcomes.

Harms (KQ3). One RCT reported approximately one additional interval cancer per 1,000 with triennial screening compared with annual screening, and data from four nonrandomized studies were limited and inconsistent. Consistently higher cumulative false positive rates were seen with shorter intervals between screenings. The probability of having at least one false positive recall and biopsy over ten years of screening was higher with annual DBT screening compared with biennial screening, with annual screening resulting in approximately 50 additional false positive biopsies per 1,000 screened over 10 years. Cumulative false positive estimates were highest among young women with dense breasts who were screened annually.

Mammography with Digital Breast Tomosynthesis. No eligible studies reported breast cancer mortality or other health outcomes to compare the effectiveness of screening with DBT versus DM only (KQ1). Intermediate outcomes that compared screening with DBT versus DM were reported in three RCTs (N = 130,196) and one nonrandomized study (N = 92,404) (KQ2). The trials screened all participants with the same screening modality at the second screening round, with DM in three trials and DBT in another. DBT was associated with increased detection of invasive cancer at the first screening round, (pooled RR 1.41, 95% CI 1.20 to 1.64, I^2 8%, 3 trials; n = 129,492); but detection was not statistically different at the second screening round

(pooled RR 0.87, 95% CI 0.73 to 1.05, I^2 0%, 3 trials; n = 105,064) and there was no evidence of a reduced risk of progression to advanced cancer in the second round with DBT compared with DM. The NRSI found higher detection at round one for the group screened with DBT, but higher detection at round two for the group screened with DM at both rounds. The three trials and nonrandomized study reported 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. Limited results stratified by age and density in two of the RCTs did not indicate differences in invasive cancer detection at a second round of screening for people who had been screened with DBT at the first screening round, but tests for interaction were not conducted and estimates were imprecise.

Harms (KQ3). Three large RCTs reported no statistically significant difference in the rates of interval cancers following screening with DBT compared with DM (pooled RR 0.87, 95% CI, 0.64 to 1.17, k = 3, n = 130,196, I^2 0%) but data from five nonrandomized studies were mixed, and interpretation was limited by differences in study design. The effects of DBT screening on recall, false-positive recalls, and biopsy rates varied between trials and by screening round, with no or small statistical differences between study groups, not consistently favoring DBT or DM. The cumulative rates of false-positive recall and false-positive biopsy were slightly lower with DBT compared with DM screening, regardless of screening interval (cumulative probability over 10 years: 50% vs 56% for annual screening, 36% vs 38% with biennial screening). An additional adverse effect of DBT reported, radiation exposure, was approximately two times higher in studies where DBT was performed in addition to DM, but exposure was similar in two studies that used DBT to generate synthetic DM images (DBT/sDM).

Supplemental screening. No eligible studies reported health outcomes when comparing supplemental screening with ultrasound or magnetic resonance imaging (MRI) to usual screening with mammography only (KQ1). No studies of supplemental screening with MRI or ultrasound were included for comparisons of benefit because the trials were incomplete and reported only one screening round (KQ2).

Harms (KQ3). In an RCT among women with dense breasts randomized to supplemental screening with MRI following a negative mammogram screening result, the risk of invasive interval cancer was reduced by approximately half (RR 0.47, 95% CI 0.29 to 0.77). Two studies of ultrasound screening in addition to mammogram did not find significant differences in the rates of interval cancers. Supplemental MRI screening for women with dense breasts with a negative mammography resulted in more recalls, false-positive recalls, and biopsies (95, 80, and 63 per 1,000 screened, respectively) than those receiving DM only. With supplemental ultrasound screening, 48 per 1,000 experienced recall in a trial among women ages 40 to 49 and in a BCSC registry analysis, referral to biopsy and false positive biopsy results were twice as high for the group screened with ultrasound compared with those receiving only mammography.

Limitations: Few published comparative effectiveness trials reported more than a single round of screening. Multiple screening rounds are necessary to identify potential intermediate effects of screening, such as stage shift, limiting conclusions about the potential health consequences of different approaches to screening. Data comparing screening outcomes for subgroups of women with different characteristics or breast cancer risk markers were limited, mainly providing

stratified results only without interaction tests. Findings from older studies included in the review may not be applicable to current programs using newer screening modalities and treatment advances.

Conclusions: We did not find evidence of lower breast cancer mortality or risk of progression to advanced cancer in eligible studies comparing different breast cancer screening strategies. There were downstream consequences (e.g., more false-positive results and biopsy) with supplemental screening. Regular mammography screening is associated with reduced breast cancer mortality for women ages 50 to 69, based on trials conducted over 20 years ago, and longer term followup from the trials has not altered these conclusions. Changes in population health, imaging technologies, and available treatments could limit the applicability of older trials. Additionally, nearly all of the trials were conducted outside of the United States and enrolled mainly White European populations. Inequities in breast cancer mortality and length of survival, especially for Black women, also warrants greater attention to health care interventions following screening, including prompt follow-up, diagnosis, and access to high quality treatment and support services, as well as more dedicated research to find effective treatments for triple negative cancers. The limited early evidence from newer comparative effectiveness trials does not yet provide answers to questions about the benefits or harms of different screening strategies, but ongoing and pending trials may further the science in coming years.

Table of Contents

Figures

Tables

Appendixes

Chapter 1. Introduction

Purpose

This comparative effectiveness evidence review and synthesis will be used by the U.S. Preventive Services Task Force (USPSTF) to inform their update to the 2016 recommendations on breast cancer screening.[1]

Condition Background

Condition Definition

Breast cancer is a proliferation of malignant cells that usually originates in the terminal ductal-lobular unit in breast tissue.[2] Invasive breast cancer extends through the basement membrane of the breast and into the adjacent stroma, allowing for potential metastatic spread. The most common sites of metastases are adjacent lymph nodes, lung, liver, and bone. The most common invasive breast cancer is invasive (infiltrating) ductal carcinoma, histologically categorized as "no special type." A smaller percentage (5-15%) are invasive (infiltrating) lobular carcinoma. The remainder of invasive cancers are less common subtypes with specific histologic features (20-30%) including tubular, papillary, apocrine, medullary, metaplastic, and mucinous.[3]

Noninvasive (in-situ) lesions are contained in the ductal-lobular unit and do not extend into the basement membrane or surrounding tissue. When confined to the duct, these lesions can be classified as usual ductal hyperplasia (UDH), flat epithelial atypia (FEA), atypical ductal hyperplasia (ADH), or ductal carcinoma in situ (DCIS).[4] DCIS is heterogenous and has varying clinical behavior and pathologic characteristics; some forms are viewed as precursor lesions for invasive ductal carcinoma and it is sometimes referred to as noninvasive or stage 0 cancer.[5] DCIS accounts for approximately 20 to 25 percent of all breast neoplasms, and incidence has increased with the widespread use of mammographic screening.[4] Lesions confined to the lobule are less common and include lobular intraepithelial neoplasia (LIN), lesions composed of benign, non-infiltrating lobular proliferations of the mammary epithelium; atypical lobular hyperplasia (ALH); and lobular carcinoma in situ (LCIS).[6, 7] LIN is an established risk marker for invasive ductal or lobular breast cancer because it is associated with bilateral invasive cancer.[6, 8]

Prevalence and Burden

Overall, among U.S. women, breast cancer is the second most common cancer (excluding non-melanoma skin cancer) and the second most common cause of cancer death after lung cancer.[9] In 2019, an estimated 3,771,795 women were living with invasive breast cancer in the United States.[10] With respect to yearly incidence, in 2022 an estimated 287,850 women in the United States were diagnosed with invasive breast cancer (representing 15% of all new cancer cases) and 43,250 were estimated to have died of breast cancer (representing 7% of all cancer deaths).[11] Although it is the second leading cause of cancer mortality for women overall,[12] it is the leading cause of mortality from cancer for Hispanic women.[13] Based on the most recent lifetime risk

estimates for the general population, approximately 12.9 percent of women will develop breast cancer during their lives, and 2.6 percent will die from the disease.[14]

An increasing trend in invasive cancer incidence has been observed with the widespread adoption of breast cancer screening programs. Overall, since 2004 there has been a 0.5 average annual percent rise in invasive cancer diagnoses.[15] A steeper increase in incidence has been observed for individuals ages 40 to 49 years from 2015 to 2019 (2.0 average annual percent change; rates were 162 per 100,000 in 2015 compared with 172 per 1000,000 in 2019). The rising incidence is mainly attribute to increases in the diagnosis of localized cancers.[16] Mortality from breast cancer has continued to decline, albeit less steeply in recent years; by approximately 1.3 percent each year on average from 2010-2019.[15]

Breast cancer incidence varies by age and race (**Table 1, Figure 1**). Incidence rates for invasive cancer are highest for women ages 65-74 (447.7 per 100,000) and decline with further increasing age.[10] Overall, average rates of invasive breast cancer from 2015-2019 were highest among Non-Hispanic White women (137.6 per 100,000 women) and second highest for Non-Hispanic Black women (129.6 per 100,000).[10] Hispanic women of any race experienced the lowest incidence (99.9 per 100,000). Rates are slightly higher for Non-Hispanic Asian/Pacific Islander women (106.9 per 100,000) and Non-Hispanic American Indian/Alaska Native (AI/AN) women (111.3 per 100,000 women),[10]although incidence rates for the Native American/Alaska Native population vary widely by region.[17]

Stage at diagnosis also varies by age and race (breast cancer staging is described in **Appendix F Table 1**). Case incidence is the lowest for younger age groups (23.7 per 100,000 for all women ages 15-39 years) but younger women are more likely to be diagnosed at a later stage (50.2% at regional or distant stage for women ages 15-39 years), than older women (35.3% for women ages 40-64 years; 27.3% for women ages 65-74 years; 28.3% for women ages 75+ years at either regional or distant stage at diagnosis).[11] Non-Hispanic Black women are more likely to be diagnosed with cancer beyond stage 1 than other race/ethnic groups, and even small tumors are more likely to present with lymph node involvement or metastases.[18]

Breast cancer mortality has steadily declined since the 1990's but remains persistently higher for Black women than for all other race and ethnicity groups (**Table 1**).[19, 20] Black women are 40 percent more likely to die from breast cancer compared with White women,[21] despite reporting similar or higher guideline concordant screening[22] and lower overall breast cancer incidence. Current estimates of breast cancer mortality (2016 to 2020) are 27.6 per 100,000 among non-Hispanic Black women compared with 19.7 per 100,000 for non-Hispanic White women.[23] Breast cancer mortality rates are lower among Hispanic (13.7 per 100,000), Asian/Pacific Islander (11.7 per 100,000), and AI/AN women (17.6 per 100,000). In terms of survival among those diagnosed with breast cancer, estimates from 2012-2018 showed an overall 5-year survival rate for breast cancer of 92.0 percent for White women and 82.6 percent for Black women.[13] Five-year breast cancer survival from the estimates for Hispanic, Asian/Pacific Islander, and AI/AN women were also higher than for Black women; 88.3 percent, 91.6 percent, and 90.1 percent, respectively. There are also disparities in incidence and survival among Black women under the age of 40. While screening is not currently recommended before age 40 because incidence rates are very low (23.7 per 100,000),[10] Black women diagnosed with breast cancer

before age 40 have the highest mortality (3.9 per 100,000)[10] and lower 5-year survival (76.9% as compared to 87.1% for White women),[24] and the breast cancer mortality rate for Black women aged 35-39 is nearly double that of White women in the same age group (11.3 versus 6.2 deaths per 100,000).[23] These inequities are discussed in detail later in this report (see **Discussion**).

Etiology and Natural History

Breast cancer develops through inherited and acquired pathogenic variants in oncogenes and tumor suppressor genes that would otherwise support normal cellular growth and replication. Inherited pathogenic variants in breast cancer susceptibility genes (e.g., *BRCA1, TP53, PTEN)* represent the minority of breast cancer cases. Most cases are sporadic arising from endogenous and exogenous environmental factors. Specific external influences, such as toxic environmental exposures, known to act on specific regulatory gene have not yet been elucidated.[25, 26] Estrogen and progestin are also implicated in tumorigenesis and growth due to the observed associations of factors such as age of menarche and menopause and parity.[26] Other potential pathways from external exposures to breast cancer continue to be investigated, including possible roles of biological aging related DNA methylation,[27] vitamin D,[28] inflammatory conditions,[29] sleep patterns,[30] and virally mediated carcinogenesis.[31]

Most breast cancers are invasive, meaning that they have infiltrated surrounding breast tissue beyond the ducts or glands where they originated. Cancer subtypes are classified according to their histology and molecular markers (e.g., ER, PR, HER2). The three main clinical subtypes of invasive breast cancer that are commonly assessed using biological markers are hormone receptor (HR) positive, human epidermal growth factor receptor 2 (HER2) positive, and triple-negative breast cancer that do not contain HR receptors (progesterone or estrogen) or HER2. Prognosis and treatment are informed by these factors.[32-35] The most common breast cancer subtype defined by receptor status is HR+/HER2-, also referred to as Luminal A tumors, and this subtype represents nearly three-quarters of invasive breast cancers (73%). These tumors are usually less aggressive than other subtypes and responsive to hormone therapy, resulting in a better prognosis. Luminal B (HR+/HER2+) tumors, representing approximately eleven percent of invasive breast cancers, are often higher grade than Luminal A tumors and have poorer outcomes. (High positivity of an additional marker for protein Ki67 that indicates actively dividing cells is also sometimes used to help define this subtype.) Twelve percent of invasive breast cancers are triple negative (HR-/HER2-).[32] Compared to the other subtypes, triple negative cancers have the worst prognosis, and reductions in mortality over the past two decades have been smaller for people affected by these cancers than for all other subtypes.[36] In the United States, triple negative cancers are twice as common among Black women (24.1 per 100,000) compared with White women (12.4 per 100,000).[37] Triple negative cancers account for 19 percent of breast cancers diagnosed among Black women compared with 9 percent of cancers diagnosed among White women and 11 percent of cancers among Hispanic and among Asian/Pacific Islander women.[14] Rates are also lower among Hispanic (11.0 per 100,000), AI/AN (10.8 per 100,000), and Asian/Pacific Islander (9.1 per 100,000) women.[37] These cancers occur more often in premenopausal women and those with BRCA1 gene mutations.[38] Regardless of the cancer subtype based on receptor status, stage at diagnosis is the strongest prognostic factor.

Ductal carcinoma in situ (DCIS) is the most common noninvasive breast condition detected with mammography.[14] Approximately 16 percent of breast neoplasms (invasive cancer and in situ conditions) diagnosed in the United States are DCIS.[32] Most DCIS cases are diagnosed with breast imaging, owing to the widespread use of screening mammography and its ability to identify microcalcifications.[39, 40] Older studies of palpable DCIS lesions indicated that 14 to 53 percent of untreated DCIS progress to invasive cancer over 8 to 22 years.[41] These studies may have limited applicability, however, since the natural history of screen-detected lesions would likely differ from clinically presenting cancers, and represent the majority of DCIS cases in the current era of widespread mammography screening. It is also not clear whether some DCIS cases regress, and the potential for overdiagnosis of breast cancer hinges heavily upon this possibility.[42] Because treatment is generally recommended, the natural history of screen-detected DCIS is unclear in terms of the percentage of cases that would have progressed to invasive cancer in the absence of treatment.[33]

DCIS is considered a precursor lesion or a risk marker for invasive cancer, especially for specific groups or when certain features are present.[4, 43] Characteristics found to be associated with subsequent invasive breast cancer include detection at a young age, clinical detection (palpation rather than screen detection), and detection in a person with Black race.[44] Lesion characteristics such as involved margins, high histologic grade, and high p16 expression are also associated with risk for subsequent invasive cancer.[45-47] A recent population-based analysis identified a three-fold increase in breast cancer mortality over 20 years after DCIS diagnosis and treatment, half due to contralateral invasive breast cancer.[44] These findings suggest that in addition to posing a risk of local progression to invasive cancer, DCIS is also a marker of elevated breast cancer risk.

Breast Cancer Risk Factors

Risk for primary breast cancer is highest among women with previous high-risk breast cancer lesions (DCIS, LCIS, atypical ductal hyperplasia, atypical lobular hyperplasia); extensive family histories of breast and ovarian cancer; clinically significant genetic markers or syndromes associated with a high risk of breast cancer (e.g., *BRCA1* or *BRCA2* pathogenic variants, Li-Fraumeni syndrome); or previous large doses of chest radiation before age 30 years. Women in these high-risk groups require different screening regimens than the general population.[48, 49]

Non-modifiable factors associated with increased risk among women eligible for routine screening include increased age;[10] at least one first- or second-degree relative with breast cancer;[50-52] and heterogeneously or extremely dense breast tissue.[53] Women in their 40s with extremely dense breast tissue or at least 1 first-degree relative with breast cancer are estimated to have a 2-fold increased risk of breast cancer.[50] Additional risk factors that have been associated with an increased risk of breast cancer include use of menopausal hormone therapy, increasing body mass index, alcohol use, nulliparity or giving birth after age 35.[54] Risk factors associated with a decreased risk of breast cancer include breastfeeding and increased physical activity.[54]

Breast density is a radiographic measure of breast tissue that is associated with increased risk for breast cancer and reduced mammography sensitivity. It describes the amount and distribution of dense fibrous and glandular tissue relative to surrounding fat tissue. Having more dense breast tissue may make it more difficult to find tumors using imaging technologies. Breast density is

currently evaluated with the BI-RADS (Breast Imaging Reporting and Data System) to standardize the interpretation of mammography results using four categories: (a) almost entirely fatty, (b) scattered fibro glandular densities, (c) heterogeneously dense, and (d) extremely dense.[55] These last two categories represent women considered to have dense breasts. Reproducibility of classifications is inconsistent and one in five women would be categorized into a different BI-RADS density category by the same radiologist during the next exam.[56] Increased breast density is more common among younger women, although it occurs in roughly one-third of women older than 65 years.[57, 58] Distributions of breast density estimated from BCSC data[59] indicate that among premenopausal women, over half are considered to have dense breasts (heterogeneously or extremely dense); proportions are highest among Asian women (80%) compared with White women (61%), Black women (56%), and Hispanic women (57%). In menopause, the proportion of women with dense breasts are lower, but remain highest among Asian women (55%) compared with White women (38%), Black women (32%), and Hispanic women (31%).

Interactions between hormonal status, breast density, and cancer risk suggest a complex relationship. Breast density can change over time, and associated risks of invasive breast cancer can vary within individuals across the lifespan.[60] Breast density is influenced by hormonal medications (such as tamoxifen or postmenopausal hormone replacement therapy), pregnancy, BMI, and age.[61, 62] Data from the BCSC suggest that having heterogeneously or extremely dense breasts (compared with scattered fibroglandular density) accounts for higher population-attributable risk proportion of invasive breast cancers for premenopausal women (24% – 35%) than for postmenopausal women (13% – 17%).[59] The study also reported differences according to race and ethnicity in the contributions of BMI, breast density, and menopausal status to invasive breast cancer risks. For example, while more Asian women are classified as having dense breasts, the magnitude of the association of extremely dense breasts with their invasive cancer risk was lower relative to Black, White, and Hispanic women. With regard to BMI, premenopausal White women with BMI >35 had increased breast cancer risk, but the association was not observed for Black, Asian, and Hispanic women whereas for postmenopausal women, increased risks for breast cancer for all groups were seen among those with BMI >35, and the risk was most elevated for Asian and Black women.[59]

A large cohort study from the US highlights the potential importance of including additional risk factors to inform supplemental screening strategies and reduce false-positive rates. Data for the years 2005 to 2014 from a prospective screening cohort of 638,856 women ages 40 to 74 years obtaining digital mammography at Breast Cancer Surveillance Consortium (BCSC) imaging facilities provides important epidemiologic information on the association of breast density with invasive cancer incidence.[63] Nearly half (47%) of women screened were identified as having dense breasts (heterogeneously or extremely dense) and overall 60 percent of advanced cancers were in these women. One-third of women with dense breasts, however, had very low rates of advanced cancer within a year of screening. The highest rates of advanced cancer were seen in women with heterogeneously dense or extremely dense breasts and at least 2.5% 5-year risk of breast cancer calculated using the BCSC risk calculator (described below).

An analysis of temporal trends in BI-RADS density readings from over 2 million mammography screenings at BCSC facilities found that despite changes in classification guidelines and the

increasing use of DBT for screening, the distribution of breast density across time and age groups has remained relatively stable.[64] Breast density over the lifespan in individual women, however, is known to change, and there is evidence that reductions in density are associated with reduced risk for invasive cancer.[65]

Multivariable Risk Prediction

Models estimating risk for breast cancer include common clinical risk factors, such as age, age at menarche, age at birth of first child, number of first-degree relatives with breast cancer, and results of previous benign breast biopsies. Additional variables differ between models including race, BMI, breast density, menopause status, use of hormone therapy, and additional family histories, among others. Risk factors are categorized and weighted differently in each model.[66-73] Risk estimation from genome wide association studies is also being used to develop polygenic risk scores.[74] Estimates of lifetime risk of breast cancer of over 15 or 20 percent are considered high.[53] Although several risk prediction models have been developed for clinical use, current versions demonstrated poor predictive performance in estimating an individual woman's risk when validated in screening populations.[72]

Rationale for Screening and Current Clinical Practices

Screening for breast cancer is a secondary prevention intervention that is initiated through primary care or other cancer prevention focused practice settings. Screening can prevent breast cancer morbidity and mortality by identifying cancer at an earlier stage than it would have presented clinically, allowing for lower intensity treatment and higher survival rates. Screening is based on mammography technology used to visually detect lesions before they become clinically apparent (**Appendix F Table 2**).

Randomized trials have established the overall effectiveness of mammography screening in reducing breast cancer mortality for women ages 50 to 69 years and were the focus of previous reviews for the USPSTF.[75, 76] Longer term followup and secondary analyses of these foundational screening effectiveness trials have been published since the 2016 review conducted for the USPSTF, but the results do not substantively change earlier conclusions (See **Discussion**).[77-81] Additionally, these older trials have limited value for identifying differential effectiveness for different subgroups of participants (apart from age differences) due to the populations enrolled and study designs. New trials of screening cannot ethically randomize participants to no screening since screening is now known to confer a mortality benefit. As for other topics that the USPSTF considers where an intervention is established, new trials seek to refine the approach using active comparators (comparative effectiveness trials).

Current practice guidelines from professional societies, guideline groups and governmental agencies recommend breast cancer screening for average risk women beginning no later than age 50 (**Table 2**). Differences across guidelines concern appropriate ages to begin and discontinue routine screening, role of risk assessment in screening decisions, appropriate screening interval, and type of screening modality. Only the American College of Obstetricians and Gynecologists (ACOG) practice guideline refers to supplemental testing in women with dense breasts.

Conventional digital mammography (DM) has essentially replaced film mammography as the primary method for breast cancer screening in the US and involves conversion of x-rays that pass through the breast tissue to electronic signals that produce a digital image. Routine screening with DM images the breast from two angles (craniocaudal and mediolateral oblique). Digital breast tomosynthesis (DBT) is a newer technology that is increasingly being used as a primary breast cancer screening strategy. The technology acquires images from multiple angles and is sometimes referred to as 3D mammography. Screening with DBT is usually accompanied by DM imaging but some imaging devices can produce 'synthetic' 2D mammography (sDM) images equivalent to the two-view screening image from standard mammography.[82, 83] An increase of 20 to 30 percent in radiation dose has been observed with DBT compared with conventional DM. When DBT is performed in combination with DM the radiation dose is at least double that of DM alone. The use of sDM constructed based on DBT images reduces the total radiation dose by 30 to 40 percent compared with DBT plus DM screening.[84] Recent systematic reviews have found no significant difference in the accuracy of DBT with DM or sDM.[85, 86] However, DBT/sDM was found to reduce the number of patient recalls ($p=0.006$) as well as improve the positive predictive value of screening ($p=0.047$) compared with DBT/DM.[86]

In addition to primary screening, the value of supplemental screening for women with dense breasts is an active area of research. The lack of a standardized and reliable assessment tool for measuring breast density, and its variability over a woman's lifespan, pose challenges for research into the optimal screening strategy for women identified as having dense breasts. The supplemental screening modalities used to screen women with dense breasts include handheld breast ultrasonography, automated whole breast ultrasound, magnetic resonance imaging (MRI), and DBT. Limited data are available for evaluating its performance in average risk populations with only breast density as a risk factor.

The potential harms of breast cancer screening include a risk of false-positive results which may lead to psychological harms, additional testing, and invasive diagnostic follow-up procedures (e.g., biopsy). In addition, overdiagnosis and overtreatment are harms from the detection, diagnosis, and treatment of DCIS or invasive cancers that would not have led to health problems without detection.[87, 88] The recurrent radiation exposure from a lifetime program of mammography screening has been proposed as increased risk for breast cancer, particularly for women with larger breasts.[89] DBT screening has been associated with an increased risk for breast biopsy compared with conventional digital mammography and higher radiation exposure when performed in conjunction with conventional digital mammography, which has been estimated to be as much as two times greater than that associated with DM (with women with dense breasts exposed to even higher doses).[90, 91] However, this level of radiation exposure falls below the FDA limit for standard mammography[92, 93] and newer, synthetic 2D DBT can lower radiation to levels comparable to or slightly above those of a conventional mammogram.[94] Potential harms of supplemental screening in women with dense breasts may include additional false-positive recall and biopsies when compared with standard screening mammography.[95]

There are several approaches to reading mammography images, including single and double reading, computer-aided detection (CAD), and artificial intelligence-supported reading. While mammography reading by two radiologists (double reading) is standard practice in parts of

Europe,[96, 97] single reading is more common in the United States. These different approaches to mammography reading can affect the test performance of the different modalities.[98-100]

Although the Task Force's 2016 recommendation statement for screening in average risk populations did not endorse use of DBT, a 2019 study found a strong increase in its utilization for breast cancer screening in the US. In 2015, DBT was used for 12.9 percent of screening examinations, and by 2017, was used in 43.2 percent of screening examinations.[101] As of December 2020, 74 percent of facilities certified by the Mammography Quality Standards Act (MQSA) program were certified for both DBT and full-field digital mammography.[102]

Many US states have enacted breast density notification laws that require insurance coverage for supplemental screening in women with dense breasts and notification directly to women regarding their breast density results and the potential effect of their breast density on the sensitivity of screening and breast cancer risk.

Interventions and Treatment Approaches

Patients with suspicious mammographic abnormalities (or palpable breast masses) may undergo additional diagnostic imaging as well as biopsy.[103, 104] The most common type of biopsy is needle biopsy (core needle biopsy or fine needle aspiration); surgical biopsy is performed if the results of the needle biopsy are unclear.[105] The pathologic stage of cancer is used to determine prognosis and inform treatment decisions.

The American Joint Committee on Cancer (AJCC) system defines cancer stages based on tumor size (T), lymph node involvement (N), and presence of metastasis (M) (**Appendix F Table 1**).[106] The most recent 8th edition of the AJCC staging guidelines incorporates histologic grade and biomarkers, including, estrogen receptor (ER) expression, progesterone receptor (PR) expression, human epidermal growth factor receptor 2 (HER2) expression, and commercially available gene-based assay results.[107] Main categories are generally defined as noninvasive cancers, such as DCIS (stage 0), localized (stage I and some stage II), locally advanced or regional (some stage II and stage III), and metastatic disease (stage IV).

Treatment regimens are highly individualized according to each patient's clinical status, cancer stage, tumor biomarkers, clinical subtype, and personal preferences, and vary in potential side effects.[33, 108] Surgical investigation is sometimes required to determine whether neighboring lymph nodes have been affected. Biopsy of sentinel nodes has demonstrated fewer long-term harms; however certain scenarios encourage investigation of the axillary nodes. Axillary lymph node dissection carries higher risk of long-term harms including numbness, swelling and pain, and is more likely to be utilized on Black and Hispanic women, as well as women without health insurance.[109, 110] Survival varies by stage, and the 5-year relative survival rates for breast cancer in the United States are 99.1 percent with localized disease, 86.1 percent with regional disease, and 30 percent with metastatic disease.[11]

Previous Evidence Reviews

Multiple evidence reviews were used to update the 2016 USPSTF breast cancer screening recommendation.[95, 111] Previous reviews addressed key questions on the effectiveness of

mammography screening in reducing breast cancer-specific and all-cause mortality, advanced breast cancer, and treatment-related morbidity compared with no screening and harms of screening. Results of key questions for the evidence review on benefits and harms of screening are summarized in this section of the report because the current evidence update does not update the foundational evidence of the effectiveness of mammography compared with no screening. A detailed summary of these findings is provided in **Appendix A**.

Screening Effectiveness

Nine fair-quality RCTs comparing mammography screening with nonscreening provided outcomes that addressed several key questions in the 2016 review.[76] Trials enrolling over 600,000 women were conducted in the United States, Canada, United Kingdom, and Sweden. Across all trials, the mean or median screening intervention time ranged from 3.5 to 14.6 years, case accrual time from 7.0 to 17.4 years, and followup time from 11.2 to 21.9 years. Meta-analyses were conducted for breast cancer mortality outcomes using the longest followup data available. These analyses estimated that over a 10-year period, screening 10,000 women ages 50 to 69 would result in 12.5 (95% CI, 5.0 to 19.5) fewer breast cancer deaths; however, estimates were not statistically significant for women aged 39 to 49 years and those aged 70 to 74 years. All-cause mortality was not reduced with screening for any age group. A statistically significant reduction in advanced disease was found for women aged 50 years or older who were randomly assigned to undergo screening, but not for women aged 39 to 49 years. In general, observational studies reported greater breast cancer mortality reduction (25% to 31% among women invited to screening) than RCTs (19% to 22% using intention-to-treat analysis) for women ages 50 to 69 years. Two observational studies of women in their 40s invited to or participating in screening indicated 26 to 44 percent reduction in breast cancer mortality.

Screening Harms

Harms of screening summarized in the 2016 evidence review[76] included false-positive and false-negative results, additional imaging, and biopsy; overdiagnosis; anxiety, distress, and other psychological responses; pain and discomfort; and radiation exposure. False-positive results were common and are higher for annual screening, younger women, and women with dense breasts. Although overdiagnosis, anxiety, pain, and radiation exposure may cause harm, their effects on individual women are difficult to estimate and vary widely.

Previous USPSTF Recommendations

In 2016, the USPSTF recommended biennial screening mammography for women aged 50 to 74 years (B recommendation) and concluded that the decision to start screening mammography in women prior to age 50 years should be an individual one (C recommendation). Additionally, the USPSTF concluded that the current evidence is insufficient to assess the balance of benefits and harms of screening mammography in women aged 75 years or older (I statement), the use of digital breast tomosynthesis (DBT) as a primary screening method (I statement), or the use of supplemental screening methods in women identified to have dense breasts (I statement).[1]

Chapter 2. Methods

Scope and Purpose

This updated review will focus new Key Questions for the USPSTF on the comparative effectiveness and harms of different screening strategies. These include strategies based on individual characteristics and risk markers of the screening population; mammography screening modalities (i.e., DBT versus digital mammography); and screening delivery approaches (e.g., intervals, use of supplemental screening). This review does not include evidence on differences in the use of technologies intended to improve the reading of mammography (e.g., computer assisted, artificial intelligence) nor evidence on interventions aimed at increasing screening uptake or adherence. The evidence synthesis follows USPSTF procedures and methods for systematic reviews.[112]

Key Questions and Analytic Framework

The review addresses three Key Questions (KQs) illustrated in an analytic framework (**Figure 2**), with predefined inclusion and exclusion criteria describing the target population, study design, intervention, and outcomes (**Appendix B Table 1).**

1. What is the comparative effectiveness of different mammography-based breast cancer screening strategies (e.g., by modality, interval, initiation age, use of supplemental imaging, or personalization based on risk factors) on breast cancer morbidity and breast cancer–specific or all-cause mortality?
    a. Does comparative effectiveness differ by population characteristics and risk markers (e.g., age, breast density, race/ethnicity, family history)?
2. What is the comparative effectiveness of different mammography-based breast cancer screening strategies (e.g., by modality, interval, initiation age, use of supplemental imaging, or personalization based on risk factors) on the incidence and progression to advanced breast cancer?
    a. Does comparative effectiveness differ by population characteristics and risk markers (e.g., age, breast density, race/ethnicity, family history)?
3. What are the comparative harms of different breast mammography-based cancer screening strategies (e.g., by modality, interval, initiation age, use of supplemental imaging, or personalization based on risk factors)?
    a. Do the comparative harms vary by population characteristics and risk markers (e.g., age, breast density, race/ethnicity, family history)?

Data Sources and Searches

Studies included in the 2016 USPSTF reviews[76, 95] were evaluated for inclusion against the updated eligibility criteria for the current review. In addition, database searches of MEDLINE, Cochrane Central Register of Controlled Clinical Trials, and the Cochrane Database of Systematic Reviews were conducted by a research librarian for relevant studies published between January 2014 and August 22, 2022. A second research librarian peer-reviewed the search strategy (**Appendix B**). Additionally, investigators examined the reference lists of other

previously published reviews, meta-analyses, and primary studies and new publications identified from table-of-contents alerts and searched ClinicalTrials.gov (https://ClinicalTrials.gov/) for ongoing trials. We supplemented these searches with suggestions from experts and articles identified through news and table-of-contents alerts. We managed all literature search results in EndNote® X9 (Thomson Reuters, New York, NY).

Study Selection

Two reviewers independently evaluated all titles and abstracts using prespecified inclusion and exclusion criteria developed for this review (**Appendix B Table 1**). The full texts of potentially relevant studies were then further evaluated by two reviewers to determine final inclusion. Disagreements regarding inclusion at both the abstract and full text review level were resolved via discussion or with the input of a third reviewer as needed. A list of studies excluded during full text review are included in **Appendix D,** along with reasons for exclusion. DistillerSR (Evidence Partners, Ottawa, Canada) was used to conduct abstract and full-text review.

Population

Routine breast cancer screening applies to adults with female sex-specific breast tissue without current symptoms of breast cancer, previous breast cancer, or high-risk breast lesions (DCIS, LCIS, atypical ductal hyperplasia, atypical lobular hyperplasia). Additional exclusions include adults with clinically significant genetic markers or syndromes associated with a high risk of breast cancer (e.g., *BRCA1* or *BRCA2* pathogenic variants, Li-Fraumeni syndrome) and those who received previous large doses of chest radiation before age 30 years because these represent high-risk conditions that may require different screening regimens. Screening in high-risk populations is outside the scope of the review.

We excluded studies conducted with populations receiving breast imaging performed for diagnostic or surveillance purposes, and studies that included a mix of screening and diagnostic populations where results were not stratified by indication. Studies of screening following gender-affirming medical treatment with exogenous estrogen, if such studies had been identified, would be excluded since care would be specific to individual clinical histories and involve specialty consultation beyond primary care. Throughout the report we incorporate gender inclusive language (people, individuals, persons with breasts) when referring to the screening population to recognize that not all people at risk of breast cancer and eligible for screening are women.[113] We use the term women primarily when referring to studies using this language to reflect the evidence base yet acknowledge that previous studies did not collect nuanced data on gender and most likely conflated biological sex characteristics with gender. While the search and eligibility criteria for this review used an inclusive definition of women, all studies referred to their populations as women and this term was used in the report to reflect the evidence base. The review was limited to studies conducted in countries with “very high” Human Development Index scores (as of 2020) as published by the United Nations Development Programme.[114]

Intervention and Comparison

We included studies that evaluated the following breast imaging screening modalities: digital mammography (DM), digital breast tomosynthesis (DBT), ultrasound, and magnetic resonance

imaging (MRI). All eligible studies included a comparator group that received screening with DM only in one of the study arms to assess the comparative effectiveness of screening relative to this established evidence-based practice. Studies were included that examined the effects of varying the primary mammography screening modality (e.g., DBT vs DM) or using supplemental/adjunctive imaging in addition to DM (e.g., DM plus ultrasound). In addition, we included studies that compared different screening strategies, including different screening intervals, ages to begin or end screening, and personalization of the screening program based on risk factors, markers, or risk assessment tools. We excluded studies of breast self-examination, clinical breast examination, and film mammography.

Outcomes

Outcomes included for KQ1 were mortality (breast cancer and all-cause), breast cancer morbidity (e.g., treatment-related morbidity, physical/functional impairment), and quality of life or subjective well-being.

For KQ2, outcomes included advanced cancer detection and the stage distribution of screen-detected invasive breast cancers from at least 2 rounds of screening and followup. For the outcome of 'advanced cancer' data were abstracted as reported by the study authors. The majority of studies reporting 'late-stage cancer' or 'advanced cancer' define this as any cancer diagnosed at stage II or later, and in a few cases as stage IIB or later to exclude localized stage II cancers from the definition of advanced cancer. Some studies provided detailed data on the stage, node, and size of detected cancers. When possible, we prioritized reporting outcomes related to invasive cancers. We have noted cases where DCIS was reported combined with invasive cancer cases.

For KQ3, eligible harms outcomes included false-positive and false-negative findings at screening and biopsy, screening recall rate (need for further evaluation following screening), psychological harms (e.g., anxiety, depression, decrease in quality of life) associated with screening or followup, and overdiagnosis and overtreatment (as defined by the study), and rates of interval cancers (occurring between screening rounds) were included for KQ3. Rates of interval cancers reported in screening studies reflect a combination of clinically presenting cancers that were missed during previous screening exams and incident cancers emerging between screening rounds or during a period of followup (ideally at least 12 months after screening).

Study Design

To minimize bias, only individual-participant data meta-analyses, randomized controlled trials, controlled clinical trials, and cohort studies analyzing the outcomes of different screening strategies between comparison groups over a concurrent time period were included. Studies of test accuracy, including where participants served as their own controls to evaluate test performance (e.g., accuracy) were excluded as were modeling studies. 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). Finally, we excluded studies that compared breast cancer mortality and detection outcomes at the population level if the study

examined differences in screening outcomes by different regional policies affecting implementation since risks of bias in such ecological studies are untenably high.[115-118]

Studies of test performance studies and studies reporting cancer detection at a single round of screening were excluded because they are not able to provide adequate evidence that a breast screening program would necessarily improve health outcomes. The natural history of breast cancer is such that a screening test could be better at detecting cancers unlikely to impact health outcomes (overdiagnosis) or detect them at an earlier time without changing the outcome (lead time bias). Cancer sojourn time describes a related concept whereby slower growing cancers are stable and detectable for a longer period before advancing to metastatic disease, and can therefore comprise a larger proportion of screen-detected cancers. While overdiagnosis is known to occur, its quantification in breast cancer screening programs is challenging, and studies have reported a wide range of estimates (11% to 22% based on individual data, higher estimates using aggregated data).[119-121] All of these factors contribute to the understanding that a small increase in screening sensitivity might lead to detection of cases at an earlier point in time or with different characteristics that without necessarily changing the risk of morbidity or mortality from breast cancer.[87] Conversely, fast growing cancers with high metastatic potential may have similarly poor outcomes even with small improvements in overall screening test sensitivity. Due to these factors, it is important to evaluate the effects of screening programs over several rounds of screening rather than at a single point in time, and to examine health outcomes to determine whether those screened benefited from the practice and the treatments received as a result.[121]

Studies were included for KQ2 only if they reported at least two rounds of screening and followup. Data from a second round of screening offer insight on the cumulative effects of screening and whether cancers detected at an earlier round were consequential or could have been detected and treated at a later point in time without adverse health consequences.[87] This is important for the reasons described above. In the absence of health outcome results, a more effective screening program would be expected to reduce detection of advanced cancers that had progressed by a second round of screening relative to a comparison screening program.

Commonly reported outcomes on potential harms of screening (KQ3) including recall (i.e., return for additional imaging), false positive recall, biopsy, and false-positive biopsy rates were obtained from the same studies included for effectiveness (KQ1/KQ2), or from studies providing data across multiple screening rounds. Studies with multiple screening rounds are essential for assessing harms where cumulative effects are important to consider. For example, a screening program with high recall at one round may show lower recall in subsequent rounds and the balance over time would be a more accurate measure of the outcome. Differences in screen-detection of DCIS lesions can be viewed as a measure of potential overdiagnosis. The incidence of DCIS lesions has increased substantially with the advent of mammography screening programs, but their natural history and contribution to health risks remain unclear,[122] as discussed above. Screen-detected DCIS may contribute to overdetection and overtreatment since patients may choose to undergo treatment (lumpectomy, mastectomy, radiation, hormone therapy).

For rates of interval cancer (including false negatives) results of studies with only a single round of screening were included because this finding has health implications at each round. Followup

data to assess interval cancers can be obtained from a variety of sources, including prospectively collected data, cancer registries, administrative data, and medical records. For other uncommonly reported outcomes (e.g., quality of life; psychological health; radiation), we also included studies reporting findings based on a single round of screening, large population-based case-control studies, and followup surveys of patient experiences from participants in large trials or cohort studies.

Where multiple publications on similar analyses from the same registry or observational cohort studies were available, such as analyses using BCSC data, the most recently analyzed data available were selected for inclusion in the review.

Quality Assessment and Data Abstraction

Two reviewers independently rated all eligible studies for potential risks of bias. Each study was given a rating of "good," "fair," or "poor" based on consensus from two reviewers. Discordant quality ratings were resolved through discussion and input from a third reviewer as needed. For randomized trials, USPSTF-specific criteria for assessing risk of bias were applied. For nonrandomized studies, we answered signaling questions from the Risk of Bias in Non-randomized Studies of Interventions (ROBIN-I) tool. **Appendix B Table 2** lists the domains and criteria applied for each study design.

Good-quality studies met nearly all design specific quality criteria indicative of good internal validity. These studies used valid randomization (for trials) or conducted appropriate statistical adjustments to create comparable study arms that were maintained throughout the study with minimal loss to follow up. Given the nature of the screening intervention, lack of allocation concealment was not considered an important risk of bias domain. Studies rated fair quality did not have serious threats to their internal validity related to design, execution, or reporting, but were found to be at risk of bias for some criteria. Studies were rated poor-quality if they had serious important limitations or a critical flaw that would likely affect the validity of study findings; these were excluded from this review. For nonrandomized studies, a rating of poor-quality often resulted from an assessment of there being a very high risk of bias due to: confounding based on imbalances in baseline characteristics (without proper statistical adjustment); a lack of 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 followup procedures based on intervention arm; high or differential rates of attrition between groups; or evidence of possible selective reporting.

One reviewer extracted key elements of included studies into standardized evidence tables in DistillerSR (Evidence Partners, Ottawa, Canada). A second reviewer checked the data for accuracy. Evidence tables were tailored to each key question, study design, and screening intervention. Tables generally included details on the study design and quality, setting and population (e.g., country, inclusion criteria, age, race/ethnicity, breast density, family history), screening features and protocol (e.g., modality, screening interval, reading procedure), and outcomes included for each key question.

Data Synthesis and Analysis

We created summary tables for all KQs to describe the study design, population, intervention characteristics, and outcomes in each included study. Individual study results were described for each KQ, and further grouped by comparison (i.e., modality, screening interval, age to start or stop screening), and outcome. When available, relative risks reported by study authors were provided in the tables, but we calculated and reported crude effect estimates and confidence intervals when studies provided only p-values, raw percentages, or other estimates of effects (e.g., odds ratios). To additionally facilitate comparisons among the studies, we converted effects to a common scale, as events per 100,000 individuals screened for mortality outcomes and per 1,000 individuals screened or examinations conducted for all others. We used summary tables and descriptive forest plots of the results to examine data for consistency, precision, and differences in effect sizes related to study and population characteristics.

Studies were considered for meta-analysis if they were sufficiently similar in terms of their study designs, populations, and reported outcomes. When very few studies were available (e.g., < 5), we presented pooled effects only if the results were relatively consistent (overlapping confidence intervals) and exhibited modest clinical diversity and statistical heterogeneity (<50%). We used a random effects model with restricted maximum likelihood estimation. A fixed effects model was also computed for sensitivity analyses exploring the potential influence of the statistical model used on results. We also recognize that estimates of statistical heterogeneity are limited when very few studies are available for pooling and therefore considered heterogeneity based on the estimated effects and study features (e.g., design, population, comparison and intervention protocols). For comparisons with few studies with clinically diversity or statistically heterogenous results, we did not generate pooled estimates using quantitative synthesis. We instead provided a narrative synthesis describing the findings separately for each Key Question. Our synthesis sought to identify the range of effects as well as sources of heterogeneity and possible explanations for similarities and differences in the findings across different studies based on their identified sources of potential bias, study settings, populations, and screening intervention protocols.

Grading the Strength of the Body of Evidence

We graded the strength of the overall body of evidence for each KQ using the Evidence-based Practice Center (EPC) approach, which is based on an adaptation of the system developed by the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group.[123] Four of five GRADE domains for assessing the strength of evidence are addressed in the Evidence-based Practice Center adaptation: consistency, precision, reporting bias, and study quality. The fifth domain, directness, is not addressed in the EPC approach since it is built into the structure of the analytic framework that underlies the key questions (i.e., link between the interventions and a health outcome).

Consistency (similarity of effect direction and size) was rated as reasonably consistent, inconsistent, or not applicable (e.g., single study). Precision (degree of certainty around an estimate) was rated as reasonably precise, imprecise, or not applicable (e.g., no evidence). Reporting bias was rated as suspected, undetected, or not applicable (e.g., when there was

insufficient evidence for a particular outcome) to address the potential for bias related to publication, selective outcome reporting, or selective analysis reporting. Study quality summarizes the degree to which the results are likely to have adequately low risk of bias (internal validity) based on individual study quality ratings. Additionally, the limitations domain highlights important constraints in answering the overall key question.

Overall strength of evidence assessments were defined as "High," "Moderate," "Low," or "Insufficient." A rating of "High" indicates high confidence that the included evidence reflects the true effect, and that further research is very unlikely to change our confidence in the estimate of effects. A "Moderate" strength of evidence ratings indicates moderate confidence that the evidence reflects the true effect and recognizes that further research may change confidence in the estimate of effect or may change the estimate itself. A "Low" strength of evidence rating indicates low confidence that the evidence reflects the true effect, and that further research is likely to change the rating and the estimate itself. A grade of "Insufficient" is used to indicate that evidence is either unavailable or does not permit estimation of an effect. The strength of evidence judgments were independently completed by at least two reviewers, with discrepancies resolved through consensus discussion involving more reviewers.

Contextual Questions

In addition to the systematically reviewed questions (KQs), we also addressed contextual questions (CQs) to aid with the broader interpretation of the evidence. Evidence for CQs was identified based on literature retrieved for the systematic search for KQs as well as targeted searches and scanning bibliographies of relevant articles. Contextual questions are not systematically reviewed. We used a 'best evidence' approach to identify the most recent, applicable, and robust evidence. Evidence related to the CQs is included throughout the Background and in dedicated sections of the Discussion to provide important context on breast cancer screening.

CQ1. How do racism, social inequalities, unequal access to high-quality healthcare, and other factors contribute to disparities in breast cancer incidence and outcomes? For example, what may account for higher breast cancer mortality among Black women in the United States?

CQ 2. How do new findings, analyses, or longer-term followup from foundational effectiveness trials of mammography screening influence conclusions about the benefits and harms of screening mammography?

CQ3. What risk assessment tools are available for use in average-risk screening populations and how well do they perform, particularly to support decisions about screening in younger women or women from racial/ethnic groups that are historically under-represented in research studies?

CQ4. How do the personal preferences of specific populations (including those that are under-represented in research) shape the ways in which they evaluate the potential harms and benefits of screening for breast cancer and decisions about whether to undergo screening?

CQ5. What are the harms of treatment associated with the detection of invasive breast cancer and ductal carcinoma in situ?

Expert Review and Public Comment

The draft Research Plan was posted for public comment on the U.S. Preventive Services Task Force (USPSTF) website from January 21, 2021, to February 18, 2021. Based on comments related to the scientific and conceptual scope of the review, the USPSTF revised the scope to require that effectiveness studies have data from at least two rounds of screening and include nonrandomized studies for the assessment of effectiveness. The USPSTF also clarified the proposed approach for including interval cancers and for defining advanced breast cancer. In addition, a proposed Contextual Question on breast density assessment was replaced with a question on treatment harms. A final Research Plan was posted on the USPSTF website on May 6, 2021.

USPSTF Involvement

The authors worked with USPSTF liaisons at key points throughout the review process to develop and refine the analytic framework and key questions and to resolve issues around scope for the final evidence synthesis. AHRQ staff provided oversight for the project, coordinated the systematic review, reviewed the draft report, and assisted in an external review of the draft evidence synthesis.

Chapter 3. Results

Literature Search

We reviewed 10,378 abstracts and assessed 419 full-text articles for inclusion (**Appendix B Figure 1**). Overall, we identified 19 studies (reported in 44 articles[79, 80, 124-165]) that met our inclusion criteria. List of included and excluded studies (with reasons for exclusion) are available in **Appendix C** and **Appendix D**. No RCTs were excluded from this review due to poor quality; however, 13 NRSI were excluded for poor quality, primarily due to confounding based on imbalances in baseline characteristics (without proper statistical adjustment), selection into study groups, and the absence of information on participant characteristic by study arm.

The numbers of studies and outcomes reported for each KQ are described in **Table 3**. Details on study design, population, and methodologies are provided in **Table 4**, **Table 5**, and **Appendix E Table 1**. The age to begin screening was not addressed in any of the included studies, the age to stop screening was addressed in one NRSI.[134] The effects of different screening intervals (i.e., annual, biennial, triennial) were addressed in five studies, 1 RCT[124] and 4 NRSI.[138, 151, 152, 157] Ten of the included studies, 4 RCTs[127, 137, 141, 158] and 6 NRSI,[79, 130, 138, 142, 145, 160] compared outcomes for screening with DBT versus DM. One RCT[162] and one NRSI[133] evaluated the effects of an invitation to supplemental screening with MRI for participants with dense breasts after receiving a negative screening mammography. One RCT[156] and one NRSI[150] addressed the use of supplemental ultrasound screening (US). No studies of interventions involving personalized screening based on predicted risk met the eligibility criteria for this review.

Health outcomes (KQ1) associated with different screening programs were reported in only two fair-quality NRSIs that addressed the age to stop screening[134] or screening interval[157] (**Table 4**). For invasive cancer detection (KQ2), two studies addressed the effect of screening frequency on the characteristics of detected cancers including one fair-quality RCT of multiple rounds of screening[124] and one fair quality cases-only analysis from the Breast Cancer Surveillance Consortium (BCSC).[152] Four studies of DBT compared with DM, 3 RCTs[127, 141, 158] [2 good- and 1 fair-quality] and one NRSI[142] reported screening outcomes from more than one round of screening and were included for KQ2. These studies reported characteristics of cancers detected at each round, necessary to assess whether screening resulted in stage shift toward less advanced cases with better prognosis. All 19 studies were included to examine potential harms of different screening approaches (KQ3).

Overall, the demographic characteristics of study participants were minimally described (**Table 5**). Most studies included participants in their 40s to 60s, with one study focusing on screening after age 70 years.[134] Only six of the 19 studies reported racial and/or ethnic characteristics. For five studies, participants were primarily White (73% to 92%), with <1 to 11 percent identified as Black, 2 to 11 percent as Asian, 5 to 7 percent as Hispanic.[130, 134, 145, 150, 152] An included electronic health record-based study included primarily Hispanic/Latina participants (76%) along with 10 percent Black and 10 percent White participants.[151]

KQ1. What Is the Comparative Effectiveness of Different Mammography-Based Screening Strategies on Breast Cancer Morbidity and Mortality?

Summary of Results

We did not identify any RCTs designed to test the comparative effectiveness of ages to start or stop screening, screening interval, or screening modality that reported morbidity, mortality, or quality of life outcomes. Two NRSIs reported mortality outcomes (breast cancer mortality, all-cause mortality) - one comparing mortality based on different ages to stop screening and another comparing annual to triennial screening intervals. One fair-quality observational study (N=264,274) was conducted in the United States using a random sample from Medicare claims data to estimate the effect of women stopping screening at age 70 compared with those that continued annual screening after age 70. Individuals included in the study had a high probability of living for 10 more years at the start of the study. The data were analyzed using statistical methods that have been developed to emulate per-protocol trials of screening. Continued screening between the ages of 70 and 74 was associated with a 22 percent decrease in the risk of breast cancer mortality compared with a cessation of screening after age 70. The difference in absolute rates was small (1 fewer death per 1000 women screened) and the confidence interval for the rate difference included null. The analysis found no difference in the hazard ratio or absolute rates of breast cancer mortality with continued versus discontinued screening from ages 75 to 84. The second NRSI was a fair-quality study (N=14,765) conducted in Finland during the years 1985 to 1995 that assigned participants ages 40 to 49 years of age to annual or triennial screening invitations using birth year. The study reported similar mortality from incident breast cancer and for all-cause mortality between the two study groups.

Detailed Results by Screening Intervention

Age to Start or Stop Screening

Study and Population Characteristics

One fair-quality NRSI study by García-Albéniz et al. used U.S. Medicare data from 1999-2008 and National Death index data to conduct an emulated trial evaluating the effect of stopping annual mammogram screening at the age of 70 versus continuing annual screening beyond this age (**Table 4**).[134] Annual screening was the most frequent pattern in the data for this time frame. An emulated trial uses statistical techniques to structure and adjust observational data in a way that can approximate a target (per-protocol) randomized trial.[166] The study was conducted using a 20 percent random sample of enrollees ages 70 to 84 years in Medicare parts A and B between 1999 and 2008 (Medicare Advantage enrollees, who comprised 13-21% of Medicare beneficiaries, were not included). Data on demographic characteristics, chronic conditions, preventive care, screening mammograms, breast cancer symptoms and signs, and breast cancer incidence and treatments were analyzed along with cause of death information obtained from the National Death Index from the National Center for Health Statistics.

A Medicare-specific comorbidity score was computed to exclude individuals that did not have a high probability of living an additional 10 years. Participants could not have a prior breast cancer diagnosis or have breast symptoms or a mammogram in the previous 9 months. The trial was emulated for two age groups, those ages 70 to 74 (n = 1,235,459) and those ages 75 to 84 (n = 1,403,735). At each year of age individuals were randomly assigned to the stop screening or continue screening strategy and the data were analyzed according to whether they had adhered to their assignment, resulting in 15 per protocol trial emulations (for each year of ages 70 to 84). Participants were followed until death, Medicare disenrollment, or the year 2008, whichever came first. A discrete hazard model was approximated using a pooled logistic regression model, and observations were cloned for analytic reasons and censored when they deviated from the randomly assigned screening strategy. Sensitivity analyses were used to evaluate the robustness of findings for a range of assumptions. The baseline characteristics for the sample were described for these two groups and showed that a majority of the sampled eligible participants in these Medicare plans were White (>90%), with 5 percent of participants reported as Black, and 3 percent as 'other' (no additional information provided) (**Table 5).** The older age group (75 to 84) had more frequent visits to the emergency room and more chronic conditions. These factors and other baseline characteristics were adjusted for in all analyses to account for possible differences that could affect assignment and adherence to the screening strategy (stop at age 70 or continue).

Outcomes

In the García-Albéniz NRSI,[134] 264,274 individuals contributed 758,127 person-years to the continued screening strategy with 118 breast cancer deaths, and 434,644 person-years to the stop screening strategy with 106 breast cancer deaths. Each woman was eligible for an average of 2.5 age-specific emulated trials. Therefore, after pooling all age groups, a total of 2,639,194 individuals contributed 4,656,465 person-years to the stop screening strategy and 7,170,142 person-years to the continue screening strategy. During followup in the emulated trials there were 1,533 breast cancer deaths under the continue screening strategy and 1,304 under the stop screening strategy.

For women ages 70 to 74 the estimated 8-year risk of breast cancer mortality with continued annual screening was 2.7 per 1,000 women (95% CI 1.8 to 3.7); it was 3.7 per 1,000 women (95% CI 2.7 to 5.0) with discontinuation after age 70 (RD -1.0 [95% CI -2.3 to 0.1]). Despite the small, statistically nonsignificant risk difference in mortality risk for the age group, the adjusted hazard ratio suggested a 22 percent lower hazard of 8-year breast cancer mortality with continued screening among those ages 70 to 74 [aHR 0.78 (95% CI 0.63 to 0.95)]. For women ages 75 to 84, the 8-year estimated risk of breast cancer mortality was 3.8 per 1,000 women (95% CI 2.7 to 5.1) with continued screening and 3.7 per 1000 women (95% CI 3.0 to 4.6) (RD 0.07 [95% CI -0.93 to 1.3]) with discontinuation, with an estimated hazard ratio of 1.00 (95% CI 0.83 to 1.19). These study results are the fully adjusted effect estimates that account for baseline demographics, chronic conditions, and health care use, as well as time-varying factors including screening history, use of health care resources, and comorbidities. Without adjustment for factors that would contribute to adherence to the continue or stop screening strategy, the risk differences are larger and more favorable for those that continued annual screening, especially in the 70 to 74 years age group. Overall, the adjusted findings did not show a statistical difference in the 8-year risk of breast mortality for women screened beyond age 75 compared with women who discontinued screening.

KQ1a. Does Comparative Effectiveness Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

The included study for KQ1 on age to stop screening did not present comparisons that tested or stratified mortality by participant characteristics or risk markers.

Screening Interval

Study and Population Characteristics

One fair-quality NRSI[157] study conducted in Turku, Finland from 1987 to 2007 compared rates of mortality associated with annual or triennial screening from ages 40 to 49 (**Table 4**). Over ten years (1985-1995) the study sent invitations to all female residents of the screening catchment area starting at age 40 as part of the national screening program (N=14,765). Study group assignment was determined based on birth year (even year birth annual, odd year birth triennial). All individuals invited to screening were followed for 10 years to assess incident cancer and an additional three years (to age 52) to assess mortality from breast cancers presenting from ages 40 to 49 as well as all-cause mortality. No data were reported on the demographics of participants, such as race, breast density, or presence of underlying risk factors (**Table 5**). Two-view, double read mammography was conducted by eight radiologists at a single screening center serving the city of Turku. The attendance rate for those invited to screening was 85 percent (not reported by study arm).

The intention to treat analysis was designed to test the effect of invitations to more or less frequent screening (2.8 versus 9.2 on average per person over the ten-year period). Data for the study outcomes was obtained through linkage with the Finnish Cancer Registry, the national Statistics Finland mortality registry, and the Turku clinical breast cancer database. All diagnoses and outcomes were cross-checked across the data sources, and medical chart review was conducted to resolve discrepancies. The analysis used person-years calculated from age 40 to 49 for breast cancer incidence outcomes and from ages 40 to 52 for mortality outcomes to compute rates per 100,000 person-years. During the study, breast cancer incidence between ages 40 and 49 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). Unadjusted Poisson regression was used to estimate the relative rate of incidence and mortality.

Outcomes

The 14,765 people invited to screening for this study contributed 100,738 person-years to the triennial screening invitation group and 88,780 person-years to the annual screening invitation group for estimation of mortality outcomes.[157] Mortality from incident breast cancer diagnoses occurring from ages 40 to 49 (with followup to age 52) was similar between groups, with 20.3 deaths per 100,000 person-years with annual screening invitations and 17.9 deaths per 100,000 person-years with triennial screening invitations (RR 1.14, 95% CI 0.59 to 1.27).

All-cause mortality (including mortality from prevalent and incident breast cancer diagnoses) was higher in the intention to treat analysis for invitation to annual screening (230.9 per 100,000 person-years) compared with invitation to triennial screening (192.6 per 100,000 person-years)

and there was a trend suggesting an estimated 20 percent increased risk due to the relative risk and a confidence interval on the margin of null (RR 1.20, 95% CI 0.99 to 1.46). An explanation or mechanism for the higher mortality rate related to more frequent screening could not be identified by the study authors. Deaths from other cancers and deaths from 'other natural causes' (not defined) were higher in the annual screening invitation group, whereas deaths from violent causes (accidents, intoxication, murder, suicide) were higher in the triennial invitation group.

KQ1a. Does Comparative Effectiveness Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

The included study for KQ1 on screening intervals did not present comparisons that tested or stratified mortality by participant characteristics or risk markers.

Digital Breast Tomosynthesis

No comparative studies reporting morbidity or mortality outcomes for screening with DBT compared with DM were identified.

Magnetic Resonance Imaging

No eligible comparative studies of MRI screening that reported mortality or morbidity health outcomes were identified.

Ultrasound

No eligible comparative studies of ultrasound screening that reported mortality or morbidity health outcomes were identified.

Personalized Screening Programs Using Risk Assessment

No eligible comparative studies of personalized screening that reported mortality or morbidity health outcomes were identified.

KQ2. What Is the Comparative Effectiveness of Different Mammography-Based Screening Strategies on the Incidence and Progression to Advanced Breast Cancer?

Summary of Results

There were no eligible comparative effectiveness studies of the age to start or stop screening that reported the outcome of cancer incidence and progression to advanced cancer across multiple screening rounds.

One older fair-quality RCT (n = 76,022) conducted between 1989 and 1996 randomized individuals to annual or triennial screening.[124] The number of screen-detected cancers was higher in the annual screening study arm (RR: 1.64 (95% CI, 1.28 to 2.09). The total number of cancers

diagnosed either clinically or with screening was similar after 3 years of screening (1 triennial incidence screen, 3 annual incidence screens). Cancers occurring in the annual screening group (including clinically diagnosed cancers) did not differ by prognostic features such as tumor size, node positivity status or histologic grade compared with those in the triennial screening group. The study did not report mortality outcomes, so it was not possible to ascertain whether the increase in the proportion of cancers detected by screening would influence health outcomes. Estimated effects based on prognostic indices did not predict statistically significant differences in mortality based on the tumor characteristics. Given the timing of the study, applicability is limited due to developments in screening technology, prognostics, and treatment effectiveness.

A fair-quality NRSI used BCSC data (N = 15,440) to compare the tumor characteristics of cancers detected following annual versus biennial screening intervals.[152] The reported tumor characteristics were presented in adjusted analyses stratified by age and menopausal status categories. The detection of stage IIB or higher cancers and cancers with less favorable characteristics did not differ by age when comparing annual to biennial screening intervals. For premenopausal individuals, however, a biennial interval preceding diagnosis was associated with having a higher stage tumor (IIB or higher) (RR: 1.28 [95% CI, 1.01 to 1.63], p=.04) and tumors with less favorable prognostic characteristics (RR: 1.11 [95% CI, 1.00 to 1.22], p=.047). For post-menopausal individuals with and without use of hormone therapy, there was no difference between cancers that were preceded by annual or biennial screening. The study did not conduct formal tests for interaction in the subgroup comparisons.

Results from three RCTs (N = 130,196)[127, 141, 158] and one NRSI (N = 92,404)[142] comparing DBT with DM screening reported invasive cancer detection and the characteristics of detected cancers from two rounds of screening (study participants were screened with a common modality at the second round). While cancer mortality results are not yet available from the trials, stage shift in the tumor characteristics across screening rounds could offer indirect evidence of potential screening benefit. Two RCTs[127, 158] and one NRSI[142] used DM for all participants at the second screening round and one RCT[141] used DBT for all participants at the second screening round. The three trials showed higher invasive cancer detection at the first round of screening in the DBT arm (pooled RR 1.41 [95% 1.20 to 1.64] I^2 7.6%, k=3, n = 129,492). Similar results were seen in the included NRSI, screening with DBT showed a higher rate of detection for invasive cancers using DBT compared with DM (RR: 1.52 [95% CI, 1.32 to 1.76]). At the second screening round (where all study participants were screened with a common modality), invasive cancer detection was similar for the group assigned to DBT at round one compared with the group assigned DM at round one in the three RCTs (pooled RR 0.87 [95% 0.73 to 1.05] I^2 0%, k=3, n = 105,244). In the NRSI when all participants were screened with DM at round two there were significantly fewer cancers detected among those originally screened with DBT at round one compared with those screened with DM at both rounds (RR: 0.71 (95% CI, 0.55 to 0.92). The three RCTs did not find a statistically significant difference in cancer stage (stage II or higher) at the second screening round. 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. Limited results stratified by age and breast density reported in the RETomo and To-Be RCTS did not suggest differences in invasive cancer

detection at a second round of screening for people who had been screened with DBT at the first screening round, but tests for interaction were not conducted and estimates were imprecise.

We did not identify any studies that reported data from more than a single screening round that could be used to compare shifts in cancer stage to assess the effectiveness of age to start or stop screening, the use of supplemental screening modalities, or personalized screening programs using risk assessment.

Detailed Results by Screening Strategy

Age to Start or Stop Screening

No eligible studies were identified that reported the cancer stage at detection across multiple screening rounds to provide evidence of a beneficial stage shift with screening when commenced earlier or continued to later ages.

Screening Interval

Study and Population Characteristics

Two studies addressed the effect of screening frequency on the characteristics of detected cancers. The fair-quality UKCCCR RCT was conducted as part of the UK National Breast Screening Program during the years 1989 to 1996 that randomized people ages 50 to 62 to annual (N= 37,530) or triennial (N=38,492) breast cancer screening (**Table 4**).[124] No characteristics other than participant age were reported (**Table 5**). The cumulative incidence of invasive cancer (including screen-detected and invasive cancers) was reported for all participants that attended a prevalence screening visit. The study was designed to compare the incidence of cancer in an annually screened group (3 screens after the prevalence screen) and in a triennially screened group (1 screen after the prevalence screen). The randomization scheme for the trial was conducted by month of birth for the first two years of the trial but thereafter used a computerized randomization scheme implemented through the national screening program. Cancer outcomes were obtained through searches of the pathology reports and databases maintained by hospitals involved in the UK National Breast Screening Program. Reports from pathologists on the prognostic factors for each cancer were obtained and reviewed by two consultants. Size, node status, and histological grade were used to code the cancers according to two different prognostic indices, the Nottingham Prognostic Index (NPI). Notably, the analysis of cancer prognostic characteristics grouped together screen-detected and interval cancers (39% of total cancer cases). The study also reported the tumor diameter, lymph node positivity, and histologic grade for all of the cancers diagnosed during the study, including interval and screen-detected cancers.

A fair-quality BCSC NRSI by Miglioretti et al. used data on cancers detected in the BCSC registries from 1996 to 2012 (**Table 4**).[152] The study compared the interval of screening relative to the characteristics of screen-detected and interval cancers. Individuals were included in the analysis if their cancer was preceded by at least two screening mammograms either 11 to 14 months apart (annual interval) or 23 to 26 months apart (biennial interval). The characteristics of women with cancers preceded by an annual screening interval (n = 12,070) and those preceded

by a biennial interval (n = 3,370) differed on some reported factors; those with an annual interval preceding a cancer diagnosis were less likely to be ages 40 to 49 (14% versus 18%) or 70-85 (29% vs 27%), and more likely to have a first-degree family history of breast cancer (23% versus 18%). The groups did not differ in race/ethnicity composition, and over three-quarters of the study population was White, non-Hispanic (78%), with the remaining participants reported as Black (5%), Asian (5%), Hispanic (5%), American Indian or Alaska Native (<1%), and 7% reported as "other" or unknown (**Table 5**). This study did not report overall effects of the screening interval on cancer detection by stage, but provided detailed results on the stage at detection stratified by age and menopausal status that are reported as KQ2a results below.

Outcomes

Screen-detected invasive cancer and prognostic characteristics by round. In the UKCCCR there were more invasive screen-detected cancers detected in the annual screening arm (4.42 per 1000 people screened, representing 71% of overall cancers) compared with the triennial screening arm (2.70 per 1000 people screened, representing 50% of overall cancers) (RR: 1.64, 95% CI: 1.28 to 2.09) (**Table 6**). After three years of screening (3 incidence screens in the annual arm, 1 incidence screen in the triennial arm) a similar number of cancers (screen-detected and interval cancers) had been diagnosed in the annual screening arm (6.26 per 1000 screened) and the triennial screening arm (5.40 per 1000 screened) (RR: 1.16, 95% CI 0.96 to 1.40). In comparisons of all of the cancers that occurred over the course of the study, including interval and screen-detected, there was no differences in tumor size, nodal status, histological grade, or the NPI prognostic index (**Table 6**). Mortality data from the study have not been reported, but based on estimates from the prognostic indices, the authors concluded that annual screening confers lead time bias (estimated to be ~6 months) but did not result in downstaging of screen-detected cancers that would influence breast cancer survival or risk of death.

KQ2a. Does Comparative Effectiveness Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

The Miglioretti BCSC NRSI reported the tumor characteristics and a prognostic characteristic variable for cancers diagnosed among individuals with an annual or biennial screening interval preceding their diagnosis, stratified by age (40 to 49, 50 to 59, 60 to 69, 70 to 85) and by menopausal status.[152] The adjusted analysis presented in the study compared those with a biennial versus an annual interval preceding their diagnosis. Screen-detected and interval cancers (12 months followup for annual screened, 24 months followup for biennial screened) were included in the comparisons (**Table 7**). The relative risk of being diagnosed with a stage IIB or higher cancer was not statistically different for biennially compared with annually screened women in any of the age categories. The composite variable indicating less favorable prognostic characteristics (stage IIB+, tumor size >15 mm, or node-positive) also was not statistically different for any age group comparing those biennially versus annually screened before their diagnosis. Analyses comparing stage (IIB or higher) and less favorable prognostic tumor characteristics stratified by menopausal status showed statistically significant effects of the screening interval. The risk of a stage IIB or higher diagnosis was higher for premenopausal women screened biennially compared with annually (RR 1.28, 95% CI 1.01 to 1.63; p=0.04). Similarly, a marginally significant increased risk of having a tumor with less favorable prognostic characteristics was seen for premenopausal women when they had been screened

biennially versus annually prior to their diagnosis (RR 1.11, 95% CI 1.00 to 1.22; p=0.047). For post-menopausal individuals (with and without hormone therapy use), tumor stage and prognosis were statistically similar when preceded by annual or biennial screening. The study did not conduct formal tests for interaction in the subgroup comparisons and did not adjust for multiple comparisons.

Digital Breast Tomosynthesis

Study and Population Characteristics

We included three RCTs[127, 141, 158] and one NRSI[142] that reported cancer detection across more than one round of screening and could therefore be used to assess invasive cancer detection and stage shift across screening rounds as an intermediate outcome to compare the effectiveness of DBT with DM screening (**Table 4**). Overall, population characteristics were sparsely reported for these trials besides participant mean age (**Table 5**).

The fair-quality Proteus Donna RCT conducted in Italy reported screening results from two rounds of screening with randomization to DBT/DM (n = 30,844) or DM (n = 43,022) for the first round of screening and DM screening for all participants at the second round of screening.[127] Participants in the national screening program from ages 46 to 49 were offered annual screening if they opted to participate in the screening program, and routine screening in the program was offered biennially for women ages 50 to 68. Recruitment began in December 2014 and was complete in December 2017. The mean age of participants was 57 years, information on breast density was not reported. Independent double reading was used with participants recalled based on the recommendation of either radiologist.

A good-quality RCT conducted in northern Italy reported on the characteristics of cancers detected at two consecutive rounds of screening. The Reggio Emilio Tomosynthesis study (RETomo) prospectively randomized women to undergo DBT/DM (n = 13,356) or DM (n = 13,521) at baseline followed by DM screening for all eligible participants one or two years later.[158] Women ages 45 to 49 were offered annual screening and those ages 50 to 69 were offered biennial screening. Follow-up is ongoing, and to date results have been reported over two rounds of screening with an additional 9 months of followup to obtain the final diagnosis for cancers detected at the second screening round. Participants were women ages 45 to 69 that had participated in the regional screening program but had never received a DBT examination. Just over one-third (38%) of participants were ages 45 to 49 at the first screening round in both study arms and the mean age was 55 (sd. 7). Breast density category distributions were similar with 9 percent of women classified as having very dense breasts. In both study arms two radiologists independently read the images and a third reader made the final judgment in cases of disagreement (usual screening program practice). Followup evaluations and final diagnosis results were obtained from screening program and cancer registry databases.

The To-Be study is a good-quality RCT conducted in Norway that randomized participants to DBT/sDM screening (n = 14,380) or DM screening (n = 14,369) and followed them for two years, or until the next screening episode.[141] The second screening round consisted of DBT/sDM for all participants. Therefore, outcomes at the second screening round with DBT/sDM were compared between those originally screened with DBT/sDM (n = 11,201) and those originally

screened with DM (n = 11,105). The study was conducted within the population-based BreastScreen Norway program, which offers all women ages 50 to 69 years of age biennial mammogram screening. The mean age of study participants was 60 years with 7 percent of women classified as having very dense breasts. In this program, independent dual reading with consensus is standard and prior mammograms, if available, are used to assist image reading. The first round of screening was conducted Jan 2016 through Dec 2017 and the second screening 2018 through Jan 2020.

A fair-quality NRSI using a geographical comparison cohort design was conducted within the BreastScreen national screening program in Norway. The Oslo-Vestfold-Vestre Viken (OVVV) cohort was used to compare cancer screening outcomes from one round of screening with DBT/sDM (n = 37,185) or DM (n = 61,742) and a second round of DM for all attending the consecutive round of screening (n = 72,017).[142] Individuals screened in Oslo received DBT at the baseline screening round and DM in the consecutive round, and in Vestfold and Vestre Viken DM screening was provided at both rounds. Those ages 50 to 69 years presenting to be screened in Oslo, Vestfold and Vestre Viken from February 2014 to December 2015 were included in the cohort. In this program, biennial screening is provided, so the second screening visit (for those not diagnosed at baseline) occurred two years later. Those participating in BreastScreen were assigned to the baseline screening modality based solely on their county of residence and were not given an option to select the screening type. The mean age of study participants was 59, data on breast density was not reported. In the BreastScreen Norway program, independent double reading of mammography images with random pairs of breast radiologists are used to determine the mammography result.

Outcomes

Screen-detected invasive cancer and prognostic characteristics by screening round. Three trials randomized participants to DBT or DM at a first round of screening, followed by a second round of screening with either DM for everyone (Proteus Donna, RETomo) or DBT for everyone (To-Be). One NRSI using a geographic comparison study design compared people receiving DBT/sDM or DM at a first screening round and DM for everyone in the second round (OVVV) (**Table 8**).

The three RCTs reported increased detection of invasive cancer with DBT at the first round of screening (pooled RR 1.41 [95% CI 1.20 to 1.64] I^2 7.6%, k=3, n = 129,492) and effects in the opposite direction, but not statistically different at second round screening (pooled RR 0.87 [95% 0.73 to 1.05] I^2 0%, k=3, n = 105,064) (**Figure 3**).[127, 141, 158] Information on the characteristics of cancers detected at each screening round can help with indirect inferences about whether the additional or earlier cancer detection at the first round of screening would affect health outcomes. Two RCTs conducted in Italy reported detection of cancers stage II or higher and the same variable was obtained via author communication from the To-Be study. There was no difference within any of the studies in the detection of Stage II or higher cancers at either round of screening, and results were inconsistent at round two with one trial nearing statistical significance for more stage II cancers and the other two trials in the direction of reduced stage II cancer in the DBT arm.

In the Proteus Donna trial, the DBT study arm detected more invasive cancers during the first round of screening (7.3 versus 5.0 per 1,000 screened; RR 1.46 [95% CI, 1.21 to 1.77]) and at the second round the detection of invasive cancers was not statistically different between arms (RR: 0.85 [95% CI, 0.64 to 1.13]).[127] For the RETomo trial detection of invasive cancer was higher for the DBT/DM study arm at round one (RR 1.60, 95% CI 1.16 to 2.22) with a rate of 6.3 versus 3.9 per 1,000 screened.[158] Detection at second round screening (all DM) did not differ by study arm (RR: 0.90 [95% CI, 0.62 to 1.30]) (**Figure 3**). There were no statistical differences in the characteristics of screen-detected cancers at either screening round, including cancers detected at stage II or higher, tumor size, histologic grade, or node status (**Table 8, Figures 4-7**).

The To-Be RCT randomized people to DBT/sDM or DM in the first round of screening followed by DBT/sDM for all at the second round of screening.[141] There was not a statistically significant difference between study arms in the detection of invasive cancer at the first round of screening using DBT or DM (5.6 versus 4.9 per 1,000 screened, respectively; RR 1.13, 95% CI 0.82 to 1.55) or the subsequent screening round using DBT for all participants (6.9 versus 7.8 per 1,000 screened, respectively; RR 0.88, 95% CI 0.65 to 1.19). No statistical differences in tumor stage, tumor size, histologic grade, or nodal status were seen for cancers detected in the DBT/sDM arm compared with DM arm (**Table 8, Figures 4-7**).

The OVVV NRSI reported on a single round of screening with DBT/sDM followed by DM at the subsequent round of screening 2 years later compared with a concurrently screened group from another region that were screened with DM at both rounds. More invasive cancers were detected at the first round of screening for those in the DBT/sDM screened region (7.6 versus 5.3 per 1,000 screened; RR 1.43, 95% CI 1.22 to 1.67). During the second round of screening, where all received DM, the incidence of screen-detected invasive cancer was lower in the arm that received DBT/sDM at the first round (3.2 versus 4.5 per 1,000 screened; RR 0.71, 95% CI 0.55, 0.92) compared with those who received DM at both screens. The study did not report cancer stage, but reported on characteristics of the screen-detected invasive cancers. No statistical differences were identified between cancers detected by either arm including tumor diameter, histologic grade, and node status (**Table 8**).

KQ2a. Does Comparative Effectiveness Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

None of the included studies were designed to enroll populations to support comparisons in the screening outcomes of DBT and DM by race ethnicity or family history.

Age and breast density stratified analysis of cancers detected at the second round of screening were reported in the RETomo RCT. As in the overall population, 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 (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. For women aged 45-49 and women with dense breasts there was no statistical difference in the detection of invasive cancers at either round of screening (**Table 9**). No test for interaction was conducted for either the age or density stratified analyses and no information on the characteristics of the screen-detected tumors was provided.

Density stratified results were presented in the To-Be RCT. No statistical difference was seen for detection of invasive cancer using DBT or DM for any breast density subgroup at both round 1 and 2 of screening (**Table 9**).

Magnetic Resonance Imaging

No eligible studies were identified that reported on cancer detection and characteristics over multiple rounds of screening comparing usual care mammography with mammography plus supplemental MRI screening.

Ultrasound

No eligible studies were identified that reported on cancer detection and characteristics over multiple rounds of screening comparing usual care mammography with mammography plus supplemental ultrasound screening.

Personalized Screening Programs Using Risk Assessment

No eligible studies were identified that reported on cancer detection and characteristics over multiple rounds of screening comparing usual care mammography personalized screening programs using risk assessment.

KQ3. What Are the Comparative Harms of Different Breast Mammography-Based Cancer Screening Strategies?

Summary of Results

One NRSI with an emulated trial design used Medicare data to estimate the effects of screening beyond age 70 compared to stopping at ages 70 or 75.[134] No difference was found in 8-year breast cancer mortality for screening beyond age 75 compared with stopping at that age. Cancers diagnosed in the stop screening strategy were more likely to receive aggressive treatment.

One RCT[124] and four NRSI[138, 151, 152, 157] reported potential harms of screening with respect to the screening interval. One RCT reported approximately one fewer interval cancer per 1,000 with annual screening compared with triennial screening. Data related to interval cancer risks were limited in the four NRSI for comparisons of different screening periods.[138, 151, 152, 157] False-positive recall was more likely to occur with annual screening compared with longer intervals between screenings. The probability of false positive recall and biopsy over ten years of screening was higher with annual screening. The highest cumulative false positive estimates occurred among young people with dense breast screened annually.

Three large RCTs found no statistically significant difference in the rates of interval cancers following screening with DBT compared with DM (pooled RR 0.87, 95% CI 0.64 to 1.17, k = 3, n = 130,196, I^2 0%) (**Figure 10**).[127, 141, 158] Data on interval cancers from 5 NRSI were mixed, and interpretation was limited by differences in study design. The effects of DBT screening on recall, false-positive recalls, and biopsy rates varied between trials and by screening round, with

no or small statistical differences between study groups, not consistently favoring DBT or DM. The cumulative rates of false-positive recall and false-positive biopsy were slightly lower with DBT compared with DM screening, regardless of screening interval (cumulative probability over 10 years: 50% vs 56% for annual screening, 36% vs 38% with biennial screening). No statistically significant differences were seen in the trials related to DCIS detection or adverse events. Rates of radiation were approximately two times higher when DBT was performed in addition to DM; however, these increases were not present in two studies using DBT to generate synthetic DM images (DBT/sDM). Data on subgroups were limited with all but one of the studies providing stratified results only, without tests for interaction.

One RCT reported on the effects of an invitation to screening MRI for women ages 50 to 75 with extremely dense breasts following a negative mammogram.[162] The risk of invasive interval cancer was reduced by approximately half (RR 0.47, 95% CI 0.29 to 0.77) after the first invitation and prior to the next screening round (2 years). MRI resulted in additional recall, false-positive recall and biopsy (95, 80, and 63 per 1,000 screened, respectively) that did not occur for the DM only group. An NRSI analysis of US insurance claims data found that health care use related to conditions that were not breast-related (a measure of possible incidental findings) was higher following screening with MRI compared with receiving mammography screening only.[133]

One RCT of women aged 40 to 49[156] and one NRSI of BCSC data[150] reported outcomes related to the potential harms of supplemental ultrasound screening. In the analyses comparing event rates presented in our review there was not a statistically significant difference in interval cancer rates between study groups in either study. In the trial, additional recalls (48 per 1000 screened) were experienced by those screened with ultrasound. In the BCSC analysis, referral to biopsy and false positive biopsy results were twice as high for the group screened with ultrasound.

No eligible studies were identified that reported on the potential harms of personalized screening programs using risk assessment.

Detailed Results by Screening Intervention

Screening Age to Start or Stop

Study and Population Characteristics

One fair-quality NRSI (n = 1,058,013) analyzed data to emulate a trial of discontinuation of mammography screening at age 70 compared with continued annual screening beyond this age (described in detail KQ1 above) (**Table 4**).[134] Additional details on study design are available in **Appendix E**.

Outcomes

Overdiagnosis and overtreatment. Overall, the 8-year cumulative risk of a breast cancer diagnosis was higher for the continued annual screening strategy after age 70 (5.5% overall; 5.3% ages 70-74, 5.8% ages 75-84) compared with the stop screening strategy (3.9% overall; same proportion for both age groups) (**Table 10**). Lumpectomy and radiotherapy were more common for cancers diagnosed in the continued annual screening strategy compared with those

that stopped screening after age 70, whereas mastectomy and chemotherapy were more common for cancers diagnosed in those that discontinued screening after age 70 (**Table 10**). Overall, because fewer cancers were diagnosed under the stop screening strategy (ages 70 to 84), there was a lower risk of undergoing follow-up and treatment. For those ages 75 to 84, additional diagnoses did not contribute to a difference in the risk of breast cancer mortality.

KQ3a. Do Comparative Harms Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

No studies of ages to start or stop screening presented data that would allow for testing of effect differences or stratification of results by different population characteristics or risk markers.

Screening Interval

Study and Population Characteristics

Three of the studies included to address potential harms of different screening intervals have been described previously in Key Questions 1 and 2.[124, 152, 157] Two additional studies examined the potential cumulative harms across multiple rounds of screening (**Table 4**). One analysis of 2005 to 2018 data from the BCSC estimates the cumulative probability of a false-positive result after 10 years of screening with DM or DBT.[138] The second additional study was the Know Your Risk: Assessment at Screening (KYRAS) study that calculated the cumulative risk of false positive screens over a median of 8.9 years at Columbia University Medical Center.[151]

Demographic characteristics were not commonly reported in the studies of screening interval (**Table 5**). The BCSC study population reported by Miglioretti was primarily White (78%) with the remaining participants reported as Black (5%), Asian (5%), Hispanic (5%), American Indian or Alaska Native (<1%), and 7% reported as "other" or unknown. In the KYRAS study the population was majority Hispanic (76%) with the remaining reported as White (10%), Black (10%), or other (4%) including Asian, Pacific Islander, Native American or Alaskan Native. Twenty-four percent of the non-Hispanic White women were of Ashkenazi Jewish descent. Additional details on study design available in **Appendix G**.

Outcomes

Interval cancers. Three studies presented data on interval cancers by participant screening interval with mixed findings (**Table 11**). The UKCCCR RCT reported the rate of interval cancers was significantly lower in the annual invitation group (1.84 per 1,000 women initially screened) than in the triennial invitation group (2.70 per 1,000 women initially screened) (RR 0.68, 95% CI 0.50 to 0.92). The Parvinen et al. quasi-randomized study found similar numbers of cases were reported in the annual screening and triennial screening groups and a statistical test for the difference was null (p=0.22). The Miglioretti et al. BCSC NRSI found that 22.2 percent of cancers diagnosed following an annual screening interval were interval cancers compared with 27.2 percent of cancers proceeded by a biennial interval. However, the study did not provide adjusted comparisons, limiting the ability to draw inferences about differences in the interval cancer rate associated with biennial and annual screening from this study.

False-positive recall. Based on two studies, false-positive recall was more likely to occur with annual screening compared with longer intervals. A NRSI of BCSC data by Ho estimated the 10-year cumulative probability of at least one false positive recall was 49.6 percent for those screened annually and 35.7 percent for those screened biennially (proportion difference: -13.9%, 95% CI -14.9% to -12.8%). The difference in cumulative FPs recall between annual and biennial screening were larger for DM (-18.2, 95% CI –18.6 to –17.7) (**Figure 8, Appendix F Tables 3, 4**). In the KYRAS study individuals screened with DM annually had 2.18 times the odds of having a false positive result compared with those who screened biennially (OR, 2.18; 95% CI, 1.70 to 2.80) after controlling for total years of follow up, age, race/ethnicity, BMI, breast density, and breast cancer risk status (**Appendix F Table 4**).

False-positive biopsy. The comparative NRSI from Ho used data from the BCSC found biennial screening compared with annual screening led to a 5 percent lower 10-year cumulative false positive biopsy rate whether the screening was conducted with DBT or DM (**Figure 9, Appendix F Tables 3, 4**). For individuals screened with DBT the estimated cumulative probability of at least one false-positive biopsy recommendation was 11.2% for those screened annually and 6.6% for those screened biennially (proportion difference: -4.6%, 95% CI -5.2 % to -3.9%). For individuals screened with DM the difference was similar (proportion difference: -5.0%, 95% CI -5.4% to -4.7%).

KQ3a. Do Comparative Harms Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

The Ho et al. BCSC NRSI reported 10-year cumulative false-positive and biopsy rates by age and breast density category. Annual screening was associated with higher cumulative FP recall and biopsy for most age and density groups (**Figures 8, 9**). There was not a strong association between and age and cumulative false positive biopsy regardless of the screening interval among those with the lowest breast density (**Figure 8, 9**, **Appendix F Tables 5, 6**).

Digital Breast Tomosynthesis

Study and Population Characteristics

We identified 10 eligible studies, 4 RCTs (3 good-quality, 1 fair-quality)[127, 137, 141, 158] and 6 fair-quality NRSIs,[79, 130, 138, 142, 145, 160] that reported on potential harms of screening associated with the use of DBT (plus DM or sDM) compared to DM only screening (**Tables 4 and 5**). Four large trials were conducted with individuals participating in organized screening programs in Germany, Italy, and Norway. Three of these trials were previously discussed in KQ2.[127, 141, 158] One additional RCT was identified that addresses the potential harms of screening with DBT compared with DM. The TOmosynthesis plus SYnthesised MAmmography Study (TOSYMA) is a good-quality RCT conducted in Germany that assigned 99,634 women ages 50 to 69 to DBT/sDM (DBT with synthetic two-view imaging) versus DM alone between July 5, 2018 and December 30, 2020. Available results from the trial report on performance at a single round of screening and for this review was included only for rare or uncommonly reported harms (adverse events, radiation exposure). The six NRSI included for KQ3 were conducted using data from populations screened with DBT and DM in the US,[130, 138, 145, 160] Sweden[79] and Norway[142] (**Tables 4 and 5**). Additional details on study design and results are available in **Appendix G.**

Outcomes

Interval cancers. Three trials reported interval cancers following screening with DBT or DM (**Table 12**).[127, 141, 158] The three RCTs did not show statistically significant differences in the risk of interval cancer following screening with DBT or DM (pooled RR 0.87 [95% CI 0.64 to 1.17], I^2 0%, k = 3, n = 130,196) (**Figure 10**). Five observational studies used data from medical systems, registries, and cancer screening and surveillance programs to compare interval cancers occurring after screening with DBT or DM (**Table 12**). Three of the NRSI found no significant difference in the rate of interval cancers diagnosed following screening with DBT or DM (including data from the BCSC, PROSPR consortium, and the OVVV comparative cohort study)[130, 142, 145] while one found a slight increased risk with DBT screening[160] and one an unadjusted decreased risk with DBT screening.[79] These studies differed in the timeline of follow up and method of identifying interval cancers (**Appendix E Table 1**) highlighting the variability in interval cancer definitions and data used to assess the outcome across the included NRSI, and the need for more standardization of definitions and study protocols.

Recall. The same three RCTs and one NRSI included for KQ2 reporting data across multiple rounds of screening were also included to assess screening recall rates and false positive recalls (**Tables 13 and 14**). Results regarding recall rates and false positive recalls were mixed across the first round of screening and inconsistency at round one resulted in high statistical heterogeneity so a pooled effect is not presented. The studies varied in their approaches to screening at round two: two RCTs used DM screening for both study groups (Proteus Donna, RETomo) and one used DBT for both study groups (To-Be) at round two. Results for round two were more consistent and did not suggest a difference in recall rates or false-positive recalls between study groups when combined using meta-analysis (**Figures 11 and 12**). The included NRSI OVVV study that used a concurrent regional comparison did not report statistically significant difference in recall rates between the DBT and DM arms at round one (**Tables 13 and 14**). At round two, when both groups received DM, and false positive recall rates were lower in the group previously screened with DBT compared with the DM group (20 versus 25 per 1,000).

Biopsy and surgical followup. Two of the included RCTs reported on the rate of referral to biopsy[141, 158] and two reported on referral to surgery following screening[127, 158] (**Table 13**). At round one when the trials compared screening with DBT and DM there were mixed results with one trial finding a significantly higher rate of referral to biopsy with DBT and another trial finding no difference in referral to biopsy or false positive biopsy rates. Two trials found that the referral to surgery was higher among those screened with DBT. The RETomo RCT reported higher referrals to surgical followup, including open biopsy following DBT screening (8.7 versus 5.0 per 1,000, RR 1.70, 95% CI 1.3 to 2.30). Findings from the Proteus Donna RCT were similarly higher for surgical referrals following DBT/DM (9.9 versus 6.4 per 1,000, RR 1.54, 95% CI 1.31 to 1.82).

The trials screened both study groups with an identical modality at the second round, and effects should be interpreted as findings from screening following previous round of screening with DBT or DM. Overall, no significant difference between arms was found for rates of biopsy at round two. The Proteus Donna trial[127] found a lower risk of surgical referrals among those originally screened with DBT (4.3 versus 5.7 per 1,000, RR 0.76, 95% CI 0.59 to 0.97);

however, this finding was not confirmed by RETomo where screening was with DBT in both study groups at round two (5.3 versus 6.4 per 1,000, RR 0.83, 95% CI 0.60 to 1.10).

Cumulative false-positive recall and biopsy. The comparative BCSC NRSI from Ho et al. reported the estimated cumulative probability of having at least one false-positive recall and biopsy over 10 years of screening with DBT or DM on an annual or biennial basis (**Figures 8, 9, Appendix F Tables 3, 4**). Probabilities were mostly lower with DBT screening compared with DM screening, regardless of the screening interval, but the difference was greater with annual screening. With annual screening, the 10-year cumulative probability of a false-positive recall was 49.6% with DBT and 56.3% with DM (Difference -6.7%, 95% CI -7.4 to -6.1). The 10-year cumulative probability of a false-positive biopsy was 11.2% with DBT and 11.7% with DM (Difference -0.5, 95% CI -1.0 to -0.1). With biennial screening, the 10-year cumulative probability was 35.7% for DBT and 38.1% for DM (Diff -2.4%, 95% CI -3.4 to -1.5) and the 10-year cumulative probability of a false-positive biopsy was 6.6% for DBT and 6.7% for DM (Difference –0.1%, 95% CI –0.5 to 0.4).

Overdetection and overtreatment. In the three RCTs rates of DCIS detected at each screening round and between study arms were similar, ranging from 0.7 to 1.3 per 1,000 screened at the first screening round and from 0.6 to 1.3 per 1,000 screened at the second screening round, with no statistical differences between the DBT and DM screened groups (**Table 15**). Meta-analysis was used to generate combined estimates that also did not show statistically significant differences 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)**. The OVVV NRSI reported higher DCIS detection at the first screening round in the DBT group compared with the DM group (1.8 versus 0.8 per 1,000 screened; RR 2.16, 95% CI 1.49 to 3.12).

Adverse events. The TOSYMA RCT reported on adverse events from a single round of screening using DBT/sDM compared with DM only. The study randomized 49,804 individuals to DBT/sDM and 49,830 to DM. Six adverse events were reported in each study arm with none categorized as serious.

Radiation exposure. Five studies (4 RCTs, 1 NRSI) reported the median, mean, or relative radiation dose by study arms from a single screening round (**Table 16**). In three of these studies participants underwent a DBT and DM screening (in one or two compressions) and in two studies participants underwent DBT with a synthetic reconstruction of a 2D DM image[137, 141]. Studies using DBT/DM screening reported radiation exposure approximately two times higher in the intervention group compared with the DM only control group.[79, 127, 158] Differences between study groups in radiation exposure were smaller in studies using DBT/sDM. The TOSYMA RCT reported median glandular radiation dose in the DBT/sDM group was 1.86 mGy (IQR 1.48 to 2.45) and in the DM group was 1.36 mGy (IQR 1.02 to 1.85). In the To-Be RCT which also used DBT/sDM, the mean radiation dose was 2.96 mGy compared with 2.95 mGy in the DM group.[141]

KQ3a. Do Comparative Harms Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

None of the included studies were designed to enroll populations to support comparisons in the screening outcomes of DBT and DM by race, ethnicity, or family history. Two RCTs[141, 158] and four NRSI[79, 138, 145, 160] that compared DBT-based screening strategies with DM only screening strategies presented results stratified by age and/or breast density. Most study did not report interaction tests and were not designed to test these subgroup comparisons making it difficult to draw conclusions about differences by age and breast density.

Age. The RETomo RCT reported the effects of DBT/sDM versus DM on recall, biopsy, and surgical procedures stratified by age category (45-49 versus 50-69) (**Tables 17 and 18**). Overall, these stratified results suggest some risk of increased biopsy or surgery with DBT screening at the first round for all, followed by lower rates at the next round for those ages 45 to 49. One trial[158] and two NRSI[79, 160] reported no significant findings related to the relationship between age and interval cancer outcomes (**Table 19**). Two of these studies study did not report interaction tests making it difficult to draw conclusions about differences by age group.

Breast density. The To-Be trial reported recall and biopsy stratified by Volpara density grade categories (VDG1-VDG4). There was lower recall at the first screening round for those screened with DBT that had lower density breasts (VDG1 and VDG2) but not for those with higher density breasts (VDG3 and VDG4) (**Table 17**). Two trials[141, 158] and one analysis of BCSC data[145] found no statistically significant differences in the incidence of interval cancer for the breast density stratified comparisons (**Table 19**).

The To-Be RCT reported mean radiation doses for the study groups, stratified by breast density in a figure. The study reported that there were not statistically significant differences in radiation dose for DBT/sDM compared with DM for any of the density categories.

Age and breast density subgroups. The Ho et al. BCSC NRSI presented 10-year cumulative false positive recall and biopsy probabilities stratified by breast density and age, comparing DBT to DM screening. Overall, the study reported lower false positive recall with DBT screening. In stratified analyses, however, there was not a statistical difference in cumulative false positive recall or biopsy among those with extremely dense breasts in any age group (**Figure 13**).

Magnetic Resonance Imaging

Study and Population Characteristics

The Dense Tissue and Early Breast Neoplasm Screening (DENSE) trial is a good-quality RCT conducted in the Netherlands that enrolled participants from December 2011 to November 2015 (N = 40,373) (**Table 4**). The aim of the study was to determine whether an invitation to supplemental MRI screening after a negative mammogram for those age 50 to 75 with extremely dense breast tissue would reduce the incidence of interval cancer. The baseline characteristics of the study groups were balanced on the reported characteristics (**Table 5**). Among those invited to MRI screening 59 percent underwent the MRI examination (n = 4,783). While this study included two rounds of screening with MRI, findings from the second round of screening in the

mammography only arm have not been published. Therefore, this study was not eligible for inclusion in KQ2, but it is included for interval cancers and potential harms of supplemental MRI imaging.

A fair-quality NRSI compared commercially insured women ages 40 to 64 years of age in the MarketScan database who had received at least one bilateral screening breast MRI (n = 9208) or mammogram (n = 9,208) between January 2017 and June 2018 (**Tables 4 and 5**). Propensity score matching was used to compare cascade events (mammary and extramammary) in the 6 months following the MRI or mammogram that were potentially attributable to having a breast MRI. Additional details on study design and results are available in **Appendix G.**

Outcomes

Interval cancers. In the DENSE RCT the ITT analysis based on invitation to MRI screening found a rate of invasive interval cancers for the DM+MRI of 2.2 per 1,000 invited to screening compared with 4.7 per 1,000 screened for the DM only control group (RR 0.47, 95% CI 0.29 to 0.77) (**Table 12**).

Adverse events. In the DENSE RCT, 8 adverse events (including 5 classified as serious adverse events) occurred during or immediately after the MRI screening. Adverse events included 2 vasovagal reactions and 3 allergic reactions to the contrast agent (serious adverse events) as well as 2 reports of extravasation (leaking) of the contrast agents and 1 shoulder subluxation. Twenty-seven individuals (0.6% of MRI arm) reported a serious adverse event within 30 days of the MRI.

Downstream consequences of supplemental imaging including incidental findings. In the first round of the DENSE trial the rate of recall among those who underwent additional imaging with MRI was 94.9 per 1000 screens and the false positive rate was 79.8 per 1,000 screened. The rate of biopsy for those undergoing supplemental MRI was 62.7 per 1000 screened (**Table 20**). Among the cancers diagnosed by MRI over 90 percent were classified as DCIS (stage 0) or stage 1 cancer. Without information for two rounds of screening from both arms of the study there is not sufficient information to weigh the relative benefit versus harms of these diagnoses and downstream imaging consequences.

In the US insurance claims NRSI, individuals that had an MRI compared to those receiving only a mammogram were more likely in the subsequent 6 months to have additional cascade events (adjusted difference between groups 19.6 per 100 screened, 95% CI 8.6 to 30.7) and were mostly comprised of additional health care visits. (**Table 20**).

KQ3a. Do Comparative Harms Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

No studies of supplemental MRI screening presented data that would allow for testing of effect differences or stratification of results by different population characteristics or risk markers.

Ultrasound

Study and Population Characteristics

The Japan Strategic Anti-cancer Randomized Trial (J-START) is a fair-quality RCT that randomly assigned asymptomatic women ages 40 to 49 years of age in 23 prefectures in Japan to breast cancer screening with mammography plus handheld ultrasound (DM/US) (n = 36,859) or mammography only (DM) (n = 36,139) over two rounds of annual screening during 2007 to 2011 (**Table 4**).[156] The two study groups were balanced across a range of characteristics (**Table 5**). The authors note that 58 percent of women were classified as having dense breasts. Only one round of screening has been reported, therefore, this study was not eligible for inclusion in KQ2, but it is discussed here for interval cancers and potential harms related to supplemental ultrasound imaging.

An NRSI by Lee et al. reported results of an analysis using data from two BCSC registries to compare screening outcomes for individuals receiving ultrasonography on the same day as a screening mammogram (DM/US) (n = 3,386, contributing 6081 screens) compared with those that received only a mammogram (DM) (n = 15,176, contributing 30,062 screens) (**Tables 4 and 5**, see **Appendix E** for detailed methods).[150] The majority of individuals included in the study were White (accounting for 80% of the screening examinations) and represent a higher risk population with a significant proportion of exams among those with a first-degree family history of breast cancer or previous breast biopsy. Additional details on study design and results are available in **Appendix G**.

Outcomes

Interval cancers. The interval cancer rates reported were not statistically significantly different in the J-START RCT when comparing the DM with ultrasound versus DM only groups (RR: 0.58, 95% CI 0.31 to 1.08). The published results from the trial were population-average effects that included DCIS and statistical adjustments for the clustered data structure. The result presented is a calculated individual-level intervention effect for invasive interval cancer without adjustment for clustering based on the reported event rates. Adjustment for clustering would result in a greater imprecision since it would statistically compensate for the correlated variances with wider confidence intervals. In the NRSI using BCSC data, the confidence interval was wide and not statistically significant (aRR 0.67, 95% CI 0.33 to 1.37) (**Table 12**).

Downstream consequences of supplemental imaging. The rate of recall based only on ultrasound was 49.7 per 1,000 in the ultrasound arm and 48.0 per 1,000 had a false positive recall (**Table 20**). Of those cancers identified only by ultrasound 76.2 percent were classified as stage 0 or 1 cancer. Without information for two rounds of screening from both arms of the study there is not sufficient information to weigh the relative benefit versus harms of these diagnoses and downstream imaging consequences. In the BCSC analysis., the rates of referral to biopsy and false positive biopsy recommendations were twice as high and short interval followup three times as high for the group screened with ultrasound (**Table 20**).

KQ3a. Do Comparative Harms Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

A secondary analysis of J-START reported results for trial participants from a single screening center in one Japanese prefecture (Miyagi) to compare interval cancer rates for DM/US and DM screening among women ages 40 to 49.[136] Analyses stratified by breast density did not show a statistically significant difference in interval cancer rates for any density category. (**Table 19**). The rates of recall based only on ultrasound were 69.7 per 1000 (95% CI 63.3 to 76.6) among those with dense breast and 39.4 per 1000 (95% CI 33.5 to 46.0) among those without dense breasts.

Personalized Screening Programs Using Risk Assessment

No eligible studies were identified that reported on the potential harms of screening comparing usual care mammography personalized screening programs using risk assessment.

Chapter 4. Discussion

Overall Summary of Evidence

We conducted this review to inform the USPSTF update to their recommendation on breast cancer screening. The 2016 review updated the evidence on screening effectiveness and provided emerging evidence on comparative effectiveness questions related to DBT and supplemental screening modalities.[76, 95] The evidence included in this review includes comparative effectiveness studies only because the evidence on mammography screening effectiveness has been reviewed and updated numerous times over the past two decades as large trials of mammography screening were completed. Based on the trials' findings of a mortality benefit for women ages 50 to 69 (**Appendix A**), new trials comparing screening versus no screening are unlikely except in groups where there is equipoise or unclear evidence of a benefit.

The results of this review are summarized in **Table 21**, with different comparisons separately considered within each key question. We included 19 studies that met the review eligibility criteria (2 included in the previous review[124, 157]) and compared active screening interventions against comparisons that differed by the timing, frequency, or modality of screening. Eligible studies using more recent registry data were included when available rather than earlier studies from the previous review.

While breast cancer screening is an active area of research, few longitudinal trials of screening have been conducted since the original effectiveness trials were completed. We included six new randomized trials in the review,[127, 137, 141, 156, 158, 162] including four comparing DBT with DM screening[127, 137, 141, 158] and two on supplemental screening compared with mammography only.[156, 162] Three of these trials are ongoing[137, 156, 162] and have only reported preliminary results, and three are completed.[127, 141, 158] Nonrandomized observational studies were also included; however, few followed a screening population over time to compare different screening approaches. Risk of bias due to confounding and selection in nonrandomized, nonexperimental studies limits the confidence in their findings.

For KQ1, two studies compared mortality outcomes for different screening strategies: one nonrandomized study of the age to stop screening[134] and one older RCT comparing annual with triennial screening.[157] For KQ2, six studies were included that reported invasive cancer detection outcomes from more than a single round of screening. Breast cancer outcomes must be assessed over a minimum of two rounds of screening to determine whether a screening approach leads to a shift toward detection at an earlier cancer stage. Studies from a single round of screening are subject to lead time bias. Two studies comparing different screening intervals (biennial or triennial versus annual) and four studies comparing mammography with DBT versus DM met this requirement. For KQ3, 19 studies provided data related to the potential relative harms of different screening strategies, including supplemental screening. No studies compared screening strategies by population characteristics and risk markers for any of the KQs, although two relevant RCTs are ongoing with estimated completion dates in 2025.[167, 168]

Overall, evidence of the relative effectiveness or harms of different breast cancer screening strategies was limited. The completion of ongoing trials will add to this evidence base in the future.

Age to Start or Stop Screening

No randomized trials that assigned individuals to different ages to start or stop screening were identified for inclusion in this review. A nonrandomized study (N = 264,274) based on data from Medicare B enrollees ages 70 to 84 suggested that continuing screening beyond age 75 did not reduce breast cancer mortality compared with stopping screening (aHR 0.78 [95% CI 0.63 to 0.95]).[124] The study used novel statistical methods to approximate a per-protocol trial effect estimate from observational data.[169] The study did not present subgroup comparisons to identify specific groups that might benefit from continued screening beyond age 75. In terms of potential harms, fewer breast cancers were diagnosed among those who stopped screening, which could indicate overdiagnosis with continued screening given the similar mortality rates in those ages 75 to 84 that continued versus stopped screening or reflect short term followup in the study (8 years). Cancers detected in those who continued screening were more likely to be treated with lumpectomy and radiotherapy than mastectomy and chemotherapy.

Screening Interval or Frequency

Two older studies that compared triennial with annual screening did not find evidence of a mortality benefit with more frequent screening. Specifically, one non-randomized experiment that assigned participants in the Finnish national screening program to annual or triennial screening did not find a difference in breast cancer mortality (RR 1.14, 95% CI 0.59 to 1.27) or all-cause mortality (RR 1.20, 95% CI 0.99 to 1. 46).[157] An RCT of annual versus triennial screening from a similar time period conducted in the UK reported more screen-detected invasive cancers over multiple rounds of screening, but no difference in invasive cancers overall (including interval cancer) or their prognostic features.[124] These studies were limited in terms of potential risk of bias related to randomization and the applicability of the studies to the current US screening population because of the study periods and settings.

No studies comparing annual to biennial screening reported breast cancer mortality or other health outcomes. Intermediate outcomes (KQ2) were reported in one non-randomized study using BCSC data to compare the progression of tumors diagnosed following an annual or biennial screening interval.[152] The study indicated no difference between annual and biennial screening by decade of age in the adjusted risk for cancer diagnosed at stage IIB+ or with less favorable prognostic characteristics (stage IIB or higher, tumor size >15mm, or positive node status).

Harms related to screening intervals were evaluated in two non-randomized studies using BCSC data[138] and a health system data source with a majority Hispanic population,[151] that provided estimates of cumulative false positive recall and false positive biopsy rates. Annual screening resulted in more false positive recall and biopsy than biennial screening, estimated to be twice as high in one study (OR 2.2 95% CI 1.7 to 2.8). The most recent analysis of BCSC data showed that at least 50 to 56% of women screened annually over 10 years would have at least one false

positive recall and approximately 12% would have at least one false positive biopsy. Among those screened biennially, 36 to 38% would experience at least one false positive recall and 7% at least one false positive biopsy. Annual screening would thereby result in approximately 50 more false positive biopsies per 1,000 women screened over a 10-year period. These estimates update previous BCSC analyses and account for the more recent increased use of DBT screening; their findings of higher rates of false positives with annual screening are consistent with those in the previous review.[143, 170]

Studies included for comparisons of annual and biennial screening were more applicable to the US screening population but were not randomized and subject to considerable risk of bias due to confounding and selection.

Our review did not identify any updated information on effect of screening interval on the lifetime impact of radiation. The 2016 review included information from models which calculated the number of deaths due to radiation-induced cancer using estimates for digital mammography is between 2 per 100,000 in women age 50 to 59 years screened biennially, and up to 11 per 100,000 in women ages 40 to 59 years screened annually.[89]

Digital Breast Tomosynthesis Screening

No studies of breast cancer screening with DBT compared with DM reported mortality outcomes. Three RCTs (N = 130,195)[127, 141, 158] and one non-randomized comparative study (N = 92,404),[142] all conducted in Europe, reported cancer detection outcomes from two rounds of screening. The DBT screening intervention group also received sDM imaging in two of the trials (synthetic views equivalent to DM) and DM imaging in one trial. The second screening was conducted after a biennial interval for most participants. The modality of screening was the same for all participants during the second round of the trial (either DM or DBT/sDM). Some trialists have proposed that a common modality at round two should be used to evaluate whether stage shift has occurred. Similarities in the study designs and effect sizes, and low statistical heterogeneity supported the estimation of pooled effects for some outcomes.

A potential benefit of a more sensitive breast cancer screening imaging technology is that it might detect small, clinically important tumors before they progress to advanced disease. Results from the trials were inconclusive as to whether the added first round of detection with DBT would reduce the incidence of advanced cancers, and thereby improve health outcomes. In three trials comparing screening with DBT versus DM, DBT was associated with increased detection in of two of the three trials at the first screening round (pooled RR 1.41, 95% CI 1.2 to 1.6, I^2 8%, 3 trials, n = 129,492), but in none of the trials at the second screening round (pooled RR 0.87, 95% CI 0.7 to 1.1, I^2 0%, 3 trials, n = 105,244). Tumor characteristics and prognostic characteristics were inconsistently reported or had heterogenous effects across the studies, precluding meta-analysis of outcomes related to breast cancer stage at detection. There was not statistically significant evidence of stage shift in the individual studies and the results were too inconsistent to pool this small number of studies. The trials primarily reported dichotomous outcomes to categorize early versus advanced disease, most commonly using stage IIB+ or tumor size greater than 20 mm as cutpoints, which may not be sensitive or meaningful for identifying clinically important differences in cancer detection.

The absence of changes in the distribution of tumor characteristics or stage at detection at round two could also be interpreted to mean that the additional detection with DBT at round one would have little to no effect on health outcomes, such as breast cancer morbidity and mortality. If the increased detection was comprised of more indolent cancers with longer sojourn times, the time of diagnosis may be shifted earlier without a change in mortality risk.

Studies describing interval cancer results were evaluated as potential harms in this review because they are due to either false-negative screening (a harm arising from low sensitivity) or missed cancers that progressed to clinical significance during the gap between screenings. The same European RCTs (n=130,196) and five non-randomized comparison studies (N = 5,327,560) were included for assessing the risk of interval cancer associated with DBT screening compared with DM only. Several studies have documented differences in the tumor characteristics of interval and screen-detected cancers, and worse prognosis; therefore, a screening program that reduced the risk of interval cancers could be more effective for prevention of mortality from breast cancer. The three large RCTs found no statistically significant difference in the rates of interval cancers following screening with DBT compared with DM. The data on interval cancers from the five NRSIs were mixed, and interpretation was limited by differences in study design. Combined with the similar cancer detection results for DBT and DM, , the findings on interval cancer additionally suggest similar screening effectiveness for the two technologies based on the available evidence.

Overdiagnosis and overdetection are important potential harms of screening. The 2016 breast cancer screening effectiveness review for the USPSTF reviewed a broad literature including the effectiveness trial evidence and modeling studies and found overdiagnosis rates ranging from 11 to 22 percent in trials; and 1 to 10 percent in observational studies.[171] These outcomes are difficult to estimate even in the setting of large effectiveness trials because of differences in definitions and data collection. Rates of DCIS are considered one measure of overdiagnosis in screening studies because DCIS is generally treated but has unclear malignant potential. The three trials with multiple screening rounds did not show statistically significant differences in DCIS detection in meta-analysis, although this outcome is only one theorized source of potential overdetection that could lead to overtreatment.

Additional harms include rates of recall for additional imaging, false-positive recall, and false-positive biopsy; however, these were inconsistent across studies comparing DBT with DM. An included study using BCSC data estimated the 10-year probability of at least one false-positive recall to be slightly lower with DBT screening when screening was conducted annually, however, rates were high for both groups, with 50 percent screened with DBT and 56 percent screened with DM experiencing at least one false-positive recall with ten years of screening. Limited evidence from other studies on less commonly reported harms included adverse events associated with screening, which were rare, and radiation exposure. In studies using DM with DBT, radiation exposure was two-fold higher than what was received in the DM group, but in two studies using DBT with synthesized DM images created from the DBT scan the dose was similar between study groups.

Current studies with more than one screening round do not provide evidence that DBT has an advantage over DM by detecting cancer at earlier stages. Breast cancer includes a range of disease features, including both indolent or slow growing tumors and rapidly progressive disease

that may have a short window for detection before metastatic disease develops. A tumor stage shift could contribute to improved health outcomes, if observed, but imprecise estimation and inconsistencies in the few studies reporting detection and tumor characteristics outcomes limit conclusions. These limitations increase uncertainty about the effect of small improvements in test performance on health outcomes.

Overall, the studies indicated no or minor differences between DBT and DM screening in effectiveness and potential harms. Very few randomized trials that have completed more than a single round of screening are available and neither RCTs nor nonrandomized studies reported mortality outcomes.

Test Performance Characteristics of DBT

A large volume of evidence on DBT comes from single round test performance studies, including paired design studies that report the detection yield for readings on the same person with DBT/DM versus DM only. The literature on the test performance of screening tests can be helpful for the evaluation of new technologies, and their potential contribution to a screening program. Three systematic reviews (including randomized trials, prospective cohorts and diagnostic accuracy studies) reported pooled estimates of PPV (i.e., percent diagnosed with cancer among those with a positive mammogram result) and false-positive recalls (i.e., proportion recalled that were not diagnosed with cancer) among average-risk women screened with DBT or DBT/sDM versus DM.[172-174] Overall, the reviews included relatively few eligible studies (k = 4-13) with fewer available for meta-analysis of most outcomes. Statistical heterogeneity was also high for most analyses, raising questions about the validity of the pooled estimates. Results of the reviews were mixed, but small increases in PPV with DBT/sDM or DM compared with DM were reported.

A 2020 review of 10 studies (3 randomized trials, 1 prospective cohort study, and 6 diagnostic accuracy studies) estimated a difference in PPV (invasive breast cancer and DCIS combined) between participants screened with DBT/sDM versus DM (pooled RR 1.26; 95% CI, 1.09 to 1.46, I^2 = 52%; k = 6, n = 213,927 screening recalls).[174] A 2022 individual participant data (IPD) meta-analysis including four prospective studies found that that PPV (invasive breast cancer and DCIS combined) improved with DBT compared to DM (pooled RR 1.31; 95% CI, 1.07 to 1.61; I^2 = 70%, n=7,274 screening recalls).[173] The 2020 meta-analysis showed no difference in false-positive recalls (invasive breast cancer and DCIS combined) between women screened with DBT versus DM (RR 1.06; 95% CI, 0.85-1.32, I^2 = 85%, k = 6, n =96,970 screening exams) or between DBT/sDM versus DM (RR 1.02; 95% CI, 0.85 to 1.23, I^2 = 90%, k = 6, n =213,927).[174] A 2022 systematic review of 13 studies (1 RCT, 12 observational cohorts) also reported meta-analyses with very high statistical heterogeneity that suggested improved PPV and recall with DBT/sDM. It was unclear whether the results were for invasive cancer detection or invasive cancer and DCIS detection.[172]

Data from the BCSC can provide estimates of screening performance from data on US populations screened in select breast cancer care systems that contribute to the registry. A 2020 publication by Lowry et al. used 2010 to 2018 data from five BCSC registries to assess the performance of digital mammography (1,273,492 screening examinations) versus DBT mammography (310,587 screening examinations) among women ages 40 to 79 years.

Improvements in cancer detection and recall with DBT were observed at baseline screening (prevalence screen) across all age groups. At subsequent screening visits (incidence screens) screening performance improvements were not uniform. Only women with heterogeneously dense breasts and women ages 50 to 79 with scattered fibroglandular breast density had reduced recall relative to cancers detected. Younger women with extremely dense breasts experienced higher recall with DBT at subsequent screens and no improvement in cancer detection. The main analyses presented were adjusted for a range of demographic and breast cancer risk characteristics, but the observational design cannot fully account for differences in the reasons women may have received DBT screening; risk of bias from selection into the study groups and potential unmeasured confounding remain even after statistical adjustments.[175]

Supplemental Screening With Ultrasound or MRI

No studies comparing women screened with mammography only with those receiving supplemental MRI screening reported health outcomes or evidence of reduced progression to advanced cancer in subsequent screening rounds. Harms were reported in one RCT (N = 40,373) that found fewer interval cancers diagnosed in the two years following the first round of screening among a group with dense breasts invited to MRI after a negative screening mammogram result (2.2 per 1,000) compared to those with dense breasts that did not receive the invitation (4.7 per 1000) in the ITT analysis (RR 0.47, 95% CI 0.29 to 0.77). The reduction in interval cancers serves as an intermediate outcome suggesting potential benefit, but the likelihood and magnitude of differences in breast cancer morbidity and mortality outcomes are not yet known. While this study was designed to consist of three MRI screening rounds, second round results for both study groups have not been published.[162]

Harms from MRI screening identified in the review included additional recalls and biopsies from the supplemental imaging. The acceptability of screening was also limited in the trial that randomized participants with dense breasts to an invitation for MRI after having a mammography screen with negative findings. Forty percent randomized to the MRI invitation did not present for screening. Data from a nonrandomized study using insurance claims data (N = 18,416) estimated compared cascade events (mammary and extramammary) in the 6 months following screening and did not find a difference between those screened with MRI or mammography.

One randomized trial conducted in Japan (N = 72,717) was designed to estimate the effectiveness of DM plus ultrasound screening compared with DM only for women ages 40 to 49, since this group tends to have higher breast density. The study has published results from the first round of screening with followup for interval cancers, and second round findings are currently being analyzed for future publication (personal communication). There was not a statistically significant difference in interval cancers following first round screening in this trial (0.4 vs 0.8 per 1000 screened) based on the event rates reported (unadjusted), but the estimate was imprecise. There was also no difference in a nonrandomized study by Lee et al. using data from two BCSC registries with propensity score matching to adjust comparisons for confounding and selection bias (1.5 vs 1.9 per 1000 screened).[150] These studies also reported additional followup testing attributed to ultrasound screening. The Japanese trial found 48 per 1000 additional false positive screens from ultrasonography. The BCSC study reported false positive

biopsy rates that were more than twice as high in the group with supplemental ultrasound compared with having only a mammogram (52.0 vs 22.2 per 1000 screened). The BCSC analysis also did not report statistically significant differences in detection or sensitivity with supplemental ultrasound screening compared to digital mammography, but these outcomes were not included in our review since the study did not report results compared across multiple screening rounds.

Differences in detection of cancer with supplemental screening in addition to mammography have been reported in studies that were not eligible for our review for the reasons outlined above (e.g., paired-studies where individuals serve as their own control through blinded readings). Two recent systematic reviews included individual paired-study designs not eligible for this SER. These reviews reported pooled estimates of sensitivity and specificity for women with dense breasts receiving supplemental screening with MRI or ultrasound.

A 2022 systematic review of 42 studies that included a wide range of study designs and settings, reported on the performance of various supplemental breast cancer screening modalities for women with dense breasts.[176] Test performance characteristics were estimated primarily based on observational studies using sequential testing where participants served as their own controls. For supplemental screening with handheld ultrasound, meta-analysis of nine studies estimated 86 percent sensitivity (95% CI, 77 to 92) and 87% specificity (95% CI, 75 to 93) for diagnosed breast cancer, with low statistical heterogeneity (k=9, n=42,242; I^2 = 0.09%). Test performance results for MRI supplemental screening were limited and inconsistent and could not be summarized using meta-analysis.

Overall, the review concluded that supplemental screening with handheld ultrasound or MRI could increase cancer detection by 2 to 3 per 1,000 women with dense breasts, but would also substantially increase recall by 73 to 134 per 1000 screens and biopsy by 33 to 73 per 1,000 screens among women without cancer. The authors noted the lack of studies reporting breast cancer mortality outcomes or intermediate outcomes that could be used to assess the health impact of the additional cancers detected.[176] A 2020 meta-analysis estimated higher pooled sensitivity and specificity for supplemental screening with ultrasound compared with DM alone for women with dense breasts based on five studies with extremely high statistical heterogeneity.[177] No systematic reviews were identified reporting pooled estimates of PPV or false-positive recalls for women receiving supplemental MRI or ultrasound.

The previous review of supplemental screening noted the shortcomings of test performance data on this topic for establishing the clinical net benefits of screening programs.[95] Comparative studies that report health outcomes are important for establishing whether supplemental or breast cancer screening tests lead to improved health outcomes or contribute to false positives, overdiagnosis and unnecessary treatments.

Screening in Different Population Subgroups

No studies evaluated potential differences in screening effectiveness and harms for population subgroups using valid rigorous methods. Subgroup comparisons were not adequately powered or assessed with statistical tests for interaction, but instead were based on presentation of stratified

results, primarily by age, breast density, breast cancer risk, and less commonly, by hormonal status. There were some consistent trends that were present in the evidence from subgroup analyses, but limitations in the study designs and analyses hindered the strength of findings (**Appendix F Table 7**). In general, breast density and younger ages were associated with higher false-positive results with screening. However, the absence of interaction tests, lack of correction for multiple comparisons, and the possibility of unmeasured confounding that can introduce bias in observational comparisons precluded conclusions. Evidence from BCSC and other registry studies generally showed findings consistent with the broader literature.[138, 145, 152]

No comparative effectiveness studies reported differences in estimates by race or ethnicity. Nearly all of the included studies were conducted in majority White, non-Hispanic populations and were not powered with adequate numbers of Black, Hispanic, Asian, or Native American/Alaska Native women for meaningful comparisons.

Inequities in Breast Cancer Incidence and Outcomes (CQ1)

A pronounced inequity in breast cancer mortality in the US is seen among non-Hispanic Black women compared with all other people. Although the incidence of breast cancer among Black women overall is not as high when compared with non-Hispanic White women, breast cancer mortality is 40 percent higher for Black women (27.6 per 100,000, compared with 19.7 per 100,000 for White women) based on the most recent US Surveillance data (2016-2020).[14] Relative risks of mortality when accounting for the age and stage at diagnosis have been estimated to be 71 percent higher for non-Hispanic Black women and 28 percent higher for American Indian/Alaska Native (AI/AN) women compared with non-Hispanic White women.[178] Mortality from breast cancer was similar between Black and White women before the 1980s, after which mortality rates abruptly diverged. The introduction of mammography screening and new treatment interventions, particularly adjuvant endocrine therapy, around the same time suggest that health care inequities underlie the emergence of the disparity and its persistence.[14, 20]

Currently, most research on health inequities compares non-Hispanic Black women to non-Hispanic White women. Many of the issues outlined below may similarly affect care and outcomes for other populations in the US, although some inequities may result from causal pathways unique to specific populations. For example, there are longstanding and substantial inequities in breast cancer survival for populations living in rural areas of the United States.[179]

The National Institute of Minority Health and Disparities (NIMHD) framework[180] was developed to guide research investigating health disparities and is helpful for examining sources of inequities in breast cancer survival, particularly higher mortality for Black women. The framework recognizes the role of the health care system,[181] the sociocultural environment, the built environment, behavioral factors, and genetic factors that contribute to health inequities. Inequities in breast cancer mortality can be examined at each step along the cancer screening, diagnosis, treatment, and survival pathway with these factors in mind.[182] The higher mortality rate seen for Black women diagnosed with breast cancer in the United States aligns with other health inequities that are attributed to the effects of structural racism, which results in inequalities in resources and exposures, including disparities in access to high quality health care.[183-185] For example, worse breast cancer survival has been associated with racialized residential segregation that has been driven by historical and ongoing discriminatory housing

policies.[186-189] Racialized and classist segregation has also been associated with exposure to cancer risk from toxic environments in terms of air pollution, industrial waste, built-environments that do not support health, and stressful life conditions.[186, 188] Although interrelated factors contribute to inequities in breast cancer mortality, the primary focus in this report is on structural, systemic, and individual factors related to health care that are in the USPSTF purview.

Research is ongoing to disentangle the factors that may contribute to the observed higher rates of cancer subtypes with worse prognoses among non-Hispanic Black women, who are more likely to present with advanced cancer compared with non-Hispanic White women.[20, 189, 190] Based on national SEER surveillance estimates (2016-2020), breast cancers having a hormone receptor (HR) negative molecular marker are more common among non-Hispanic Black women compared with non-Hispanic White women (30.6 vs 17.4 per 100,000). The higher incidence of negative hormonal receptor status leads to worse outcomes since these subtypes are less readily detected through screening and less responsive to adjuvant endocrine therapy.[191] Triple negative cancers (i.e., ER-, PR-, HER2-) are also more likely to be diagnosed at younger ages and among Black women (24.1 per 100,000) compared with White women (12.4 per 100,000) based on data from 2015 to 2019. These cancers tend to be particularly aggressive and more likely to be diagnosed at later stages than other subtypes. Sub-Saharan African ancestry may contribute a genetic component to this difference, but HR negative cancers have decreased for all racial and ethnic groups in the United States, and variability in rates of decline by region suggests a more complex etiology.[192] Observed regional differences in the incidence of HR negative cancer within and between racial groups suggest that environmental and social determinants of health may contribute to the risk of developing HR-negative cancer.[20, 192] Although differences in the incidence of different cancer subtypes explain some of the differences in breast cancer mortality (estimated 56%), race differences in mortality within subtypes point to barriers to obtaining high quality health care and disparities in screening followup and treatment initiation.[20]

Differences in recent trends in breast cancer incidence are difficult to attribute to specific factors due to the complex interactions of structural and environmental conditions, health care, and individual health mediated processes that can be associated with cancer detection and diagnosis. Breast cancer incidence trends show slight increases from 2005 to 2019 for Non-Hispanic Black women and Non-Hispanic White women ages 50 to 74 (0.9 and 0.4 average annual percent change, respectively) and similar increases among those ages 40 to 49 (0.6 average annual percent change for both groups).[16] Other race and ethnicity group have experienced steeper increases in incidence since 2015. Average percent increases in incidence were higher and similar among Asian/Pacific Islander women (2.0 average annual percent change [AAPC]) and Hispanic women (1.7 AAPC) ages 50 to 74. Incidence among American Indian/Alaska Native women has also risen by at least 1.7 percent on average each year, but the trend is not precisely estimated for all age groups (ages 40-64, 1.7 AAPC; ages 50-74, 6.1 AAPC [p = 0.14]; ages 75+ 1.8 AAPC). At younger ages, 40 to 49, increasing trends have been steepest among Asian/Pacific Islander women (4.0 AAPC), followed by AI/AN women (2.0 AAPC), and Hispanic women (1.6 AAPC).[16] Overall, however, among women below age 40, Black women have the highest breast cancer incidence (27.6 per 100,000 women).[10]

Structural and contextual factors affect the well-being, health, and resources (e.g., financial, health literacy) of individuals when they enter the health care system, and factor into their

experiences obtaining care.[193] The next sections focus on inequities that accumulate along the health care pathway that contribute to mortality disparities, drawing on a conceptual framework presented by Nelson et al. for a systematic review on interventions to address inequities in preventive health services.[193]

Inequities in Access to Screening

Despite having a higher rate of breast cancer mortality, non-Hispanic Black women report the highest rates of mammography screening. Based on self-reported BRFSS data from 2020, 78 percent of all women aged 50 to 74 reported having a mammography in the past two years. For non-Hispanic Black women, the rate was 84.5 percent, followed by Hispanic women (79.8%), Native-Hawaiian/Pacific Islander women (79.7%), Hispanic women (79.8%), non-Hispanic White women (77.8%), and American Indian/Alaskan Native women (68.7%). Non-Hispanic Black women also reported higher levels of screening than non-Hispanic White women from age 40 to 44 (60% vs. 54%) and age 45 to 49 (76% vs. 68%).[194] Self-report data from the 2015 and 2018 National Health Interview Survey indicate lower, but similar, rates of breast cancer screening for non-Hispanic Black and non-Hispanic White women (72.9% and 71.7%, respectively).[195]

Although evidence remains unclear regarding the relative benefit of DBT compared with DM screening, adoption of DBT occurred most rapidly in regions with proportionally larger White non-Hispanic populations.[196] In addition, even as the availability of DBT increased, Black, Asian, and Hispanic women remained less likely to be screened with DBT compared with White women. Analysis of data from the BCSC indicates that when both technologies were available at the screening site, over half of White women (53%), and smaller percentages of Black (38%), Hispanic (44%), and Asian women (43%) were screened with DBT.[197] Out of pocket costs often required for DBT screening may contribute to these differences, as well as inequities in the geographic distribution of health resources and clinician behaviors.[198-200]

Although there are not currently recommendations for supplemental screening in the general screening population, barriers to access for individuals at increased risk of breast cancer could contribute to mortality risks. Uneven access to supplemental screening modalities (e.g., MRI, ultrasound) has been documented in the US, and is most likely to impact Native American women, and those living in rural areas.[201]

Inequities in Diagnostic Follow-Up and Access to Evidence-Based Cancer Treatments

Health outcome benefits from mammography screening require initiation and completion of appropriate and effective followup and treatment. Microsimulation modeling and other population-based studies have suggested that treatment advances have had a greater impact on reducing breast cancer mortality than screening.[36] These advances have been most pronounced for hormone receptor positive cancer subtypes. Delays and inadequacies in the diagnostic and treatment pathway likely contribute to increased mortality relative to those receiving prompt, effective care.[193]

Disparities in followup after screening have been observed for Black, Hispanic, and Asian women compared with White women.[182, 202-209] Interventions to address delays in followup of abnormal screening results, treatment initiation, and treatment completion, especially for Black women for whom delays and reduced access to timely care are most pronounced, could address disparities in the care pathway following a positive screening mammogram. The use of navigators, shown to improve cancer screening rates, deserves investigation for potential effects on reducing inequities in followup and treatment.[182]

Adjuvant endocrine therapy reduces the risk of cancer recurrence among individuals with hormone receptor positive cancers by up to 30 percent, but long-term adherence can be difficult. Adherence has been associated with factors such as health literacy, comorbidities, depression, cognitive function, and social support, as well as the types of side effects experienced with therapy.[210] Black women are more likely to discontinue adjuvant endocrine therapy compared with White women, in part due to greater physical symptom (vasomotor, musculoskeletal, cardiorespiratory) and psychological symptom (distress, despair) burdens and owing to structural and contextual factors such as neighborhood and community resources and supports.[211, 212] Improved symptom management and social support could improve adherence and help reduce cancer outcome inequities. Improvements in access to effective health care, removal of financial barriers, and use of support services for followup and treatment of breast cancer could reduce mortality risks for individuals experiencing disparities related to their race or ethnicity, rural location, low income, or other factors associated with lower breast cancer survival.

Additional Findings From Original Effectiveness Trials (CQ2)

A detailed overview of the findings of the original effectiveness trials of mammography screening from the 2016 evidence review can be found in **Appendix A**. These trials include the Canadian Breast Cancer Screening Studies (CNBSS-1 and CNBSS-2), the United Kingdom Age trial, and four trials from Sweden, including the Stockholm trial, Malmö Mammographic Screening Trial (referred to separately as MMST I and MMST II), Gothenburg (Göteburg) trial, and Swedish Two-County Study (referred to separately as Östergötland and Kopparberg).[111] We conducted a literature scan that identified updated estimates of effectiveness for four of the trials reporting on extended followup.[77, 78, 213]

A single 2017 publication presented an updated analysis of mammography effectiveness from a series of Swedish screening trials (the Malmo (MMST I and MMST II), Stockholm, and Göteburg [Gothenburg] trials) with over 20 years of followup data (30, 22, 25, and 24 years, respectively).[213] These analyses focus on the difference in breast cancer mortality between screening and control groups among women with breast cancers diagnosed between randomization and completion of the first screening round of the control group (time varied by trial from 4.3 to 12.4 years). The previous review classified these analyses as using the 'short case accrual method' (sometimes referred to as the 'evaluation method' in trial publications). This method of analysis reduces the risk of contamination in the control group after the screening phase of a trial is completed but includes fewer cases in the analysis. Overall, the combined results from the Swedish trials retained the originally reported statistically significant effect of screening. The updated estimate from these three trials showed a 15 percent relative reduction in breast cancer mortality for women ages 40 to 74 years (RR: 0.85 [95% CI, 0.73 to 0.98]). When the age-stratified results were compared with the study-specific estimates for short case accrual

from the previous review, the point estimates were similar, although confidence intervals included 1.0 for all age groups: ages 40 to 49 years at randomization (RR: 0.79 [95% CI, 0.62 to 1.0]), ages 50 to 59 years at randomization (RR: 0.89 [95% CI, 0.71 to 1.1)], and ages 60 to 70 years at randomization (RR: 0.73 [95% CI, 0.58 to 1.2]).

The UK Age trial of mammography effectiveness among women ages 40 to 49 years published final results incorporating nearly 23 years of participant followup data.[77, 78] In addition to the short-case accrual method, utilized in the Swedish trials, the UK Age trial also presented results using the long case accrual method, which counts all breast cancer cases contributing to breast cancer deaths diagnosed over the course of the screening intervention period and the followup period. The long accrual method is considered least biased because it accounts for lead time and detection bias inherent in studies of cancer mortality.

The UK Age trial recruited women ages 39 to 41 years for random assignment to yearly screening up to and including the calendar year that they reached age 48 years (intervention group), or to usual care that included no screening until entering the National Health Service Breast Screening Program (NHSBSP) at approximately 50 years of age (control group). The primary endpoint was mortality from breast cancer diagnosed in the intervention period for both groups (all breast cancer diagnosed after randomization but before first NHSBSP invitation).

Based on a median of 22.8 years of followup, the final primary analysis showed no statistically significant difference in breast cancer mortality from starting screening at ages 39 to 41 (RR: 0.88 [95% CI, 0.74 to 1.03]). An analysis based on long-term case accrual also resulted in no statistically significant impact on breast cancer mortality (RR: 0.90 [95% CI, 0.79 to 1.03)] or all-cause mortality (RR: 1.01 [95% CI, 0.96 to 1.05)]. In addition to the protocol specified primary analyses, the publication provided findings from several secondary post-hoc analyses stratified by followup periods. These analyses suggested a reduction in breast cancer mortality when followup was limited to the first 10 years of the trial (RR: 0.75 [95% CI, 0.58 to 0.97]), but no differences with followup from 10 years post-randomization and beyond or overall. These stratified analyses were not prespecified for the trial and use different definitions of the intervention period than previous analyses from the trial.

New publications reporting long-term outcomes are consistent with findings summarized in the 2016 evidence review. Results of nine RCTs individually and collectively indicate no statistically significant reduction in breast cancer mortality for women screened at ages 40 to 49. Breast cancer mortality is reduced in trials of women ages 50 to 69, although results of individual trials are mixed, and the magnitude of effect is small. Results for women ages 70 to 74 are inconclusive because few women in this age group were enrolled in the screening trials. Application of these findings to current practice remains questionable, although few other preventive health services offer trials of effectiveness with mortality outcomes, and clinical practice assumes benefits of screening regardless of the trial limitations.

Risk Assessment Tools to Personalize Breast Cancer Screening (CQ3)

Models estimating risk for breast cancer generally include common clinical risk factors, such as age, age at menarche, age at birth of first child, number of first-degree relatives, and number of previous breast biopsies. Additional variables differ between models including race, BMI, breast density, menopause status, use of hormone therapy, additional family histories, and others. Risk factors are categorized and weighted differently in each model. While all models published to date include age and number of first-degree relatives with breast cancer in their calculations, they vary in their complexity. These include the Gail,[66] Claus,[67] and Breast Cancer Surveillance Consortium (BCSC v2) models.[214]

A systematic review for the USPSTF published in 2019 included 25 studies of the diagnostic accuracy of 18 risk assessment methods to predict risk for breast cancer based on data from more than 5 million women.[48] The most studied methods include the Gail model and its variations including versions specific to Black and Asian women, and versions that include breast density. Studies also evaluated four versions of the Breast Cancer Surveillance Consortium (BCSC) model; two versions of the Rosner-Colditz model, two versions of the Tyrer-Cuzick model; a model based on data from Italian women; the Chlebowski model; and a model to predict estrogen receptor–positive and estrogen receptor–negative breast cancer.

Results of studies indicated modest discriminatory accuracy in predicting incidence of breast cancer in individual women with AUC values ranging from 0.55 to 0.65.[48] Studies of models specific to Black or Asian women showed similar results.[215-217] These values are generally considered too low for clinical applications, although they have been used as entry criteria and for risk stratification in research studies. The only study reporting AUC values above 0.70 for both the Gail-2 model (AUC, 0.74) and the Tyrer-Cuzick model (AUC, 0.76) was small and did not include a primary care population, limiting its clinical applicability.[218] Studies also indicated that adding variables, such as breast density, race, or BMI, to simple models had little effect on improving accuracy. [219, 220] Performance characteristics of individual models varied when applied across different validation samples.[48] In studies where multiple models were validated in the same population, different models predicted different results.[221, 222]

Diagnostic accuracy studies published since the 2019 systematic review further confirm the limited accuracy of risk models. These include studies of the BCSC (AUC, 0.63-0.68) [190, 223]; Gail (AUC, 0.59) [224]; International Breast Cancer Intervention Study (IBIS) (AUC, 0.66) [225]; and Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) (AUC, 0.56) [225] risk models; and a new model incorporating family history of breast cancer, proliferative benign breast disease, and previous breast calcifications (AUC, 0.587 to 0.647). [226] A study comparing the performance of commonly used models in predicting breast cancer risk among 35,921 women ages 40 to 84 years in a U.S. community screening population indicated AUC values of 0.61 for BRCAPRO, 0.64 for Gail, 0.64 for BCSC, and 0.62 for the Tyrer-Cuzick models. [227] These values are consistent with previous validation studies.

Additional new models include approaches or technologies not currently clinically available. These include a model with image-derived risk factors combined with clinical risk factors (AUC,

0.67) [228]; an image-based model (AUC, 0.73) [229]; a model using artificial intelligence and thermal radiomics (AUC, 0.89) [230]; models enhanced by machine learning (AUC, 0.88-0.90) [231]; mammography-based deep learning models (AUC, 0.79-0.81) [232]; and models incorporating single-nucleotide polymorphisms. [233, 234] However, these models have not been applied to routine population screening and require further validation.

There is not yet evidence from trials of screening informed by validated risk assessment instruments to tailor screening initiation, intervals, or modalities. Such evidence would be important to inform changes to clinical practice. Risk estimation from genome wide association studies is also being used to develop polygenic risk scores that may be used to further personalize screening, however the clinical utility of these models is unknown.[74]

Patient Perspectives on Balance of Benefits and Harms of Screening (CQ4)

Few studies directly examine how persons at average risk for breast cancer and eligible for routine screening incorporate personal preferences into their decisions about screening for breast cancer. Informed decisions about whether to undergo mammography may draw upon a range of factors including values, cultural influences, personal experience and risk factors, and awareness of screening benefits and harms. A 2019 systematic review of 22 studies[235] cited logistical challenges, psychological distress associated with the screening process, fear of a positive result, embarrassment, and not receiving services that align with cultural and/or religious beliefs as factors that influenced screening use. A 2022 systematic review including 66 studies focused on a broad set of individual social and structural factors that could influence screening use and access.[236] Several social and structural determinants of health were associated with reduced screening attendance in the review, including lacking access to a vehicle, living in crowded housing conditions, living further from a screening center, and being unemployed. Very few included studies in these reviews were conducted in the US, however, limiting generalizability and comparisons between different US population groups.

Experiencing screening harms, such as a false positive result, has been identified as a potential deterrent to screening.[236] A 2021 systematic review that pooled six NRSIs conducted in very high development index settings (none in the US), estimated reduced return to subsequent screening among women that had been previously screened and then experienced a false-positive result (pooled OR 0.77, 95% CI, 0.68 to 0.88). Older research on the phenomenon had mixed findings.[237] A more recent study conducted in a Chicago area health system (n = 261,767), found that women receiving a false positive screening result were less likely to return for their next scheduled mammogram (22.1% vs 15.0%, adjusted HR 0.74, p<.001) compared with women who had experienced a true negative result; findings were more pronounced when a biopsy was conducted (adjusted HR 0.66, p <.001). Women who experienced a false positive result were more likely to be non-Hispanic Black, younger, premenopausal, or to have dense breasts.[238]

Little is known about patient awareness of the possible benefits and harms of mammographic screening, and how it may differ among different patient populations. A small nationally representative survey of US women ages 30 to 59 (n = 557),[239] found that most women were

familiar with the benefits of mammography (>85%); fewer were aware of some of the potential harms. Nearly three-quarters were aware of psychological anxiety and the risk of false-positive results, but less than one-third were aware of the possibility of overdiagnosis. Personal preferences vary with respect to communication about the benefits and harms of mammography. A series of in-depth interviews (n= 58)[240] with older Latina, Non-Hispanic White, and Non-Hispanic Black women found that most participants (regardless of age, race/ethnicity, or education) preferred to hear about the benefits and harms of mammographic screening, including information about overdiagnosis, when deciding whether to continue screening beyond the age of 75 years. Highlighting the personal nature of such preferences, however, the study found that some participants preferred being encouraged to continue screening without discussion of possible harms, and some of the participants felt it was important to avoid discussing the prospect of overdiagnosis with older women, as it might deter them from getting mammograms.

To improve breast cancer screening programs, develop decision aids, and inform screening implementation, studies have investigated how mammography screening is perceived by various groups. Qualitative studies and survey research provide evidence on factors that influence how people weigh their personal risks and the potential benefits and harms of screening, which factor into the decision to be screened. For instance, Ro et al. conducted a series of interviews with Asian (n=4), Black (n=4), Hispanic (n=7), and Non-Hispanic White (n=10) patients at average risk of developing breast cancer.[241] Most reported having an annual mammography schedule, primarily influenced by their physician's recommendation, and several described receiving automatic annual reminders. Other factors that influenced their decisions included having a family history of breast cancer (n = 9), an interest in early detection (n = 5) and age (n = 5). For some, biennial screening intervals were considered acceptable if recommended by their physician and when they did not consider themselves to be at high risk for breast cancer. Others considered two years to be too long between screening visits regardless of physician recommendations or risk status. The study also identified a theme related to confusion about screening due to conflicting and frequently changing guidelines. Similar themes were noted in other qualitative studies regarding confusion about screening recommendations and desires for additional information to help make more informed decisions.[199, 242, 243]

Decision aids have been developed to help inform the shared decision-making process, which may help address some of the confusion described by patients in qualitative studies. A 2021 systematic review by Esmaeli et al.[244] reviewed 16 unique breast cancer screening decision aids that had been developed and tested in the US, Australia, Germany, Spain, UK, France, Taiwan, Italy, and The Netherlands. The review found that the decision aids improved patient knowledge and decreased decisional conflict, but had little or no influence on mammography participation rates, attitude, perceived risk of breast cancer, or anticipated regret.

Knowledge of inequities in breast cancer risks, mortality, and in access to treatment could influence some individuals' preferences for and decisions about breast cancer screening. Unfortunately, relatively few studies focused on the populations that experience inequities in screening access and breast cancer morality, such as non-Hispanic Black women, recent immigrants, and people living in rural settings. A focus group study (n = 39) including Black and Latina participants described diverse perspectives on breast, colon, and cervical cancer screening.[243] Some participants had strong, positive feelings towards preventive screenings,

considering them a critical tool for staying healthy. Others were more inclined to wait until a health issue became visible or problematic before seeking care, citing cultural norms, cost barriers, or personal history of challenges with affording care. Women who were born outside the US described feeling less acquainted with preventive health care, such as screening, as it was less likely to be offered or considered a cultural norm in the country in which they were raised. Trust in health care systems was also influenced by personal experiences with culturally insensitive or incompetent care, or awareness of historical concerns involving medical maltreatment (e.g., USPHS Syphilis Study at Tuskegee and forced sterilization in the early to mid-20th century). These factors can serve as barriers to receiving preventive screenings; however, having a positive, trusting relationship with a physician that encourages screening was described as helpful in rebuilding trust.

Perceptions and awareness of personal breast cancer risk can inform decisions about mammographic screening. In a study with Black and Latina focus group participants, lack of knowledge of family medical history served as a challenge in assessing individual risk for cancer.[243] In a focus group study[199] with Asian American (n=3), Non-Hispanic Black (n=8), Hispanic (n=2), and Non-Hispanic White (n=30) women with dense breasts, many were not aware that they were identified as having dense breasts, and almost none were aware that having dense breasts was an independent risk factor for breast cancer. Some study participants were receiving supplemental screening (such as ultrasound or MRI) in addition to DM but described having little knowledge of any specific benefit or possible harms of additional screening.

More research is needed to better understand to whether individuals are aware of the benefits and harms when determining whether or when to pursue breast cancer screening. It is particularly important to understand how race, ethnicity, gender identity, and cultural influences shape these decisions, to better inform shared decision-making practices and provide culturally competent care.

Breast Cancer and DCIS Treatment Harms (CQ5)

Breast cancer treatment regimens are highly individualized according to each patient's clinical status, cancer stage, tumor biomarkers, clinical subtype, and personal preferences, and vary in terms of potential side effects and morbidity.[33] For individuals with early stage (stage 1, IIA, and some stage IIB cancers) treatment generally involves lumpectomy with radiotherapy or mastectomy with or without radiotherapy.[245] Depending on patient and tumor characteristics, adjuvant systemic therapy may be used to reduce the risk of recurrence. Locally advanced cancers (stage IIB and stage IIIA to IIIC disease) will generally receive neoadjuvant systemic therapy prior to surgery with some cases receiving additional adjuvant therapy following surgical treatment.[245] Most patients with metastatic breast cancer receive systemic medical therapy along with supportive care measures.[246]

Complications following breast surgery include seroma formation, infection, pain, and arm morbidity (either directly attributable to the surgery or through a combination of surgery and adjuvant radiation).[247] The risk of postoperative complications increases with age[247] and is greater with mastectomy than with lumpectomy.[245, 247, 248] Additional adverse events associated with mastectomy may include skin flap necrosis (in 10 to 18 percent of cases) which may require

additional surgery or delays in adjuvant treatment, nipple necrosis (in 3 to 22 percent of cases), and phantom breast syndrome (the sensation of residual breast tissue).[248]

Whole breast radiation is associated with uncommon acute toxicities (e.g., severe breast pain, moist desquamation) involving the treatment area. In addition, radiation may result in longer-term complications of cardiotoxicity, lung injury, or secondary malignancies. Improvements in radiotherapy techniques have reduced these risks over time.[249] Chemotherapy is associated with acute toxicity resulting in side effects that usually resolve after treatment and differ based on the individual agents used; they most often include motor and sensory neuropathy, nausea, vomiting, hair loss, fatigue, vasomotor symptoms, and depression.[250] Longer term adverse effects of chemotherapy (including trastuzumab and hormonal therapy) vary by agent, but may include neuropathy,[251] cardiovascular disease,[252-254] osteoporosis, cognitive dysfunction, and secondary malignancies.[255]

Long-term complications of primary treatment of breast cancer can include recurrent pain and skin infections in the chest wall, musculoskeletal issues (particularly reduced arm mobility), neurologic morbidity (including nerve injury, peripheral neuropathy, and cognitive dysfunction), cardiovascular disease, menopausal symptoms, psychological effects, fatigue, and an increased risk of second cancers associated with breast irradiation, chemotherapy, or tamoxifen.[256]

Given the uncertainty regarding the prognostic importance of DCIS, there is clinical variability in the treatment approach taken when DCIS is identified at screening. DCIS treatment (which may include surgery, radiation, and endocrine treatment) is intended to reduce the risk for future invasive ipsilateral (same side) breast cancer and consequent breast cancer mortality, but is associated with harms. Prevention of future invasive cancer does not seem to be greater among those who undergo mastectomy in lieu of less invasive DCIS treatments.[257, 258] Despite lacking evidence of improved health outcomes, an analysis of SEER data from women diagnosed with unilateral primary DCIS between 2000 and 2014 found that over one-quarter of those referred for surgery chose mastectomy and the remaining 73 percent chose lumpectomy. Among those selecting mastectomy, most (75%) opted for removal of the affected breast, while the remaining opted for removal of both breasts.[259] Treatment of DCIS with mastectomy was associated with younger age, having health insurance, and living in a region with fewer radiation oncologists.[44] Research is ongoing to identify biomarkers and risk factors for progression, and to understand differences in the effectiveness of management and treatment options for reducing the risk of invasive cancer.[260] Three clinical trials of active surveillance without surgery as a management strategy for low-risk DCIS are being evaluated within the international PREvent ductal Carcinoma In Situ Invasive Overtreatment Now (PRECISION) collaboration. These include two randomized controlled trials in the United States (COMET)[261] and United Kingdom (LORIS).[262] and a patient preference trial in the Netherlands (LORD).[263] Until these trials are complete (estimated 2029-2030), the effectiveness of treatment of screen-detected DCIS to reduce breast cancer mortality remains unclear, and the extent to which it represents overdiagnosis and overtreatment is unknown.[122]

Treatment harms are of greatest concern when occurring among people who would not have otherwise experienced negative health consequences had their cancer not been screen-detected and treated. For some proportion of individuals participating in a screening program, the program may pose a greater risk to health than the breast cancer that was diagnosed.

Unfortunately, it is very difficult to estimate the extent to which a screening program contributes to overdiagnosis. Based on the effectiveness studies from the 2016 evidence review, estimates of overdiagnosis ranged from non-existent to nearly 50 percent of diagnosed breast cancer cases. Methods for estimating overdiagnosis varied in many ways, particularly by the type of comparison groups, assumptions about lead time, and the denominator used to calculate rates. In general, most adjusted estimates of overdiagnosis based on trials ranged from 11 to 22 percent. Estimates from observational and aggregated data range more widely, from nearly zero to over half of cases being overdiagnosed.[76, 121] Estimates from statistical models ranged from 0.4 to 50 percent. In the context of these findings, a recently published analysis using a statistical model based on BCSC data estimated that 15.4% (95% uncertainty interval, 9.4% to 26.5%) of screen-detected cancer cases would be overdiagnosed in a program of biennial screening from ages 50 to 74 years.[120]

Limitations of Our Review

Our review scope was developed following USPSTF procedures for assessing the comparative effectiveness of screening for eligible populations (not high risk) seen in settings reasonably comparable in terms of technology and practice to the U.S. health care environment (very high HDI settings). Comparative studies were included to inform USPSTF refinement of their guidelines on screening intervals, ages to begin and end screening, screening modalities, and supplemental or personalized screening strategies. The literature on breast cancer screening is vast. We conducted a comprehensive search of the literature, reviewed the reference lists of key studies and review articles, and sought expert input. Although unlikely, it is possible that our review could have missed relevant eligible studies published in English or in a different language.

Some included studies did not report complete data or provided results that were coded or described in ways inconsistent with other included studies. We sent inquiries to trial authors seeking additional information or data on key outcomes, but not all authors responded and were able to provide needed results.

The study design inclusion criteria for this review contributed to the low number of included studies and may be considered a limitation of our approach, despite its adherence to the USPSTF procedures. The included NRSI literature was limited to studies that compared screening approaches in at least two study groups either assigned or selected into different screening programs. This criterion meant that our review excluded single-arm studies often used to examine screening test performance. The review also did not include questions about the accuracy of screening for detection of invasive cancer and therefore did not include data on the commonly reported metric of cancer detection rate or positive predictive value.

Detection rates from a single screening round were not an included outcome for this review since improvements in detection would not necessarily reduce cancer mortality. This was because of the potential for bias introduced by studies considering only a single screening round. Additional detection, especially of DCIS and early-stage cancers, might extend lead time without altering health outcomes or contribute to overdiagnosis. In studies considering more than one round of screening, reduced mortality could be inferred if subsequent screening rounds had fewer advanced cancers in the intervention group. This would suggest that the intervention was

effective for better detection of early cancers that would have otherwise progressed, impacting treatment morbidity and breast cancer mortality. Similarly, commonly reported potential harms, such as recall rates and biopsy, were not taken from studies reporting only a single round of screening.

Limitations of the Evidence and Future Research Needs

Inherent challenges limit research and the availability of evidence on breast cancer screening. The majority of literature related to screening mammography comes from trials conducted in the 1970s through 1990s. The availability of more effective treatments and changes to screening technology could have implications for the estimated benefits and harms of screening obtained from earlier cohorts. Estimates of mortality benefits from historical trials could be greater or smaller than what is obtained with present day screening programs in the United States. While new trials on approaches to breast cancer screening could help inform screening programs, mortality from breast cancer is low at the population level, and therefore large sample sizes and long follow-up times would be needed to evaluate screening program effectiveness. Because of these challenges, much of the newer literature on breast cancer screening is focused on single-round comparative or diagnostic accuracy studies. Such studies have limitations for estimating the ultimate health effects of screening in light of potential overdiagnosis and improvements in survival in recent decades for cancer regardless of whether it is screen detected or clinically presenting with symptoms or a palpable mass.

Very few trials evaluate the comparative effectiveness of screening with different screening tests, intervals, or at different ages, and none have been conducted in the United States. Much of the recent literature on mammography screening has been aimed at estimating the test performance characteristics of different screening modalities, and especially whether the use of DBT screening alongside mammography might be more sensitive (for detecting cancers early) and specific (reducing the likelihood of false positive results), or whether certain subgroups may especially benefit from the new technology. Such studies can be informative for determining whether a new test is as good as or better than an older test. In the case of breast cancer, however, the advantages of earlier detection may be mitigated by the fact that treatment has grown increasingly effective for cancers detected at later stages, and small indolent or slow growing cancers could have similar outcomes if detected later. This makes it difficult to determine from test performance alone whether a modest gains in detection of smaller, early-stage cancers would necessarily lead to improved health or simply lengthen the time women live with a diagnosis. Studies reporting health outcomes are needed to resolve these questions. Finally, more robust measures and data collection on potential screening harms, including the patient perspective on false-positive experiences and harms from treatment.

The ongoing trials comparing DBT to DM from Europe included in this review has not reported results suggestive of stage shift, which would be anticipated if a health or mortality benefit were to be obtained. Such prospective intervention studies that use randomization or quasi-randomization to help overcome confounding and selection bias common to nonrandomized observational studies are needed to evaluate the effects of screening with DBT/DM or DBT/sDM compared with DM in the United States. Importantly, such studies should actively recruit enough Black, Asian, Hispanic, American Indian/Alaska Native and Pacific Islander participants to investigate how differences in screening, diagnosis, and treatment vary and affect outcomes.

DBT has increasingly been adopted for routine screening in the United States, and there are disparities in access to this technology seen for Black women, rural women, and others. Even if DBT itself has not yet been shown to confer a screening advantage over DM, limited access to this newer technology may be a marker for broader inequities in followup and treatment that contribute to higher breast cancer mortality for Black women. Studies comparing the health outcomes of DBT and DM screening often are biased by selection and confounding by indication – meaning populations that suffer lower access to comprehensive evidence-based health care are also less likely to be screened with DBT. Studies employing randomization are critical for obtaining unbiased effects.

As discussed above, research is needed to identify the underlying causes of inequities in breast cancer mortality along the clinical pathway. Screening rates are similar when comparing national data between Black women and White women, although for some vulnerable groups living in resource limited areas rates are lower and inequities greater. In addition to supporting guideline concordant screening, research is needed to identify and address factors other than access to screening. The importance of inequities along the clinical pathway following screening including diagnostic followup, treatment, and support services is increasingly recognized. Research is needed to identify where inequities exist and to develop interventions that close the care gaps following a positive screening result.

A consistent definition of advanced cancer has not been established in the literature, but stages II+ and stages IIB+ are the most common distinctions. Greater uniformity of reporting would benefit the comparability and interpretation of breast cancer screening studies. Since stage II includes localized cancers with average survival rates of 99.1 percent, their inclusion in study-reported definitions of advanced cancer may limit conclusions; treatment approaches and clinical outcomes differ for localized cancers. Including descriptions of whether cancers were staged according to an anatomic or prognostic staging system would add additional insight, as predicted mortality rates can vary slightly between the two.[264]

Additional studies with longer-term followup, preferably extended randomized trials allowing for comparisons across multiple rounds of screening, are needed to understand the impact of supplemental testing in women with dense breasts or other factors associated with increased risk on important breast cancer outcomes, including morbidity and mortality. Only RCTs and longer term followup can address risks of bias due to length time bias (earlier detection of cancer not resulting in improved outcomes) as well as the impact of overdiagnosis (leading to unnecessary treatment).

Our review did not identify any completed studies comparing outcomes for people with different screening initiation ages that met the review inclusion criteria. Study design challenges limit rigorous research on this topic. Studies comparing a group screened in their 40s with a cohort initiating screening at age 50 ten years later are subject to risk of bias since cancers detected and treated a decade apart experience different screening and treatment protocols. In the United States, many people commence screening at age 40, in part due to the discordant screening recommendations among leading guideline groups. This further reduces opportunities to randomize people in this age group to begin or delay screening. Newer methods for analyzing

observational data, such as those using emulated trials,[166] propensity scoring, or Mendelian randomization,[265] may be able to better address confounding and selection biases.

Ongoing Studies

We identified several ongoing studies relative to this review that are examining individualized risk-based screening, screening interval, and use of DBT with DM (**Appendix H**).

The current review did not identify any completed studies that incorporated a personalized approach to decisions about when to begin screening using an experimental design. The ongoing WISDOM trial should provide new evidence to improve our understanding of the effect of practical implementation of personalized screening on cancer detection, health outcomes, patient satisfaction, and screening adherence. Recruitment is ongoing with a target of enrolling 100,000 women ages 40 to 74 consenting to be randomized to either annual screening or individualized, risk-based screening. The trial is expected to be completed in March 2025. Another ongoing trial will contribute data on breast cancer-specific survival to a combined analysis with the WISDOM trial. The My Personalized Breast Screening study (MyPeBS, expected completion in 2025) is randomizing 85,000 women in Europe and Israel to standard screening (based on current national or regional guidelines) or screening with DM and/or DBT every 1 to 4 years (with or without ultrasound depending on breast density) based on estimated five-year risk of developing breast cancer. These two trials will provide valuable data to address research gaps identified in the current review.

The comparative effectiveness of different screening intervals will be assessed in the ongoing MISS trial (expected completion in 2026). The trial will randomize 60,000 women ages 45 to 49 years presenting for their first or second mammography screening to one of three arms – annual screening according to Italian screening program guidelines, biennial screening with DBT/sDM, or a tailored screening interval based on breast density (women with dense breasts being screened annually and women with non-dense breasts screened biennially). Participants will be followed for six years to compare the cumulative incidence of advanced breast cancer (stage 2 or higher), recall from screening, and interval cancers between screening intervals. A second Italian trial (Tailored Screening for Breast Cancer in Premenopausal Women, or TBST) planned to randomize 33,000 women ages 44 to 45 years to annual screening or tailored screening based on breast density; the results of this study will be part of a pooled analysis with the MISS trial.

Two ongoing Italian trials are comparing use of DBT/sDM versus DM. The MAITA trial is randomizing 8,000 women ages 45 to 65 years to one round of screening with DBT/sDM or DM. After one year for women ages 45 to 49 years and two years for women ages 50 to 65, all participants will be re-screened with DM. The similarly designed IMPETO trial aims to randomize 6,000 women ages 45 to 46 years to one round of screening with DBT/sDM or DM; after one year, all women will be re-screened with DM. The primary outcome is the cumulative incidence of advanced breast cancer (stage 2 or higher). Recall rates and benign biopsy rates will also be assessed. The MAITA trial is expected to be completed in 2026; enrollment in the IMPETO trial was postponed due to COVID-19 and the completion date has not been updated. Additionally, the PROSPECTS trial, set in the United Kingdom, is randomizing 100,000 women ages 49 to 71 years to one round of screening with DBT plus sDM or DM versus DM alone. The

primary outcome is invasive cancer detection rates and interval cancer rates. Recall rates and benign biopsy rates will also be assessed.

Finally, a trial to evaluate the comparative effectiveness of DBT and DM mammography among women with dense breasts is currently underway in the United States and Canada (expected completion 2030).[266] The Tomosynthesis Mammographic Imaging Screening Trial (TMIST) is randomizing 128,905 women ages 45 to 74 years with dense breasts (BIRADS density C or D). Women will be screened annually for four years and followed for four additional years (total followup eight years). The primary outcome is the incidence of advanced breast cancer (defined according to combinations of tumor size; ER, PR, and HER2 status; and tumor spread).[264] Secondary outcomes include breast cancer-specific mortality, test performance, interval cancers, and recall and biopsy rates.

Future comparative effectiveness reviews will benefit from the publication of additional followup from the included trials and of new trials currently underway. Studies using existing registry and cohort data analyzed using advanced statistical methods may also contribute to addressing current evidence gaps.

Conclusions

Previous reviews of breast cancer screening for the USPSTF, and the basis for its current screening recommendations, were grounded in evidence from effectiveness trials that showed decreased breast cancer mortality with mammography screening for women ages 50 to 69. Newer publications with long term followup to trial endpoints would not change previous conclusions based on these trials indicating a screening benefit for this age group. No new trials of the effectiveness of breast cancer screening are forthcoming, yet unanswered questions remain with respect to features of an optimal screening program designed to save the most lives while not subjecting healthy people to screening-related harms.

Comparative effectiveness trials comparing different screening modalities have not reported mortality outcomes, but among those with results from multiple rounds of different screening interventions an effect on mortality might be inferred if subsequent screening rounds had fewer advanced cancers. The ongoing trials comparing DBT to DM from Europe included in this review do not show a signal suggestive of stage shift, however, which would be anticipated if a health benefit is ultimately to be obtained. Overall, the studies indicated no or minor differences between DBT and DM screening in effectiveness and potential harms. Results from studies comparing screening programs involving supplemental imaging were too limited to evaluate potential benefits that could be inferred from signs of stage shift, but increased false-positive and biopsy harms occurred with supplemental screening.

The current evidence synthesis reflects a progression of the science from questions of effectiveness towards questions of comparative effectiveness. Also, while related questions on test performance were examined in previous reviews, the current review uses different selection parameters to include studies. Applying the USPSTF review procedures and evidence requirements to the comparative effectiveness literature on breast cancer screening intentionally narrowed the focus, resulting in fewer included studies, relative to prior reviews. Changes in screening recommendations could arise from evidence on the effectiveness of new screening

technologies or improved understanding of differential effects of screening starting and stopping age, or evidence on supplemental screening for women based on their breast cancer risks and personal preferences. Our review found little evidence to guide these refinements in breast cancer screening. Ongoing trials and future comparative studies may help fill the research gaps we have outlined, ideally in populations including people reflective of the United States demographic composition with respect to race and ethnicity. Notably, nearly all breast cancer screening trials have been conducted outside of the United States, most enrolling mainly White European populations. Studies are needed that focus on and enroll adequate numbers of underrepresented populations that face increased risk of breast cancer mortality. Finally, research and programs to identify and address factors underlying inequities in breast cancer survival, especially for Black women, are needed to improve interventions along the clinical pathway, including screening, timely diagnostic evaluation, and high-quality treatment programs, that could lead to better health and survival.

References

1. Siu AL. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2016;164(4):279-96. https://doi.org/10.7326/m15-2886
2. The Breast : Comprehensive Management of Benign and Malignant Disorders, 3rd Edition. Bland K, Copeland E, editors. St. Louis: WB Saunders; 2004.
3. Dillon D, Guidi A, Schnitt S. Ch 25: Pathology of Invasive Breast Cancer. In: Harris J, Lippman M, Morrow M, Osborne C, editors. Diseases of the Breast. 5th ed. Philadelphia: Lippincott-Williams & Wilkins; 2014.
4. Corben A, Brogi E. Ch 21: Ductal Carcinoma In Situ and Other Intraductal Lesions: Pathology, Immunochemistry, and Molecular Alterations. In: Harris J, Lippman M, Morrow M, Osborne C, editors. Diseases of the Breast. 5th ed. Philadelphia: Lippincott-Williams & Wilkins; 2014.
5. Burstein HJ, Polyak K, Wong JS, et al. Ductal carcinoma in situ of the breast. N Engl J Med. 2004;350(14):1430-41. https://doi.org/10.1056/nejmra031301
6. King T, Reis-Filho J. Ch 22: Lobular Carcinoma In Situ: Biology and Management. In: Harris J, Lippman M, Morrow M, Osborne C, editors. Diseases of the Breast. 5th ed. Philadelphia: Lippincott-Williams & Wilkins; 2014.
7. Houghton J, George WD, Cuzick J, et al. Radiotherapy and tamoxifen in women with completely excised ductal carcinoma in situ of the breast in the UK, Australia, and New Zealand: randomised controlled trial. Lancet. 2003;362(9378):95-102. https://www.doi.org/10.1016/s0140-6736(03)13859-7
8. Schwartz G. Biology and Management of Lobular Carcinoma In Situ of the Breast. In: Bland K, Copeland E, editors. The Breast : Comprehensive Management of Benign and Malignant Disorders, 3rd Edition. 3rd ed. St Louis: Saunders; 2004.
9. U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Visualizations Tool, based on 2020 submission data (1999-2018). https://gis.cdc.gov/Cancer/USCS/#/AtAGlance/ Accessed: May 22, 2022.
10. SEER Explorer. Breast Cancer: SEER 5-year Age-Adjusted Incidence Rates, 2015-2019. https://seer.cancer.gov/explorer/application.html. Accessed: February 16, 2023.
11. National Cancer Institute (NCI) Surveillance Epidemiology and End Results (SEER) Program. Cancer Stat Facts: Female Breast Cancer. https://seer.cancer.gov/statfacts/html/breast.html. Accessed: March 3, 2023.
12. American Cancer Society. Cancer Facts & Figures 2019. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf. Accessed: March 10, 2023.
13. U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on 2019 submission data (1999-2017). www.cdc.gov/cancer/dataviz. Accessed: September 24, 2020.
14. Giaquinto AN, Sung H, Miller KD, et al. Breast Cancer Statistics, 2022. CA Cancer J Clin. 2022;72(6):524-41. https://doi.org/10.3322/caac.21754
15. Cronin KA, Scott S, Firth AU, et al. Annual report to the nation on the status of cancer, part 1: National cancer statistics. Cancer. 2022;128(24):4251-84. https://doi.org/10.1002/cncr.34479

16. Surveillance Epidemiology and End Results (SEER) Program. SEER*Stat Database: Incidence - SEER Research Limited-Field Data with Delay-Adjustment, 22 Registries, Malignant Only, Nov 2021 Sub (2000-2019) - Linked To County Attributes - Time Dependent (1990-2019) Income/Rurality, 1969-2020 Counties. National Cancer Institute, DCCPS, Surveillance Research Program; Released April 2022, based on the November 2021 submission.
17. Melkonian SC, Jim MA, Haverkamp D, et al. Disparities in Cancer Incidence and Trends among American Indians and Alaska Natives in the United States, 2010-2015. Cancer Epidemiol Biomarkers Prev. 2019;28(10):1604-11. https://www.doi.org/10.1158/1055-9965.EPI-19-0288
18. Iqbal J, Ginsburg O, Rochon PA, et al. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. JAMA. 2015;313(2):165-73. https://www.doi.org/10.1001/jama.2014.17322
19. DeSantis CE, Ma J, Gaudet MM, et al. Breast cancer statistics, 2019. CA Cancer J Clin. 2019;69(6):438-51. https://doi.org/10.3322/caac.21583
20. Jatoi I, Sung H, Jemal A. The Emergence of the Racial Disparity in U.S. Breast-Cancer Mortality. N Engl J Med. 2022;386(25):2349-52. https://doi.org/10.1056/NEJMp2200244
21. American Association for Cancer Research. AACR Cancer Disparities Progress Report 2020. http://www.CancerDisparitiesProgressReport.org. Accessed: September 23, 2020.
22. National Center for Health Statistics. Table 033: Use of mammography among women aged 40 and over, by selected characteristics: United States, selected years 1987–2015. https://www.cdc.gov/nchs/hus/contents2018.htm#Table_033. Accessed: Oct 27, 2020.
23. SEER Explorer. Breast Cancer: U.S Mortality Rates by Age at Death, 2016-2020. https://seer.cancer.gov/explorer/application.html. Accessed: February 16, 2023.
24. SEER Explorer. Breast Cancer: SEER 5-Year Relative Survival Rates, 2012-2018. https://seer.cancer.gov/explorer/application.html. Accessed: February 16, 2023.
25. Rodgers KM, Udesky JO, Rudel RA, et al. Environmental chemicals and breast cancer: An updated review of epidemiological literature informed by biological mechanisms. Environ Health.160:152-82. https://doi.org/10.1016/j.envres.2017.08.045
26. National Cancer Institute. Breast Cancer and the Environment: Controversial and Emerging Exposures Workshop Summary. https://www.cancer.gov/research/areas/causes/breast-cancer-environment. Accessed: June 2, 2022.
27. Kresovich JK, Xu Z, O'Brien KM, et al. Methylation-based biological age and breast cancer risk. Journal of the National Cancer Institute. 2019;111(10):1051-8. https://doi.org/10.1093/jnci/djz020
28. O'Brien KM, Sandler DP, Taylor JA, et al. Serum vitamin D and risk of breast cancer within five years. Environ Health Perspect. 2017;125(7):077004. https://doi.org/10.1289/ehp943
29. Park YM, O'Brien KM, Zhao S, et al. Gestational diabetes mellitus may be associated with increased risk of breast cancer. Br J Cancer. 2017;116(7):960-3. https://doi.org/10.1038/bjc.2017.34
30. White AJ, Weinberg CR, Park YM, et al. Sleep characteristics, light at night and breast cancer risk in a prospective cohort. Int J Cancer. 2017;141(11):2204-14. https://doi.org/10.1002/ijc.30920
31. Lawson JS, Günzburg WH, Whitaker NJ. Viruses and human breast cancer. Future Microbiol. 2006;1(1):33-51. https://www.doi.org/10.2217/17460913.1.1.33
32. American Cancer Society. Breast Cancer Facts & Figures 2019-2020. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-

cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf. Accessed: March 8, 2023.
33. National Comprehensive Cancer Network. NCCN Guidelines: Breast Cancer Screening and Diagnosis. https://www.nccn.org/guidelines/nccn-guidelines/guidelines-detail?category=2&id=1421. Accessed: March 10, 2023.
34. Allison KH, Hammond MEH, Dowsett M, et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP Guideline update. J Clin Oncol. 2020;38(12):1346-66. https://www.doi.org/10.1200/jco.19.02309
35. Zhu X, Chen L, Huang B, et al. The prognostic and predictive potential of Ki-67 in triple-negative breast cancer. Sci Rep. 2020;10(1):225. https://doi.org/10.1038/s41598-019-57094-3
36. Plevritis SK, Munoz D, Kurian AW, et al. Association of Screening and Treatment With Breast Cancer Mortality by Molecular Subtype in US Women, 2000-2012. JAMA. 2018;319(2):154-64. 10.1001/jama.2017.19130
37. Surveillance Epidemiology and End Results Program. Cancer Stat Facts: Female Breast Cancer Subtypes. https://seer.cancer.gov/statfacts/html/breast.html. Accessed: March 3, 2023.
38. Howlader N, Cronin KA, Kurian AW, et al. Differences in Breast Cancer Survival by Molecular Subtypes in the United States. Cancer Epidemiol Biomarkers Prev. 2018;27(6):619-26. https://www.doi.org/10.1158/1055-9965.Epi-17-0627
39. Evans A, Pinder S, Wilson R, et al. Ductal carcinoma in situ of the breast: correlation between mammographic and pathologic findings. AJR Am J Roentgenol. 1994;162(6):1307-11. https://doi.org/10.2214/ajr.162.6.8191988
40. Neal CH, Joe AI, Patterson SK, et al. Digital Mammography Has Persistently Increased High-Grade and Overall DCIS Detection Without Altering Upgrade Rate. AJR Am J Roentgenol. 2021;216(4):912-8. https://dx.doi.org/10.2214/AJR.20.23314
41. Erbas B, Provenzano E, Armes J, et al. The natural history of ductal carcinoma in situ of the breast: a review. Breast Cancer Res Treat. 2006;97(2):135-44. https://www.doi.org/10.1007/s10549-005-9101-z
42. Chootipongchaivat S, van Ravesteyn NT, Li X, et al. Modeling the natural history of ductal carcinoma in situ based on population data. Breast Cancer Res. 2020;22(1):53. https://dx.doi.org/10.1186/s13058-020-01287-6
43. Gorringe KL, Fox SB. Ductal carcinoma in situ biology, biomarkers, and diagnosis. Front Oncol. 2017;7:248. https://doi.org/10.3389/fonc.2017.00248
44. Giannakeas V, Sopik V, Narod SA. Association of a Diagnosis of Ductal Carcinoma In Situ With Death From Breast Cancer. JAMA Netw Open. 2020;3(9):e2017124. https://www.doi.org/10.1001/jamanetworkopen.2020.17124
45. Hwang ES, Miglioretti DL, Ballard-Barbash R, et al. Association between breast density and subsequent breast cancer following treatment for ductal carcinoma in situ. Cancer Epidemiol Biomarkers Prev. 2007;16(12):2587-93. https://www.doi.org/10.1158/1055-9965.Epi-07-0458
46. Allegra CJ, Aberle DR, Ganschow P, et al. NIH state-of-the-science conference statement: diagnosis and management of ductal carcinoma in situ (DCIS). NIH Consens State Sci Statements. 2009;26(2):1-27. https://consensus.nih.gov/2009/dcisstatement.htm
47. Visser LL, Groen EJ, van Leeuwen FE, et al. Predictors of an invasive breast cancer recurrence after DCIS: a systematic review and meta-analyses. Cancer Epidemiol Biomarkers Prev. 2019;28(5):835-45. https://doi.org/10.1158/1055-9965.Epi-18-0976
48. Nelson HD, Pappas M, Cantor A, et al. Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer in Women: Updated Evidence Report and Systematic

Review for the US Preventive Services Task Force. JAMA. 2019;322(7):666-85. https://doi.org/10.1001/jama.2019.8430
49. Owens DK, Davidson KW, Krist AH, et al. Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2019;322(7):652-65. https://doi.org/10.1001/jama.2019.10987
50. Nelson HD, Zakher B, Cantor A, et al. Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. Ann Intern Med. 2012;156(9):635-48. https://doi.org/10.7326/0003-4819-156-9-201205010-00006
51. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet. 2001;358(9291):1389-99. https://www.doi.org/10.1016/s0140-6736(01)06524-2
52. Braithwaite D, Miglioretti DL, Zhu W, et al. Family History and Breast Cancer Risk Among Older Women in the Breast Cancer Surveillance Consortium Cohort. JAMA internal medicine. 2018;178(4):494-501. https://www.doi.org/10.1001/jamainternmed.2017.8642
53. 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;15(6):1159-69. https://doi.org/10.1158/1055-9965.Epi-06-0034
54. National Cancer Institute. Breast Cancer Prevention (PDQ®)–Health Professional Version 06/10/2022. https://www.cancer.gov/types/breast/hp/breast-screening-pdq#section/all. Accessed: August 27, 2022.
55. D'Orsi C, Sickles E, Mendelson E, et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology; 2013.
56. Melnikow J, Fenton J, DL M, et al. Screening for Breast Cancer with Digital Breast Tomosynthesis Rockville, MD: Agency for Healthcare Research and Quality; 2016.
57. Sprague BL, Gangnon RE, Burt V, et al. Prevalence of mammographically dense breasts in the United States. Journal of the National Cancer Institute. 2014;106(10). https://doi.org/10.1093/jnci/dju255
58. Advani SM, Zhu W, Demb J, et al. Association of breast density with breast cancer risk among women aged 65 years or older by age group and body mass index. JAMA Netw Open. 2021;4(8):e2122810-e. https://www.doi.org/10.1001/jamanetworkopen.2021.22810
59. Bissell MCS, Kerlikowske K, Sprague BL, et al. Breast cancer population attributable risk proportions associated with body mass index and breast density by race/ethnicity and menopausal status. Cancer Epidemiol Biomarkers Prev. 2020;29(10):2048-56. https://www.doi.org/10.1158/1055-9965.Epi-20-0358
60. Mokhtary A, Karakatsanis A, Valachis A. Mammographic density changes over time and breast cancer risk: a systematic review and meta-analysis. Cancers (Basel). 2021;13(19):4805. https://doi.org/10.3390/cancers13194805
61. Burton A, Maskarinec G, Perez-Gomez B, et al. Mammographic density and ageing: A collaborative pooled analysis of cross-sectional data from 22 countries worldwide. PLoS Med. 2017;14(6):e1002335. https://www.doi.org/10.1371/journal.pmed.1002335
62. National Cancer Institute. Dense Breasts: Answers to Commonly Asked Questions. https://www.cancer.gov/types/breast/breast-changes/dense-breasts. Accessed: September 25, 2020.

63. Kerlikowske K, Sprague BL, Tosteson ANA, et al. Strategies to Identify Women at High Risk of Advanced Breast Cancer During Routine Screening for Discussion of Supplemental Imaging. JAMA internal medicine. 2019;[Epub ahead of print]. PMID: 31260054 https://doi.org/10.1001/jamainternmed.2019.1758
64. Sprague BL, Kerlikowske K, Bowles EJA, et al. Trends in Clinical Breast Density Assessment From the Breast Cancer Surveillance Consortium. Journal of the National Cancer Institute. 2019. PMID: 30624682. https://doi.org/10.1093/jnci/djy210
65. Azam S, Sjölander A, Eriksson M, et al. Determinants of Mammographic Density Change. JNCI Cancer Spectr. 2019;3(1):pkz004-pkz. https://www.doi.org/10.1093/jncics/pkz004
66. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. Journal of the National Cancer Institute. 1989;81(24):1879-86. https://doi.org/10.1093/jnci/81.24.1879
67. Claus EB, Risch N, Thompson WD. The calculation of breast cancer risk for women with a first degree family history of ovarian cancer. Breast Cancer Res Treat. 1993;28(2):115-20. https://doi.org/10.1007/bf00666424
68. Tice JA, Cummings SR, Smith-Bindman R, et al. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med. 2008;148(5):337-47. https://doi.org/10.7326/0003-4819-148-5-200803040-00004
69. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction modelincorporating genetic and nongenetic risk factors. Genet Med. 2019;21(8):1708-18. 10.1038/s41436-018-0406-9
70. Euhus DM, Smith KC, Robinson L, et al. Pretest prediction of BRCA1 or BRCA2 mutation by risk counselors and the computer model BRCAPRO. Journal of the National Cancer Institute. 2002;94(11):844-51. https://doi.org/10.1093/jnci/94.11.844
71. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA Oncol. 2016;2(10):1295-302. https://doi.org/10.1001/jamaoncol.2016.1025
72. Nelson HD, Fu R, Zakher B, et al. Medication Use for the Risk Reduction of Primary Breast Cancer in Women: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA. 2019;322(9):868-86. https://doi.org/10.1001/jama.2019.5780
73. Louro J, Posso M, Hilton Boon M, et al. A systematic review and quality assessment of individualised breast cancer risk prediction models. Br J Cancer. 2019;121(1):76-85. https://doi.org/10.1038/s41416-019-0476-8
74. Yanes T, Young M-A, Meiser B, et al. Clinical applications of polygenic breast cancer risk: a critical review and perspectives of an emerging field. Breast Cancer Res. 2020;22(1):21. https://doi.org/10.1186/s13058-020-01260-3
75. Nelson HD, Tyne K, Naik A, et al. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Screening for Breast Cancer: Systematic Evidence Review Update for the US Preventive Services Task Force. Rockville (MD): Agency for Healthcare Research and Quality (US); 2009.
76. Nelson HD, Fu R, Cantor A, et al. Effectiveness of Breast Cancer Screening: Systematic Review and Meta-analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med. 2016;164(4):244-55. https://dx.doi.org/10.7326/M15-0969

77. Duffy SW, Vulkan D, Cuckle H, et al. Effect of mammographic screening from age 40 years on breast cancer mortality (UK Age trial): final results of a randomised, controlled trial. Lancel Oncol. 2020;21(9):1165-72. https://dx.doi.org/10.1016/S1470-2045(20)30398-3
78. Duffy S, Vulkan D, Cuckle H, et al. Annual mammographic screening to reduce breast cancer mortality in women from age 40 years: long-term follow-up of the UK Age RCT. Health Technol Assess. 2020;24(55):1-24. https://dx.doi.org/10.3310/hta24550
79. 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
80. Johnson K, Zackrisson S, Rosso A, et al. Tumor Characteristics and Molecular Subtypes in Breast Cancer Screening with Digital Breast Tomosynthesis: the Malmö Breast Tomosynthesis Screening Trial. Radiology. 2019;293(2):273-81. https://doi.org/10.1148/radiol.2019190132
81. Canelo-Aybar C, Posso M, Montero N, et al. Benefits and harms of annual, biennial, or triennial breast cancer mammography screening for women at average risk of breast cancer: a systematic review for the European Commission Initiative on Breast Cancer (ECIBC). Br J Cancer. 2022;126(4):673-88. https://www.doi.org/10.1038/s41416-021-01521-8
82. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancel Oncol. 2013;14(7):583-9. https://doi.org/10.1016/s1470-2045(13)70134-7
83. Skaane P, Bandos AI, Eben EB, et al. Two-View Digital Breast Tomosynthesis Screening with Synthetically Reconstructed Projection Images: Comparison with Digital Breast Tomosynthesis with Full-Field Digital Mammographic Images. Radiology. 2014;271(3):655-63. https://doi.org/10.1148/radiol.13131391
84. Durand MA. Synthesized Mammography: Clinical Evidence, Appearance, and Implementation. Diagnostics (Basel). 2018;8(2):04. https://dx.doi.org/10.3390/diagnostics8020022
85. Abdullah P, Alabousi M, Ramadan S, et al. Synthetic 2D Mammography Versus Standard 2D Digital Mammography: A Diagnostic Test Accuracy Systematic Review and Meta-Analysis. AJR Am J Roentgenol. 2021;217(2):314-25. https://www.doi.org/10.2214/ajr.20.24204
86. Alabousi M, Wadera A, Kashif Al-Ghita M, et al. Performance of Digital Breast Tomosynthesis, Synthetic Mammography, and Digital Mammography in Breast Cancer Screening: A Systematic Review and Meta-Analysis. J Natl Cancer Cent. 2021;113(6):680-90. https://dx.doi.org/10.1093/jnci/djaa205
87. Welch HG, Prorok PC, O'Malley AJ, et al. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. N Engl J Med. 2016;375(15):1438-47. https://doi.org/10.1056/nejmoa1600249
88. Davies L, Petitti DB, Woo M, et al. Defining, Estimating, and Communicating Overdiagnosis in Cancer Screening. Ann Intern Med. 2018;169(11):824. https://dx.doi.org/10.7326/L18-0517
89. Miglioretti DL, Lange J, van den Broek JJ, et al. Radiation-Induced Breast Cancer Incidence and Mortality From Digital Mammography Screening: A Modeling Study. Ann Intern Med. 2016;164(4):205-14. https://dx.doi.org/10.7326/M15-1241
90. Ma AK, Darambara DG, Stewart A, et al. Mean glandular dose estimation using MCNPX for a digital breast tomosynthesis system with tungsten/aluminum and tungsten/aluminum+silver

x-ray anode-filter combinations. Med Phys. 2008;35(12):5278-89. https://doi.org/10.1118/1.3002310
91. Pattacini P, Nitrosi A, Giorgi Rossi P, et al. Digital Mammography versus Digital Mammography Plus Tomosynthesis for Breast Cancer Screening: The Reggio Emilia Tomosynthesis Randomized Trial. Radiology. 2018;288(2):375-85. https://doi.org/10.1148/radiol.2018172119
92. Niklason LT, Christian BT, Niklason LE, et al. Digital tomosynthesis in breast imaging. Radiology. 1997;205(2):399-406. 10.1148/radiology.205.2.9356620
93. Niklason LT, Kopans DB, Hamberg LM. Digital breast imaging: tomosynthesis and digital subtraction mammography. Breast Dis. 1998;10(3-4):151-64. https://doi.org/10.3233/bd-1998-103-415
94. Caumo F, Romanucci G, Hunter K, et al. Comparison of breast cancers detected in the Verona screening program following transition to digital breast tomosynthesis screening with cancers detected at digital mammography screening. Breast Cancer Res Treat. 2018;20:20. https://dx.doi.org/10.1007/s10549-018-4756-4
95. Melnikow J, Fenton JJ, Whitlock EP, et al. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the U.S. Preventive Service Task Force. Rockville (MD): Agency for Healthcare Research and Quality (US); January 2016.
96. Yankaskas BC, Klabunde CN, Ancelle-Park R, et al. International comparison of performance measures for screening mammography: can it be done? J Med Screen. 2004;11(4):187-93. https://doi.org/10.1258/0969141042467430
97. Thurfjell EL, Lernevall KA, Taube AA. Benefit of independent double reading in a population-based mammography screening program. Radiology. 1994;191(1):241-4. https://doi.org/10.1148/radiology.191.1.8134580
98. Azavedo E, Zackrisson S, Mejàre I, et al. Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. BMC Med Imaging. 2012;12:22. https://www.doi.org/10.1186/1471-2342-12-22
99. Taylor P, Potts HW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. Eur J Cancer. 2008;44(6):798-807. https://doi.org/10.1016/j.ejca.2008.02.016
100. Bennett RL, Blanks RG, Moss SM. Does the accuracy of single reading with CAD (computer-aided detection) compare with that of double reading?: A review of the literature. Clin Radiol. 2006;61(12):1023-8. https://www.doi.org/10.1016/j.crad.2006.09.006
101. Richman IB, Hoag JR, Xu X, et al. Adoption of Digital Breast Tomosynthesis in Clinical Practice. JAMA internal medicine. 2019;179(9):1292-5. https://doi.org/10.1001/jamainternmed.2019.1058
102. Mammography Quality Standards Act and Program (MQSA). MQSA National Statistics. https://www.fda.gov/radiation-emitting-products/mqsa-insights/mqsa-national-statistics. Accessed: December 10, 2020.
103. Waks AG, Winer EP. Breast cancer treatment: a review. JAMA. 2019;321(3):288-300. https://doi.org/10.1001/jama.2018.19323
104. Czajka M, Pfeifer C. Breast Cancer Surgery. Treasure Island (FL): StatPearls Publishing; [cited NA - Report Only (MK). Available from: https://www.ncbi.nlm.nih.gov/books/NBK553076/.

105. American Cancer Society. Breast Biopsy. https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/breast-biopsy.html. Accessed: October 27, 2020.
106. Amin MB, Edge SB, Greene F, et al. AJCC cancer staging manual, 8th Edition. 2017.

107. Giuliano AE, Connolly JL, Edge SB, et al. Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin. 2017;67(4):290-303. https://doi.org/10.3322/caac.21393
108. Crane-Okada R, Wascher RA, Elashoff D, et al. Long-term morbidity of sentinel node biopsy versus complete axillary dissection for unilateral breast cancer. Ann Surg Oncol. 2008;15(7):1996-2005. https://doi.org/10.1245/s10434-008-9909-y
109. Chen JC, Li Y, Fisher JL, et al. Neighborhood socioeconomic status and low-value breast cancer care. J Surg Oncol. 2022;126(3):433-42. https://doi.org/10.1002/jso.26901
110. Nayyar A, Strassle PD, Schlottmann F, et al. Disparities in the Use of Sentinel Lymph Node Dissection for Early Stage Breast Cancer. J Surg Res. 2020;254:31-40. https://dx.doi.org/10.1016/j.jss.2020.03.063
111. Nelson HD, Cantor A, Humphrey L, et al. Screening for Breast Cancer: A Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. 2016.
112. U.S. Preventive Services Task Force. U.S. Preventive Services Task Force Procedure Manual. Rockville, MD: U.S. Preventive Services Task Force; 2015.
113. Caughey AB, Krist AH, Wolff TA, et al. USPSTF Approach to Addressing Sex and Gender When Making Recommendations for Clinical Preventive Services. JAMA. 2021;326(19):1953-61. https://www.doi.org/10.1001/jama.2021.15731
114. United Nations Development Programme. 2020 Statistical Update: Human Development Indices and Indicators. New York: 2020.
115. Hellquist BN, Duffy SW, Abdsaleh S, et al. Effectiveness of population-based service screening with mammography for women ages 40 to 49 years: evaluation of the Swedish Mammography Screening in Young Women (SCRY) cohort. Cancer. 2011;117(4):714-22. https://doi.org/10.1002/cncr.25650
116. Mao Z, Nystrom L, Jonsson H. Effectiveness of Population-Based Service Screening with Mammography for Women Aged 70-74 Years in Sweden. Cancer Epidemiol Biomarkers Prev. 2020;29(11):2149-56. https://dx.doi.org/10.1158/1055-9965.EPI-20-0523
117. Seely JM, Peddle SE, Yang H, et al. Breast Density and Risk of Interval Cancers: The Effect of Annual Versus Biennial Screening Mammography Policies in Canada. Can Assoc Radiol J. 2021:8465371211027958. https://dx.doi.org/10.1177/08465371211027958
118. Castellano CR, Aguilar Angulo PM, Hernandez LC, et al. Breast cancer mortality after eight years of an improved screening program using digital breast tomosynthesis. J Med Screen. 2021:9691413211002556. https://dx.doi.org/10.1177/09691413211002556
119. Canelo-Aybar C, Carrera L, Beltran J, et al. Digital breast tomosynthesis compared to diagnostic mammographic projections (including magnification) among women recalled at screening mammography: a systematic review for the European Commission Initiative on Breast Cancer (ECIBC). Cancer Med. 2021;10(7):2191-204. https://dx.doi.org/10.1002/cam4.3803
120. Ryser MD, Lange J, Inoue LYT, et al. Estimation of Breast Cancer Overdiagnosis in a U.S. Breast Screening Cohort. Ann Intern Med. 2022;175(4):471-8. https://doi.org/10.7326/m21-3577
121. Dunn BK, Woloshin S, Xie H, et al. Cancer overdiagnosis: A challenge in the era of screening. J Natl Cancer Cent. 2022;2(4):235-42. https://doi.org/10.1016/j.jncc.2022.08.005

122. van Ravesteyn NT, van den Broek JJ, Li X, et al. Modeling Ductal Carcinoma In Situ (DCIS): An Overview of CISNET Model Approaches. Med Decis Making. 2018;38(1_suppl):126S-39S. https://dx.doi.org/10.1177/0272989X17729358
123. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383-94. 10.1016/j.jclinepi.2010.04.026
124. 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
125. Aase HS, Danielsen AS, Hoff SR, et al. Mammographic features and screening outcome in a randomized controlled trial comparing digital breast tomosynthesis and digital mammography. Eur J Radiol. 2021;141:109753. https://dx.doi.org/10.1016/j.ejrad.2021.109753
126. Aase HS, Holen AS, Pedersen K, et al. A randomized controlled trial of digital breast tomosynthesis versus digital mammography in population-based screening in Bergen: interim analysis of performance indicators from the To-Be trial. Eur Radiol. 2019;29(3):1175-86. https://dx.doi.org/10.1007/s00330-018-5690-x
127. 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
128. Bakker MF, de Lange SV, Pijnappel RM, et al. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. N Engl J Med. 2019;381(22):2091-102. https://dx.doi.org/10.1056/NEJMoa1903986
129. Braithwaite D, Zhu W, Hubbard RA, et al. Screening outcomes in older US women undergoing multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor characteristics or false positive rates? Journal of the National Cancer Institute. 2013;105(5):334-41. https://www.doi.org/10.1093/jnci/djs645
130. 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
131. de Lange SV, Bakker MF, Monninkhof EM, et al. Reasons for (non)participation in supplemental population-based MRI breast screening for women with extremely dense breasts. Clin Radiol. 2018;73(8):759.e1-.e9. https://dx.doi.org/10.1016/j.crad.2018.04.002
132. Dittus K, Geller B, Weaver DL, et al. Impact of mammography screening interval on breast cancer diagnosis by menopausal status and BMI. J Gen Intern Med. 2013;28(11):1454-62. https://doi.org/10.1007/s11606-013-2507-0
133. 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
134. 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

135. Goel A, Littenberg B, Burack RC. The association between the pre-diagnosis mammography screening interval and advanced breast cancer. Breast Cancer Res Treat. 2007;102(3):339-45. https://doi.org/10.1007/s10549-006-9334-5
136. Harada-Shoji N, Suzuki A, Ishida T, et al. Evaluation of Adjunctive Ultrasonography for Breast Cancer Detection Among Women Aged 40-49 Years With Varying Breast Density Undergoing Screening Mammography: A Secondary Analysis of a Randomized Clinical Trial. JAMA Netw Open. 2021;4(8):e2121505. https://dx.doi.org/10.1001/jamanetworkopen.2021.21505
137. 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
138. 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
139. Hofvind S, Holen AS, Aase HS, et al. Two-view digital breast tomosynthesis versus digital mammography in a population-based breast cancer screening programme (To-Be): a randomised, controlled trial. Lancel Oncol. 2019;20(6):795-805. https://dx.doi.org/10.1016/S1470-2045(19)30161-5
140. Hofvind S, Hovda T, Holen AS, et al. Digital Breast Tomosynthesis and Synthetic 2D Mammography versus Digital Mammography: Evaluation in a Population-based Screening Program. Radiology. 2018;287(3):787-94. https://dx.doi.org/10.1148/radiol.2018171361
141. 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
142. 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
143. Hubbard RA, Kerlikowske K, Flowers CI, et al. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med. 2011;155(8):481-92. 10.7326/0003-4819-155-8-201110180-00004
144. Ishida T, Suzuki A, Kawai M, et al. A randomized controlled trial to verify the efficacy of the use of ultrasonography in breast cancer screening aged 40-49 (J-START): 76 196 women registered. Jpn J Clin Oncol. 2014;44(2):134-40. https://dx.doi.org/10.1093/jjco/hyt199
145. 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
146. Kerlikowske K, Zhu W, Hubbard RA, et al. Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. JAMA internal medicine. 2013;173(9):807-16. https://doi.org/10.1001/jamainternmed.2013.307
147. Klemi PJ, Toikkanen S, Räsänen O, et al. Mammography screening interval and the frequency of interval cancers in a population-based screening. Br J Cancer. 1997;75(5):762-6. https://doi.org/10.1038/bjc.1997.135

148. Lang K, Andersson I, Rosso A, et al. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmo Breast Tomosynthesis Screening Trial, a population-based study. Eur Radiol. 2016;26(1):184-90. https://dx.doi.org/10.1007/s00330-015-3803-3
149. Lang K, Nergarden M, Andersson I, et al. False positives in breast cancer screening with one-view breast tomosynthesis: An analysis of findings leading to recall, work-up and biopsy rates in the Malmo Breast Tomosynthesis Screening Trial. Eur Radiol. 2016;26(11):3899-907. https://doi.org/10.1007/s00330-016-4265-y
150. 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
151. 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
152. 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
153. Moger TA, Swanson JO, Holen AS, et al. Cost differences between digital tomosynthesis and standard digital mammography in a breast cancer screening programme: results from the To-Be trial in Norway. Eur J Health Econ. 2019;20(8):1261-9. https://dx.doi.org/10.1007/s10198-019-01094-7
154. Moshina N, Aase HS, Danielsen AS, et al. Comparing Screening Outcomes for Digital Breast Tomosynthesis and Digital Mammography by Automated Breast Density in a Randomized Controlled Trial: Results from the To-Be Trial. Radiology. 2020;297(3):522-31. https://dx.doi.org/10.1148/radiol.2020201150
155. Nelson HD, O'Meara ES, Kerlikowske K, et al. Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data. Ann Intern Med. 2016;164(4):226-35. https://dx.doi.org/10.7326/M15-0971
156. 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
157. 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
158. 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
159. Pattacini P, Nitrosi A, Giorgi Rossi P, et al. Digital Mammography versus Digital Mammography Plus Tomosynthesis for Breast Cancer Screening: The Reggio Emilia Tomosynthesis Randomized Trial. Radiology. 2018;288(2):375-85. https://dx.doi.org/10.1148/radiol.2018172119

160. 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
161. Rosso A, Lang K, Petersson IF, et al. Factors affecting recall rate and false positive fraction in breast cancer screening with breast tomosynthesis - A statistical approach. Breast. 2015;24(5):680-6. https://dx.doi.org/10.1016/j.breast.2015.08.007
162. 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
163. Waade G, Holen Å, Sebuødegård S, et al. Breast compression parameters among women screened with standard digital mammography and digital breast tomosynthesis in a randomized controlled trial. Acta Radiol. 2020;61(3):321-30. https://www.doi.org/10.1177/0284185119863989
164. White E, Miglioretti DL, Yankaskas BC, et al. Biennial versus annual mammography and the risk of late-stage breast cancer. Journal of the National Cancer Institute. 2004;96(24):1832-9. https://doi.org/10.1093/jnci/djh337
165. Zackrisson S, Lang K, Rosso A, et al. One-view breast tomosynthesis versus two-view mammography in the Malmo Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. Lancel Oncol. 2018;19(11):1493-503. https://dx.doi.org/10.1016/S1470-2045(18)30521-7
166. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64. https://doi.org/10.1093/aje/kwv254
167. Acerbi I, Fiscalini A, Che M, et al. Personalized breast cancer screening in a population-based study: women informed to screen depending on measures of risk (WISDOM). Cancer Res. 2021;81(4 suppl). https://www.doi.org/10.1158/1538-7445.SABCS20-OT-21-01
168. Doubovetzky J, Bour C. Use of screening research to boost overdiagnosis: the mypebs trial. BMJ Evid Based Med. 2019;24:A43-. https://www.doi.org/10.1136/bmjebm-2019-POD.88
169. García-Albéniz X, Hsu J, Bretthauer M, et al. Effectiveness of Screening Colonoscopy to Prevent Colorectal Cancer Among Medicare Beneficiaries Aged 70 to 79 Years: A Prospective Observational Study. Ann Intern Med. 2017;166(1):18-26. https://doi.org/10.7326/m16-0758
170. Cook AJ, Elmore JG, Miglioretti DL, et al. Decreased accuracy in interpretation of community-based screening mammography for women with multiple clinical risk factors. J Clin Epidemiol. 2010;63(4):441-51. https://doi.org/10.1016/j.jclinepi.2009.06.008
171. Nelson HD. Mammography Screening and Overdiagnosis. JAMA Oncol. 2016;2(2):261-2. https://dx.doi.org/10.1001/jamaoncol.2015.4096
172. Heywang-Köbrunner S-H, Jänsch A, Hacker A, et al. Tomosynthesis with synthesised two-dimensional mammography yields higher cancer detection compared to digital mammography alone, also in dense breasts and in younger women: A systematic review and meta-analysis. Eur J Radiol. 2022;152. https://doi.org/10.1016/j.ejrad.2022.110324
173. Libesman S, Zackrisson S, Hofvind S, et al. An individual participant data meta-analysis of breast cancer detection and recall rates for digital breast tomosynthesis versus digital mammography population screening. Clin Breast Cancer. 2022;22(5):e647-e54. https://doi.org/10.1016/j.clbc.2022.02.005
174. Giampietro RR, Cabral MVG, Lima SAM, et al. Accuracy and Effectiveness of Mammography versus Mammography and Tomosynthesis for Population-Based Breast Cancer

Screening: A Systematic Review and Meta-Analysis. Sci Rep. 2020;10(1):7991. https://doi.org/10.1038/s41598-020-64802-x
175. Lowry KP, Coley RY, Miglioretti DL, et al. Screening Performance of Digital Breast Tomosynthesis vs Digital Mammography in Community Practice by Patient Age, Screening Round, and Breast Density. JAMA Netw Open. 2020;3(7):e2011792-e. https://doi.org/10.1001/jamanetworkopen.2020.11792
176. Mizzi D, Allely C, Zarb F, et al. Examining the effectiveness of supplementary imaging modalities for breast cancer screening in women with dense breasts: A systematic review and meta-analysis. Eur J Radiol. 2022;154:110416. https://doi.org/10.1016/j.ejrad.2022.110416
177. Yuan WH, Hsu HC, Chen YY, et al. Supplemental breast cancer-screening ultrasonography in women with dense breasts: a systematic review and meta-analysis. Br J Cancer. 2020;123(4):673-88. https://doi.org/10.1038/s41416-020-0928-1
178. Jemal A, Ward EM, Johnson CJ, et al. Annual Report to the Nation on the Status of Cancer, 1975-2014, Featuring Survival. Journal of the National Cancer Institute. 2017;109(9). https://doi.org/10.1093/jnci/djx030
179. Lewis-Thames MW, Langston ME, Khan S, et al. Racial and Ethnic Differences in Rural-Urban Trends in 5-Year Survival of Patients With Lung, Prostate, Breast, and Colorectal Cancers: 1975-2011 Surveillance, Epidemiology, and End Results (SEER). JAMA Netw Open. 2022;5(5):e2212246-e. https://doi.org/10.1001/jamanetworkopen.2022.12246
180. Alvidrez J, Castille D, Laude-Sharp M, et al. The National Institute on Minority Health and Health Disparities Research Framework. Am J Public Health. 2019;109(S1):S16-s20. https://www.doi.org/10.2105/ajph.2018.304883
181. Nong P, Raj M, Creary M, et al. Patient-Reported Experiences of Discrimination in the US Health Care System. JAMA Netw Open. 2020;3(12):e2029650. https://doi.org/10.1001/jamanetworkopen.2020.29650
182. Nelson HD, Cantor A, Wagner J, et al. Effectiveness of Patient Navigation to Increase Cancer Screening in Populations Adversely Affected by Health Disparities: a Meta-analysis. J Gen Intern Med. 2020;35(10):3026-35. https://doi.org/10.1007/s11606-020-06020-9
183. Williams DR, Priest N, Anderson NB. Understanding associations among race, socioeconomic status, and health: Patterns and prospects. Health Psychol. 2016;35(4):407-11. 10.1037/hea0000242
184. Bailey ZD, Krieger N, Agénor M, et al. Structural racism and health inequities in the USA: evidence and interventions. Lancet. 2017;389(10077):1453-63. https://www.doi.org/10.1016/s0140-6736(17)30569-x
185. Zavala VA, Bracci PM, Carethers JM, et al. Cancer health disparities in racial/ethnic minorities in the United States. Br J Cancer. 2020;124(2):315-32. https://www.doi.org/10.1038/s41416-020-01038-6
186. Bemanian A, Beyer KM. Measures Matter: The Local Exposure/Isolation (LEx/Is) Metrics and Relationships between Local-Level Segregation and Breast Cancer Survival. Cancer Epidemiol Biomarkers Prev. 2017;26(4):516-24. https://www.doi.org/10.1158/1055-9965.Epi-16-0926
187. Collin LJ, Gaglioti AH, Beyer KM, et al. Neighborhood-Level Redlining and Lending Bias Are Associated with Breast Cancer Mortality in a Large and Diverse Metropolitan Area. Cancer Epidemiol Biomarkers Prev. 2021;30(1):53-60. https://doi.org/10.1158/1055-9965.Epi-20-1038

188. Goel N, Westrick AC, Bailey ZD, et al. Structural Racism and Breast Cancer-specific Survival: Impact of Economic and Racial Residential Segregation. Ann Surg. 2022;275(4):776-83. https://doi.org/10.1097/sla.0000000000005375
189. Siegel SD, Brooks MM, Lynch SM, et al. Racial disparities in triple negative breast cancer: toward a causal architecture approach. Breast Cancer Res. 2022;24(1):37. https://doi.org/10.1186/s13058-022-01533-z
190. Kerlikowske K, Chen S, Golmakani MK, et al. Cumulative advanced breast cancer risk prediction model developed in a screening mammography population. Journal of the National Cancer Institute. 2022;114(5):676-85. https://doi.org/10.1093/jnci/djac008
191. Niraula S, Biswanger N, Hu P, et al. Incidence, Characteristics, and Outcomes of Interval Breast Cancers Compared With Screening-Detected Breast Cancers. JAMA Netw Open. 2020;3(9):e2018179. https://dx.doi.org/10.1001/jamanetworkopen.2020.18179
192. Davis Lynn BC, Chernyavskiy P, Gierach GL, et al. Decreasing Incidence of Estrogen Receptor-Negative Breast Cancer in the United States: Trends by Race and Region. Journal of the National Cancer Institute. 2022;114(2):263-70. https://doi.org/10.1093/jnci/djab186
193. Nelson HD, Cantor A, Wagner J, et al. Achieving Health Equity in Preventive Services: A Systematic Review for a National Institutes of Health Pathways to Prevention Workshop. Ann Intern Med. 2020;172(4):258-71. https://doi.org/10.7326/m19-3199
194. Centers for Disease Control and Prevention. BRFSS Web Enabled Analysis Tool. https://nccd.cdc.gov/weat/#/analysis. Accessed: August 29, 2022.
195. Gorina Y, Elgaddal N. Patterns of Mammography, Pap Smear, and Colorectal Cancer Screening Services Among Women Aged 45 and Over. Natl Health Stat Report. 2021(157):1-18.
196. Lee CI, Lehman CD. Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. J Am Coll Radiol. 2013;10(12):913-7. https://doi.org/10.1016/j.jacr.2013.09.010
197. Lee CI, Zhu W, Onega T, et al. Comparative Access to and Use of Digital Breast Tomosynthesis Screening by Women's Race/Ethnicity and Socioeconomic Status. JAMA Netw Open. 2021;4(2):e2037546. https://doi.org/10.1001/jamanetworkopen.2020.37546
198. Clark CR, Tosteson TD, Tosteson ANA, et al. Diffusion of digital breast tomosynthesis among women in primary care: associations with insurance type. Cancer Med. 2017;6(5):1102-7. https://dx.doi.org/10.1002/cam4.1036
199. Schifferdecker KE, Tosteson ANA, Kaplan C, et al. Knowledge and Perception of Breast Density, Screening Mammography, and Supplemental Screening: in Search of "Informed". J Gen Intern Med. 2020;35(6):1654-60. https://dx.doi.org/10.1007/s11606-019-05560-z
200. Warnecke RB, Campbell RT, Vijayasiri G, et al. Multilevel Examination of Health Disparity: The Role of Policy Implementation in Neighborhood Context, in Patient Resources, and in Healthcare Facilities on Later Stage of Breast Cancer Diagnosis. Cancer Epidemiol Biomarkers Prev. 2019;28(1):59-66. https://dx.doi.org/10.1158/1055-9965.EPI-17-0945
201. Onega T, Hubbard R, Hill D, et al. Geographic access to breast imaging for US women. J Am Coll Radiol. 2014;11(9):874-82. https://dx.doi.org/10.1016/j.jacr.2014.03.022
202. Fayanju OM, Ren Y, Stashko I, et al. Patient-reported causes of distress predict disparities in time to evaluation and time to treatment after breast cancer diagnosis. Cancer. 2021;127(5):757-68. https://doi.org/10.1002/cncr.33310
203. Selove R, Kilbourne B, Fadden MK, et al. Time from Screening Mammography to Biopsy and from Biopsy to Breast Cancer Treatment among Black and White, Women Medicare

Beneficiaries Not Participating in a Health Maintenance Organization. Womens Health Issues. 2016;26(6):642-7. https://dx.doi.org/10.1016/j.whi.2016.09.003
204. Nguyen KH, Pasick RJ, Stewart SL, et al. Disparities in abnormal mammogram follow-up time for Asian women compared with non-Hispanic white women and between Asian ethnic groups. Cancer. 2017;123(18):3468-75. https://dx.doi.org/10.1002/cncr.30756
205. Warner ET, Tamimi RM, Hughes ME, et al. Time to diagnosis and breast cancer stage by race/ethnicity. Breast Cancer Res Treat. 2012;136(3):813-21. https://doi.org/10.1007/s10549-012-2304-1
206. Kovar A, Bronsert M, Jaiswal K, et al. The Waiting Game: How Long Are Breast Cancer Patients Waiting for Definitive Diagnosis? Ann Surg Oncol. 2020;27(10):3641-9. https://dx.doi.org/10.1245/s10434-020-08484-9
207. Elmore JG, Nakano CY, Linden HM, et al. Racial inequities in the timing of breast cancer detection, diagnosis, and initiation of treatment. Med Care. 2005;43(2):141-8. https://doi.org/10.1097/00005650-200502000-00007
208. Emerson MA, Golightly YM, Aiello AE, et al. Breast cancer treatment delays by socioeconomic and health care access latent classes in Black and White women. Cancer. 2020;126(22):4957-66. https://www.doi.org/10.1002/cncr.33121
209. Lawson MB, Bissell MCS, Miglioretti DL, et al. Multilevel Factors Associated With Time to Biopsy After Abnormal Screening Mammography Results by Race and Ethnicity. JAMA Oncol. 2022;8(8):1115-26. https://doi.org/10.1001/jamaoncol.2022.1990
210. Yussof I, Mohd Tahir NA, Hatah E, et al. Factors influencing five-year adherence to adjuvant endocrine therapy in breast cancer patients: A systematic review. Breast. 2022;62:22-35. https://doi.org/10.1016/j.breast.2022.01.012
211. Hu X, Walker MS, Stepanski E, et al. Racial Differences in Patient-Reported Symptoms and Adherence to Adjuvant Endocrine Therapy Among Women With Early-Stage, Hormone Receptor-Positive Breast Cancer. JAMA Netw Open. 2022;5(8):e2225485. https://doi.org/10.1001/jamanetworkopen.2022.25485
212. Hu X, Chehal PK, Kaplan C, et al. Characterization of Clinical Symptoms by Race Among Women With Early-Stage, Hormone Receptor-Positive Breast Cancer Before Starting Chemotherapy. JAMA Netw Open. 2021;4(6):e2112076. https://doi.org/10.1001/jamanetworkopen.2021.12076
213. Nystrom L, Bjurstam N, Jonsson H, et al. Reduced breast cancer mortality after 20+ years of follow-up in the Swedish randomized controlled mammography trials in Malmo, Stockholm, and Goteborg. J Med Screen. 2017;24(1):34-42. https://dx.doi.org/10.1177/0969141316648987
214. Tice JA, Miglioretti DL, Li CS, et al. Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer. J Clin Oncol. 2015;33(28):3137-43. https://doi.org/10.1200/jco.2015.60.8869
215. Adams-Campbell LL, Makambi KH, Palmer JR, et al. Diagnostic accuracy of the Gail model in the Black Women's Health Study. Breast J. 2007;13(4):332-6. https://www.doi.org/10.1111/j.1524-4741.2007.00439.x
216. Gail MH, Anderson WF, Garcia-Closas M, et al. Absolute risk models for subtypes of breast cancer. Journal of the National Cancer Institute. 2007;99(22):1657-9. https://doi.org/10.1093/jnci/djm228
217. Matsuno RK, Costantino JP, Ziegler RG, et al. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. Journal of the National Cancer Institute. 2011;103(12):951-61. https://doi.org/10.1093/jnci/djr154

218. Amir E, Evans DG, Shenton A, et al. Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. J Med Genet. 2003;40(11):807-14. https://www.doi.org/10.1136/jmg.40.11.807
219. Chlebowski RT, Anderson GL, Lane DS, et al. Predicting risk of breast cancer in postmenopausal women by hormone receptor status. Journal of the National Cancer Institute. 2007;99(22):1695-705. https://www.doi.org/10.1093/jnci/djm224
220. Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. Journal of the National Cancer Institute. 2006;98(17):1204-14. https://www.doi.org/10.1093/jnci/djj331
221. Vacek PM, Skelly JM, Geller BM. Breast cancer risk assessment in women aged 70 and older. Breast Cancer Res Treat. 2011;130(1):291-9. https://doi.org/10.1007/s10549-011-1576-1
222. Brentnall AR, Harkness EF, Astley SM, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. Breast Cancer Res. 2015;17(1):147. https://dx.doi.org/10.1186/s13058-015-0653-5
223. Tice JA, Bissell MCS, Miglioretti DL, et al. Validation of the breast cancer surveillance consortium model of breast cancer risk. Breast Cancer Res Treat. 2019;175(2):519-23. https://dx.doi.org/10.1007/s10549-019-05167-2
224. Nickson C, Procopio P, Velentzis LS, et al. Prospective validation of the NCI Breast Cancer Risk Assessment Tool (Gail Model) on 40,000 Australian women. Breast Cancer Res. 2018;20(1):155. https://dx.doi.org/10.1186/s13058-018-1084-x
225. MacInnis RJ, Knight JA, Chung WK, et al. Comparing 5-Year and Lifetime Risks of Breast Cancer using the Prospective Family Study Cohort. J Natl Cancer Cent. 2021;113(6):785-91. https://dx.doi.org/10.1093/jnci/djaa178
226. Louro J, Roman M, Posso M, et al. Developing and validating an individualized breast cancer risk prediction model for women attending breast cancer screening. PLoS ONE. 2021;16(3):e0248930. https://dx.doi.org/10.1371/journal.pone.0248930
227. McCarthy AM, Guan Z, Welch M, et al. Performance of Breast Cancer Risk-Assessment Models in a Large Mammography Cohort. J Natl Cancer Cent. 2020;112(5):489-97. https://dx.doi.org/10.1093/jnci/djz177
228. Abdolell M, Payne JI, Caines J, et al. Assessing breast cancer risk within the general screening population: developing a breast cancer risk model to identify higher risk women at mammographic screening. Eur Radiol. 2020;30(10):5417-26. https://dx.doi.org/10.1007/s00330-020-06901-x
229. Eriksson M, Czene K, Strand F, et al. Identification of Women at High Risk of Breast Cancer Who Need Supplemental Screening. Radiology. 2020;297(2):327-33. https://dx.doi.org/10.1148/radiol.2020201620
230. Kakileti ST, Madhu HJ, Manjunath G, et al. Personalized risk prediction for breast cancer pre-screening using artificial intelligence and thermal radiomics. Artif Intell Med. 2020;105:101854. https://dx.doi.org/10.1016/j.artmed.2020.101854
231. Ming C, Viassolo V, Probst-Hensch N, et al. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast Cancer Res [serial on the Internet]. 2019 [cited Search 1 - Cochrane Central Register of Controlled Clinical Trials Dual Screen]; 21(1): Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01958376/full.

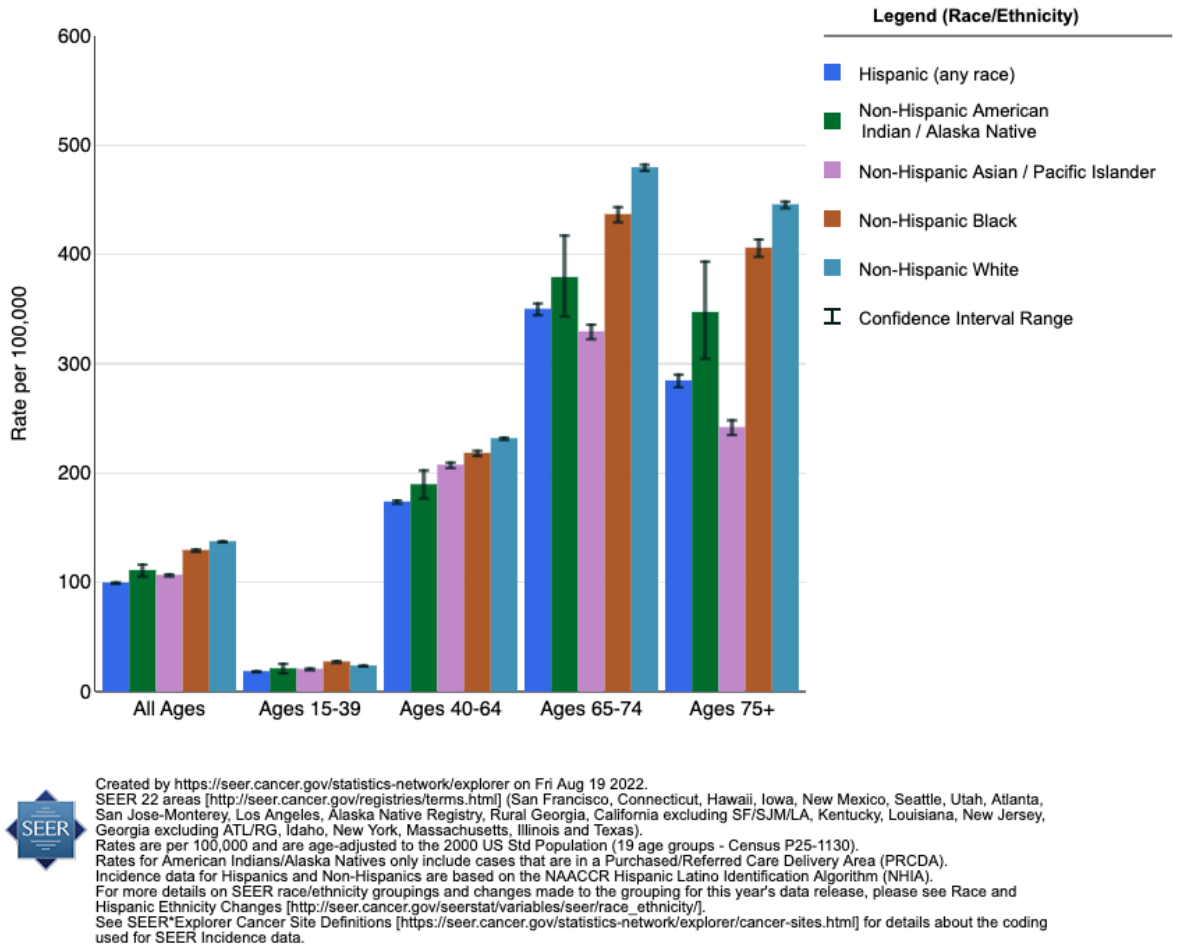
232. Yala A, Mikhael PG, Strand F, et al. Toward robust mammography-based models for breast cancer risk. Sci Transl Med. 2021;13(578):27. https://dx.doi.org/10.1126/scitranslmed.aba4373
233. van Veen EM, Brentnall AR, Byers H, et al. Use of Single-Nucleotide Polymorphisms and Mammographic Density Plus Classic Risk Factors for Breast Cancer Risk Prediction. JAMA Oncol. 2018;4(4):476-82. https://dx.doi.org/10.1001/jamaoncol.2017.4881
234. Lall K, Lepamets M, Palover M, et al. Polygenic prediction of breast cancer: comparison of genetic predictors and implications for risk stratification. BMC Cancer. 2019;19(1):557. https://dx.doi.org/10.1186/s12885-019-5783-1
235. Mathioudakis AG, Salakari M, Pylkkanen L, et al. Systematic review on women's values and preferences concerning breast cancer screening and diagnostic services. Psychooncology. 2019;28(5):939-47. https://dx.doi.org/10.1002/pon.5041
236. Mottram R, Knerr WL, Gallacher D, et al. Factors associated with attendance at screening for breast cancer: a systematic review and meta-analysis. BMJ Open. 2021;11(11):e046660. https://doi.org/10.1136/bmjopen-2020-046660
237. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med. 2007;146(7):502-10. https://www.doi.org/10.7326/0003-4819-146-7-200704030-00006
238. Dabbous FM, Dolecek TA, Berbaum ML, et al. Impact of a False-Positive Screening Mammogram on Subsequent Screening Behavior and Stage at Breast Cancer Diagnosis. Cancer Epidemiol Biomarkers Prev. 2017;26(3):397-403. https://dx.doi.org/10.1158/1055-9965.EPI-16-0524
239. Shi W, Nagler RH, Fowler EF, et al. Predictors of Women's Awareness of the Benefits and Harms of Mammography Screening and Associations with Confusion, Ambivalence, and Information Seeking. Health Commun. 2021;36(3):303-14. https://dx.doi.org/10.1080/10410236.2019.1687129
240. Hoover DS, Pappadis MR, Housten AJ, et al. Preferences for Communicating about Breast Cancer Screening Among Racially/Ethnically Diverse Older Women. Health Commun. 2019;34(7):702-6. https://dx.doi.org/10.1080/10410236.2018.1431026
241. Ro V, Jones T, Silverman T, et al. Patient, primary care provider, and stakeholder perspectives on mammography screening frequency: lessons learned from a qualitative study. BMC Cancer. 2022;22(1):819. https://doi.org/10.1186/s12885-022-09900-x
242. Thomson MD, Siminoff LA. Perspectives on mammography after receipt of secondary screening owing to a false positive. Womens Health Issues. 2015;25(2):128-33. 10.1016/j.whi.2014.11.003
243. Brandzel S, Chang E, Tuzzio L, et al. Latina and Black/African American women's perspectives on cancer screening and cancer screening reminders. J Racial Ethn Health Disparities. 2016. https://www.doi.org/10.1007/s40615-016-0304-2
244. Esmaeili M, Ayyoubzadeh SM, Javanmard Z, et al. A systematic review of decision aids for mammography screening: Focus on outcomes and characteristics. Int J Med Inf. 2021;149:104406. https://dx.doi.org/10.1016/j.ijmedinf.2021.104406
245. Taghian A, Merajver SD. Overview of the treatment of newly diagnosed, invasive, non-metastatic breast cancer. In: Post TW, editor. UpToDate; 2022. PMID.
246. Sharma P. Overview of the approach to metastatic breast cancer. In: Post TW, editor. UpToDate; 2022.
247. Sabel MS. Breast-conserving therapy. In: Post TW, editor. UpToDate; 2022.

248. Kwong A, Sabel MS. Mastectomy. In: Post TW, editor. UpToDate; 2022.
249. Taghian A. Adjuvant radiation therapy for women with newly diagnosed, non-metastatic breast cancer. In: Post TW, editor. UpToDate; 2022.
250. Shapiro CL. Acute side effects of adjuvant chemotherapy for early-stage breast cancer. In: Post TW, editor. UpToDate; 2022..
251. Rivera DR, Ganz PA, Weyrich MS, et al. Chemotherapy-Associated Peripheral Neuropathy in Patients With Early-Stage Breast Cancer: A Systematic Review. Journal of the National Cancer Institute. 2018;110(2). https://doi.org/10.1093/jnci/djx140
252. Elghazawy H, Venkatesulu BP, Verma V, et al. The role of cardio-protective agents in cardio-preservation in breast cancer patients receiving Anthracyclines ± Trastuzumab: a Meta-analysis of clinical studies. Crit Rev Oncol Hematol. 2020;153:103006. https://doi.org/10.1016/j.critrevonc.2020.103006
253. Upshaw JN, Finkelman B, Hubbard RA, et al. Comprehensive Assessment of Changes in Left Ventricular Diastolic Function With Contemporary Breast Cancer Therapy. JACC Cardiovasc Imaging. 2020;13(1 Pt 2):198-210. https://doi.org/10.1016/j.jcmg.2019.07.018
254. Chen DH, Tyebally S, Malloupas M, et al. Cardiovascular Disease Amongst Women Treated for Breast Cancer: Traditional Cytotoxic Chemotherapy, Targeted Therapy, and Radiation Therapy. Curr Cardiol Rep. 2021;23(3):16. https://www.doi.org/10.1007/s11886-021-01446-x
255. Stan D, Loprinzi CL, Ruddy KJ. Breast cancer survivorship issues. Hematol Oncol Clin North Am. 2013;27(4):805-27, ix. https://doi.org/10.1016/j.hoc.2013.05.005
256. Come SE. Overview of long-term complications of therapy in breast cancer survivors and patterns of relapse. In: Post TW, editor. UpToDate; 2022.
257. Mannu GS, Wang Z, Broggio J, et al. Invasive breast cancer and breast cancer mortality after ductal carcinoma in situ in women attending for breast screening in England, 1988-2014: population based observational cohort study. BMJ. 2020;369:m1570. https://dx.doi.org/10.1136/bmj.m1570
258. Narod SA, Iqbal J, Giannakeas V, et al. Breast cancer mortality after a diagnosis of ductal carcinoma in situ. JAMA Oncol. 2015;1(7):888-96. https://doi.org/10.1001/jamaoncol.2015.2510
259. Zhang B, Coopey SB, Gadd MA, et al. Trends in unilateral and contralateral prophylactic mastectomy use in ductal carcinoma in situ of the breast: patterns and predictors. Ann Surg Oncol. 2019;26(12):3863-73. https://doi.org/10.1245/s10434-019-07628-w
260. Bouskill K, Hempel S, Richardson A, et al. Evidence map of ductal carcinoma in situ management options. Menopause. 2019;26(11):1250-8. https://www.doi.org/10.1097/gme.0000000000001397
261. Hwang ES, Hyslop T, Lynch T, et al. The COMET (Comparison of Operative versus Monitoring and Endocrine Therapy) trial: a phase III randomised controlled clinical trial for low-risk ductal carcinoma in situ (DCIS). BMJ Open. 2019;9(3):e026797. https://doi.org/10.1136/bmjopen-2018-026797
262. Francis A, Fallowfield L, Rea D. The LORIS Trial: Addressing overtreatment of ductal carcinoma in situ. Clin Oncol (R Coll Radiol). 2015;27(1):6-8. https://dx.doi.org/10.1016/j.clon.2014.09.015
263. Solin L. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. Breast diseases [serial on the Internet]. 2016. 27(1):

Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01199838/full.
264. Kerlikowske K, Bissell MCS, Sprague BL, et al. Advanced Breast Cancer Definitions by Staging System Examined in the Breast Cancer Surveillance Consortium. J Natl Cancer Cent. 2021;113(7):909-16. https://dx.doi.org/10.1093/jnci/djaa176
265. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601. https://doi.org/10.1136/bmj.k601
266. ECOG-ACRIN Cancer Research Group. EA1151 / TMIST: Digital Tomosynthesis Mammography and Digital Mammography in Screening Patients for Breast Cancer. https://ecog-acrin.org/clinical-trials/ea1151-tmist-breast-cancer-screening-mammography/. Accessed: March 8, 2023.
267. SEER Explorer. Breast Cancer: Recent Trends in SEER Age-Adjusted Incidence Rates, 2000-2019. https://seer.cancer.gov/explorer/application.html. Accessed: February 16, 2023.
268. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA. 2015;314(15):1599-614. https://dx.doi.org/10.1001/jama.2015.12783
269. American Academy of Family Physicians. Clinical Preventive Service Recommendation Web site. https://www.aafp.org/family-physician/patient-care/clinical-recommendations/all-clinical-recommendations/breast-cancer.html. Accessed: Accessed August 26, 2020.
270. Anonymous. Practice Bulletin Number 179: Breast Cancer Risk Assessment and Screening in Average-Risk Women. Obstet Gynecol. 2017;130(1):e1-e16. https://dx.doi.org/10.1097/AOG.0000000000002158
271. Monticciolo DL, Malak SF, Friedewald SM, et al. Breast Cancer Screening Recommendations Inclusive of All Women at Average Risk: Update from the ACR and Society of Breast Imaging. J Am Coll Radiol. 2021;18(9):1280-8. https://dx.doi.org/10.1016/j.jacr.2021.04.021
272. Women's Preventive Services Initiative (WPSI). Breast Cancer Screening for Average-Risk Women. https://www.womenspreventivehealth.org/recommendations/breast-cancer-screening-for-average-risk-women/. Accessed: Accessed December 4, 2020.
273. NHS Breast Cancer Screening Programme. Breast screening: programme overview. https://www.gov.uk/guidance/breast-screening-programme-overview. Accessed: September 3, 2020.
274. European Commission Initiative on Breast Cancer (ECIBC). European guidelines on breast cancer screening and diagnosis. https://healthcare-quality.jrc.ec.europa.eu/ecibc/european-breast-cancer-guidelines. Accessed: August 18, 2022.
275. Klarenbach S, Sims-Jones N, Lewin G, et al. Recommendations on screening for breast cancer in women aged 40-74 years who are not at increased risk for breast cancer. CMAJ. 2018;190(49):E1441-E51. https://dx.doi.org/10.1503/cmaj.180463
276. Cancer Australia. Early detection of breast cancer. https://www.canceraustralia.gov.au/publications-and-resources/position-statements/early-detection-breast-cancer. Accessed: August 31, 2020.
277. Habbema JD, van Oortmarssen GJ, van Putten DJ, et al. Age-specific reduction in breast cancer mortality by screening: an analysis of the results of the Health Insurance Plan of Greater New York study. Journal of the National Cancer Institute. 1986;77(2):317-20.

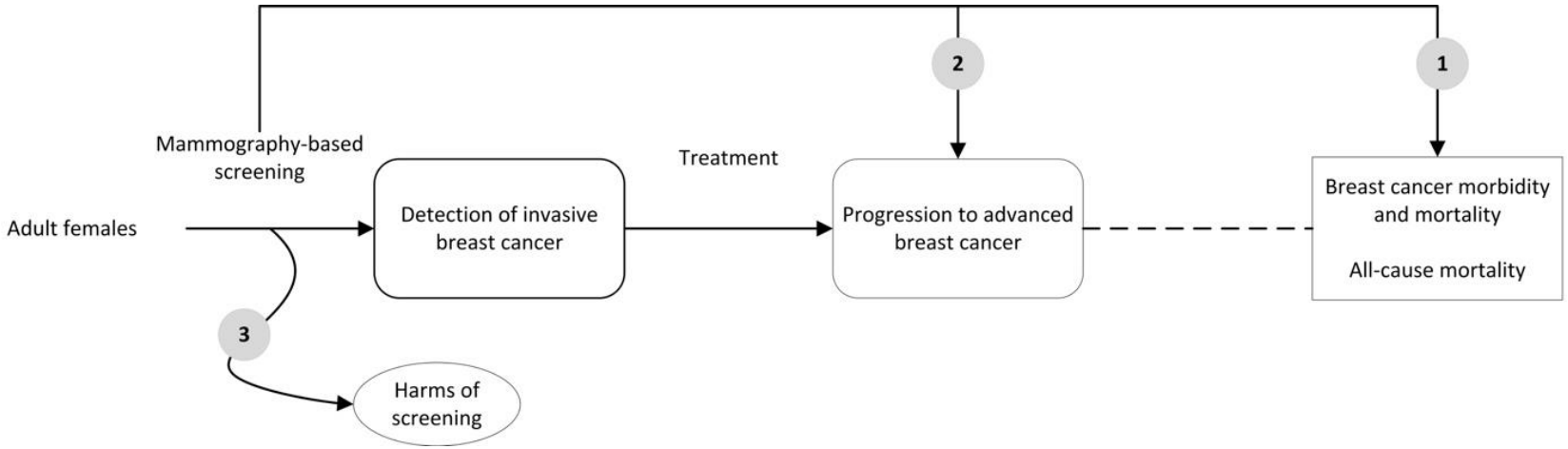
278. Miller A, Wall C, Baines C, et al. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. BMJ 2014; 348: Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-00986506/full.
279. Moss SM, Cuckle H, Evans A, 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;368(9552):2053-60. https://www.doi.org/10.1016/s0140-6736(06)69834-6.
280. Bjurstam N, Björneld L, Warwick J, et al. The Gothenburg Breast Screening Trial. Cancer. 2003;97(10):2387-96. 10.1002/cncr.11361
281. Nyström L, Andersson I, Bjurstam N, et al. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet. 2002;359(9310):909-19. 10.1016/s0140-6736(02)08020-0
282. Tabar L, Fagerberg G, Chen HH, et al. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. Cancer. 1995;75(10):2507-17. 10.1002/1097-0142(19950515)75:10<2507::aid-cncr2820751017>3.0.co;2-h
283. Nelson HD, Pappas M, Cantor A, et al. Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Annals of Internal Medicine. 2016;164(4):256-67. https://dx.doi.org/10.7326/M15-0970
284. Yaffe MJ, Mainprize JG. Risk of radiation-induced breast cancer from mammographic screening. Radiology. 2011;258(1):98-105. 10.1148/radiol.10100655
285. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Bmj. 2016;355:i4919. 10.1136/bmj.i4919
286. Dale A, Slade E, Heneghan C. Adjunctive ultrasonography for breast cancer screening. The Lancet. 2016;387(10036):2379-80. 10.1016/S0140-6736(16)30735-8

Figure 1. Breast Cancer Incidence Rates by Age at Diagnosis, 2015 to 2019, by Race/Ethnicity



Source: SEER Explorer[10]

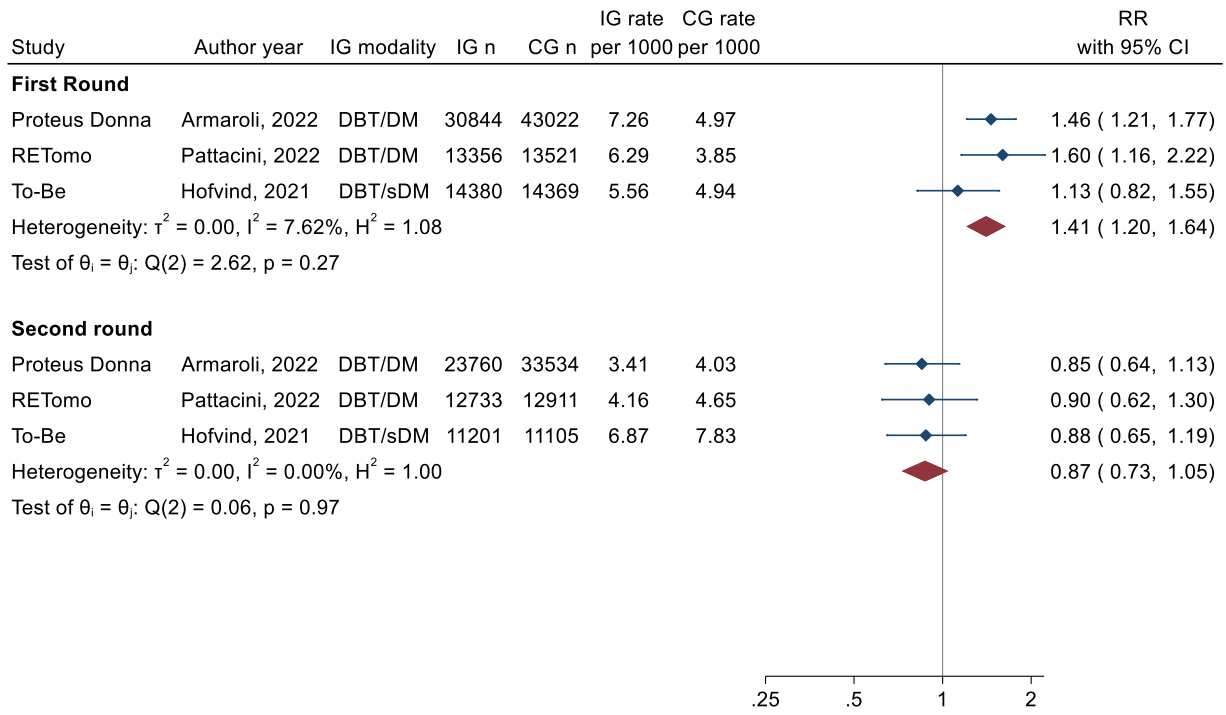
Figure 2. Analytic Framework



Key Questions

1. What is the comparative effectiveness of different mammography-based breast cancer screening strategies (e.g., by modality, interval, initiation age, use of supplemental imaging, or personalization based on risk factors) on breast cancer morbidity and breast cancer–specific or all-cause mortality?
    a. Does comparative effectiveness differ by population characteristics and risk markers (e.g., age, breast density, race/ethnicity, family history)?
2. What is the comparative effectiveness of different mammography-based breast cancer screening strategies (e.g., by modality, interval, initiation age, use of supplemental imaging, or personalization based on risk factors) on the incidence and progression to advanced breast cancer?
    a. Does comparative effectiveness differ by population characteristics and risk markers (e.g., age, breast density, race/ethnicity, family history)?
3. What are the comparative harms of different breast mammography-based cancer screening strategies (e.g., by modality, interval, initiation age, use of supplemental imaging, or personalization based on risk factors)?
    a. Do the comparative harms vary by population characteristics and risk markers (e.g., age, breast density, race/ethnicity, family history)?

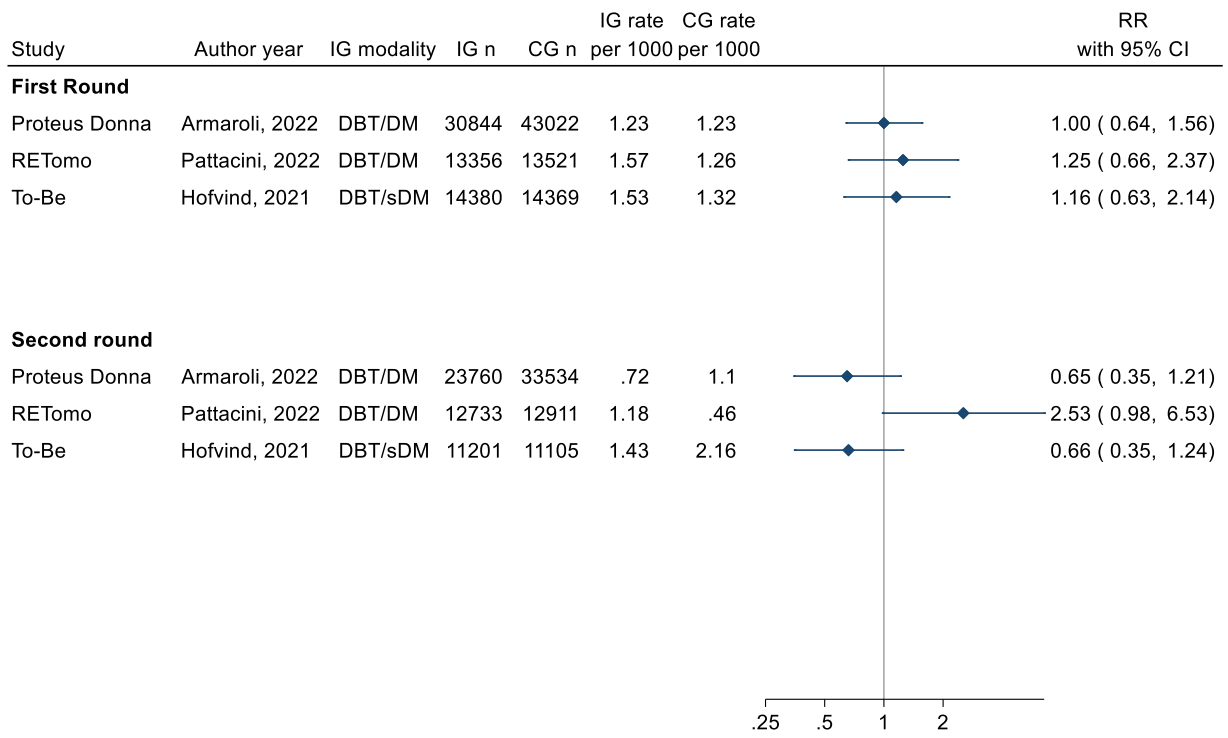
Figure 3. Pooled Analysis of Screen-Detected Invasive Cancers Diagnosed in Trials Comparing Digital Breast Tomosynthesis and Digital Mammography



Random-effects REML model

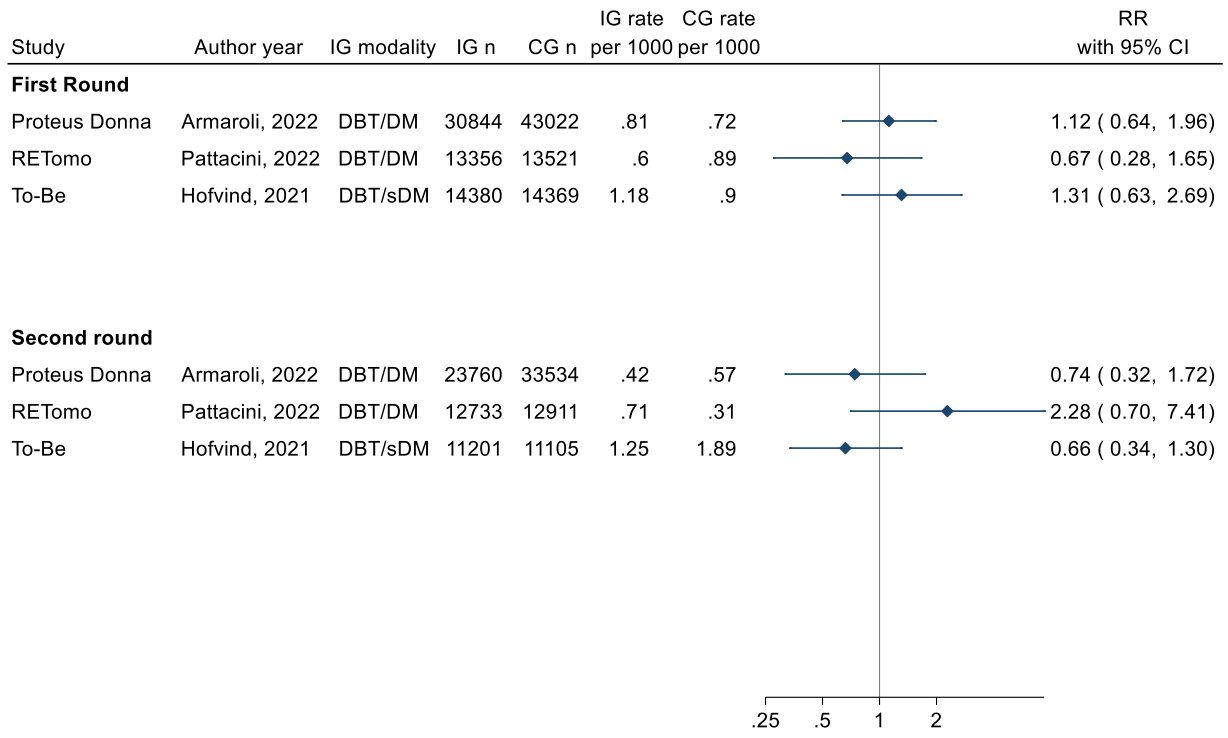
Abbreviations: CG = control group; DBT=digital breast tomosynthesis; DM = digital mammography; IG = intervention group; RETomo=Reggio Emilia Tomosynthesis Trial; RR = relative risk; sDM = synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen

Figure 4. Proportion of Screen-Detected Invasive Cancers Diagnosed at Stage II or Higher in Trials Comparing Digital Breast Tomosynthesis and Digital Mammography



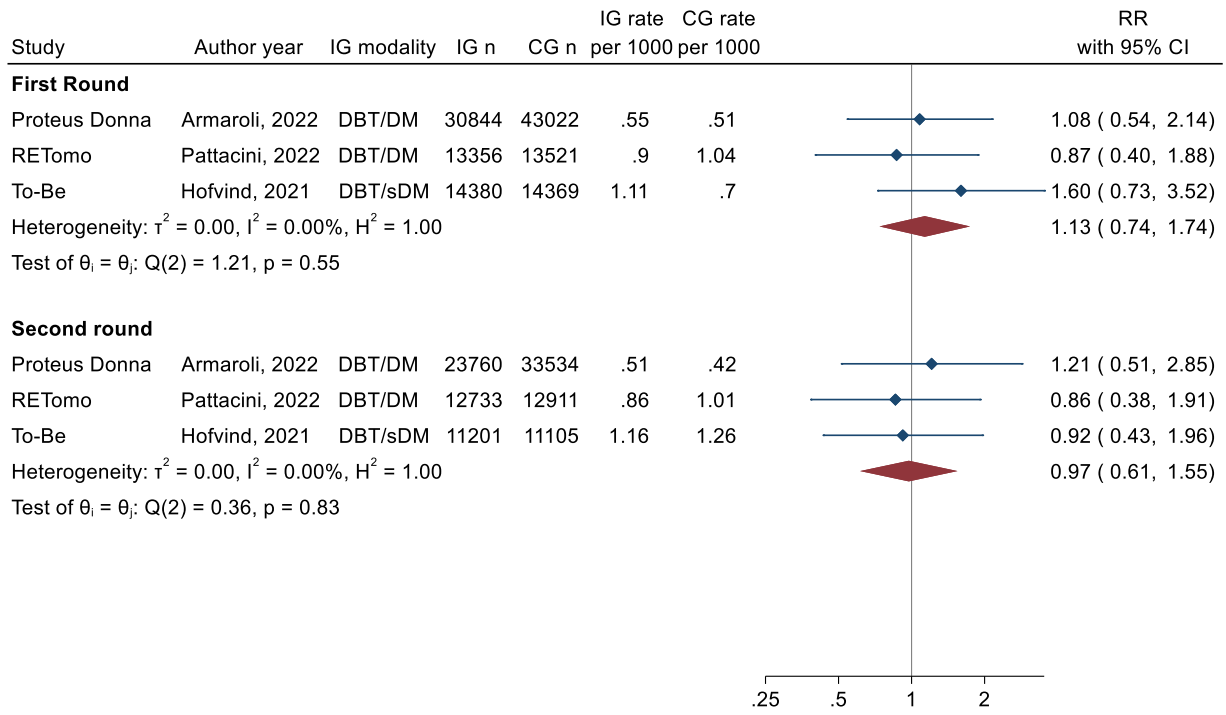
Abbreviations: CG = control group; DBT=digital breast tomosynthesis; DM = digital mammography; IG = intervention group; RETomo=Reggio Emilia Tomosynthesis Trial; RR = relative risk; sDM = synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen

Figure 5. Proportion of Screen-Detected Invasive Cancers Diagnosed With Tumor Size >20 mm in Trials Comparing Digital Breast Tomosynthesis and Digital Mammography



Abbreviations: CG = control group; DBT=digital breast tomosynthesis; DM = digital mammography; IG = intervention group; mm = millimeters; RETomo=Reggio Emilia Tomosynthesis Trial; RR = relative risk; sDM = synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen

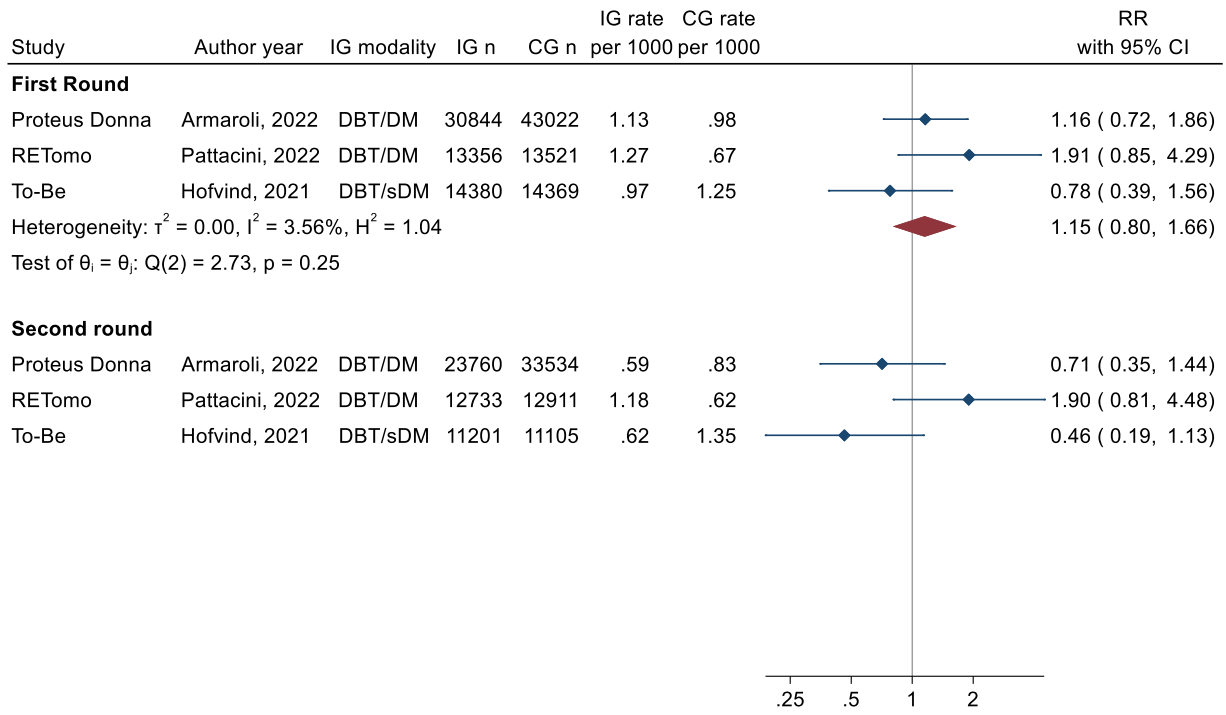
Figure 6. Proportion of Screen-Detected Invasive Cancers Diagnosed as Grade 3 in Trials Comparing Digital Breast Tomosynthesis and Digital Mammography



Random-effects REML model

Abbreviations: CG = control group; DBT=digital breast tomosynthesis; DM = digital mammography; IG = intervention group; RETomo=Reggio Emilia Tomosynthesis Trial; RR = relative risk; sDM = synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen

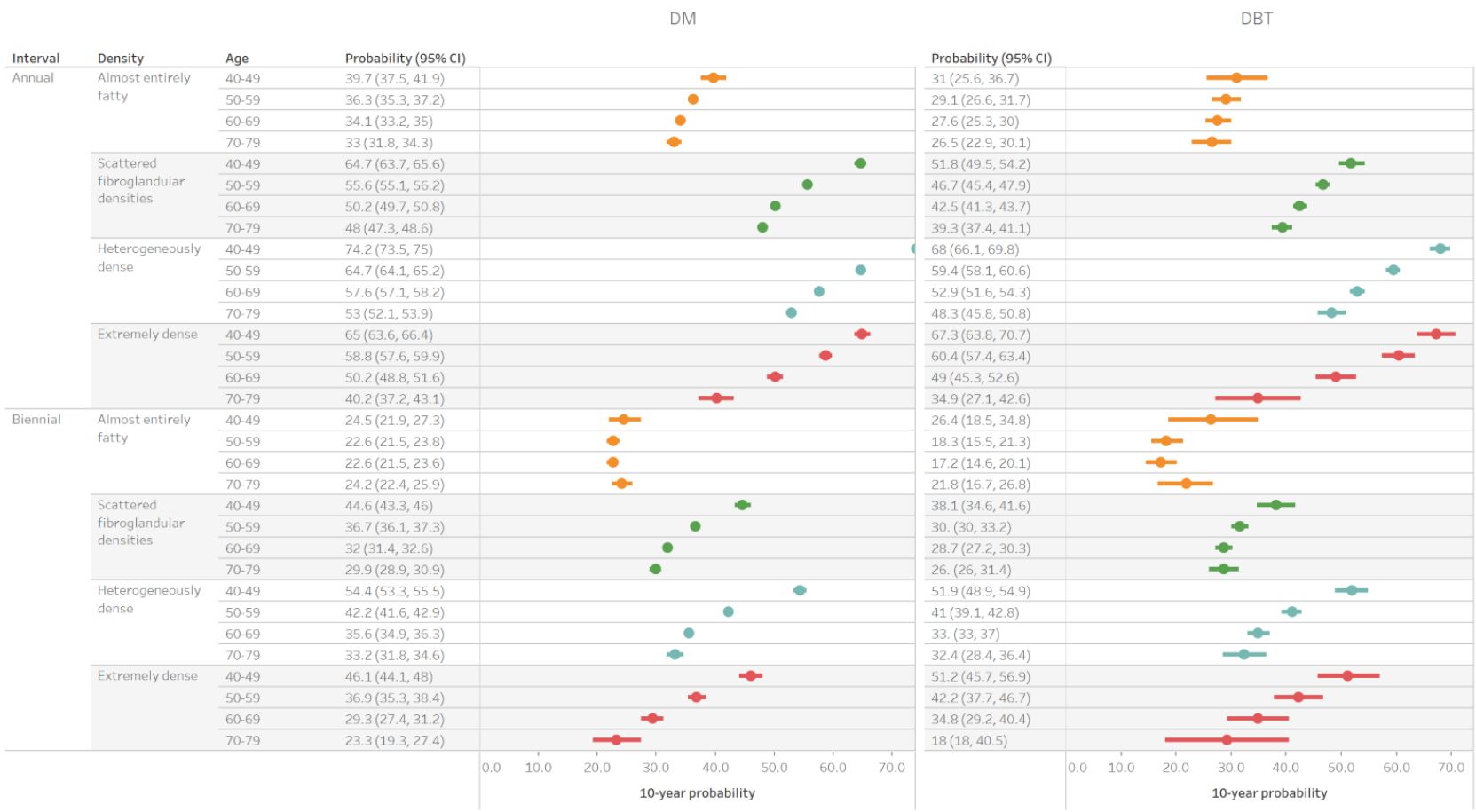
Figure 7. Proportion of Screen-Detected Invasive Cancers Diagnosed as Node Positive in Trials Comparing Digital Breast Tomosynthesis and Digital Mammography



Random-effects REML model

Abbreviations: CG = control group; DBT=digital breast tomosynthesis; DM = digital mammography; IG = intervention group; RETomo=Reggio Emilia Tomosynthesis Trial; RR = relative risk; sDM = synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen

Figure 8. Cumulative Probability of False-Positive Recall in One NSRI* Using BCSC Data Comparing Annual vs. Biennial Screening With DBT or DM



*Based on data from Ho et al. (2022)[138]

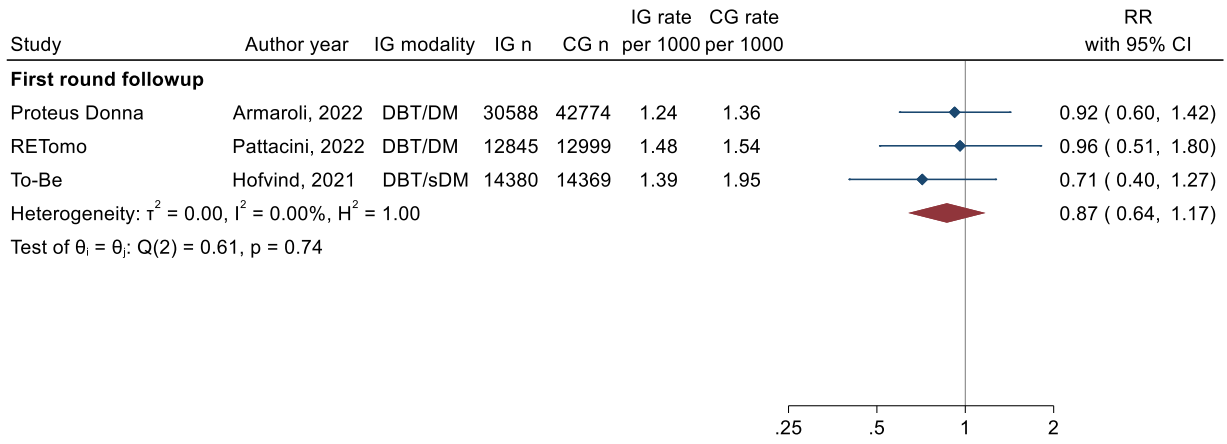
Abbreviations: BCSC = Breast Cancer Screening Consortium; DBT=digital breast tomosynthesis; DM = digital mammography

Figure 9. Cumulative Probability of False-Positive Biopsy in One NSRI* Using BCSC Data Comparing Annual vs. Biennial Screening With DBT or DM

*Based on data from Ho et al. (2022)[138]

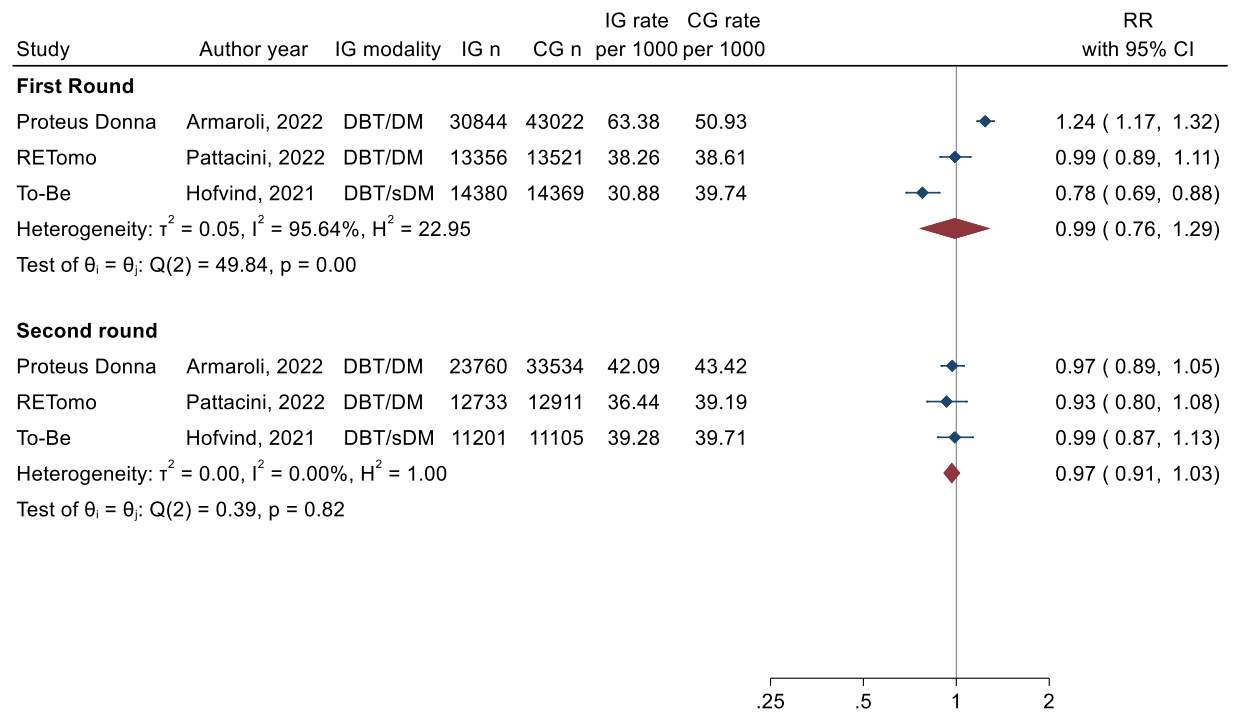
Abbreviations: BCSC = Breast Cancer Screening Consortium; DBT=digital breast tomosynthesis; DM = digital mammography

Figure 10. Pooled Analysis of Interval Cancers Diagnosed in Trials Comparing Digital Breast Tomosynthesis and Digital Mammography



Abbreviations: CG = control group; DBT=digital breast tomosynthesis; DM = digital mammography; IG = intervention group; RETomo=Reggio Emilia Tomosynthesis Trial; RR = relative risk; sDM = synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen

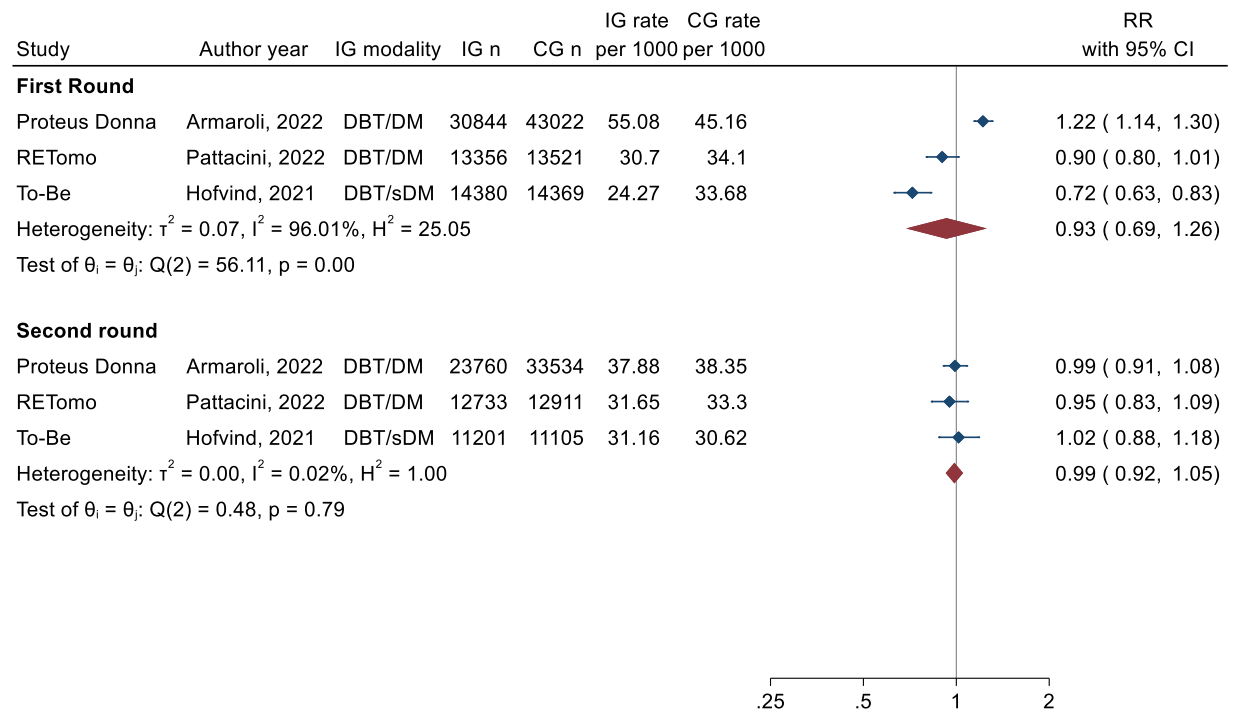
Figure 11. Pooled Analysis of Recall Rates Reported in Trials Comparing Digital Breast Tomosynthesis and Digital Mammography



Random-effects REML model

Abbreviations: CG = control group; DBT=digital breast tomosynthesis; DM = digital mammography; IG = intervention group; RETomo=Reggio Emilia Tomosynthesis Trial; RR = relative risk; sDM = synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen

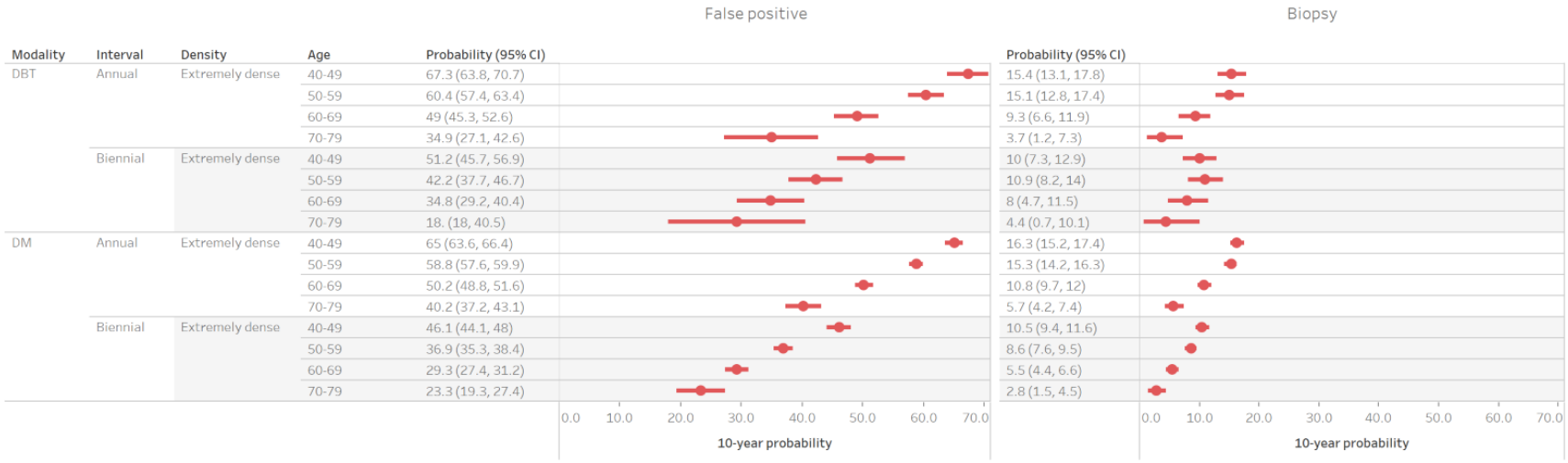
Figure 12. Pooled Analysis of False-Positive Recalls Reported in Trials Comparing Digital Breast Tomosynthesis and Digital Mammography



Random-effects REML model

Abbreviations: CG = control group; DBT=digital breast tomosynthesis; DM = digital mammography; IG = intervention group; RETomo=Reggio Emilia Tomosynthesis Trial; RR = relative risk; sDM = synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen

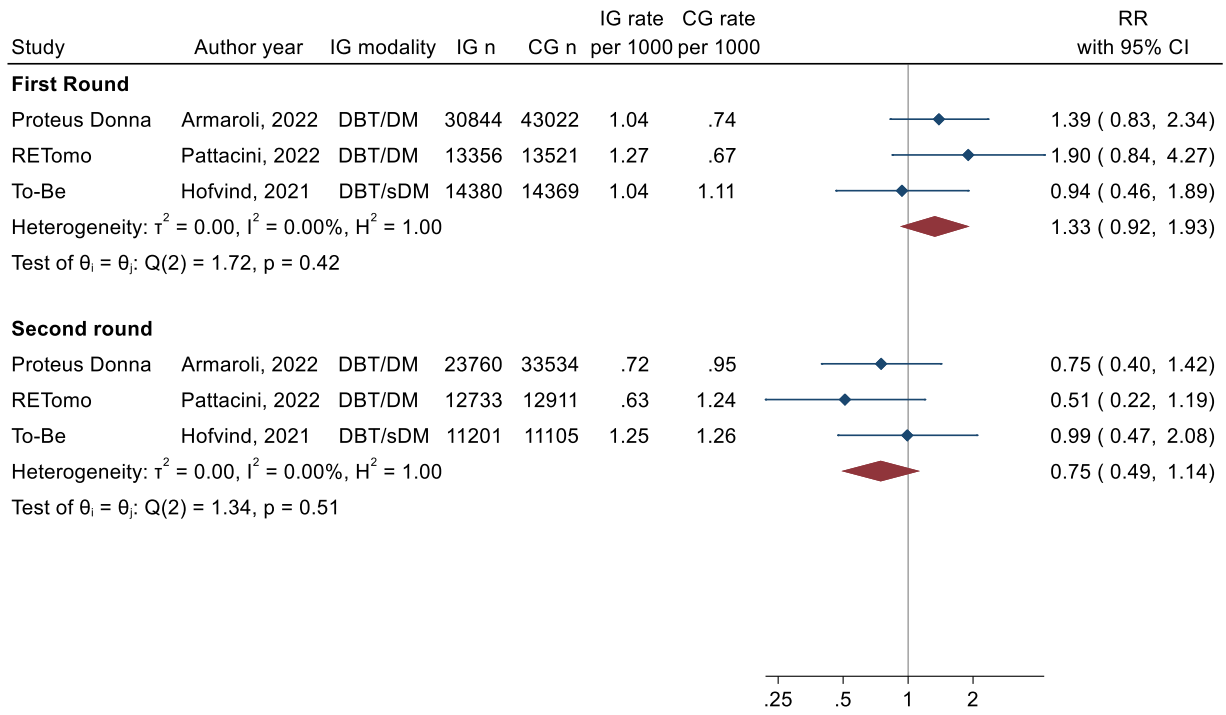
Figure 13. Cumulative Probability of False-Positive Recall or Biopsy in One NSRI* Using BCSC Data Comparing Annual vs. Biennial Screening With DBT or DM, Among Women With Extremely Dense Breasts



*Based on data from Ho et al. (2022)[138]

Abbreviations: BCSC = Breast Cancer Screening Consortium; DBT=digital breast tomosynthesis; DM = digital mammography

Figure 14. Pooled Analysis of DCIS Diagnosed in Trials Comparing Digital Breast Tomosynthesis and Digital Mammography



Random-effects REML model

Abbreviations: CG = control group; DBT=digital breast tomosynthesis; DCIS=ductal carcinoma in situ**;** DM = digital mammography; IG = intervention group; RETomo=Reggio Emilia Tomosynthesis Trial; RR = relative risk; sDM = synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen;

Table 1. Breast Cancer Incidence and Mortality, by Age and Race/Ethnicity

	Incidence of New Cases* (per 100,000 women, per year)	Mortality Rate† (per 100,000 women, per year)
Age (years)		
30-34	30.9	2.9
35-39	64.9	6.6
40-44	132.9	11.5
45-49	199.9	18.0
50-54	238.1	27.1
55-59	273.0	36.0
60-64	343.9	45.3
65-69	428.9	56.8
70-74	477.7	71.4
75-79	460.2	90.0
80-84	416.5	115.0
≥85	325.0	174.3
Race/Ethnicity		
Non-Hispanic White	137.6	19.7
Non-Hispanic Black	129.6	27.6
Non-Hispanic Asian/Pacific Islander	106.9	11.7
Non-Hispanic American Indian/Alaska Native	111.3	17.6
Hispanic	99.9	13.7

*U.S. 5-year Age-Adjusted incidence (2015-2019), all stages
† U.S. 5-Year Age-Adjusted Mortality rates by age at death (2016-2020), all stages

Source: SEER 5-year Age-Adjusted Incidence Rates (2015-2019)[267] and U.S. Mortality Rates by Age at Death (2016-2020)[23]

Table 2. Breast Cancer Screening Guidelines From Other Organizations

	Society or Professional Organization, Year	Age to Begin Screening – Recommended	Screening Frequency - Recommended	Age to Stop Screening	Screening Test	
					DM/FM	DBT
United States-Based Society or Professional Organization	The American Cancer Society (ACS), 2015[268]	45*	45-54: Annual 55+: Biennial	As long as in good health	Y	-
	The American Academy of Family Physicians (AAFP), 2020[269]	50*	Biennial	74	Y	N
	The American Congress of Obstetrics and Gynecology (ACOG), 2017[270]	50*	Annual or Biennial	As long as in good health	Y	-
	The American College of Radiology (ACR), 2017[271]	40	Annual	No limit	Y	O
	The National Comprehensive Cancer Network (NCCN), 2019[33]	40	Annual	No limit	Y	Y
	Women's Preventive Services Initiative (WPSI), 2020[272]	40-50*	Annual or Biennial	As long as in good health	Y	-
International Society or Professional Organization	United Kingdom (UK) National Health Service, 2020[273]	50	Triennial	71	Y	-
	The European Commission Initiative for Breast Cancer (ECIBC), 2022 [274]	45	45-49: Biennial or Triennial 50-69: Biennial 70-74: Triennial	74	Y†	Y†
	The Canadian Task Force on Preventive Health Care, 2018[275]	50	Biennial or Triennial	74	Y	N
	Cancer Australia, 2020[276]	50*	Biennial	74 (*Optional: no limit*)	Y	-

*These organization include recommendations to begin screening starting at age 40 years for some women based on shared decision-making.
†The ECIBC recommends using either DM or DBT, but not both. For women with dense breasts, it is recommended that DBT is used.

Abbreviations: DBT=digital breast tomosynthesis; DM=digital mammography; FM=film mammography; N=does not recommend; O=optional, based on shared decision making between the patient and her provider; Y=recommended

Table 3. Health Outcomes and Harms Reported by Included Trials and Non-Randomized Studies, by Intervention Category (k = 19*)

Intervention Category	Number of Included Studies	Population	Health outcomes (KQ1) (# of included studies)	Intermediate outcomes (KQ2) (# of included studies)	Harms (KQ3) (# of included studies)
Age to Stop	1 (NSRI=1[134])	General screening population	Breast cancer mortality (1)		Overtreatment (1) Overdiagnosis (1)
Frequency	5 (RCT=1[124]; NSRI=4[138, 151, 152, 157])	General screening populations starting screening at ages 40 or 50.	Breast cancer mortality (1) All-cause mortality (1)	Screen-detected invasive cancers (2) Tumor characteristics (2)	Interval cancers (2) Cumulative false-positive rates (1) False-positive recalls (1) False-positive biopsy recommendations (1)
Digital Breast Tomosynthesis	10 (RCT=4[127, 137, 141, 158]; NRSI=6[79, 124, 130, 138, 142, 145, 160])	General screening populations starting screening at ages 40 years, 45 years, and 50 years.		Screen-detected cancers (4) Tumor characteristics (4)	Interval cancers (false-negative and new cancers combined) (7) False-negative cancers (1) Recall rates (4) Biopsies (2) False-positive recalls (3) False-positive biopsy recommendations (1) Overtreatment (1) Adverse events (1)
Supplemental MRI	2 (RCT=1[162]; NRSI=1[133])	NRSI in general screening population (ages 40 to 64 years); RCT among individuals with negative mammography and extremely dense breasts			Interval cancers (1) Adverse events (1) Incidental findings/overtreatment (1)
Supplemental Ultrasound	2 (RCT=1[156]; NRSI=1[150])	NRSI in general screening populations; RCT among individuals ages 40 to 49 years			Interval cancers (2) Recall rates (1) Biopsies (1)

*One study (Ho et al., 2022) is reflected in both the interval and DBT intervention categories

Abbreviations: DBT=digital breast tomosynthesis; DCIS=ductal carcinoma in situ; DM=digital mammography; k=number of included studies; MRI=magnetic resonance imaging; NSRI=nonrandomized study of intervention; US=ultrasound

Table 4. Study Characteristics of Included Trials and Non-Randomized Studies of Screening Approaches and Modalities

Intervention Category	Study Design	Author, Year Study/Trial Name Quality	Country	N Screened (Round 1)	Brief population description	Study Years	Screening Intervention	Screening Control
Age to Stop	NRSI	Garcia-Albeniz, 2020[134] Fair	US	264274	Women aged 70 to 84, enrolled in Medicare Parts A and B between 1999 and 2008, with a high probability of living 10 additional years (based on calculated Medicare-specific comorbidity score <1)	2000 to 2008	Continuing annual DM beyond 70 years of age	Stopping annual DM at 70 years of age
Screening Frequency	RCT	Blamey, 2002[124] UKCCCR Fair	UK	76022	Women aged 50 to 62 attending a population-based screening program	1989 to 1996	Annual DM	Triennial DM
	NRSI	Ho, 2022[138] BCSC Fair	US	903495	Women aged 40 to 79 years	2005 to 2018	Annual DBT/sDM or DM	Biennial DBT/sDM or DM
		McGuinness, 2018[151] KYRAS Fair	US	2019	Women aged 18 years or older attending screening at one academic medical center	2014 to 2015	Annual DM	Biennial DM
		Miglioretti, 2015[152] BCSC Fair	US	15440	Women aged 40 to 85 years diagnosed with a screen-detected or interval invasive breast cancer or DCIS and at least 2 screening mammography examinations 11-14 or 23-26 months apart before diagnosis	1996 to 2012	Annual DM	Biennial DM
		Parvinen, 2011[157] Fair	Finland	14765	Women aged 40-49 years attending a population-based screening program	1987 to 2007	Annual DM	Triennial DM
Digital Breast Tomosynthesis (DBT)	RCT	Armaroli, 2022[127] Proteus Donna Fair	Italy	73866	Women aged 46 to 68 years attending a population-based screening program	2004 to 2017	DBT/DM (round 1), DM (round 2)	DM

Table 4. Study Characteristics of Included Trials and Non-Randomized Studies of Screening Approaches and Modalities

Intervention Category	Study Design	Author, Year Study/Trial Name Quality	Country	N Screened (Round 1)	Brief population description	Study Years	Screening Intervention	Screening Control
		Heindel, 2022[137] TOSYMA Good	Germany	99634	Women aged 50-69 years attending a population-based screening program	2018 to 2020	DBT/sDM	DM
		Pattacini, 2022[158] RETomo Good	Italy	26877	Women aged 45 to 69 years attending screening in one of three clinics equipped with DBT who had already participated in at least one round of the Reggio Emilia screening program	2014 to 2017	DBT/DM (round 1), DM (round 2)	Annual DM (age 45-49), biennial DM (age 50-69)
		Hofvind, 2021[141] To-Be Good	Norway	28749	Women aged 50-69 years attending a population-based screening program	2016 to 2020	DBT/sDM	DM (round 1), DBT/sDM (round 2)
	NRSI	Ho, 2022[138] BCSC-2022a Fair	US	903495	Women aged 40 to 79 years	2005 to 2018	DBT	DM
		Kerlikowske, 2022[145] BCSC-2022b Fair	US	504427	Women aged 40 to 79 years with no history of breast cancer or mastectomy who had a screening mammogram and/or DBT	2011 to 2018	DBT	DM
		Johnson, 2021[79] MBTST Fair	Sweden	40107	Women enrolled in a breast cancer screening trial and population-based match controls	2010 to 2015	DBT/DM	DM

Table 4. Study Characteristics of Included Trials and Non-Randomized Studies of Screening Approaches and Modalities

Intervention Category	Study Design	Author, Year Study/Trial Name Quality	Country	N Screened (Round 1)	Brief population description	Study Years	Screening Intervention	Screening Control
		Richman, 2021[160] Fair	US	4580698	Women aged 40-64 years with at least 1 screening mammogram between January 1, 2015, and December 31, 2017	2015 to 2017	DBT/DM	DM
		Hovda, 2020[142] OVVV Fair	Norway	92404	Women aged 50 to 69 years participating in population-based screening program	2014 to 2017	DBT/sDM (round 1), DM (round 2)	DM
		Conant, 2016[130] PROSPR Fair	US	103401	Women aged 40 to 74 years attending screening at academic medical centers participating in screening consortium	2011 to 2014	DBT/DM	DM
Supplemental MRI	RCT	Veenhuizen, 2021[162] DENSE Good	Netherlands	40373	Women aged 50-75 years of age with negative mammography results (BI-RADS radiographic score of 1 or 2) and extremely dense breast tissue	2011 to 2015	DM plus MRI	DM
	NRSI	Ganguli, 2022[133] Fair	US	18416	Women aged 40-64 years who had a bilateral breast MRI or bilateral screening mammogram claim	2016 to 2018	MRI	DM
Supplemental Ultrasound	RCT	Ohuchi, 2016[156] J-START Fair	Japan	72717	Women aged 40 to 49 years	2007 to 2011	DM plus US	DM
	NRSI	Lee, 2019[150] BCSC Fair	US	18562	Women undergoing screening at eligible BCSC sites	2000 to 2013	DM plus US	DM

Abbreviations: BCSC=Breast Cancer Surveillance Consortium; BI-RADS=Breast Imaging Reporting and Data System; DBT=Digital breast tomosynthesis; DCIS=ductal carcinoma in situ; DM=Digital mammography; DENSE=Dense Tissue and Early Breast Neoplasm Screening; J-START= Japan Strategic Anti-cancer

Table 4. Study Characteristics of Included Trials and Non-Randomized Studies of Screening Approaches and Modalities

Randomized Trial; MBTST=Malmo Breast Tomosynthesis Screening Trial; MRI=magnetic resonance imaging; RETomo=Reggio Emilia Tomosynthesis Trial; sDM=synthetic mammography; PROSPR=Population-based Research Optimizing Screening through Personalized Regimens; To-Be=Tomosynthesis Trial in Bergen; TOSYMA=TOmosynthesis plus SYnthesized MAmmography study; OVVV=Oslo-Vestfold-Vestre Viken; UKCCR=United Kingdom Coordinating Committee on Cancer Research trial; US=ultrasound

Table 5. Population Characteristics of Included Trials and Non-Randomized Studies, by Intervention Category

Intervention Category	Study Design	Author, Year Study/Trial Name	Age (years)	Study-described Race/ethnicity	Breast Density*	First-Degree Family History of Breast Cancer	Hormonal Status
Age to Stop	NRSI	Garcia-Albeniz, 2020[134]	70 to 74: 47% ≥75: 53%	White: 92% Black 5% Other: 3%	NR	NR	NR
Interval	RCT	Blamey, 2002[124] UKCCCR	50 to 62 (range)	NR	NR	NR	NR
	NRSI	Ho, 2022[138] BCSC-2022a	NR	NR	NR	NR	NR
		McGuinness, 2018[151] KYRAS	59 (median)	White: 10% Black: 10% Hispanic: 76% Other: 4%	BI-RADS A, B: 70% BI-RADS C, D: 30%	NR	NR
		Miglioretti, 2015[152] BCSC	40 to 85 (range)	White: 78% Black 5% Asian: 5% AI/AN: <1% Hispanic: 5% Other: 1% Unknown: 6%	NR	22%†	Premenopausal: 13% Menopausal: 64% Current HRT use: 22%
		Parvinen, 2011[157]	40 to 49: 100%	NR	NR	NR	NR
Digital Breast Tomosynthesis (DBT)	RCT	Armaroli, 2022[127] Proteus Donna	57 (mean; SD, 6)	NR	NR	NR	NR
		Heindel, 2022[137] TOSYMA	50 to 59: 62% 60 to 69: 38%	NR	NR	NR	NR
		Pattacini, 2022[158] RETomo	55 (mean)	NR	BI-RADS A: 8% BI-RADS B: 38% BI-RADS C: 35% BI-RADS D: 9% NR: 9%	NR	NR

Table 5. Population Characteristics of Included Trials and Non-Randomized Studies, by Intervention Category

Intervention Category	Study Design	Author, Year Study/Trial Name	Age (years)	Study-described Race/ethnicity	Breast Density*	First-Degree Family History of Breast Cancer	Hormonal Status
		Hofvind, 2021[141] To-Be	60 (mean)	NR	Volpara grade 1: 25% Volpara grade 2: 43% Volpara grade 3: 24% Volpara grade 4: 7% NR: 7%	NR	NR
	NRSI	Ho, 2022[138] BCSC-2022a	40 to 79 (range)	NR	NR	NR	NR
		Kerlikowske, 2022[145] BCSC-2022b	40 to 49: 23% 50 to 59: 33% 60 to 69: 29% 70 to 79: 15%	White: 73% Black 11% Asian: 9% Hispanic: 5% Other: 2%	BI-RADS A: 11% BI-RADS B: 45% BI-RADS C: 37% BI-RADS D: 7%	19†	Premenopausal: 28% Postmenopausal or surgical menopause: 72%
		Johnson, 2021[79] MBTST	56 (mean)	NR	NR	NR	NR
		Richman, 2021[160]	40 to 49: 31% 50 to 59: 47% 60 to 69: 22%	NR	NR	7%‡	NR
		Hovda, 2020[142] OVVV	59 (mean)	NR	NR	NR	NR
		Conant, 2016[130] PROSPR	40 to 49: 28% 50 to 59: 36% 60 to 74: 36%	White: 79% Black 10% Asian: 2% AI/AN: <1% Hispanic: 6% Other: 3%	BI-RADS A: 14% BI-RADS B: 45% BI-RADS C: 29% BI-RADS D: 4% NR: 8%	NR	NR
Supplemental MRI	RCT	Veenhuizen, 2021[162] DENSE	54 (median; IQR 51-61)	NR	Volpara grade 4: 100%	NR	NR
	NRSI	Ganguli, 2022[133]	51 (mean)	NR	17%§	50%‖	NR

Table 5. Population Characteristics of Included Trials and Non-Randomized Studies, by Intervention Category

Intervention Category	Study Design	Author, Year Study/Trial Name	Age (years)	Study-described Race/ethnicity	Breast Density*	First-Degree Family History of Breast Cancer	Hormonal Status
Supplemental Ultrasound	RCT	Ohuchi, 2016[156] J-START	44 (mean)	NR	Dense breasts (BI-RADS 3 or 4): 58%	5%†	Premenopausal: 76% Menopausal: 24%
	NRSI	Lee, 2019[150] BCSC	<40: 4% 40 to 49: 42% 50 to 59: 35% 60 to 69: 14% ≥70: 40%	White: 80% Black: 0.4% Asian: 11% Hispanic: 7% Other: 2%	BI-RADS A: 2% BI-RADS B: 29% BI-RADS C: 57% BI-RADS D: 8% NR: 4%	31%†	Premenopausal: 38% Menopausal: 41%

*Breast density defined using the BI-RADS system (which uses visual assessent to categorize breast density as (A) almost entirely fatty; (B) scattered areas of fibroglandular density; (C) heterogeneously dense, and (D) extremely dense) or the Volpara system (which uses a quantitative measure ofvolumetric breast density and assigns density to one of four categories (Volpara density grade [VDG] 1 to 4), which are analogous to BI-RADS A to D).
†First-degree family history of breast cancer
‡ Family history of breast cancer
§ Breast density definition based on ICD-10-CM diagnosis code (R92.2 – inconclusive mammogram) which indicates dense breast findings
‖Family history of breast cancer or genetic susceptibility

Abbreviations: AI/AN=American Indian/Alaskan Native; BCSC=Breast Cancer Surveillance Consortium; BI-RADS=Breast Imaging Reporting and Data System; DENSE=Dense Tissue and Early Breast Neoplasm Screening; HRT=hormone replacement therapy; IQR=interquartile range; J-START= Japan Strategic Anti-cancer Randomized Trial; MBTST=Malmo Breast Tomosynthesis Screening Trial; NRSI=nonrandomized study of intervention; RCT=randomized controlled trial; RETomo=Reggio Emilia Tomosynthesis Trial; SD=standard deviation; sDM=synthetic mammography; PROSPR=Population-based Research Optimizing Screening through Personalized Regimens; To-Be=Tomosynthesis Trial in Bergen; TOSYMA=TOmosynthesis plus SYnthesized MAmmography study; OVVV=Oslo-Vestfold-Vestre Viken; UKCCR=United Kingdom Coordinating Committee on Cancer Research trial

Table 6. Characteristics of Screen-Detected Invasive Cancers Diagnosed Following an Annual vs. Triennial Screening Frequency

Author, Year Study/Trial Name	Population	Followup	Frequency	Invasive Cancer Detection (Rate per 1000)	≥ Stage II (Rate per 1000)	Tumor Diameter, mm (sd)	Tumor >20mm (Rate per 1000)	Lymph Node Positive (Rate per 1000)	Histologic Grade 3 (Rate per 1000)	Poor Prognostic Index
Blamey, 2002[124] UKCCCR	Women aged 50 to 62 attending a population-based screening program	Cumulative cancer incidence (3 years)	Annual DM	166/37530 (4.4)	NR	NR	63 (1.7)*	63 (1.7)*	96 (2.6)*	20 (0.5)*†
			Triennial DM	104/38492 (2.7)	NR	NR	69 (1.8)*	61 (1.6)*	86 (2.2)*	22 (0.6)*†

*Includes both screen-detected invasive cancers and interval cancers but excludes cancers with missing information.
†Nottingham Prognostic Index score.

Abbreviations: DM=digital mammography; mm=millimeter; NR=not reported; UKCCR=United Kingdom Coordinating Committee on Cancer Research trial.

Table 7. Characteristics of Cancers Diagnosed Following an Annual vs. Biennial Screening Frequency, by Population Subgroup

Author, Year Study/ Trial Name	Comparison (IG vs. CG)	Outcome Definition	Subgroup	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Effect (95% CI)*
Miglioretti, 2015[152] BCSC	Annual DM preceding diagnosis vs. biennial DM preceding diagnosis†	Stage IIB or higher	40-49 years	246/1155 (213.0)	103/425 (242.4)	RR: 1.17 (95% CI, 0.93 to 1.46)
			50-59 years	499/2532 (197.1)	129/680 (189.8)	RR: 0.98 (95% CI, 0.80 to 1.21)
			60-69 years	429/2616 (164.0)	98/666 (147.2)	RR: 0.99 (95% CI, 0.79 to 1.24)
			70-85 years	341/2506 (136.1)	95/782 (121.5)	RR: 0.98 (95% CI, 0.76 to 1.27)
			Premenopausal	217/1095 (198.2)	89/346 (257.2)	RR: 1.28 (95% CI, 1.01 to 1.63)
			Postmenopausal without HRT use	588/3720 (158.1)	141/1071 (131.7)	RR: 0.95 (95% CI, 0.79 to 1.15)
			Postmenopausal with HRT use	355/1982 (169.0)	96/547 (175.5)	RR: 1.14 (95% CI, 0.89 to 1.47)
		Less favorable characteristic (stage IIB or higher, tumor size greater than 15 mm, or positive node status)	40-49 years	692/1171 (591.0)	268/425 (630.6)	RR: 1.04 (95% CI, 0.94 to 1.14)
			50-59 years	1374/2545 (539.9)	368/685 (537.2)	RR: 1.03 (95% CI, 0.94 to 1.12)
			60-69 years	1277/2627 (486.1)	329/662 (497.0)	RR: 1.07 (95% CI, 0.97 to 1.19)
			70-85 years	1102/2505 (439.9)	345/774 (445.7)	RR: 1.05 (95% CI, 0.94 to 1.18)
			Premenopausal	660/1105 (597.3)	229/349 (656.2)	RR: 1.11 (95% CI, 1.00 to 1.22) ‡
			Postmenopausal without HRT use	1737/3735 (465.1)	494/1062 (465.2)	RR: 1.03 (95% CI, 0.95 to 1.12)
			Postmenopausal with HRT use	1005/1995 (503.8)	278/547 (508.2)	RR: 1.12 (95% CI, 1.00 to 1.25) §

*Adjusted for race/ethnicity, first-degree family history of breast cancer, and Breast Cancer Surveillance Consortium registry using log-binomial regression unless otherwise specified.
† Annual cancers diagnosed within 12 months of screening examination performed 11 to 14 months after prior mammogram; biennial cancers diagnosed within 24 months of screening examination performed 23 to 26 months after prior mammogram.
‡ p=0.047.
§ p=0.05.

Abbreviations: BCSC= Breast Cancer Surveillance Consortium; DM=digital mammography; HRT= hormone replacement therapy; mm= millimeter

Table 8. Characteristics of Screen-Detected Invasive Cancers Diagnosed in Studies Comparing Digital Breast Tomosynthesis and Digital Mammography

Author, Year Study/Trial Name Study Design	Population	Follow-up	Modality (previous round modality)	Invasive Cancer Detection (Rate per 1000)	≥Stage II (Rate per 1000)	Tumor Diameter, mm (SD)	Tumor >20mm (Rate per 1000)	Lymph Node Positive (Rate per 1000)	Histologic Grade 3 (Rate per 1000)	Poor Prognostic Index
Armaroli, 2022[127] Proteus Donna RCT	Women aged 46 to 68 years	First Round	DBT/DM	224/30844 (7.3)	38 (1.2)	NR	25 (0.8)*	35 (1.1)	17 (0.6)	NR
			DM	214/43022 (5.0)	53 (1.2)	NR	31 (0.7)*	42 (1.0)	22 (0.5)	NR
		Second Round	DM (DBT/DM)	81/23760 (3.4)	17 (0.7)	NR	10 (0.4)*	14 (0.6)	12 (0.5)	NR
			DM	135/33354 (4.0)	37 (1.1)	NR	19 (0.6)*	28 (0.8)	14 (0.4)	NR
Pattacini, 2022[158] RETomo RCT	Women aged 45 to 69 years	First Round	DBT/DM	84/13356 (6.3)	21 (1.6)†	NR	8 (0.6)	17 (1.3)	12 (0.9)	NR
			DM	52/13521 (3.8)	17 (1.3)‡	NR	12 (0.9)	9 (0.7)	14 (1.0)	NR
		Second Round	DBT/DM	53/12733 (4.2)	15 (1.2)§	NR	9 (0.7)	15 (1.2)	11 (0.9)	NR
			DM	60/12911 (4.6)	6 (0.5)‖	NR	4 (0.3)	8 (0.6)	13 (1.0)	NR
Hofvind, 2021[141] To-Be RCT	Women aged 50 to 69 years	First Round	DBT/sDM	80/14380 (5.6)	22 (1.5)¶	16.0 (8.4)	17 (1.2)	14 (1.0)	16 (1.1)	NR
			DM	71/14369 (4.9)	19 (1.3) ¶	14.5 (8.8)	13 (0.9)	18 (1.3)	10 (0.7)	NR
		Second Round	DBT/sDM	77/11201 (6.9)	16 (1.4) ¶	15.2 (9.3)	14 (1.2)¶	7 (0.6)	13 (1.2)	NR
			DBT/sDM (DM)	87/11105 (7.8)	24 (2.2) ¶	15.8 (8.6)	21 (1.9)¶	15 (1.4)	14 (1.3)	NR
Hovda, 2020[142] OVVV NRSI	Women aged 50 to 69 years	First Round	DBT/sDM	283/37815 (7.6)	NR	NR	38 (1.0)	36 (1.0)	28 (0.8)	NR
			DM	329/61742 (5.3)	NR	NR	57 (0.9)	45 (0.7)	56 (0.9)	NR
		Second Round	DM (DBT/sDM)	84/26474 (3.2)	NR	15.4 (13.0)	NR	16 (0.6)	19 (0.7)	NR
			DM	203/45543 (4.5)	NR	14.3 (8.2)	NR	30 (0.7)	30 (0.7)	NR

Note: Unless otherwise indicated, data are number of patients. Rates are per 1,000 women screened.
*Tumor diameter >10mm
†Stage III or higher: 3 (0.2)
‡Stage III or higher: 2 (0.1)
§Stage III or higher: 2 (0.2)
‖Stage III or higher: 4 (0.3)
¶ Provided through author communication

Abbreviations: DBT=Digital breast tomosynthesis; DM=Digital mammography; mm=millimeter; NR=not reported; RETomo=Reggio Emilia Tomosynthesis Trial; SD=standard deviation; sDM=synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen; OVVV=Oslo-Vestfold-Vestre Viken

Table 9. Incidence of Screen-Detected Invasive Cancers Diagnosed in Trials Comparing Digital Breast Tomosynthesis and Digital Mammography, by Population Subgroup

Author, Year Study/Trial Name	Follow-up	Subgroup	Modality (previous round modality)	Invasive Cancer Detection*	Effect (95% CI)
Pattacini, 2022[158] RETomo	First Round	Ages 45 to 49 years	DBT/DM	19/5053 (3.8)	RR=1.9 (95% CI, 0.89 to 4.1)
			DM	10/5103 (2.0	
		Ages 50 to 69 years	DBT/DM	65/8303 (7.8)	RR=1.6 (95% CI, 1.1 to 2.3)
			DM	42/8418 (5.0)	
		Nondense breasts (BI-RADS A or B)	DBT/DM	39/6261 (6.2)	RR=1.8 (95% CI, 1.1 to 3.0)
			DM	22/6286 (3.5)	
		Dense breasts (BI-RADS C or D)	DBT/DM	40/5970 (6.7)	RR=1.5 (95% CI, 0.94 to 2.5)
			DM	26/5978 (4.3)	
	Second Round	Ages 45 to 49 years	DBT/DM	7/4813 (1.5)	RR=0.50 (95% CI, 0.20 to 1.2)
			DM	14/4855 (2.9)	
		Ages 50 to 69 years	DBT/DM	46/7920 (5.8)	RR=1.0 (95% CI, 0.68 to 1.5)
			DM	46/8056 (5.7)	
		Nondense breasts (BI-RADS A or B)	DBT/DM	31/5970 (5.2)	RR=0.97 (95% CI, 0.60 to 1.6)
			DM	32/6002 (5.3)	
		Dense breasts (BI-RADS C or D)	DBT/DM	16/5686 (2.8)	RR=0.64 (95% CI, 0.34 to 1.2)
			DM	25/5706 (4.4)	
Hofvind, 2021[141] To-Be	First Round	VDG1 Density	DBT/sDM	17/3929 (4.3)	RR=1.07 (95% CI, 0.52 to 2.20)
			DM	13/3212 (4.0)	
		VDG2 Density	DBT/sDM	38/6216 (6.1)	RR=1.16 (95% CI, 0.73 to 1.85)
			DM	33/6280 (5.3)	
		VDG3 Density	DBT/sDM	20/3152 (6.3)	RR=1.10 (95% CI, 0.60 to 2.03)
			DM	21/3655 (5.7)	
		VDG4 Density	DBT/sDM	5/962 (5.2)	RR=1.97 (95% CI, 0.47 to 8.21)
			DM	3/1136 (2.6)	
	Second Round	VDG1 Density	DBT/sDM	18/3214 (5.6)	RR=1.04 (95% CI, 0.53 to 2.03)
			DBT/sDM (DM)	16/2960 (5.4)	
		VDG2 Density	DBT/sDM	29/4353 (6.7)	RR=0.94 (95% CI, 0.57 to 1.56)
			DBT/sDM (DM)	31/4395 (7.1)	
		VDG3 Density	DBT/sDM	23/2656 (8.7)	RR=0.82 (95% CI, 0.47 to 1.41)
			DBT/sDM (DM)	29/2736 (10.6)	
		VDG4 Density	DBT/sDM	7/900 (7.8)	RR=0.66 (95% CI, 0.26 to 1.70)
			DBT/sDM (DM)	11/934 (11.8)	

*Rate per 1,000 women screened

Abbreviations: BI-RADS=Breast Imaging Reporting and Data Systems; DBT=Digital breast tomosynthesis; DM=Digital mammography; mm=millimeter; RETomo=Reggio Emilia Tomosynthesis Trial; sDM=synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen; VDG=Volpara Density Grade

Table 10. Overdiagnosis and Overtreatment in Studies of Age to Stop Screening in an Emulated Trial, by Population Subgroup

Author, Year Study/ Trial Name	Comparison (IG vs. CG)	Population	Outcome category	Outcome definition	IG proportion (95% CI)	CG proportion (95% CI)
Garcia-Albeniz, 2020[134]	Continued screening after age 70 versus cessation of screening	Women aged 70 to 74 years with a life expectancy of at least 10 years	Incidence	8-year cumulative risk of breast cancer diagnosis	5.3 (NR)	3.9 (NR)
			Treatments received by women diagnosed with breast cancer*	Lumpectomy	52.6 (51.8–53.4)	36.5 (35.2–38.0)
				Simple mastectomy	11.3 (10.8–11.8)	10.4 (9.5–11.3)
				Radical mastectomy	13.9 (13.4–14.5)	18.2 (17.0–19.4)
				Radiotherapy	51.0 (50.3–51.8)	39.9 (38.6–41.3)
				Chemotherapy	15.2 (14.7–15.8)	15.2 (14.7–15.8)
	Continued screening after age 75 versus cessation of screening	Women aged 75 to 84 years with a life expectancy of at least 10 years	Incidence	8-year cumulative risk of breast cancer diagnosis	5.8 (NR)	3.9 (NR)
			Treatments received by women diagnosed with breast cancer*	Lumpectomy	48.8 (47.9–49.5)	32.6 (31.5–33.8)
				Simple mastectomy	10.8 (10.3–11.2)	10.1 (9.4–10.9)
				Radical mastectomy	14.2 (13.7–14.6)	17.0 (16.0–17.9)
				Radiotherapy	41.2 (40.4–41.9)	31.9 (30.7–33.1)
				Chemotherapy	8.6 (8.3–9.1)	11.5 (10.6–12.3)

*Percentages are standardized to the age group–specific distribution of age; comorbidity score; new diagnosis of Alzheimer disease, acute myocardial infarction, chronic heart failure, chronic kidney disease, chronic obstructive pulmonary disease, hip fracture, stroke, or cancer (lung, endometrial, or colorectal); and institutionalization in a long-term care center.

Abbreviations: CG=control group; CI=confidence interval; IG=intervention group; NR=not reported

Table 11. Interval Cancer Rates in Studies Comparing Breast Cancer Screening Frequencies

Study Design	Author, Year Study/Trial Name	Comparison (IG vs. CG)	Population	Outcome Definition	Histologic Type	IG n/N (rate per 1000 screened), or IG Proportion (95% CI)	CG n/N (rate per 1000 screened), or CG Proportion (95% CI)	Effect (95% CI)
RCT	Blamey, 2002[124] UKCCCR	Annual DM vs. Triennial DM	Women aged 50 to 62	Interval cancers Annual screening group: detected in the three intervals between screening visits during the three years of followup. Triennial screening group: detected before the consecutive screen following the baseline screen.	Invasive	69/37530 (1.8)	104/38492 (2.7)	RR: 0.68 (95% CI, 0.50 to 0.92)*
NRSI	Parvinen, 2011[157]	Annual DM vs. Triennial DM	Women aged 40 to 49	Interval cancers occurring after a negative mammogram and between two subsequent screening visits	Invasive	NR	NR	P=0.22
	Miglioretti, 2015[152]	Annual DM vs. Biennial DM	Women ages 40 to 85, diagnosed with an incident invasive breast cancer or DCIS	Proportion of cancers detected that were diagnosed clinical as interval cancers. Defined as occurring with 12 months following an annual screening interval and within 24 months following a biennial screening interval	Invasive or DCIS	22.2% (21.5% to 23.0%)	27.2% (25.7% to 28.8%)	NR

*Relative risk calculated from Ns.

Abbreviations: DCIS= ductal carcinoma in situ; DM=Digital mammography; NRSI=nonrandomized study of intervention; UKCCR=United Kingdom Coordinating Committee on Cancer Research trial

Table 12. Rates of Interval Cancers (Invasive Cancer and DCIS) in Studies Comparing Breast Cancer Screening Modalities, per 1,000 Screened*

Modality	Study Design	Author, Year Study/ Trial Name	Comparison (IG vs. CG)	Age	Follow up following screening	Histologic type	IG n/N (rate per 1000 screened)	CG n/N (rate per 100 screened)	Effect (95% CI)
DBT	RCT	Armaroli, 2022[127] Proteus Donna	DBT/DM (round 1), DM (round 2) vs. DM*	46 to 68	Median 25 months (range 0 to 36 months)	Invasive	38/30588 (1.2)	58/42774 (1.4)	RR: 0.92 (95% CI, 0.59 to 1.40)
						DCIS	4/30588 (0.1)	5/42774 (0.1)	RR: 1.12 (95% CI, 0.22 to 5.20)
		Pattacini, 2022[158] RETomo	DBT/DM vs. DM	45 to 70	12-months (40 to 49 years) or 24-months (50 to 69 years)	Invasive	19/12845 (1.5)	20/12999 (1.5)	RR: 0.96 (95% CI, 0.51 to 1.80)
					Interval cancers (DCIS only) at 12-months followup (women aged 40 to 49 years) or 24-months followup (women aged 50 to 69 years) among women with negative findings at previous screening	DCIS	2/12845 (0.2)	2/12999 (0.2)	RR: 1.0 (95% CI, 0.14 to 7.20)
		Hofvind, 2021[141] To-Be	DBT/sDM vs. DM	50 to 69	24-months †	Invasive	20/14380 (1.4)	28/14369 (1.9)	RR: 0.71 (95% CI, 0.40 to 1.27)‡
					Interval cancers (DCIS only) at 24-months followup among women with negative findings at previous screening or 6 to 24-months followup among women with a false-positive result at previous screening	DCIS	0/14380	1/14369 (0.07)	p=0.32
	NRSI	Kerlikowske, 2022[145] BCSC-2022b	DBT/sDM vs. DM	40 to 79	12 months	Invasive	NR (0.57)§	NR (0.61)§	Adj rate difference: -0.04 (95% CI, -0.14 to 0.06)§
		Johnson, 2021[79] MBTST‖	DBT/DM vs. DM	40 to 74	18-months (40 to 54 years) or 24 (ages 55 to 74 years)	Invasive and DCIS	21/13369 (1.6)	76/26738 (2.8)	OR: 0.6 (95% CI, 0.3 to 0.9)
			DBT/DM vs. DM	Wome n aged	Interval cancers diagnosed after a	Invasive	19/13369 (1.4)	72/26738 (2.7)	RR: 0.53 (95% CI, 0.32 to 0.87)

Table 12. Rates of Interval Cancers (Invasive Cancer and DCIS) in Studies Comparing Breast Cancer Screening Modalities, per 1,000 Screened*

Modality	Study Design	Author, Year Study/ Trial Name	Comparison (IG vs. CG)	Age	Follow up following screening	Histologic type	IG n/N (rate per 1000 screened)	CG n/N (rate per 100 screened)	Effect (95% CI)
				40 to 74 years	negative DM screening but before the next scheduled screening round at 18-months (ages 40 to 54 years), or within 24 months of screening for women who have reached the upper age limit (ages 55 to 74 years)	DCIS	2/13369 (0.1)	4/26738 (0.1)	RR: 1.0 (95% CI, 0.18 to 5.46)
		Richman, 2021[160]	DBT/DM vs. DM	40 to 64	12 months ¶	Invasive	NR (0.52)#	NR (0.45)#	Ajd. proportion difference: 0.07 (99% CI, 0.01 to 0.12)#
		Hovda, 2020[142] OVVV	DBT/sDM (round 1), DM (round 2) vs. DM	50 to 69	6 to 24-months followup among women with a false-positive result at baseline screening	Invasive and DCIS	68/34641 (2.0)	88/57763 (1.5)	Adj RR: 1.30 (95% CI, 0.95 to 1.78)
		Hovda, 2020 OVVV	DBT/sDM (round 1), DM (round 2) vs. DM	50 to 69	24-months	Invasive	63/34641 (1.8)	83/57763 (1.4)	RR: 1.27 (95% CI, 0.91 to 1.76)‖
						DCIS	5/34641 (0.1)	5/57763 (0.1)	RR: 1.67 (95% CI, 0.48 to 5.76)
		Conant, 2016[130] PROSPR	DBT/DM vs. DM	40 to 74	12 months	Invasive and DCIS	68/113061 (0.6)	12/25268 (0.5)	Adj OR: 0.55 (95% CI, 0.13 to 2.26)
Suppl. MRI	RCT	Veenhuizen, 2021[128, 162] DENSE	DM plus MRI vs.DM	50 to 75**	24-months††	Invasive	18/8061 (2.2)	152/32312 (4.7)	RR: 0.47 (95% CI, 0.29 to 0.77)‡
						DCIS	2/8061 (0.2)‡‡	9/32312 (0.3)‡‡	RR: 0.89 (95% CI, 0.19 to 4.12)‡, ‡‡
Suppl. US	RCT	Ohuchi, 2016[156] J-START	DM plus US vs. DM	40 to 49	12-month §§	Invasive	16/36752 (0.4)‡‡	27/35965 (0.8)‡‡	RR: 0.58 (95% CI, 0.31 to 1.08)‡, ‡‡
						DCIS and LCIS	2/36752 (0.1)‡‡	8/35965 (0.2)‡‡	RR: 0.24 (95% CI: 0.05 to 1.15)‡, ‡‡
	NRSI	Lee, 2019[150] BCSC	DM plus US vs. DM	NR	12 months‖‖	Invasive and DCIS	9/6081 (1.5) ‡‡	56/30062 (1.9) ‡‡	Adj RR: 0.67 (95% CI, 0.33 to 1.37) ‡‡, ¶¶

Note: Unless otherwise noted, rates are per 1,000 women screened.
*Screening interval varied by age group (annual 45-49, biennial 50-69)
†6 to 24-months followup among women with a false-positive result at previous screening

Table 12. Rates of Interval Cancers (Invasive Cancer and DCIS) in Studies Comparing Breast Cancer Screening Modalities, per 1,000 Screened*

‡Relative risk calculated from Ns.
§Adjusted for at examination, BCSC registry, facility academic or not, calendar year, race and ethnicity, breast density, first-degree family history, time since last mammogram, and the most severe prior benign biopsy result. Rates are per 1,000 screening examinations.
‖This study is a nonrandomized study based on MBTST participants compared with a contemporary age-matched population cohort. Data are presented per 1,000 mammograms.
¶ Excluded cancers within 5 months of screening
This mammogram-level, multivariate logistic regression analysis was
#Adjusted for use of screening ultrasound, age, time period of index mammogram, time since last mammogram, metro location, hospital referral region, and family history of breast cancer. The analysis cluster-robust standard errors at the person level to account for the correlation of mammogram. Rates are per 1,000 screening examinations.
**with extremely dense breasts
††Before the next scheduled mammogram if less than 24-month interval
‡‡Rates are per 1,000 screening examinations.
§§ Limited to individuals with screening results of no findings or benign findings at round 1
‖‖Before next scheduled mammograph if less than 12-month interval
¶¶Adjusted for site, age, menopausal status, first-degree family history of breast cancer, year of examination, prior benign breast biopsy result, and correlation among women within the same matched set using generalized estimated equations.

Abbreviations: BCSC=Breast Cancer Surveillance Consortium; BI-RADS=Breast Imaging Reporting and Data System; CI=confidence interval; DBT=Digital breast tomosynthesis; DCIS=ductal carcinoma in situ; DM=Digital mammography; DENSE=Dense Tissue and Early Breast Neoplasm Screening; J-START= Japan Strategic Anti-cancer Randomized Trial; LCIS=lobular carcinoma in situ; MBTST=Malmo Breast Tomosynthesis Screening Trial; MRI=magnetic resonance imaging; NRSI=nonrandomized study of intervention; PROSPR=Population-based Research to Optimize the Screening Process through Personalized Regimens; RETomo=Reggio Emilia Tomosynthesis Trial; RR=relative risk; sDM=synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen; TOSYMA=TOmosynthesis plus SYnthesized MAmmography study; OVVV=Oslo-Vestfold-Vestre Viken; UKCCR=United Kingdom Coordinating Committee on Cancer Research trial; US=ultrasound; VDG=Volpara Density Grade

Table 13. Followup of Abnormal Screening in Studies Comparing Digital Breast Tomosynthesis and Digital Mammography

Outcome	Study Design	Author, Year Study/Trial Name	Comparison (IG vs. CG)	Age	Followup	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Effect (95% CI)
Recalled for further assessment	RCT	Armaroli, 2022[127] Proteus Donna*	DBT/DM (round 1), DM (round 2) vs. DM	46 to 68	First Round	1995/30844 (63.4)	2191/43022 (50.9)	RR: 1.24 (95% CI, 1.17 to 1.32)
					Second Round	1000/23760 (42.1)	1456/33534 (43.4)	RR: 0.97 (95% CI, 0.89 to 1.05)
		Pattacini, 2022[158]	RETomo† DBT/DM (round 1), DM (round 2) vs. DM	45 to 69	First Round	511/13356 (38.3)	522/13521 (38.6)	RR: 0.99 (95% CI, 0.88 to 1.10)
					Second Round	464/12733 (36.4)	506/12911 (39.2)	RR: 0.93 (95% CI, 0.82 to 1.10)
		Hofvind, 2021[141] To-Be †	DBT/sDM vs. DM (round 1), DBT/sDM (round 2)	50 to 69	First Round	444/14380 (30.9)	571/14369 (39.7)	RR: 0.78 (95% CI, 0.69 to 0.88)‡
					Second Round	440/11201 (39.3)	441/11105 (39.7)	RR: 0.99 (95% CI, 0.87 to 1.13)‡
	NRSI	Hovda, 2020[142] OVVV †	DBT/sDM (round 1), DM (round 2) vs. DM	50 to 69	First Round	1253/37185 (33.7)	2037/61742 (33.0)	RR: 1.02 (95% CI, 0.95 to 1.09)‡
					Second Round	621/26474 (23.5)	1408/45543 (30.9)	RR: 0.76 (95% CI, 0.69 to 0.83)‡
Percutaneous needle biopsy	RCT	Pattacini, 2022[158] RETomo	DBT/DM (round 1), DM (round 2) vs. DM	45 to 69	First Round	159/13356 (11.9)	110/13521 (8.1)	RR: 1.50 (95% CI, 1.10 to 1.90)
					Second Round	78/12733 (6.1)	104/12911 (8.1)	RR: 0.76 (95% CI, 0.57 to 1.00)
Biopsy§	RCT	Hofvind, 2021[141] To-Be	DBT/sDM vs. DM (round 1), DBT/sDM (round 2)	50 to 69	First Round	252/14380 (17.5)	271/14369 (18.9)	RR: 0.93 (95% CI, 0.78 to 1.10)‡
					Second Round	248/11201 (22.1)	258/11105 (23.2)	RR: 0.95 (95% CI, 0.80 to 1.13)‡
Surgical referrals	RCT	Armaroli, 2022[127] Proteus Donna	DBT/DM (round 1), DM (round 2) vs. DM	46 to 68	First Round	305/30844 (9.9)	276/43022 (6.4)	RR: 1.54 (95% CI, 1.31 to 1.82)‡
					Second Round	103/23760 (4.3)	191/33534 (5.7)	RR: 0.76 (95% CI, 0.59 to 0.97)‡
Surgical procedures (including open biopsy)	RCT	Pattacini, 2022[158] RETomo	DBT/DM vs. DM	45 to 69	First Round	116/13356 (8.7)	68/13521 (5.0)	RR: 1.70 (95% CI, 1.30 to 2.30)
					Second Round	68/12733 (5.3)	83/12911 (6.4)	RR: 0.83 (95% CI, 0.60 to 1.10)

* Recalled for an assessment after double reading based on positive or suspicious screening result by either radiologist (without consensus or arbitration)
† Recalled for an assessment (after double reading and arbitration) based on positive or suspicious screening results.
‡ Relative risk calculated from Ns.
§ Type of biopsy not defined.

Table 13. Followup of Abnormal Screening in Studies Comparing Digital Breast Tomosynthesis and Digital Mammography

Abbreviations: CG=control group; CI=confidence interval; DBT=Digital Breast Tomosynthesis; DM=Digital mammography; IG=intervention group; NRSI=nonrandomized study of intervention; RCT=randomized controlled trial; RETomo=Reggio Emilia Tomosynthesis Trial; RR=relative risk; sDM=synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen; OVVV=Oslo-Vestfold-Vestre Viken

Table 14. False Positives in Studies Comparing Digital Breast Tomosynthesis and Digital Mammography

Outcome	Study Design	Author, Year Study/Trial Name	Comparison (IG vs. CG)	Age	Followup	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Effect (95% CI)
False positive recall†	RCT	Armaroli, 2022[127] Proteus Donna‡	DBT/DM vs. DM (round 1), DM vs. DM‡ (round 2)	46 to 68	First Round	1699/30844 (55.1)	1943/43022 (45.2)	RR: 1.22 (95% CI, 1.14 to 1.30)*
					Second Round	900/23760 (37.9)	1286/33534 (38.3)	RR: 0.99 (95% CI, 0.91 to 1.08)*
		Pattacini, 2022[158] RETomo§	DBT/DM (round 1), DM (round 2) vs. DM	45 to 69	First Round	410/13356 (30.7)	461/13521 (34.1)	RR: 0.90 (95% CI, 0.79 to 1.00)*
					Second Round	403/12733 (31.7)	430/12911 (33.3)	RR: 0.95 (95% CI, 0.83 to 1.09)*
		Hofvind, 2021[141] To-Be§	DBT/sDM vs. DM (round 1), DBT/sDM (round 2)	50 to 69	First Round	349/14380 (24.3)	484/14369 (33.7)	RR: 0.72 (95% CI, 0.63 to 0.83)*
					Second Round	349/11201 (31.2)	340/11105 (30.6)	RR:1.02 (95% CI, 0.88 to 1.18)*
	NRSI	Hovda, 2020[142] OVVV §	DBT/sDM vs. DM (round 1), DM vs. DM (round 2)	50 to 69	First Round	905/37185 (24.3)	1658/61742 (26.9)	RR: 0.91 (95% CI, 0.84 to 0.98)*
					Second Round	518/26474 (19.6)	1154/45543 (25.3)	RR: 0.77 (95% CI, 0.70 to 0.86)*
False positive biopsy‖	RCT	Hofvind, 2021[141] To-Be	DBT/sDM vs. DM (round 1), DBT/sDM (round 2)	50 to 69	First Round	157/14380 (10.9)	184/14369 (12.8)	RR: 0.85 (95% CI, 0.69 to 1.05)*
					Second Round	157/11201 (14.0)	157/11105 (14.1)	RR: 0.99 (95% CI: 0.80 to 1.24)*

*Relative risk calculated from Ns.
† Recalled for assessment without a finding of invasive cancer or DCIS
‡ Recalled for an assessment after double reading based on positive or suspicious screening result by either radiologist (without consensus or arbitration)
§ Recalled for an assessment (after double reading and arbitration) based on positive or suspicious screening results.
‖ Underwent biopsy without a finding of invasive cancer or DCIS

Abbreviations: CG=control group; CI=confidence interval; DBT=Digital Breast Tomosynthesis; DM=Digital mammography; IG=intervention group; NRSI=nonrandomized study of intervention; RCT=randomized controlled trial; RETomo=Reggio Emilia Tomosynthesis Trial; RR=relative risk; sDM=synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen; OVVV=Oslo-Vestfold-Vestre Viken

Table 15. Screen-Detected DCIS Diagnosed in Studies Comparing Digital Breast Tomosynthesis and Digital Mammography

Study Design	Author, Year Study/Trial Name	Follow-up	Modality (previous round modality)	Invasive Cancer Detection*	Effect (95% CI)
RCT	Armaroli, 2022[127] Proteus Donna	First Round	DBT/DM	32/30844 (1.04)	RR=1.39 (95% CI, 0.83 to 2.35)
			DM	32/43022 (0.74)	
		Second Round	DM (DBT/DM)	17/23760 (0.72)	RR=0.75 (95% CI, 0.39 to 1.39)
			DM	32/33534 (0.95)	
	Pattacini, 2022[158] RETomo	First Round	DBT/DM	17/13356 (1.27)	RR=1.91 (95% CI, 0.85 to 4.30)
			DM	9/13521 (0.67)	
		Second Round	DBT/DM	8/12733 (0.63)	RR=0.51 (95% CI, 0.22 to 1.20)
			DM	16/12911 (1.24)	
	Hofvind, 2021[141] To-Be	First Round	DBT/sDM	15/14380 (1.04)	RR=0.94 (95% CI, 0.46 to 1.89)
			DM	16/14369 (1.11)	
		Second Round	DBT/sDM	14/11201 (1.25)	RR=0.99 (95% CI, 0.47 to 2.08)
			DBT/sDM (DM)	14/11105 (1.26)	
NRSI	Hovda, 2020[142] OVVV	First Round	DBT/sDM	65/37185 (1.75)	RR=2.16 (95% CI, 1.49 to 3.12)
			DM	50/61742 (0.81)	
		Second Round	DM (DBT/sDM)	19/26474 (0.72)	RR=0.64 (95% CI, 0.38 to 1.09)
			DM	51/45543 (1.12)	

*Rate per 1,000 women screened

Abbreviations: CI=confidence inerval; DBT=Digital breast tomosynthesis; DCIS=ductal carcinoma in situ; DM=Digital mammography; mm=millimeter; NR=not reported; RETomo=Reggio Emilia Tomosynthesis Trial; RR=relative risk; sDM=synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen; OVVV=Oslo-Vestfold-Vestre Viken

Table 16. Radiation Exposure in Studies Comparing Digital Breast Tomosynthesis and Digital Mammography

Study Design	Author, Year Study/Trial Name	Modality (previous round modality)	Mean (SD) or Median (IQR) glandular dose
RCT	Armaroli, 2022[127] Proteus Donna	DBT/DM vs. DM (round 1); DM vs. DM (round 2)	Combined DBT and DM was approximately 2.5 times higher than that of a standard DM alone*
	Heindal, 2022[137] TOSYMA	DBT/sDM	1.86 mGy (IQR, 1.48 to 2.45 mGy)
		DM	1.36 mGy (IQR, 1.02 to 1.85 mGy)
	Pattacini, 2022[158] RETomo	DBT/DM	6.40 mGy (IQR, 5.68 to 7.36 mGy)
		DM	4.84 mGy (IQR, 4.24 to 5.72 mGy)
	Hofvind, 2021[141] To-Be †	DBT/sDM	2.96 mGy (SD, NR)
		DM (round 1), DBT/sDM (round 2)	2.95 mGy (SD, NR) ‡
NRSI	Johnson, 2021[79] MBTST	DBT/DM	2.3 mGy (SD, 0.7 mGy)
		DM	2.7 mGy (SD, 0.8 mGy)

*Radiation dose description as reported by the study
†Study also noted that radiation dose did not differ with mammographic density, or within the density groups.
‡Test for mean difference between groups p = 0.43

Abbreviations: DBT=Digital breast tomosynthesis; DM=Digital mammography; IQR=interquartile range; MBTST=Malmo Breast Tomosynthesis Screening Trial; mGy=milligray; RETomo=Reggio Emilia Tomosynthesis Trial; sDM=synthetic 2-view mammography; SD=standard deviation; To-Be=Tomosynthesis Trial in Bergen; TOSYMA=TOmosynthesis plus SYnthesized MAmmography study

Table 17. Followup of Abnormal Screening in Randomized Trials Comparing Digital Breast Tomosynthesis and Digital Mammography, by Population Subgroup

Outcome	Author, Year Study/Trial Name	Comparison (IG vs. CG)	Followup	Subgroup	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Effect (95% CI)
Recalled for further assessment	Pattacini, 2022[158] RETomo†	DBT/DM (round 1), DM (round 2) vs. DM	Round 1	Ages 45 to 49	200/5053 (39.6)	206/5103 (40.4)	RR: 0.98 (95% CI, 0.81 to 1.20)
				Ages 50 to 69	311/8303 (37.5)	316/8418 (37.5)	RR: 1.00 (95% CI, 0.86 to 1.20)
			Round 2	Ages 45 to 49	163/4813 (33.9)	195/4855 (40.2)	RR: 0.84 (95% CI, 0.69 to 1.00)
				Ages 50 to 69	301/7920 (38.0)	311/8056 (38.6)	RR: 0.98 (95% CI, 0.84 to 1.10)
	Hofvind, 2021[141] To-Be†	DBT/sDM vs. DM (round 1), DBT/sDM (round 2)	Round 1	VDG1	83/3929 (21.1)	106/3212 (33.0)	RR: 0.64 (95% CI, 0.48 to 0.85)*
				VDG2	199/6216 (32.0)	270/6280 (43.0)	RR: 0.74 (95% CI, 0.62 to 0.89)*
				VDG3	129/3152 (40.9)	146/3655 (39.9)	RR: 1.02 (95% CI, 0.81 to 1.29)*
				VDG4	30/962 (31.2)	45/1136 (39.6)	RR: 0.79 (95% CI, 0.50 to 1.24)*
			Round 2	VDG1	74/3214 (23.0)	79/2960 (26.7)	RR: 0.86 (95% CI, 0.63 to 1.18)*
				VDG2	168/4353 (38.6)	177/4395 (40.3)	RR: 0.96 (95% CI, 0.78 to 1.18)*
				VDG3	141/2656 (53.1)	139/2736 (50.8)	RR: 1.04 (95% CI, 0.83 to 1.31)*
				VDG4	53/900 (58.9)	44/934 (47.1)	RR: 1.25 (95% CI, 0.85 to 1.84)*
Percutaneous biopsy	Pattacini, 2022[158] RETomo	DBT/DM (round 1), DM (round 2) vs. DM	Round 1	Ages 45 to 49	47/5053 (9.3)	31/5103 (6.1)	RR: 1.50 (95% CI, 0.97 to 2.40)
				Ages 50 to 69	112/8303 (13.5)	79/8418 (9.4)	RR: 1.40 (95% CI, 1.10 to 1.90)
			Round 2	Ages 45 to 49	15/4813 (3.1)	30/4855 (6.2)	RR: 0.50 (95% CI, 0.27 to 0.94)
				Ages 50 to 69	63/7920 (8.0)	74/8056 (9.2)	RR: 0.87 (95% CI, 0.62 to 1.20)
Biopsy	Hofvind, 2021[141] To-Be	DBT/sDM vs. DM (round 1), DBT/sDM (round 2)	Round 1	VDG1	47/3929 (12.0)	45/3212 (14.0)	RR: 0.85 (95% CI, 0.57 to 1.28)*
				VDG2	106/6216 (17.1)	132/6280 (21.0)	RR: 0.81 (95% CI, 0.63 to 1.05)*
				VDG3	79/3152 (25.1)	69/3655 (18.9)	RR: 1.33 (95% CI: 0.96 to 1.83)*
				VDG4	19/962 (19.8)	25/1136 (22.0)	RR: 0.90 (95% CI: 0.50 to 1.62)*
			Round 2	VDG1	49/3214 (15.2)	52/2960 (17.6)	RR: 0.87 (95% CI: 0.59 to 1.28)*
				VDG2	89/4353 (20.4)	96/4395 (21.8)	RR: 0.94 (95% CI, 0.70 to 1.25)*
				VDG3	79/2656 (29.7)	84/2736 (30.7)	RR: 0.97 (95% CI, 0.72 to 1.31)*
				VDG4	29/900 (32.2)	25/934 (26.8)	RR: 1.20 (95% CI, 0.71 to 2.04)*
Surgical procedures (including open biopsy)	Pattacini, 2022[158] RETomo	DBT/DM (round 1), DM (round 2) vs. DM	Round 1	Ages 45 to 49	29/5053 (5.7)	14/5103 (2.7)	RR: 2.10 (95% CI, 1.10 to 4.00)
				Ages 50 to 69	87/8303 (10.5)	54/8418 (6.4)	RR: 1.60 (95% CI, 1.20 to 2.30)
			Round 2	Ages 45 to 49	11/4813 (2.3)	22/4855 (4.5)	RR: 0.50 (95% CI, 0.24 to 1.00)
				Ages 50 to 69	57/7920 (7.2)	61/8056 (7.6)	RR: 0.95 (95% CI, 0.66 to 1.40)

*Relative risk calculated from Ns.
†Recalled for an assessment (after double reading and arbitration) based on positive or suspicious screening results.

Abbreviations: CG=control group; CI=confidence interval; DBT=Digital Breast Tomosynthesis; DM=Digital mammography; IG=intervention group; RETomo=Reggio Emilia Tomosynthesis Trial; RR=relative risk; sDM=synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen; OVVV=Oslo-Vestfold-Vestre Viken; VDG=Volpara Density Grade

Table 18. False Positives in Randomized Trials Comparing Digital Breast Tomosynthesis and Digital Mammography, by Population Subgroup

Outcome	Author, Year Study/Trial Name	Comparison (IG vs. CG)	Followup	Subgroup	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Effect (95% CI)
False positive recall†	Pattacini, 2022[158] RETomo	DBT/DM vs. DM	Round 1	Ages 45 to 49	179/5053 (35.4)	194/5103 (38.0)	RR: 0.93 (95% CI, 0.76 to 1.10)*
				Ages 50 to 69	231/8303 (27.8)	267/8418 (31.7)	RR: 0.88 (95% CI, 0.74 to 1.00)*
			Round 2	Ages 45 to 49	154/4813 (32.0)	177/4855 (36.5)	RR: 0.88 (95% CI, 0.71 to 1.09)*
				Ages 50 to 69	249/7920 (31.4)	253/8056 (31.4)	RR: 1.00 (95% CI, 0.84 to 1.19)*
	Hofvind, 2021[141] To-Be	DBT/sDM vs. DM	Round 1	VDG1	65/3929 (16.5)	91/3212 (28.3)	RR: 0.58 (95% CI, 0.43 to 0.80)*
				VDG2	151/6216 (24.3)	231/6280 (36.8)	RR: 0.66 (95% CI, 0.54 to 0.81)*
				VDG3	106/3152 (33.6)	121/3655 (33.1)	RR: 1.02 (95% CI, 0.79 to 1.31)*
				VDG4	24/962 (24.9)	38/1136 (33.5)	RR: 0.75 (95% CI, 0.45 to 1.23)*
			Round 2	VDG1	55/3214 (17.1)	62/2960 (20.9)	RR: 0.81 (95% CI, 0.57 to 1.17)*
				VDG2	132/4353 (30.3)	142/4395 (32.3)	RR: 0.94 (95% CI, 0.74 to 1.19)*
				VDG3	114/2656 (42.9)	105/2736 (38.4)	RR: 1.12 (95% CI, 0.86 to 1.45)*
				VDG4	44/900 (48.9)	29/934 (31.0)	RR: 1.57 (95% CI, 0.99 to 2.49)*
False positive biopsy‡	Hofvind, 2021[141] To-Be	DBT/sDM vs. DM	Round 1	VDG1	21/3929 (5.3)	30/3212 (9.3)	RR: 0.57 (95% CI, 0.33 to 1.00)*
				VDG2	59/6216 (9.5)	93/6280 (14.8)	RR: 0.64 (95% CI, 0.46 to 0.89)*
				VDG3	68/3152 (21.6)	44/3655 (12.0)	RR: 1.79 (95% CI, 1.23 to 2.61)*
				VDG4	17/962 (17.7)	18/1136 (15.8)	RR: 1.12 (95% CI, 0.58 to 2.15)*
			Round 2	VDG1	30/3214 (9.3)	35/2960 (11.8)	RR: 0.79 (95% CI, 0.49 to 1.28)*
				VDG2	53/4353 (12.2)	61/4395 (13.9)	RR: 0.88 (95% CI, 0.61 to 1.26)*
				VDG3	52/2656 (19.6)	50/2736 (18.3)	RR: 1.07 (95% CI, 0.72 to 1.57)*
				VDG4	20/900 (22.2)	10/934 (10.7)	RR: 2.08 (95% CI, 0.98 to 4.41)*

*Relative risk calculated from Ns.
† Recalled for assessment without a finding of invasive cancer or DCIS
‡ Underwent biopsy without a finding of invasive cancer or DCIS

Abbreviations: CG=control group; CI=confidence interval; DBT=Digital Breast Tomosynthesis; DM=Digital mammography; IG=intervention group; RETomo=Reggio Emilia Tomosynthesis Trial; RR=relative risk; sDM=synthetic 2-view mammography; To-Be=Tomosynthesis Trial in Bergen; OVVV=Oslo-Vestfold-Vestre Viken; VDG=Volpara Density Grade

Table 19. Rates of Interval Cancers (Invasive Cancers and DCIS) in Studies Comparing Breast Cancer Screening Modalities, by Population Subgroup

Modality	Study Design	Author, Year Study/ Trial Name	Comparison (IG vs. CG)	Timepoint	Histologic type	Subgroup	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Effect (95% CI)
DBT	RCT	Pattacini, 2022[158] RETomo	DBT/DM vs. DM	12-months (40 to 49 years) or 24-months (50 to 69 years)	Invasive	Age 45 to 49	3/4853 (0.6)	7/4897 (1.4)	RR: 0.43 (95% CI, 0.11 to 1.70)
						Age 50 to 69	16/7992 (2.0)	13/8102 (1.6)	RR: 1.20 (95% CI, 0.60 to 2.60)
					DCIS	Age 45 to 49	0/4853	1/4897 (0.2)	-
						Age 50 to 69	2/7992 (0.3)	1/8102 (0.1)	RR: 2.0 (95% CI, 0.18 to 22.0)
					Invasive	Nondense breasts (BI-RADS A, B)*	6/6051 (1.0)	4/6094 (0.7)	RR: 1.50 (95% CI, 0.43 to 5.30)
						Dense breasts (BI-RADS C, D)*	13/5704 (2.3)	15/5691 (2.6)	RR: 0.86 (0.41 to 1.80)
					DCIS	Nondense breasts (BI-RADS A, B)*	1/6051 (0.2)	2/6094 (0.3)	RR: 0.50 (95% CI, 0.05 to 5.60)
						Dense breasts (BI-RADS C, D)*	1/5704 (0.2)	0/5691	-
		Hofvind, 2021[141] To-Be	DBT/sDM vs. DM	24-months†	Invasive	VDG1 Density‡	2/3930 (0.5)	0/3212	-
						VDG2 Density‡	6/6215 (1.0)	7/6279 (1.1)	RR: 0.87 (95% CI, 0.29 to 2.58)§
						VDG3 Density‡	8/3146 (2.5)	16/3654 (4.4)	RR: 0.58 (95% CI, 0.25 to 1.36)§
						VDG4 Density‡	4/967 (4.1)	5/1136 (4.4)	RR: 0.94 (95% CI: 0.25 to 3.49)§
					DCIS	VDG1 Density‡	0/3930	0/3212	-
						VDG2 Density‡	0/6215	0/6279	-
						VDG3 Density‡	0/3146	1/3654 (0.3)	-
						VDG4 Density‡	0/967	0/1136	-
		Johnson, 2021[79] MBTST‖	DBT/DM vs. DM	18-months (40 to 54 years) or 24 (ages 55 to 74 years)	Invasive and DCIS	Age 40 to 54	8/6289 (1.3)	33/12541 (2.6)	OR=0.5 (95% CI, 0.2 to 1.1)¶
						Age 55 to 74	13/7080 (1.8)	43/14197 (3.0)	OR=0.6 (95% CI, 0.3 to 1.1)¶

Table 19. Rates of Interval Cancers (Invasive Cancers and DCIS) in Studies Comparing Breast Cancer Screening Modalities, by Population Subgroup

Modality	Study Design	Author, Year Study/ Trial Name	Comparison (IG vs. CG)	Timepoint	Histologic type	Subgroup	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Effect (95% CI)
	NSRI	Kerlikowske, 2022 BCSC-2022b	DBT/sDM vs. DM	12 months	Invasive	BI-RADS A*	NR (0.12)	NR (0.24)	Adj. rate difference: -0.12 (95% CI, -0.31 to 0.07)
						BI-RADS B*	NR (0.31)	NR (0.39)	Adj. rate difference: -0.08 (95% CI, -0.24 to 0.07)
						BI-RADS C*	NR (0.99)	NR (0.87)	Adj. rate difference: 0.11 (95% CI, -0.11 to 0.34)
						BI-RADS D*	NR (0.87)	NR (1.21)	Adj. rate difference: -0.34 (95% CI, -0.76 to 0.07)
		Richman, 2021[160]	DBT/DM vs. DM	12 months **	Invasive	Age 40 to 44	NR (0.6)††	NR (0.5)††	Adj. proportion difference: 0.04 (99% CI, -0.13 to 0.21)‡‡
						Age 45 to 49	NR (0.6)††	NR (0.5)††	Adj. proportion difference: 0.03 (99% CI, -0.12 to 0.17)‡‡
						Age 50 to 54	NR (0.7)††	NR (0.5)**	Adj. proportion difference: 0.18 (99% CI, 0.04 to 0.32)‡‡
						Age 55 to 59	NR (0.5)††	NR (0.5)††	Adj. proportion difference: 0.02 (99% CI, -0.09 to 0.13)‡‡
						Age 60 to 64	NR (0.6)††	NR (0.5)††	Adj. proportion difference: 0.08 (99% CI, -0.05 to 0.22)‡‡
Suppl. US	RCT	Ohuchi, 2016[136, 156] J-START	DM plus US vs. DM	12-month	Invasive and DCIS	Nondense breasts (BI-RADS A, B)*	2/3908 (0.5)	9/3915 (2.3)	RR: 0.22 (95% CI, 0.05 to 1.03)§
						Dense breasts (BI-RADS C, D)*	3/5797 (0.5)	10/5593 (1.8)	RR: 0.29 (95% CI, 0.08 to 1.05)§

*BI-RADS uses visual assessment of to categorize breast density as (A) almost entirely fatty; (B) scattered areas of fibroglandular density; (C) heterogeneously dense, and (D) extremely dense.
†6 to 24-months followup among women with a false-positive result at previous screening
‡ The Volpara system uses a quantitative measure of volumetric breast density and assigns density to one of four categories (Volpara density grade [VDG] 1 to 4), which are analogous to BI-RADS A to D.
§ Relative risk calculated from Ns.
|| This study is a nonrandomized study based on MBTST participants compared with a contemporary age-matched population cohort.c

Table 19. Rates of Interval Cancers (Invasive Cancers and DCIS) in Studies Comparing Breast Cancer Screening Modalities, by Population Subgroup

¶ Age-adjusted odds ratio
** Excluded cancers diagnosed within 5 months of screening
†† Models adjusted for use of screening ultrasound, time period of index mammogram, time since last mammogram, metro location, hospital referral region, and family history of breast cancer. Rates and risk differences reported in table are adjusted per 100,000 from original study reported rates of a per 1,000 mammogram scale.
‡‡ Interaction between screening type and all age group from multivariable logistic regression model was p=0.54

Abbreviations: BCSC=Breast Cancer Surveillance Consortium; BI-RADS=Breast Imaging Reporting and Data System; CI=confidence interval; DBT=Digital breast tomosynthesis; DCIS=ductal carcinoma in situ; DM=Digital mammography; DENSE=Dense Tissue and Early Breast Neoplasm Screening; J-START=Japan Strategic Anti-cancer Randomized Trial; MBTST=Malmo Breast Tomosynthesis Screening Trial; MRI=magnetic resonance imaging; RETomo=Reggio Emilia Tomosynthesis Trial; NRSI=nonrandomized study of intervention; RR=relative risk; sDM=synthetic mammography; PROSPR=Population-based Research Optimizing Screening through Personalized Regimens; To-Be=Tomosynthesis Trial in Bergen; TOSYMA=TOmosynthesis plus SYnthesized MAmmography study; OVVV=Oslo-Vestfold-Vestre Viken; UKCCR=United Kingdom Coordinating Committee on Cancer Research trial; US=ultrasound; VDG=Volpara Density Grade

Table 20. Downstream Consequences of Supplemental Screening With MRI or Ultrasound

Modality	Author, Year Study/Trial Name Study Design	Comparison	Population	Outcome	Followup	IG n/N (rate per 1000 screened)	Details
Suppl. MRI	Veenhuizen, 2021[27] DENSE RCT	DM+MRI vs. DM	Women aged 50-75 years of age with negative mammography results (BI-RADS radiographic score of 1 or 2) and extremely dense breast tissue	Serious adverse event	0-days	5/4783 (1.0)	2 vasovagal reactions, 3 allergic reactions to contrast agent
					30-days	27/4783 (5.6)	Events reported regardless of relatedness to screening MRI: 27 serious adverse events (required emergency department visit or unplanned hospital admission: 5 nervous system disorders, 2 GI disorders, 2 skin/subcutaneous tissue disorders, 2 cardiovascular disorders, 7 musculoskeletal disorders, 3 ear/nose/throat/eye disorders, 1 general disorder, 1 respiratory disorder, 1 urologic disorder, 1 reproductive system and breast disorder/complaint, 2 medical procedures (complications during after biopsy procedure)
				Adverse event	0-days	3/4783 (0.6)	2 extravasation of contrast agent, 1 subluxation shoulder
					30-days	1233/4783 (257.8)	Events reported regardless of relatedness to screening MRI. Most common listed were nervous system disorder, GI disorder, psychosocial/psychiatric disorder, musculoskeletal disorder, respiratory disorder.
				Recall	0-days	454/4783 (94.9)	
				False-positive recall	0-days	375/4700 (80.0)	
				Biopsy	0-days	300/4783 (62.7)	
	Ganguli, 2022[133]	MRI vs. DM	Women aged 40-64 years who had	Laboratory tests due to	6-months	NR/9208 (12)	Rate difference: 3.2 (95% CI, -2.2 to 8.5)

Table 20. Downstream Consequences of Supplemental Screening With MRI or Ultrasound

Modality	Author, Year Study/Trial Name Study Design	Comparison	Population	Outcome	Followup	IG n/N (rate per 1000 screened)	Details
	NRSI		a bilateral breast MRI or bilateral screening mammogram claim	extramammary findings			
				Imaging tests due to extramammary findings	6-months	NR/9208 (3)	Rate difference: 0.5 (95% CI, -2.7 to 1.8)
				Procedures following extramammary findings	6-months	NR/9208 (1)	Rate difference: 1.0 (95% CI, -0.1 to 2.0)
				New diagnoses following extramammary findings	6-months	NR/9208 (0.5)	Rate difference: 0.3 (95% CI, -0.5 to 1.1)
				All extramammary cascade events	6-months	NR/9208 (31)	Rate difference: 19.6 (95% CI, 8.6 to 30.7)
Suppl. US	Ohuchi, 2016[156] J-START RCT	DM+US vs. DM	Women aged 40 to 49 years	Recall rate for positive ultrasound only	First round	1826/36752 (49.7)	
				False-positive recall for positive ultrasound only	First round	1765/36752 (48.0)	
			Women aged 40 to 49 years with nondense breasts	Recall rate for positive ultrasound only	First round	154/3908 (39.4)	
			Women aged 40 to 49 years with dense breasts	Recall rate for positive ultrasound only	First round	404/5797 (69.7)	
	Lee, 2019[150] BCSC NRSI	DM+US vs. DM	Women undergoing screening at eligible BCSC sites	Biopsy	First round	NR (57)	RR=2.05 (95% CI, 1.79 to 2.34)
				False-positive biopsy recommendation	First round	NR (52)	RR=2.23 (95% CI, 1.93 to 2.58)
				Short-interval imaging follow-up	First round	NR (0.4)	RR=3.1 (95% CI, 2.6 to 3.7)

Abbreviations: BCSC=Breast Cancer Surveillance Consortium; DM=Digital mammography; DENSE=Dense Tissue and Early Breast Neoplasm Screening; J-START= Japan Strategic Anti-cancer Randomized Trial; MRI=magnetic resonance imaging; NR=not reported; NRSI=non-randomized study of intervention; US=ultrasound

Table 21. Summary of Evidence*

Key Question (KQ) Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
KQ1 Age to start or stop screening	Age to start: k = 0 Age to stop: k = 1 (n = 264,274) NRSI, Fair quality	Age to start: NA Age to stop: Screening from age 70 to 74: 8-year risk of breast cancer mortality was 1 fewer death per 1,000 women who continued screening (RD: -1.0 [95% CI -2.3 to 0.1]). Adjusted hazard ratio suggested a 22 percent lower hazard of 8-year breast cancer mortality with continued screening (aHR: 0.78 [95% CI 0.63 to 0.95]) Screening beyond age 74: No difference in 8-year estimated risk in breast cancer mortality (RD: 0.07 [95% CI -0.93 to 1.3]; aHR: 1.00 [95% CI 0.83 to 1.19]) with continued screening.	Age to start: NA Age to stop: NA (for consistency) Imprecise	Advanced statistical methods to emulate per protocol trial; differences in estimates of effects depending on adjustments used. Risk of bias from unmeasured confounding and selection.	**Insufficient**	US Medicare A, B enrollees ages 70 to 84 in the years 1999 to 2008 with high probability of living >10 years; Population over 90% White non-Hispanic
KQ1 Screening interval	Annual vs. Triennial: k=1 (n = 14,765) NRSI, Fair quality Annual vs. Biennial: k = 0	Annual vs. triennial: No difference in breast cancer mortality (RR 1.14, 95% CI 0.59 to 1.27) or all-cause mortality (RR 1.20, 95% CI 0.99 to 1.46) at 13 years. Annual vs. biennial: NA	Annual vs Triennial: NA (for consistency) Imprecise Annual vs. biennial: NA	Assignment based on birth year, limited information baseline characteristics, potential risk of bias due to unmeasured confounding or selection.	**Insufficient**	Invitation to annual or triennial film mammography for ages 40 to 49 in Finnish national screening program; treatment advances since the study conducted (1985-1995). No reporting of participant characteristics.
KQ1 DBT vs DM	k = 0	NA	NA	NA	**Insufficient**	NA
KQ1 Supplemental screening with MRI	k = 0	NA	NA	NA	**Insufficient**	NA
KQ1 Supplemental screening with ultrasound	k = 0	NA	NA	NA	**Insufficient**	NA

Table 21. Summary of Evidence*

Key Question (KQ) Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
KQ1 Personalized screening	k = 0	NA	NA	NA	**Insufficient**	NA
KQ2 Age to start or stop screening	k = 0	NA	NA	NA	**Insufficient**	NA
KQ2 Screening Interval	Annual vs. Triennial: k = 1 (n = 76,022) RCT, Fair quality Annual vs. Biennial: k = 1 (n = 15,440) NRSI Fair quality	Annual vs. Triennial: More invasive cancers screen-detected over 3 years with annual screening screen (RR: 1.64 [95% CI, 1.28 to 2.09]) Total number of invasive cancers similar between groups (RR: 1.16, 95% CI 0.96 to 1.40); no statistical differences by screening interval in tumor size, nodal status, grade, or prognostic index for all cancers diagnosed. Annual vs. Biennial: No difference in risk of stage IIB+ or less favorable prognosis cancers diagnosed after a biennial compared with annual interval for any age group.	Annual vs. Triennial: NA (for consistency) Imprecise Annual vs. Biennial: NA (for consistency) Imprecise	Annual vs. Triennial: Birth month used to assign intervention group first two years of trial, which could introduce bias, no reporting of participant characteristics. Study never reported mortality outcome as planned. Annual vs. Biennial: Risk of bias due to limited adjustment for confounding and potential unmeasured confounding and selection into study groups.	Annual vs. Triennial: **Low** for greater detection of invasive cancer and no difference in tumor characteristics with annual screening Annual vs biennial: **Insufficient**	Annual vs. Triennial: People ages 50 to 62 screened in UK screening program 1989 to 1996; changes in population health, cancer treatment, screening modalities. No reporting of participant characteristics. Annual vs. Biennial: Conducted using BCSC data linked with US SEER and other tumor registry sources; Ages 40 to 85; >77% population White non-Hispanic

Table 21. Summary of Evidence*

Key Question (KQ) Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
KQ2 DBT vs DM	k = 3 (n = 130,196) RCT 2 Good quality, 1 Fair quality k = 1 (n = 98,927) NRSI 1 Fair quality	Three RCTs reported higher invasive cancer detection in first screening round with DBT (RCT pooled RR 1.41, 95% CI 1.20 to 1.64, I^2 8%, n = 129,492) with absolute differences in the trials ranging from 0.6 to 2.4 additional cancers per 1000 screened. Similar results were seen in the NRSI (2.3 per 1000 screened). No difference was found at round two (RCT pooled RR 0.87, 95% CI 0.73 to 1.05, I^2 0%, n = 105,064). The NRSI found higher detection at round two for the study group screened with DM at round one (1.3 per 1000 screened). No clear evidence of stage shift (i.e., reduction in more advanced cancer at subsequent screening). 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.	Detection of invasive cancer: Consistent Precise Stage Shift: Consistent, Imprecise	At round 2 screening three studies used DM for both arms and one RCT used DBT for both arms. The NRSI used a concurrent geographic comparison group design within a national organized screening program; results were unadjusted. The fair-quality RCT did not describe randomization procedures and balance in baseline characteristics could not be assessed due to limited reporting.	**Moderate** for increased detection with DBT at an initial screening with the modality **Low** for absence of stage shift	All studies conducted in European countries with national organized screening programs (Italy, Sweden, Norway) that use independent dual reading and consensus procedures different from US practice. Some studies used DBT paired with DM and some used DBT paired with sDM. Prior readings were generally available. All studies had limited reporting of participant characteristics with no data on racial and/or ethnic characteristics.
KQ2 Supplemental screening with MRI	k = 0	NA	NA	NA	**Insufficient**	NA

Table 21. Summary of Evidence*

Key Question (KQ) Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
KQ2 Supplemental screening with ultrasound	k = 0	NA	NA	NA	**Insufficient**	NA
KQ2 Personalized screening	k = 0	NA	NA	NA	**Insufficient**	NA
KQ3 Age to start or stop screening	<u>Age to start:</u> k = 0 <u>Age to stop:</u> k = 1 (n = 264,274) NRSI, Fair quality	Fewer cancers diagnosed in stop screening strategy; possible overdiagnosis with continued screening. Cancers diagnosed in stop screening strategy more likely to receive aggressive treatments (radical mastectomy and chemotherapy versus lumpectomy and radiotherapy)	Age to start: NA Age to stop: NA (for consistency) Imprecise	Advanced statistical methods to emulate per protocol trial; differences in estimates of effects depending on adjustments used. Risk of bias from unmeasured confounding and selection.	**Insufficient**	US Medicare A, B enrollees ages 70 to 84 in the years 1999 to 2008 with high probability of living >10 years; Population over 90% White non-Hispanic

Table 21. Summary of Evidence*

Key Question (KQ) Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
KQ3 Screening Interval	Annual vs. Triennial: k = 1 (n = 76,022) RCT, Fair-quality; k = 1 (n = 14,765) NRSI, Fair-quality Annual vs. Biennial: k = 3 (n = 920,954) NRSI, Fair-quality	Annual versus triennial: Interval cancers: One RCT (n= 76,022) estimated one fewer invasive interval cancer in the annual screening arm (1.8 vs 2.7 per 1,000 screened; RR: 0.68 [95% CI 0.50 to 0.92]). One NRSI (n = 14,765) using birth year to assign screening intervals found no difference in interval cancer incidence (p = 0.22). False positives: NR Annual vs biennial: Interval cancers: One NRSI using BCSC data (n = 15,440) reported the unadjusted percent with interval cancer for people screened negative after an annual (22.2%) or biennial screening (27.2%) interval. False positive recall and biopsy: FP recall and biopsy higher with annual compared with biennial screening. One analysis of BCSC data (n = 903.495) reported that over 10-years of DBT screening approximately 50% of those undergoing annual screening had at least one false positive recall, compared with approximately 35% of those undergoing biennial 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). One NRSI (n = 2,019) reported >2 times higher odds (OR 2.2, 95% CI 1.7 to 2.8) of a false positive result over a median of 8.9 years.	Annual vs triennial: Interval cancer: Inconsistent, Precise False positives: NA Annual vs. Biennial: Interval cancer: NA (for consistency) Imprecise False positives: Consistent, Precise	Annual vs. Triennial: RCT did not use random allocation first 2 years of study (birth month) and NRSI assigned interval based on birth year (odd, even), lack of information on group baseline characteristics in both studies; potential risk of bias due to unmeasured confounding or selection. Annual vs Biennial: NRSIs used EMR, registry, and self-report data; the largest study (n = 903,945) did not provide information participant characteristics; data on interval cancers is unadjusted for participant characteristics; risk of bias from potential selection and confounding bias, including time varying factors. BCSC NRSI with cumulative FP did not include prevalence screens, may underestimate cumulative FP from start of screening	Annual vs triennial: **Low** for a small difference interval cancer with annual screening; **Insufficient** for other harms Annual vs. Biennial: **Insufficient** for interval cancers; **Moderate** for higher recall, biopsy, and false positives with annual screening	Annual vs triennial: Both triennial screening interval studies conducted in Europe in 1990s; RCT women screened ages 50 to 62; NRSI among women screened ages 40 to 49. No information on participant characteristics other than age. Annual vs biennial: screening studies conducted in US, one NRSI conducted in a single academic medical center that reported FP recall had majority Hispanic population (76%). BCSC data occurred in primarily non-Hispanic White participants (78%).

Table 21. Summary of Evidence*

KQ3 DBT versus DM	k = 4 (n = 229,830) RCT 3 good-quality 1 fair quality k = 6 (n = 6,231,055) NRSI fair-quality	Interval cancers: Three RCTs did not find difference in interval cancer rates (pooled RR = 0.87, 95% CI 0.64 to 1.17, k = 3 RCT, n = 130,196, I^2 = 0%); five NRSI had inconsistent results - three did not find differences, one commercial claims registry study reported more interval cancers with DBT (adj difference: 0.07 per 1000 screens, 99% CI 0.01 to 0.12), and one comparing trial participants to an age-matched population reported fewer interval cancers with DBT (2.7 v 1.4 per 1,000, RR 0.53, 95% CI 0.32 to 0.87). Recall and false positive recalls: Three RCTs and one NRSI reported on recall rates and false positive recall rates at two rounds of screening. Results were mixed, with heterogeneity between studies. One NRSI reported the 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%). Biopsy and false positive biopsy: Two RCTs reported biopsy rates from 2 rounds of screening with no cumulative differences. One trial reported no significant difference in false positive biopsy. One NRSI reported the cumulative probability of at least one false-positive biopsy over 10 years of screening and suggested no difference in cumulative FP biopsy for DBT v DM regardless of screening interval (11-12% annual, 7-8% biennial). Over-detection: Three 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	Interval cancer: Consistent, Imprecise Recall and false-positive recall: Inconsistent, Precise Biopsy and false-positive biopsy: Consistent, Imprecise Overdetection/ overtreatment: Consistent, Imprecise Adverse events: NA (for consistency), Imprecise Radiation: Consistent, Imprecise	None of the RCT maintained the same IG and CG modality both rounds. Intervention changed for IG or CG at round 2 om RCTs - two screened all with DM and one all with DBT. NRSI had substantial risk of bias, limited adjustment for potential confounding and selection. Most NRSI included retrospective assessment screening from records; limited adjust for all factors that may contribute to DBT vs. DM screening, including time dependent factors. One NRSI age-matched trial participants with controls from general screening population lacked adjustment for any factor other than age. NRSI with geographical comparator did not describe characteristics by study group, only minor statistical adjustment elevated selection and confounding risk of bias concerns.	Interval cancer: **Moderate** for no difference Recall and false-positive recall: **Low** for no difference Biopsy and false positive biopsy: **Low** for no difference Overdetection/overtreatment: **Low** for no difference Adverse events: **Insufficient** Radiation: **Moderate** for increased radiation with DBT/DM and no increased radiation with DBT/sDM	No US-based RCTs; European RCTs and NRSI based in organized screening programs and use independent dual mammography reading, consensus. Limited reporting on population characteristics, including no racial/ethnic data. Of the 6 NRSI, four were conducted in the US and 1 each in Sweden and Norway. The US-based studies included data from the BCSC, medical administrative claims, and the multisite PROSPR study. Participants characteristics were only reported in two US studies with ~76% non-Hispanic White participants. The Swedish and Norwegian studies reported no baseline characteristics.

Table 21. Summary of Evidence*

Key Question (KQ) Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
		0.49 to 1.14, k = 3 RCT, n = 130,196, I^2 = 0%) Adverse events: One RCT (n = 99,634) reported the same number of adverse events for both screening tests (DBT/sDM vs DM) (n=6), all nonserious. Radiation exposure: In 3 studies using DBT/DM the dose was ~2x mGy higher than DM; for 2 studies using DBT/sDM the dose was similar to DM only.		.		

Table 21. Summary of Evidence*

Key Question (KQ) Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
KQ3 Supplemental screening with MRI	k = 1 (n = 40,373) RCT Good-quality k = 1 (n = 18,416) NRSI Fair-qualit	Interval cancer: One RCT reported reduced invasive interval cancer 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) Adverse events: RCT reported 8 adverse events (5 serious) during or immediately after MRI - vasovagal reactions, allergic reactions to contrast agent, leaking of contrast agent (extravasation), shoulder subluxation Downstream consequences of supplemental imaging (including incidental findings): MRI resulted in additional recall (95 per 1,000 screened), FP recall (80 per 1,000) and biopsy (63 per 1,000 screened) that did not occur for the DM only group. RCT did not report on incidental findings from MRI. One NRSI reported no difference in new diagnoses unrelated to breast conditions. Events unrelated to breast diagnostic codes were higher in the MRI group (304.5 per 100) than in the mammography group (284.8 per 100), and the adjusted difference between groups (19.6 per 100, 95% CI 8.6 to 30.7) was mostly comprised of additional health care visits.	Interval cancer: NA (for consistency) Precise Adverse events: NA (for consistency), Imprecise Downstream consequences: Consistent, Imprecise	In the trial, 59 percent of those invited to MRI screening attended; possible unmeasured differences between population invited to screening and those attending (e.g., breast cancer risk, concerns about FP and overdiagnosis). Usual care DM outcomes for round two not available limiting interpretation of any screening results The NRSI was based on US insurance claims with no clinical data to determine if followup was causally linked to breast screening.	Interval cancer: **Low** for reduced interval cancers with invitation to MRI Adverse events: **Insufficient** Downstream consequences: **Low** for increased followup	RCT conducted in The Netherlands through organized biennial breast screening program. Limited to women with extremely dense breasts identified using Volpara (category D). Study randomized people with extremely dense breasts to MRI screening *invitation* – provides estimates of likely response and effects of invitation to MRI No data on race or ethnicity for either study population. In the NRSI 50% of individuals had a family history of breast cancer or genetic susceptibility.

Table 21. Summary of Evidence*

Key Question (KQ) Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
KQ3 Supplemental screening with ultrasound	k = 1 (n = 72,717) RCT Fair-quality k = 1 (n = 18,562) NRSI Fair-quali	Interval cancer: RCT of supplemental US did not find statistical difference in invasive interval cancer (0.4 (DM/US) versus 0.8 (DM) per 1,000 screened; RR 0.58, 95% CI 0.31 to 1.08) nor did NRSI using BCSC data (1.5 (DM/US) vs. 1.9 (DM) per 1,000 screened; aRR 0.67, 95% CI 0.33 to 1.37) Downstream consequences of supplemental imaging (including incidental findings): The RCT reported recall attributable to positive findings only on US resulting in an additional 50 recalls per 1000 screened of which 48 were false positives. Incidental findings were not reported. A BCSC NRSI found that referral to biopsy and false positive biopsy were twice as high for those who underwent ultrasound.	Interval cancer: Consistent, Imprecise Downstream consequences: Consistent, Imprecise	Interval cancers rare in young women enrolled in RCT (age 40-49), limited power to detect differences. Population-averaged GEE effect estimate for interval cancer reported in RCT including DCIS lesions (23% of CG interval tumors) was statistically significant; second round results not yet published NRSI used propensity score matching to adjust for potential confounding by indication for screening; unmeasured confounding may still affect results. Ultrasound and mammography results not reported separately, therefore attribution of follow up specifically to US screening not possible.	Interval cancer: **Low** for no difference Downstream consequences: **Low** for increased follow up with ultrasound	RCT conducted in Japan; included people ages 40 to 49; 23% of study population prevalence screened; 58% reported to have dense breasts, distribution not reported; US and DM results interpreted independently; performance could differ if considered together BCSC NRSI included population representative of US overall; age included 30 to 80+ years; inadequate numbers for comparisons of effects by race and ethnicity (80% White, non-Hispanic); 31% had a first-degree family history of breast cancer.

*Summary of evidence for subgroup populations are available in Appendix F Table 7.

Screening Effectiveness

Nine fair-quality RCTs comparing mammography screening with nonscreening provided outcomes that addressed several key questions in the 2016 review. Trials enrolling over 600,000 women were conducted in the United States,[277] Canada,[278] United Kingdom,[279] and Sweden.[280-282] Across all trials, the mean or median screening intervention time ranged from 3.5 to 14.6 years, case accrual time from 7.0 to 17.4 years, and followup time from 11.2 to 21.9 years. To account for clinical heterogeneity and obtain clinically meaningful estimates, the analyses were stratified by age group (39 to 49 years, 50 to 59 years, 60 to 69 years, 70 to 74 years, or ≥75 years). Meta-analyses of breast cancer mortality outcomes using the longest followup data available indicated:

- For women ages 39 to 49 years, the combined RR for breast cancer mortality was 0.92 (95% CI, 0.75 to 1.02; 9 trials); absolute breast cancer mortality reduction was 2.9 (95% CI, – 0.6 to 8.9) deaths prevented per 10,000 women over 10 years. None of the trials indicated statistically significantly reduced breast cancer mortality with screening, including the UK Age trial, the largest (*N*=160,921) and most recent RCT designed specifically to determine the effectiveness of screening women in their 40s. Results of the UK Age trial included in the meta-analysis reflected 17.5 years of followup.[279] A subsequent publication in 2020 indicated a similar lack of breast cancer mortality reduction after 22.8 years of followup (RR, 0.88; 95% CI, 0.74 to 1.03).[77]
- For age 50 to 59 years, the combined RR for breast cancer mortality was 0.86 (95% CI, 0.68 to 0.97; 7 trials); absolute breast cancer mortality reduction was 7.7 (95% CI, 1.6 to 17.2) deaths prevented per 10,000 women over 10 years.
- For age 60 to 69 years, the combined RR for breast cancer mortality was 0.67 (95% CI, 0.54 to 0.83; 5 trials); absolute breast cancer mortality reduction was 21.3 (CI, 10.7 to 31.7) deaths prevented per 10,000 women over 10 years.
- Combining results across women aged 50 to 69 years indicated a RR of 0.78 (95% CI, 0.68 to 0.90; I^2=41.0%; p=0.118).
- For age 70 to 74 years the combined RR for breast cancer mortality was 0.80 (CI, 0.51 to 1.28; 3 trials); absolute breast cancer mortality reduction was 12.5 (CI, –17.2 to 32.1). However, these estimates were limited by low numbers of events from only 3 trials that had smaller sample sizes of women in this age group.
- All-cause mortality was not reduced with screening for any age group.
- Mortality results by risk factors other than age or by screening intervals were not provided.

The screening trials reported several measures of intermediate breast cancer outcomes, however, most comparisons between screening and control groups provided results for relatively early stages of disease, rather than advanced stages. Meta-analyses of outcomes included:

- Combining estimates based on definitions corresponding to Stage II disease or higher (Stage II+, size ≥20 mm, 1+ positive lymph node) indicated no significant reductions in advanced disease for women ages 39 to 49 or 50 years and older.
- When thresholds were defined by the most severe disease categories available from the trials (Stage III + IV disease, size ≥50 mm, 4+ positive lymph nodes), meta-analysis indicated no reductions for age 39 to 49 years (RR 0.98; 95% CI, 0.74 to 1.37; 4 trials); but reduced risk

of advanced cancer in the screening group for age 50 years and older (RR, 0.62; 95% CI, 0.46 to 0.83; 3 trials).
- No RCTs evaluated the incidence of advanced breast cancer outcomes and treatment based on risk factors or screening intervals.

Observational studies of population-based mammography screening reported a wide range of reductions in breast cancer death. Most studies were conducted in Europe or the United Kingdom and included women ages 50 to 69 years. In general, observational studies reported greater breast cancer mortality reduction (25% to 31% among women invited to screening) than RCTs (19% to 22% using intention-to-treat analysis) for women ages 50 to 69 years. Two observational studies of women in their 40s invited to or participating in screening indicated 26 to 44 percent reduction in breast cancer mortality. Observational studies also reported mixed and varied results regarding detection of earlier versus later stage breast cancer with screening and were considered inconclusive. Similarly, studies of screening and treatment morbidity were inconclusive.

Studies comparing different screening modalities (digital mammography, tomosynthesis, ultrasound, or MRI) for women not at high risk for breast cancer did not report breast cancer specific or all-cause mortality outcomes. In studies comparing tomosynthesis and digital mammography versus mammography alone, detection rates were higher with tomosynthesis, but there were no differences in tumor size, stage, or node status.

Screening Harms

Harms of screening summarized in the 2016 evidence review included false-positive and false-negative results, additional imaging, and biopsy; overdiagnosis; anxiety, distress, and other psychological responses; pain and discomfort; and radiation exposure.

Rates of false-positive and false-negative results, additional imaging, and biopsy were determined from a primary analysis of data from the BCSC specifically for the USPSTF that included regularly screened women in the United States using digital mammography based on results from a single screening round:[155]

- False-positive mammography rates were highest among women ages 40 to 49 years (121.2 per 1,000 women; 95% CI, 105.6 to 138.7) and declined with age. False-negative rates were low across all age groups (age 40 to 49 years; 1.0 per 1,000 women; 95% CI, 0.9 to 1.2).
- Rates of recommendations for additional diagnostic imaging were highest among women age 40 to 49 years (124.9 per 1,000 women; 95% CI, 109.3 to 142.3) and decreased with age, while rates of recommendations for biopsy did not differ between age groups (age 40 to 49 years; 16.4 per 1,000 women; 95% CI, 13.2 to 20.3).
- Rates of invasive breast cancer were lowest among women aged 40 to 49 years (2.2 per 1000 women; 95% CI, 1.8 to 2.6) and increased across age groups ($P < 0.001$). For every case of invasive breast cancer detected by mammography screening in women age 40 to 49 years, 464 women had screening mammography, 58 were recommended for additional diagnostic imaging, and 10 were recommended for biopsies. These estimates declined with age for all three outcomes, indicating lower NNS for older women.

- False-positive and negative rates and recommendations for additional imaging and biopsy were generally higher, and measures of NNS lower, for women with risk factors, although these varied slightly across age groups (first-degree relatives with breast cancer; heterogeneous breast density; previous benign biopsy; white or Hispanic race; premenopausal; low BMI).
- Rates of false-positive results, false-negative results, and recommendations for additional imaging did not differ in comparisons of time since the last mammography screening (9 to 18 months versus 19 to 30 months).

A published study of BCSC data that provided results of screening over a 10-year period indicated that when screening began at age 40 years, cumulative rates of false-positive mammography and benign biopsy results were higher for annual than biennial screening (mammography, 61% versus 42%; biopsy, 7% versus 5%).[143] A second analysis of BCSC data reported that 10-year cumulative rates of false-positive mammography results and biopsy were highest among women with a family history of breast cancer, heterogeneously dense or extremely dense breasts, and combination hormone therapy use.[170]

In the few studies comparing screening modalities, four of five observational studies demonstrated statistically significantly lower rates of recall for tomosynthesis and mammography compared with mammography alone. A U.S. study comparing tomosynthesis and mammography with mammography alone reported a reduction of 16 recalls per 1,000 women and an increase in cancer detection of 1.2 cases per 1,000 women, but also an increase of 1.3 biopsies per 1,000 women. Another U.S. study reported a 38 percent reduction in recall rates when tomosynthesis was added to digital mammography versus mammography alone.

In an extensive literature described in the 2016 evidence review,[283] estimates of overdiagnosis ranged from non-existent to nearly 50 percent of diagnosed breast cancer cases. Methods for estimating overdiagnosis varied in many ways, particularly by the type of comparison groups, assumptions about lead time, and the denominator used to calculate rates. In general, most adjusted estimates of overdiagnosis based on trials ranged from 11 to 22 percent, while estimates based on observational studies ranged from 1 to 10 percent. Estimates from statistical models ranged from 0.4 to 50 percent.

No studies directly measured the association between radiation exposure from mammography screening and the incidence of breast cancer and death for film, digital, or tomosynthesis. Models calculate the number of deaths due to radiation induced cancer using estimates for digital mammography is between 2 per 100,000 in women age 50 to 59 years screened biennially, and up to 11 per 100,000 in women ages 40 to 59 years screened annually.[284]

Appendix A Table 1. Summary of Mortality Reductions From Mammography Screening Compared With No-Screening, by Age (Data From Nelson, 2016)

Age	Breast Cancer Mortality Reduction: Relative Risk (95% CI)	Deaths Prevented With Screening 10,000 Women Over 10 Years (95% CI)
39-49 years	0.92 (0.75 to 1.02)	2.9 (–0.6 to 8.9)
50-59 years	0.86 (0.68 to 0.97)	7.7 (1.6 to 17.2)
60-69 years	0.67 (0.54 to 0.83)	21.3 (10.7 to 31.7)
70-74 years	0.80 (0.51 to 1.28)	12.5 (–17.2 to 32.1)
50-69 years	0.78 (0.68 to 0.90)	12.5 (5.9 to 19.5)

Appendix A Table 2. Summary of Results: Estimated Benefits of Screening (Data From Nelson, 2016)

Age, years	Reduction in breast cancer deaths from RCTs; RR (95% CI)*	Breast cancer deaths prevented per 10,000 over 10 years (95% CI)*	Reduction in breast cancer deaths from observational studies; RR (95% CI)	Reduction in all-cause deaths from RCTs; RR (95% CI)*	Reduction in advanced breast cancer from RCTs; RR (95% CI)	Reduction in treatment morbidity from RCTs; RR (95% CI)†
40-49	0.88 (0.73 to 1.003) 0.84 (0.70 to 1.002)	4 (0 to 9)	0.74 (0.66 to 0.83); 0.56 (0.45 to 0.67)‡	0.99 (0.94 to 1.06)	0.98 (0.74 to 1.37)	Screening results in more mastectomies 1.20 (1.11 to 1.30) and radiation 1.32 (1.16 to 1.50); the majority of cases from screening are DCIS and early stage.
50-59	0.86 (0.68 to 0.97) 0.86 (0.69 to 1.007)	5 to 8 (0 to 17)		1.02 (0.94 to 1.10)		
60-69	0.67 (0.54 to 0.83) 0.67 (0.55 to 0.91)	12 to 21 (3 to 32)		0.97 (0.90 to 1.04)		
70-74	0.80 (0.51 to 1.28) 0.90 (0.46 to 1.78)	12 to 13 (0 to 32)		0.98 (0.86 to 1.14)		
50-69	0.78 (0.68 to 0.90) 0.81 (0.69 to 0.95)	6 to 13 (1 to 20)	0.75 (0.69 to 0.81)§ 0.69 (0.57 to 0.83)‖		0.62 (0.46 to 0.83)	

*From meta-analyses of screening trials using two different methods of case accrual; long case accrual results are provided first, then short case accrual results.
†Based on trials of screening included in the meta-analysis.
‡Based on a study in Sweden, and a study in Canada (standardized mortality ratio), respectively.
§Based on seven incidence-based mortality studies.
‖Based on eight case-control studies.

Abbreviations: CI=confidence interval; RR=relative risk.

Appendix A Table 3. Summary of Results: Estimated Harms of Screening (Data From Nelson, 2016)

Age, years	FP mammo-graphy*	Additional imaging recomm-ended*	Biopsy recomm-ended*	10-yr FP mammo-graphy rates (annual; biennial)	10-yr FP biopsy rates (annual; biennial)	Over-diagnosis estimates from RCTs % (95% CI)†	Over-diagnosis estimates from screening programs‡	Radiation exposure
40-49	121.2	124.9	16.4	61%;42%	7%; 5%	10.7 (9.3 to 12.2) 19.0 (15.2 to 22.7)	0 to 54% unadjusted 1 to 10% adjusted	Annual screening 40-55 years, biennial to 74 years: 86 cases, 11 deaths§
50-59	93.2	98.5	15.9	61%;42%	9%; 6%			
60-69	80.8	88.7	16.5					
70-74	69.6	79.0	17.5					

*Number per 1,000 screened per screening round.
†From meta-analysis of screening trials using two different methods of case accrual; long case accrual results are provided first, then short case accrual results.
‡From EUROSCREEN review based on 13 studies overall and 6 studies adjusted for breast cancer risk and lead time.
§From a model of digital mammography.

Abbreviations: CI=confidence interval; FP=false positive.

Literature Search Strategies for Primary Literature

Key:
/ = MeSH subject heading
$ = truncation
ti = word in title
ab = word in abstract
pt = publication type
* = truncation
kw = keyword
kf = keyword (author attributed keyword)

MEDLINE
Database: Ovid MEDLINE(R) ALL <1946 to September 08, 2021>
Search Strategy:
---
1 exp Mammography/ or mammogra$.ti,ab,kf. (42796)
2 (digital breast tomosynthesis or (breast and dbt)).ti,ab,kf. (1024)
3 exp Breast Neoplasms/dg, di or *Breast/dg or *Breast Diseases/dg (61902)
4 ((breast adj2 (adenocarcinoma$ or cancer$ or carcinoma$ or neoplasm$ or tumo?r$ or malignan$ or metasta$)) or (dense adj2 breast$)).ti,kf. (232922)
5 1 or 2 or 3 or 4 (276872)
6 mass screening/ or "Early Detection of Cancer"/ (132362)
7 (screen$ or detect$).ti. (577900)
8 ((routine or detect$ or cancer or ultrasound or multimodal or program$) adj3 (mammogra$ or screen$)).ti,ab. (109023)
9 or/6-8 (680574)
10 5 and 9 (32197)
11 "Delivery of Health Care"/mt, og, sn (31361)
12 (strateg$ or modalit$ or pattern$).ti,ab,kf. (2759513)
13 Time factors/ (1214828)
14 (interval$ or mean time or frequency or frequent or biannual$ or biennual$ or annual$).ti,ab,kf. (2239088)
15 Age factors/ (465601)
16 (initiation or initiated).ti,ab,kf. (410612)
17 Precision Medicine/ (21498)
18 Personali$.ti,ab,kf. (147594)
19 ((supplemental or supplemented) adj2 (imaging or screening)).ti,ab,kf. (472)
20 ((follow$ or after or plus or prior or versus) adj2 (mammogram$ or dbt or tomosynthesis)).ti,ab,kf. (880)
21 Survival rate/ (183317)
22 Morbidity/ (31781)
23 Life tables/ (6511)
24 Mortality/ (47448)
25 (morbidity or mortality or impairment or impaired).ti,ab,kf. (1615191)

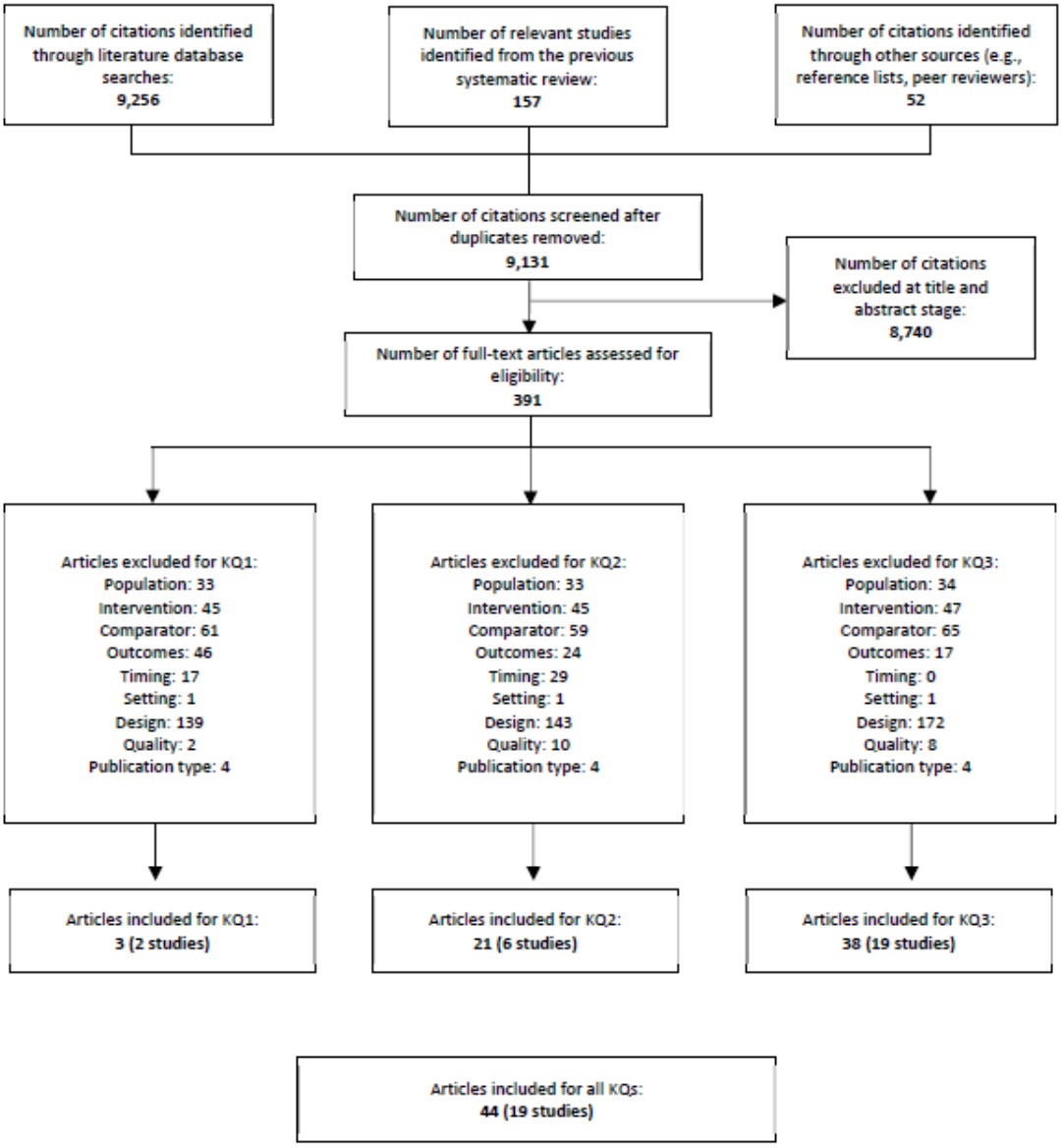
26 Risk factors/ (885488)
27 Risk reduction behavior/ (13682)
28 Incidence/ (281166)
29 Disease progression/ (177438)
30 (Incidence or incident or progression or detection rate$ or recall rate$).ti,ab,kf. or (risk or risks).ti. (1910562)
31 False Positive Reactions/ (28344)
32 (False positive or false negative).ti,ab,kf. (68831)
33 ((incorrect$ or false$ or wrong$ or bias$ or mistake$ or error$ or erroneous$) adj3 (result$ or finding$ or outcome$ or test$ or diagnos$)).ti,ab,kf. (90383)
34 medical overuse/ or unnecessary procedures/ (7907)
35 (overdiagnos$ or over-diagnos$ or misdiagnos$).ti,ab,kf. (42993)
36 (overtreat$ or over treat$).ti,ab,kf. (7976)
37 Anxiety/ (90741)
38 (anxiety or anxious).ti,ab,kf. (225786)
39 (pain or painful).ti,ab,kf. (715605)
40 (embarrassment or psychological or distress or stigma or fatalism or fatalistic).ti,ab,kf. (374937)
41 (harm or harms or harmful or harmed).ti,ab. (132817)
42 (adverse effects or mortality).fs. (2337681)
43 death/ (18406)
44 (death or deaths).ti,ab. (888163)
45 (adverse adj (effect$ or event$ or outcome$ or reaction$)).ti,ab. (402597)
46 complication$.ti,ab. (983995)
47 side effect$.ti,ab. (266086)
48 safety.ti,ab. (568105)
49 (radiation or exposure).ti,ab,kf. (1232997)
50 Long Term Adverse Effects/ (704)
51 "Quality of Life"/ (220401)
52 ("quality of life" or well-being).ti,ab,kf. (391275)
53 or/11-52 (12563328)
54 10 and 53 (21014)
55 limit 54 to (english language and yr="2014 -Current") (7617)

Cochrane Central Register of Controlled Clinical Trials (CENTRAL) via Wiley
Date Run: 09/09/2021 18:40:11

ID Search Hits
#1 mammogra*:ti,ab,kw 2606
#2 ("digital breast tomosynthesis" or (breast and dbt)):ti,ab,kw 52
#3 (breast NEAR/2 (adenocarcinoma* or cancer* or carcinoma* or neoplasm* or tumo?r* OR malignan* OR metasta*)):ti,kw or (dense NEAR/2 breast*):ti,kw 35760
#4 #1 OR #2 OR #3 36543
#5 (screen* or detect*):ti,kw 30087
#6 ((routine or detect* or cancer or ultrasound or multimodal or program*) NEAR/3 (mammogra* or screen*)):ti,ab,kw 9764

#7 #5 or #6 32468
#8 #4 AND #7 with Cochrane Library publication date from Jan 2014 to present, in Trials 1698
#9 #8 AND (clinicaltrials or trialsearch):so 300
#10 #8 AND conference:pt 443
#11 #8 NOT (#9 OR #10) 955
#12 #4 AND #7 with Cochrane Library publication date from Jan 2014 to present, in Cochrane Reviews 4

Appendix B Figure 1. Literature Flow Diagram



Appendix B Table 1. Inclusion and Exclusion Criteria

Category	Include	Exclude
Population*	Adult females	History of breast cancer or high-risk breast lesions (DCIS, LCIS, ADH, ALH) Clinically significant genetic markers or syndromes associated with high risk (e.g., *BRCA1* or *BRCA2* gene mutations, Li-Fraumeni syndrome, Cowden syndrome, hereditary diffuse gastric cancer, or other familial breast cancer syndromes) Previous large doses of chest radiation (≥20 Gy) before age 30 years
Interventions	Any mammography screening modality (i.e., film or digital mammography, digital breast tomosynthesis [3D mammography]) Screening strategy (e.g., screening interval, age to start or stop screening, personalized screening based on risk and other characteristics) Any mammography screening modality plus supplemental screening (e.g., ultrasound, MRI) Any mammography screening modality plus supplemental screening for a defined population (e.g., negative mammography, dense breasts, age group)	Breast imaging or clinical examinations conducted for diagnosis or surveillance Screening strategies that do not include mammography
Comparisons	Standard population-based screening with film or digital mammography	Breast imaging or clinical examinations conducted for diagnosis or surveillance Screening that does not include mammography
Outcomes	KQ 1: Breast cancer morbidity (e.g., adverse effects of treatment, physical/functional impairment) Quality of life or subjective well-being Breast cancer mortality All-cause mortality KQ 2: Detection and stage distribution of screen-detected invasive breast cancer Detection of advanced cancer and stage distribution of any invasive breast cancer at time of screening and across followup, including interval cancers Cancer subtypes will be defined by receptor status (e.g., ER/PR, HER2) since these are associated with prognosis Advanced cancer definitions are not standardized and available outcomes will likely vary across studies (e.g., metastatic breast cancer, different stage and tumor size cutpoints)† KQ 3: False-positive and false-negative findings at screening and biopsy Recall rate (need for further evaluation) Overdiagnosis and overtreatment Psychological harms (e.g., anxiety, depression)	

Appendix B Table 1. Inclusion and Exclusion Criteria

Category	Include	Exclude
	Quality of life and subjective well-being Radiation exposure	
Timing	KQs 1, 2: Followup from at least two rounds of screening, duration of followup KQ 3: Per round of screening Over multiple rounds of screening Lifetime	
Setting	Settings and populations of women applicable to U.S. primary care settings Studies conducted in countries categorized as "Very High" on the Human Development Index (as defined by the United Nations Development Programme)	
Study Design	KQs 1, 2: Individual participant data meta-analyses Randomized, controlled trials; controlled clinical trials Prospective cohort studies with contemporaneous comparison groups selected using unbiased criteria (e.g., screening modality used does not vary based on risk factor or marker) KQ 3: Above, plus, population-based nested case-control studies, and cross-sectional studies from included trials or large population-based studies	Narrative reviews, case reports, case series, editorials Observational studies using paired designs (i.e., within-person comparisons)
Language	English-language abstracts and articles (includes English-language abstracts of non–English-language papers)	

*Most breast cancer cases occur in cis-gender women with breast tissue that developed during puberty when rising endogenous estrogen hormone stimulated the proliferation of duct and lobule tissue. Throughout the report we incorporate gender inclusive language (people, individuals, persons with breasts) when referring to the screening population to recognize that not all people at risk of breast cancer and eligible for screening are women. Transgender men and non-binary or gender non-conforming people that have breasts are also important to consider when talking about screening for breast cancer, especially since they face unique preventive health care access barriers. In addition to using gender-inclusive terminology throughout the review, we at times refer to women as the study population, especially when citing existing studies, recommendations, registries, and data sources that did not collect nuanced data on gender. The included population for this review does not include studies of screening for transgender women, nonbinary individuals, and others that have developed breast tissue following gender-affirming medical treatment with exogenous estrogen. This population should receive specialty care that can attend to their specific clinical history (length, type of hormone use, etc.), and would rely on different, not yet available evidence for assessing breast cancer risk and screening outcomes.
†Where possible we will report outcomes for cancers diagnosed at American Joint Committee on Cancer (AJCC) stage IIA or higher (Kerlikowske K, Bissell MCS, Sprague BL, et al. Advanced Breast Cancer Definitions by Staging System Examined in the Breast Cancer Surveillance Consortium. Journal of the National Cancer Institute. 2021;113(7):909-16)

Abbreviations: ADH=atypical ductal hyperplasia; ALH=atypical lobular hyperplasia; BI-RADS=Breast Imaging-Reporting and Data System; BRCA=breast cancer gene; DCIS=ductal carcinoma in situ; ER/PR=estrogen receptor/progesterone receptor; LCIS=lobular carcinoma in situ; HER2=human epidermal growth factor receptor 2; MRI=magnetic resonance imaging.

Appendix B Table 2. Quality Assessment Criteria*

Study Design	Adapted Quality Criteria
Randomized clinical trials, adapted from U.S. Preventive Services Task Force Manual[112]	**Bias arising in the randomization process or due to confounding** • Valid random assignment/random sequence generation method used • Allocation concealed • Balance in baseline characteristics **Bias in selecting participants into the study** • CCT only: No evidence of biased selection of sample **Bias due to departures from intended interventions** • Fidelity to the intervention protocol • Low risk of contamination between groups • Participants were analyzed as originally allocated **Bias from missing data** • No, or minimal, post-randomization exclusions • Outcome data are reasonably complete and comparable between groups • Reasons for missing data are similar across groups • Missing data are unlikely to bias results **Bias in measurement of outcomes** Blinding of outcome assessors • Outcomes are measured using consistent and appropriate procedures and instruments across treatment groups • No evidence of biased use of inferential statistics **Bias in reporting results selectively** • No evidence that the measures, analyses, or subgroup analyses are selectively reported
Adapted Risk of Bias Assessment (ROBINS-I)[285]	**Bias due to confounding** • No baseline confounding • No time-varying confounding **Bias in selecting participants into the study** • No evidence of biased selection of sample • Start of followup and start of intervention coincide **Bias in classifying interventions** • Intervention groups are clearly defined • Information used to define intervention groups was recorded at the start of the intervention • Classification of intervention status is unaffected by knowledge of the outcome or risk of the outcome **Bias due to deviations from intended interventions** • No deviations from intended intervention • Important co-interventions are balanced across intervention groups • Analysis adjusts for deviations from intended intervention that could have affected outcomes **Bias from missing data** • Outcome data are available for all, or nearly all, participants • Proportion of participants and reasons for missing data are similar across groups • Appropriate statistical methods used to account for missing data or there was evidence that results were robust to the presence data **Bias in measurement of outcomes** • Blinding of participants • Blinding of outcome assessors • Methods of outcome assessment are comparable across intervention groups • No systematic errors in measurement of the outcome related to intervention received

Appendix B Table 2. Quality Assessment Criteria*

	Bias in reporting results selectively • No evidence that the measures, analyses, or subgroup analyses are selectively reported

*All randomized clinical trials and non-randomized studies of intervention were classified as good, fair, or poor.[112, 285] Good quality studies generally meet all quality criteria. Fair quality studies do not meet all the criteria but do not have critical limitations that could invalidate study findings. Poor quality studies have a single fatal flaw or multiple important limitations that could invalidate study findings. Critical appraisal of studies using *a priori* quality criteria are conducted independently by at least two reviewers. Disagreements in final quality assessment are resolved by consensus, and, if needed, consultation with a third independent reviewer.

Appendix C. Included Studies

Below is a list of included studies and their ancillary publications (indented below main results publication):

1. 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://dx.doi.org/10.1016/s0959-8049(01)00397-5

2. 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://dx.doi.org/10.1002/ijc.34161

3. 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 Research & Treatment. 2016;156(1):109-16. https://dx.doi.org/10.1007/s10549-016-3695-1

4. 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://dx.doi.org/10.1001/jamanetworkopen.2022.7234

5. Garcia-Albeniz X, Hernan MA, Logan RW, et al. Continuation of Annual Screening Mammography and Breast Cancer Mortality in Women Older Than 70 Years. Annals of Internal Medicine. 2020;172(6):381-9. https://dx.doi.org/10.7326/M18-1199

6. 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://dx.doi.org/10.1016/s1470-2045(22)00194-2

7. 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://dx.doi.org/10.1001/jamanetworkopen.2022.2440

8. 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

    a. Aase HS, Danielsen AS, Hoff SR, et al. Mammographic features and screening outcome in a randomized controlled trial comparing digital breast tomosynthesis and digital mammography. Eur J Radiol. 2021;141:109753. https://dx.doi.org/10.1016/j.ejrad.2021.109753

    b. Aase HS, Holen AS, Pedersen K, et al. A randomized controlled trial of digital breast tomosynthesis versus digital mammography in population-based screening in Bergen: interim analysis of performance indicators from the To-Be trial. Eur Radiol. 2019;29(3):1175-86. https://dx.doi.org/10.1007/s00330-018-5690-x

    c. Moshina N, Aase HS, Danielsen AS, et al. Comparing Screening Outcomes for Digital Breast Tomosynthesis and Digital Mammography by Automated Breast Density in a Randomized Controlled Trial: Results from the To-Be Trial. Radiology. 2020;297(3):522-31. https://dx.doi.org/10.1148/radiol.2020201150

d. Moger TA, Swanson JO, Holen AS, et al. Cost differences between digital tomosynthesis and standard digital mammography in a breast cancer screening programme: results from the To-Be trial in Norway. Eur J Health Econ. 2019;20(8):1261-9. https://dx.doi.org/10.1007/s10198-019-01094-7

e. Hofvind S, Holen AS, Aase HS, et al. Two-view digital breast tomosynthesis versus digital mammography in a population-based breast cancer screening programme (To-Be): a randomised, controlled trial. Lancet Oncology. 2019;20(6):795-805. https://dx.doi.org/10.1016/S1470-2045(19)30161-5

f. Waade G, Holen Å, Sebuødegård S, et al. Breast compression parameters among women screened with standard digital mammography and digital breast tomosynthesis in a randomized controlled trial. Acta Radiologica. 2020; 61(3). https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01980122/full

9. 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

    a. Hofvind S, Hovda T, Holen AS, et al. Digital Breast Tomosynthesis and Synthetic 2D Mammography versus Digital Mammography: Evaluation in a Population-based Screening Program. Radiology. 2018;287(3):787-94. https://dx.doi.org/10.1148/radiol.2018171361

10. 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

    a. Zackrisson S, Lang K, Rosso A, et al. One-view breast tomosynthesis versus two-view mammography in the Malmo Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. Lancet Oncology. 2018;19(11):1493-503. https://dx.doi.org/10.1016/S1470-2045(18)30521-7

    b. Lang K, Nergarden M, Andersson I, et al. False positives in breast cancer screening with one-view breast tomosynthesis: An analysis of findings leading to recall, work-up and biopsy rates in the Malmo Breast Tomosynthesis Screening Trial. Eur Radiol. 2016;26(11):3899-907.

    c. Lang K, Andersson I, Rosso A, et al. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmo Breast Tomosynthesis Screening Trial, a population-based study. Eur Radiol. 2016;26(1):184-90. https://dx.doi.org/10.1007/s00330-015-3803-3

    d. Johnson K, Zackrisson S, Rosso A, et al. Tumor Characteristics and Molecular Subtypes in Breast Cancer Screening with Digital Breast Tomosynthesis: the Malmö Breast Tomosynthesis Screening Trial. Radiology. 2019; 293(2): Available from: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-02083775/full

    e. Rosso A, Lang K, Petersson IF, et al. Factors affecting recall rate and false positive fraction in breast cancer screening with breast tomosynthesis - A statistical approach. Breast. 2015;24(5):680-6. https://dx.doi.org/10.1016/j.breast.2015.08.007

11. 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://dx.doi.org/10.1001/jama.2022.7672

12. 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

13. 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 Epidemiology, Biomarkers & Prevention. 2018;27(4):446-53. https://dx.doi.org/10.1158/1055-9965.EPI-17-0009

14. Miglioretti DL, Zhu W, Kerlikowske K, et al. Breast Tumor Prognostic Characteristics and Biennial vs Annual Mammography, Age, and Menopausal Status. JAMA Oncology. 2015;1(8):1069-77. https://dx.doi.org/10.1001/jamaoncol.2015.3084

    a. Dittus K, Geller B, Weaver DL, et al. Impact of mammography screening interval on breast cancer diagnosis by menopausal status and BMI. J Gen Intern Med. 2013;28(11):1454-62. https://dx.doi.org/10.1007/s11606-013-2507-0

    b. Goel A, Littenberg B, Burack RC. The association between the pre-diagnosis mammography screening interval and advanced breast cancer. Breast Cancer Res Treat. 2007;102(3):339-45. https://dx.doi.org/10.1007/s10549-006-9334-5

    c. Hubbard RA, Kerlikowske K, Flowers CI, et al. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med. 2011;155(8):481-92. https://dx.doi.org/10.7326/0003-4819-155-8-201110180-00004

    d. Kerlikowske K, Zhu W, Hubbard RA, et al. Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. JAMA internal medicine. 2013;173(9):807-16. https://dx.doi.org/10.1001/jamainternmed.2013.307

    e. White E, Miglioretti DL, Yankaskas BC, et al. Biennial versus annual mammography and the risk of late-stage breast cancer. Journal of the National Cancer Institute. 2004;96(24):1832-9. https://dx.doi.org/10.1093/jnci/djh337

    f. Nelson HD, O'Meara ES, Kerlikowske K, et al. Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data. Annals of Internal Medicine. 2016;164(4):226-35. https://dx.doi.org/10.7326/M15-0971

    g. Braithwaite D, Zhu W, Hubbard RA, et al. Screening outcomes in older US women undergoing multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor characteristics or false positive rates? Journal of the National Cancer Institute. 2013;105(5):334-41. https://dx.doi.org/10.1093/jnci/djs645

15. 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

a. Harada-Shoji N, Suzuki A, Ishida T, et al. Evaluation of Adjunctive Ultrasonography for Breast Cancer Detection Among Women Aged 40-49 Years With Varying Breast Density Undergoing Screening Mammography: A Secondary Analysis of a Randomized Clinical Trial. JAMA netw. 2021;4(8):e2121505. https://dx.doi.org/10.1001/jamanetworkopen.2021.21505

b. Ishida T, Suzuki A, Kawai M, et al. A randomized controlled trial to verify the efficacy of the use of ultrasonography in breast cancer screening aged 40-49 (J-START): 76 196 women registered. Jpn J Clin Oncol. 2014;44(2):134-40. https://dx.doi.org/10.1093/jjco/hyt199

16. 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://dx.doi.org/10.1038/bjc.2011.372

    a. Klemi PJ, Toikkanen S, Räsänen O, et al. Mammography screening interval and the frequency of interval cancers in a population-based screening. Br J Cancer. 1997;75(5):762-6. https://dx.doi.org/10.1038/bjc.1997.135

17. 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://dx.doi.org/10.1148/radiol.211132

    a. Pattacini P, Nitrosi A, Giorgi Rossi P, et al. Digital Mammography versus Digital Mammography Plus Tomosynthesis for Breast Cancer Screening: The Reggio Emilia Tomosynthesis Randomized Trial. Radiology. 2018;288(2):375-85. https://dx.doi.org/10.1148/radiol.2018172119

18. Richman IB, Long JB, Hoag JR, et al. Comparative Effectiveness of Digital Breast Tomosynthesis for Breast Cancer Screening among Women 40-64 Years Old. Journal of the National Cancer Institute. 2021;03:03. https://dx.doi.org/10.1093/jnci/djab063

19. 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

    a. Bakker MF, de Lange SV, Pijnappel RM, et al. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. New England Journal of Medicine. 2019;381(22):2091-102. https://dx.doi.org/10.1056/NEJMoa1903986

    b. de Lange SV, Bakker MF, Monninkhof EM, et al. Reasons for (non)participation in supplemental population-based MRI breast screening for women with extremely dense breasts. Clinical Radiology. 2018;73(8):759.e1-.e9. https://dx.doi.org/10.1016/j.crad.2018.04.002

Appendix D. Excluded Studies

Reason for Exclusion	
E1	Population
E1a	Population: Limited to high-risk women
E1b	Population: Limited to women referred for diagnosis/surveillance or combines screening and referred populations but does not stratify outcomes
E2	Intervention
E2a	Intervention: Limited to excluded screening methods (BSE, CBE)
E2b	Intervention: Limited to DM or FM and is not assessing a screening strategy that differs from usual care
E3	Comparator
E3a	Comparator: Comparison arm limited to excluded screening methods (BSE, CBE)
E4	Outcomes: No relevant outcomes
E5	Timing (single round of screening only [KQ1, KQ2 exclusion])
E6	Setting
E6a	Setting: Screening not done in primary care or setting with a primary care-comparable population
E6b	Setting: Study was not conducted in a very high HDI country
E7	Study design
E7a	Study design: (systematic) reviews, editorials, opinions
E7b	Study design: Case series
E7c	Study design: Cohort study using non-contemporaneous selection of study groups
E7d	Study design: Studies using paired designs (i.e., within-person comparison)
E7e	Study design: Case control study, but not a large population-based nested case-control (KQ3 only)
E7f	Study design: Cross sectional study, but not from an included trial (KQ3 only)
E7g	Study design: Large population-based registry or surveillance system using selected, nonrepresentative subset of data
E7h	Study design: Modeling study
E7i	Study design: Biased selection into study groups
E7j	Study design: Does not report cumulative harms by number of completed screens (KQ3 only)
E8	Non-English language publication
E9	Unable to locate article
E10	Abstract only or study ongoing, no outcomes published
E11	Poor quality

1. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. *Cancer Epidemiol Biomarkers Prev*. 15(1): 45-51. 2006. https://dx.doi.org/10.1158/1055-9965.Epi-05-0349. **KQ1E2b, KQ2E2b, KQ3E2b**

2. The benefits and harms of breast cancer screening: an independent review. *Lancet*. 380(9855): 1778-86. 2012. https://dx.doi.org/10.1016/s0140-6736(12)61611-0. **KQ1E7a, KQ2E7a, KQ3E7a**

3. Advani, S, Abraham, L, et al. Breast biopsy patterns and findings among older women undergoing screening mammography: The role of age and comorbidity. *J Geriatr Oncol*. 13(2): 161-169. 2022. https://doi.org/10.1016/j.jgo.2021.11.013. **KQ1E4, KQ2E4, KQ3E7j**

4. Advani, SM, Zhu, W, et al. Association of Breast Density With Breast Cancer Risk Among Women Aged 65 Years or Older by Age Group and Body Mass Index. *JAMA Netw Open*. 4(8): e2122810. 2021. https://doi.org/10.1001/jamanetworkopen.2021.22810. **KQ1E3, KQ2E3, KQ3E3**

5. Ahmadiyeh, N, Mendez, MA, et al. Factors Associated with Late-Stage Breast Cancer Diagnosis in an Urban Safety-net Hospital. *J Health Care Poor*

Underserved. 31(3): 1152-1165. 2020. https://doi.org/10.1353/hpu.2020.0087. **KQ1E3, KQ2E3, KQ3E3**

6. Alabousi, A, Patlas, MN. Annual Mammographic Screening Reduces the Risk of Interval or Higher Stage Invasive Breast Cancers: Lessons for Today and Tomorrow. *Can Assoc Radiol J*. 73(3): 446-447. 2022. https://doi.org/10.1177/08465371211070555. **KQ1E7a, KQ2E7a, KQ3E7a**

7. Allgood, PC, Warwick, J, et al. A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality. *Br J Cancer*. 98(1): 206-9. 2008. https://dx.doi.org/10.1038/sj.bjc.6604123. **KQ1E2b, KQ2E2b, KQ3E2b**

8. Alsheik, N, Blount, L, et al. Outcomes by Race in Breast Cancer Screening With Digital Breast Tomosynthesis Versus Digital Mammography. *J Am Coll Radiol*. 18(7): 906-918. 2021. https://dx.doi.org/https://dx.doi.org/10.1016/j.jacr.2020.12.033. **KQ1E4, KQ2E4, KQ3E7j**

9. 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**

10. Ambinder, EB, Oluyemi, E, et al. Disparities in the uptake of digital breast tomosynthesis for breast cancer screening: A retrospective cohort study. *Breast Journal*. 27(12): 872-876. 2021. https://doi.org/10.1111/tbj.14292. **KQ1E4, KQ2E4, KQ3E7j**

11. Ambinder, EB, Harvey, SC, et al. Synthesized Mammography: The New Standard of Care When Screening for Breast Cancer with Digital Breast Tomosynthesis?. *Acad Radiol*. 25(8): 973-976. 2018. https://dx.doi.org/https://dx.doi.org/10.1016/j.acra.2017.12.015. **KQ1E7c, KQ2E7c, KQ3E7c**

12. Amir, T, Ambinder, EB, et al. Benefits of digital breast tomosynthesis: A lesion-level analysis. *J Med Screen*. 28(3): 311-317. 2021. https://dx.doi.org/https://dx.doi.org/10.1177/0969141320978267. **KQ1E5, KQ2E5, KQ3E7j**

13. Aron, JL, Prorok, PC. An analysis of the mortality effect in a breast cancer screening study. *Int J Epidemiol*. 15(1): 36-43. 1986. https://dx.doi.org/10.1093/ije/15.1.36. **KQ1E2b, KQ2E2b, KQ3E2b**

14. Ascunce, EN, Moreno-Iribas, C, et al. Changes in breast cancer mortality in Navarre (Spain) after introduction of a screening programme. *J Med Screen*. 14(1): 14-20. 2007. https://dx.doi.org/10.1258/096914107780154558. **KQ1E2b, KQ2E2b, KQ3E2b**

15. Aujero, MP, Gavenonis, SC, et al. Clinical Performance of Synthesized Two-dimensional Mammography Combined with Tomosynthesis in a Large Screening Population. *Radiology*. 283(1): 70-76. 2017. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2017162674. **KQ1E4, KQ2E5, KQ3E7j**

16. Autier, P, Boniol, M, et al. Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. *BMJ*. 343:d4411. 2011. https://dx.doi.org/10.1136/bmj.d4411. **KQ1E3, KQ2E3, KQ3E3**

17. Autier, P, Boniol, M, et al. Effectiveness of and overdiagnosis from mammography screening in the Netherlands: population based study. *BMJ*. 359: j5224. 2017. https://dx.doi.org/https://dx.doi.org/10.1136/bmj.j5224. **KQ1E2b, KQ2E2b, KQ3E2b**

18. Bae, MS, Sung, JS, et al. Survival Outcomes of Screening with Breast MRI in Women at Elevated Risk of Breast Cancer. *J Breast Imaging*. 2(1): 29-35. 2020. https://dx.doi.org/https://dx.doi.org/10.1093/jbi/wbz083. **KQ1E1a, KQ2E1a, KQ3E1a**

19. Bahl, M, Gaffney, S, et al. Breast Cancer Characteristics Associated with 2D Digital Mammography versus Digital Breast Tomosynthesis for Screening-detected and Interval Cancers. *Radiology*. 287(1): 49-57. 2018. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2017171148. **KQ1E7c, KQ2E7c, KQ3E7c**

20. Bahl, M, Mercaldo, S, et al. Breast Cancer Screening with Digital Breast Tomosynthesis: Are Initial Benefits Sustained?. *Radiology*. 295(3): 529-539. 2020. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2020191030. **KQ1E7c, KQ2E7c, KQ3E7c**

21. Bahl, M, Mercaldo, S, et al. Comparison of performance metrics with digital 2D versus tomosynthesis mammography in the diagnostic setting. *Eur Radiol*. 29(2): 477-484. 2019. https://dx.doi.org/https://dx.doi.org/10.1007/s00330-018-5596-7. **KQ1E1b, KQ2E1b, KQ3E1b**

22. Bahl, M, Pinnamaneni, N, et al. Digital 2D versus Tomosynthesis Screening Mammography among Women Aged 65 and Older in the United States. *Radiology*. 291(3): 582-590. 2019. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2019181637. **KQ1E7c, KQ2E7c, KQ3E7c**

23. Balamou, C, Koivogui, A, et al. Impact of tomosynthesis on the evolution of the cancer detection rate in the French National Breast Cancer Screening Program. *Public Health*. 210: 65-73. 2022. https://doi.org/10.1016/j.puhe.2022.06.011. **KQ1E4, KQ2E4, KQ3E4**

24. Ban, K, Tsunoda, H, et al. Comparative study of the usefulness of adjunctive tomosynthesis in breast cancer screening by mammography and ultrasound in Japan. *Breast Cancer*. 29(5): 790-795. 2022. https://doi.org/10.1007/s12282-022-01358-w. **KQ1E3, KQ2E3, KQ3E3**

25. Ban, K, Tsunoda, H, et al. Breast cancer screening using digital breast tomosynthesis compared to digital mammography alone for Japanese women. *Breast Cancer*. 28(2): 459-464. 2021. https://dx.doi.org/https://dx.doi.org/10.1007/s12282-020-01180-2. **KQ1E4, KQ2E5, KQ3E7j**

26. Barchielli, A, Paci, E. Trends in breast cancer mortality, incidence, and survival, and mammographic screening in Tuscany, Italy. *Cancer Causes Control*. 12(3): 249-55. 2001. https://dx.doi.org/10.1023/a:1011280204842. **KQ1E2b, KQ2E2b, KQ3E2b**

27. Beckmann, K, Duffy, SW, et al. Estimates of over-diagnosis of breast cancer due to population-based mammography screening in South Australia after adjustment for lead time effects. *J Med Screen*. 22(3): 127-35. 2015. https://dx.doi.org/https://dx.doi.org/10.1177/0969141315573978. **KQ1E2b, KQ2E2b, KQ3E2b**

28. Berg, WA, Bandos, AI, et al. Ultrasound as the Primary Screening Test for Breast Cancer: Analysis From ACRIN 6666. *J Natl Cancer Inst*. 108(4)**.** 2016. https://dx.doi.org/https://dx.doi.org/10.1093/jnci/djv367. **KQ1E7d, KQ2E7d, KQ3E7d**

29. Berg, WA, Zhang, Z, 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*. 307(13): 1394-404. 2012. https://dx.doi.org/10.1001/jama.2012.388. **KQ1E1, KQ2E1, KQ3E1**

30. Bernardi, D, Caumo, F, et al. Effect of integrating 3D-mammography (digital breast tomosynthesis) with 2D-mammography on radiologists' true-positive and false-positive detection in a population breast screening trial. *Eur J Cancer*. 50(7): 1232-8. 2014. https://dx.doi.org/https://dx.doi.org/10.1016/j.ejca.2014.02.004. **KQ1E7d, KQ2E7d, KQ3E7d**

31. Bernardi, D, Ciatto, S, et al. Application of breast tomosynthesis in screening: incremental effect on mammography acquisition and reading time. *Br J Radiol*. 85(1020): e1174-8. 2012. https://dx.doi.org/10.1259/bjr/19385909. **KQ1E3, KQ2E3, KQ3E3**

32. Bernardi, D, Gentilini, MA, et al. Effect of implementing digital breast tomosynthesis (DBT) instead of mammography on population screening outcomes including interval cancer rates: Results of the Trento DBT pilot evaluation. *Breast*. 50: 135-140. 2020. https://dx.doi.org/https://dx.doi.org/10.1016/j.breast.2019.09.012. **KQ1E7c, KQ2E7c, KQ3E7c**

33. Bernardi, D, Li, T, et al. Effect of integrating digital breast tomosynthesis (3D-mammography) with acquired or synthetic 2D-mammography on radiologists' true-positive and false-positive detection in a population screening trial: A descriptive study. *Eur J Radiol*. 106: 26-31. 2018. https://dx.doi.org/https://dx.doi.org/10.1016/j.ejrad.2018.07.008. **KQ1E7d, KQ2E7d, KQ3E7d**

34. Bernardi, D, Macaskill, P, et al. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. *Lancet Oncol*. 17(8): 1105-1113. 2016. https://dx.doi.org/https://dx.doi.org/10.1016/S1470-2045(16)30101-2. **KQ1E7d, KQ2E7d, KQ3E7d**

35. Bernardi, G, Cavallaro, G, et al. Usefulness of ultrasounds in the management of breast phyllodes tumors. *G Chir*. 33(3): 81-5. 2012. **KQ1E1b, KQ2E1b, KQ3E1b**

36. Bian, T, Lin, Q, et al. Digital Breast Tomosynthesis: A New Diagnostic Method for Mass-Like Lesions in Dense Breasts. *Breast Journal*. 22(5): 535-40. 2016. https://dx.doi.org/https://dx.doi.org/10.1111/tbj.12622. **KQ1E6b, KQ2E6b, KQ3E6b**

37. Biesheuvel, C, Barratt, A, et al. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. *Lancet Oncol*. 8(12): 1129-1138. 2007. https://dx.doi.org/10.1016/s1470-2045(07)70380-7. **KQ1E7a, KQ2E7a, KQ3E7a**

38. Bjurstam, N, Björneld, L, et al. The Gothenburg Breast Screening Trial. *Cancer*. 97(10): 2387-96. 2003. https://dx.doi.org/10.1002/cncr.11361. **KQ1E3, KQ2E3, KQ3E3**

39. Blackmore, KM, Chiarelli, AM, et al. Annual Mammographic Screening Reduces the Risk of Interval or Higher Stage Invasive Breast Cancers Among Postmenopausal Women in the Ontario Breast Screening Program. *Can Assoc Radiol J*. 73(3): 524-534. 2022. https://doi.org/10.1177/08465371211062883. **KQ1E7i, KQ2E7i, KQ3E7i**

40. Blamey, RW Duffy S. The frequency of breast cancer screening results of a randomised trial. 2004. **KQ1E10, KQ2E10, KQ3E10**

41. Blanks, RG, Moss, SM, et al. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990-8: comparison of observed with predicted mortality. *BMJ*. 321(7262): 665-9. 2000. https://dx.doi.org/10.1136/bmj.321.7262.665. **KQ1E2b, KQ2E2b, KQ3E2b**

42. Blue, Cross, Blue Shield Association, et al. Use of digital breast tomosynthesis with mammography for breast cancer screening or diagnosis. *Technol Eval Cent Assess Program Exec Summ*. 28(6): 1-6. 2014. **KQ1E7a, KQ2E7a, KQ3E7a**

43. Bolejko, A, Hagell, P, et al. Prevalence, Long-term Development, and Predictors of Psychosocial Consequences of False-Positive Mammography among Women Attending Population-Based Screening. *Cancer Epidemiol Biomarkers Prevent*. 24(9): 1388-97. 2015. https://dx.doi.org/https://dx.doi.org/10.1158/1055-9965.EPI-15-0060. **KQ1E3, KQ2E3, KQ3E3**

44. Bond, M, Pavey, T, et al. Psychological consequences of false-positive screening mammograms in the UK. *Evid Based Med*. 18(2): 54-61. 2013. https://dx.doi.org/10.1136/eb-2012-100608. **KQ1E7a, KQ2E7a, KQ3E7a**

45. Bottorff, JL, Ratner, PA, et al. Women's responses to information on mammographic breast density. *Can J Nurs Res*. 39(1): 38-57. 2007. **KQ1E2, KQ2E2, KQ3E2**

46. Brancato, B, Bonardi, R, et al. Negligible advantages and excess costs of routine addition of breast ultrasonography to mammography in dense breasts. *Tumori*. 93(6): 562-6. 2007. https://doi.org/10.1177/030089160709300608. **KQ1E7, KQ2E7, KQ3E7**

47. Brem, RF, Tabar, L, et al. Assessing improvement in detection of breast cancer with three-dimensional

automated breast US in women with dense breast tissue: the SomoInsight Study. *Radiology*. 274(3): 663-73. 2015. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.14132832. **KQ1E7d, KQ2E7d, KQ3E7d**

48. Brem, RF, Tabár, László, et al. Assessing Improvement in Detection of Breast Cancer with Three-dimensional Automated Breast US in Women with Dense Breast Tissue: The SomoInsight Study. *Radiology*. 274(3): 663-673. 2015. https://dx.doi.org/10.1148/radiol.14132832. **KQ1E7d, KQ2E7d, KQ3E7d**

49. Brett, J, Austoker, J. Women who are recalled for further investigation for breast screening: psychological consequences 3 years after recall and factors affecting re-attendance. *J Public Health Med*. 23(4): 292-300. 2001. https://dx.doi.org/10.1093/pubmed/23.4.292. **KQ1E2b, KQ2E2b, KQ3E2b**

50. Brett, J, Bankhead, C, et al. The psychological impact of mammographic screening. A systematic review. *Psychooncology*. 14(11): 917-38. 2005. https://dx.doi.org/10.1002/pon.904. **KQ1E7a, KQ2E7a, KQ3E7a**

51. Brewer, NT, Salz, T, et al. Systematic review: the long-term effects of false-positive mammograms. *Ann Intern Med*. 146(7): 502-10. 2007. https://dx.doi.org/10.7326/0003-4819-146-7-200704030-00006. **KQ1E7a, KQ2E7a, KQ3E7a**

52. Brodersen, J, Siersma, VD. Long-term psychosocial consequences of false-positive screening mammography. *Ann Fam Med*. 11(2): 106-15. 2013. https://dx.doi.org/10.1370/afm.1466. **KQ1E2b, KQ2E2b, KQ3E2b**

53. Broeders, M, Moss, S, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen*. 19 Suppl 1: 14-25. 2012. https://dx.doi.org/10.1258/jms.2012.012078. **KQ1E7a, KQ2E7a, KQ3E7a**

54. Bucchi, L, Ravaioli, A, et al. Annual mammography at age 45-49 years and biennial mammography at age 50-69 years: comparing performance measures in an organised screening setting. *Eur Radiol*. 29(10): 5517-5527. 2019. https://dx.doi.org/https://dx.doi.org/10.1007/s00330-019-06050-w. **KQ1E7c, KQ2E7c, KQ3E7c**

55. Buchberger, W, Geiger-Gritsch, S, et al. Combined screening with mammography and ultrasound in a population-based screening program. *Eur J Radiol*. 101: 24-29. 2018. https://dx.doi.org/https://dx.doi.org/10.1016/j.ejrad.2018.01.022. **KQ1E4, KQ2E5, KQ3E7d**

56. Buchberger, W, Niehoff, A, et al. Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. *Semin Ultrasound CT MR*. 21(4): 325-36. 2000. https://dx.doi.org/10.1016/s0887-2171(00)90027-1. **KQ1E7d, KQ2E7d, KQ3E7d**

57. Bull, AR, Campbell, MJ. Assessment of the psychological impact of a breast screening programme. *Br J Radiol*. 64(762): 510-5. 1991. https://dx.doi.org/10.1259/0007-1285-64-762-510. **KQ1E2b, KQ2E2b, KQ3E2b**

58. Buseman, S, Mouchawar, J, et al. Mammography screening matters for

young women with breast carcinoma: evidence of downstaging among 42-49-year-old women with a history of previous mammography screening. *Cancer*. 97(2): 352-8. 2003. https://dx.doi.org/10.1002/cncr.11050. **KQ1E3, KQ2E3, KQ3E3**

59. Cabanes, A, Vidal, E, et al. Age-specific breast, uterine and ovarian cancer mortality trends in Spain: changes from 1980 to 2006. *Cancer Epidemiol*. 33(3-4): 169-75. 2009. https://dx.doi.org/10.1016/j.canep.2009.08.010. **KQ1E7, KQ2E7, KQ3E7**

60. Castellano, CR, Aguilar Angulo, PM, et al. Breast cancer mortality after eight years of an improved screening program using digital breast tomosynthesis. *J Med Screen*. 9691413211002556. 2021. https://dx.doi.org/https://dx.doi.org/10.1177/09691413211002556. **KQ1E2, KQ2E2, KQ3E2**

61. Caumo, F, Bernardi, D, et al. Incremental effect from integrating 3D-mammography (tomosynthesis) with 2D-mammography: Increased breast cancer detection evident for screening centres in a population-based trial. *Breast*. 23(1): 76-80. 2014. https://dx.doi.org/https://dx.doi.org/10.1016/j.breast.2013.11.006. **KQ1E7d, KQ2E7d, KQ3E7d**

62. 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**

63. Caumo, F, Romanucci, G, et al. Comparison of breast cancers detected in the Verona screening program following transition to digital breast tomosynthesis screening with cancers detected at digital mammography screening. *Breast Cancer Res Treat*. 170(2): 391-397. 2018. https://dx.doi.org/https://dx.doi.org/10.1007/s10549-018-4756-4l. **KQ1E5, KQ2E7c, KQ3E7c**

64. Caumo, F, Zorzi, M, et al. Digital Breast Tomosynthesis with Synthesized Two-Dimensional Images versus Full-Field Digital Mammography for Population Screening: Outcomes from the Verona Screening Program. *Radiology*. 287(1): 37-46. 2018. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2017170745. **KQ1E7c, KQ2E7c, KQ3E7c**

65. 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**

66. Chan, HP, Helvie, MA, et al. Characterization of Breast Masses in Digital Breast Tomosynthesis and Digital Mammograms: An Observer Performance Study. *Acad Radiol*. 24(11): 1372-1379. 2017. https://dx.doi.org/10.1016/j.acra.2017.04.016. **KQ1E1b, KQ2E1b, KQ3E1b**

67. Chen, SQ, Huang, M, et al. Abbreviated MRI Protocols for Detecting Breast Cancer in Women with Dense Breasts. *Korean J Radiol*. 18(3): 470-475. 2017. https://dx.doi.org/10.3348/kjr.2017.18.3.470. **KQ1E7d, KQ2E7d, KQ3E7d**

68. Chiarelli, AM, Blackmore, KM, et al. Annual vs Biennial Screening: Diagnostic Accuracy Among Concurrent Cohorts Within the Ontario Breast Screening Program. *J Natl Cancer Inst*. 112(4): 400-409. 2020. https://dx.doi.org/https://dx.doi.org/10.1093/jnci/djz131. **KQ1E4, KQ2E7i, KQ3E7i**

69. Chiarelli, AM, Blackmore, KM, et al. Performance Measures of Magnetic Resonance Imaging Plus Mammography in the High Risk Ontario Breast Screening Program. *J Natl Cancer Inst*. 112(2): 136-144. 2020. https://dx.doi.org/https://dx.doi.org/10.1093/jnci/djz079. **KQ1E1a, KQ2E1a, KQ3E1a**

70. Chiarelli, AM, Majpruz, V, et al. The contribution of clinical breast examination to the accuracy of breast screening. *J Natl Cancer Inst*. 101(18): 1236-43. 2009. https://dx.doi.org/10.1093/jnci/djp241. **KQ1E3a, KQ2E3a, KQ3E3a**

71. Cho, KR, Seo, BK, et al. Breast Cancer Detection in a Screening Population: Comparison of Digital Mammography, Computer-Aided Detection Applied to Digital Mammography and Breast Ultrasound. *J Breast Cancer*. 19(3): 316-323. 2016. https://doi.org/10.4048/jbc.2016.19.3.316. **KQ1E7d, KQ2E7d, KQ3E7d**

72. Choi, E, Jun, JK, et al. Effectiveness of the Korean National Cancer Screening Program in reducing breast cancer mortality. *NPJ Breast Cancer*. 7(1): 83. 2021. https://dx.doi.org/https://dx.doi.org/10.1038/s41523-021-00295-9. **KQ1E3, KQ2E3, KQ3E3**

73. Chough, Denise M, Berg, Wendie A, et al. A Prospective Study of Automated Breast Ultrasound Screening of Women with Dense Breasts in a Digital Breast Tomosynthesis-based Practice. *J Breast Imaging*. 2(2): 125-133. 2020. https://doi.org/10.1093/jbi/wbaa006. **KQ1E7d, KQ2E7d, KQ3E7d**

74. Chu, KC, Smart, CR, et al. Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan clinical trial. *J Natl Cancer Inst*. 80(14): 1125-32. 1988. https://dx.doi.org/10.1093/jnci/80.14.1125. **KQ1E3, KQ2E3, KQ3E3**

75. Ciatto, S, Houssami, N, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol*. 14(7): 583-9. 2013. https://dx.doi.org/10.1016/s1470-2045(13)70134-7. **KQ1E7d, KQ2E7d, KQ3E7d**

76. Cochon, LR, Giess, CS, et al. Comparing Diagnostic Performance of Digital Breast Tomosynthesis and Full-Field Digital Mammography. *J Am Coll Radiol*. 17(8): 999-1003. 2020. https://dx.doi.org/https://dx.doi.org/10.1016/j.jacr.2020.01.010. **KQ1E7c, KQ2E7c, KQ3E7c**

77. Cohen, EO, Weaver, OO, et al. Breast Cancer Screening via Digital Mammography, Synthetic Mammography, and Tomosynthesis. *Am J Prev Med*. 58(3): 470-472. 2020. https://doi.org/10.1016/j.amepre.2019.09.016. **KQ1E4, KQ2E5, KQ3E7j**

78. Cohen, EO, Perry, RE, et al. Breast cancer screening in women with and without implants: retrospective study

comparing digital mammography to digital mammography combined with digital breast tomosynthesis. *Eur Radiol*. 20: 20. 2021. https://dx.doi.org/https://dx.doi.org/10.1007/s00330-021-08040-3. **KQ1E4, KQ2E5, KQ3E7j**

79. Cohen, EO, Tso, HH, et al. Screening Mammography Findings From One Standard Projection Only in the Era of Full-Field Digital Mammography and Digital Breast Tomosynthesis. *AJR Am J Roentgenol*. 211(2): 445-451. 2018. https://dx.doi.org/https://dx.doi.org/10.2214/AJR.17.19023. **KQ1E5, KQ2E5, KQ3E7j**

80. Cohen, EO, Weaver, OO, et al. Breast Cancer Screening via Digital Mammography, Synthetic Mammography, and Tomosynthesis. *Am J Prev Med*. 58(3): 470-472. 2020. https://dx.doi.org/10.1016/j.amepre.2019.09.016. **KQ1E5, KQ2E5, KQ3E7j**

81. Coldman, A, Phillips, N, et al. Pan-Canadian study of mammography screening and mortality from breast cancer. *J Natl Cancer Inst*. 106(11)**.** 2014. https://dx.doi.org/10.1093/jnci/dju261. **KQ1E2b, KQ2E2b, KQ3E2b**

82. Coldman, A, Phillips, N. Incidence of breast cancer and estimates of overdiagnosis after the initiation of a population-based mammography screening program. *CMAJ*. 185(10): E492-8. 2013. https://dx.doi.org/10.1503/cmaj.121791. **KQ1E4, KQ2E4, KQ3E2b**

83. Coldman, AJ, Phillips, N, et al. Impact of changing from annual to biennial mammographic screening on breast cancer outcomes in women aged 50-79 in British Columbia. *J Med Screen*. 15(4): 182-7. 2008. https://dx.doi.org/10.1258/jms.2008.008064. **KQ1E7c, KQ2E7c, KQ3E7c**

84. Comstock, CE, Gatsonis, C, et al. Comparison of Abbreviated Breast MRI vs Digital Breast Tomosynthesis for Breast Cancer Detection Among Women With Dense Breasts Undergoing Screening. *JAMA*. 323(8): 746-756. 2020. https://dx.doi.org/https://dx.doi.org/10.1001/*JAMA*.2020.0572. **KQ1E7d, KQ2E7d, KQ3E7d**

85. Conant, EF, Barlow, WE, et al. Association of Digital Breast Tomosynthesis vs Digital Mammography With Cancer Detection and Recall Rates by Age and Breast Density. *JAMA Oncology*. 5(5): 635-642. 2019. https://dx.doi.org/https://dx.doi.org/10.1001/*JAMA*oncol.2018.7078. **KQ1E4, KQ2E7, KQ3E7j**

86. Conant, EF, Zuckerman, SP, et al. Five Consecutive Years of Screening with Digital Breast Tomosynthesis: Outcomes by Screening Year and Round. *Radiology*. 295(2): 285-293. 2020. https://dx.doi.org/10.1148/radiol.2020191751. **KQ1E7c, KQ2E7c, KQ3E7c**

87. Corsetti, V, Houssami, N, 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*. 47(7): 1021-6. 2011. https://dx.doi.org/10.1016/j.ejca.2010.12.002. **KQ1E3, KQ2E3, KQ3E3**

88. Crystal, P, Strano, SD, et al. Using sonography to screen women with mammographically dense breasts. *AJR*

Am J Roentgenol. 181(1): 177-82. 2003. https://dx.doi.org/10.2214/ajr.181.1.1810177. **KQ1E7d, KQ2E7d, KQ3E7d**

89. Dang, PA, Wang, A, et al. Comparing Tumor Characteristics and Rates of Breast Cancers Detected by Screening Digital Breast Tomosynthesis and Full-Field Digital Mammography. *AJR Am J Roentgenol*. 214(3): 701-706. 2020. https://dx.doi.org/https://dx.doi.org/10.2214/AJR.18.21060. **KQ1E5, KQ2E5, KQ3E4**

90. de Gelder, R, Fracheboud, J, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. *Prev Med*. 53(3): 134-40. 2011. https://dx.doi.org/10.1016/j.ypmed.2011.06.009. **KQ1E2b, KQ2E2b, KQ3E2b**

91. de Gelder, R, Heijnsdijk, EA, et al. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev*. 33(1): 111-21. 2011. https://dx.doi.org/10.1093/epirev/mxr009. **KQ1E7h, KQ2E7h, KQ3E7h**

92. de Glas, NA, de Craen, AJ, et al. Effect of implementation of the mass breast cancer screening programme in older women in the Netherlands: population based study. *BMJ*. 349: g5410. 2014. https://dx.doi.org/https://dx.doi.org/10.1136/bmj.g5410. **KQ1E2b, KQ2E2b, KQ3E2b**

93. de Koning, HJ, Draisma, G, et al. Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data. *Breast Cancer Res*. 8(1): 202. 2006. https://dx.doi.org/10.1186/bcr1369. **KQ1E7h, KQ2E7h, KQ3E7h**

94. Destounis, S, Arieno, A, et al. Comparison of Cancers Detected by Screening Breast Ultrasound and Digital Breast Tomosynthesis. *Acad Radiol*. 12: 12. 2021. https://dx.doi.org/https://dx.doi.org/10.1016/j.acra.2021.01.008. **KQ1E7d, KQ2E7d, KQ3E7d**

95. 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**

96. Destounis, S, Arieno, A, et al. New York State Breast Density Mandate: Follow-up Data With Screening Sonography. *J Ultrasound Med*. 36(12): 2511-2517. 2017. https://dx.doi.org/https://dx.doi.org/10.1002/jum.14294. **KQ1E3, KQ2E3, KQ3E3**

97. Dibble, EH, Singer, TM, et al. BI-RADS 3 on dense breast screening ultrasound after digital mammography versus digital breast tomosynthesis. *Clin Imaging*. 80: 315-321. 2021. https://dx.doi.org/https://dx.doi.org/10.1016/j.clinimag.2021.07.030. **KQ1E4, KQ2E4, KQ3E4**

98. Dibble, EH, Singer, TM, et al. Dense Breast Ultrasound Screening After Digital Mammography Versus After Digital Breast Tomosynthesis. *AJR Am J Roentgenol*. 213(6): 1397-1402. 2019. https://dx.doi.org/https://dx.doi.org/10.2214/AJR.18.20748. **KQ1E1b, KQ2E1b, KQ3E1b**

99. Ding, L, Greuter, MJW, et al. Irregular screening participation increases advanced stage breast cancer at diagnosis: A population-based study. *Breast*. 65: 61-66. 2022. https://doi.org/10.1016/j.breast.2022.07.004. **KQ1E3, KQ2E3, KQ3E3**

100. DiPrete, O, Lourenco, AP, et al. Screening Digital Mammography Recall Rate: Does It Change with Digital Breast Tomosynthesis Experience?. *Radiology*. 286(3): 838-844. 2018. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2017170517. **KQ1E7c, KQ2E7c, KQ3E7c**

101. Dodelzon, K, Starikov, A, et al. Breast cancer in women under age 40: A decade of trend analysis at a single institution. *Clin Imaging*. 78: 165-170. 2021. https://dx.doi.org/https://dx.doi.org/10.1016/j.clinimag.2021.03.031. **KQ1E1, KQ2E1, KQ3E1**

102. Domingo, L, Hofvind, S, et al. Cross-national comparison of screening mammography accuracy measures in U.S., Norway, and Spain. *Eur Radiol*. 26(8): 2520-8. 2016. https://dx.doi.org/https://dx.doi.org/10.1007/s00330-015-4074-8. **KQ1E2b, KQ2E2b, KQ3E2b**

103. Duffy, S, Vulkan, D, et al. Annual mammographic screening to reduce breast cancer mortality in women from age 40 years: long-term follow-up of the UK Age RCT. *Health Technol Assess*. 24(55): 1-24. 2020. https://dx.doi.org/https://dx.doi.org/10.3310/hta24550. **KQ1E7, KQ2E7, KQ3E7**

104. Duffy, SW, Agbaje, O, et al. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Res*. 7(6): 258-65. 2005. https://dx.doi.org/10.1186/bcr1354. **KQ1E3, KQ2E3, KQ3E3**

105. Duffy, SW, Tabar, L, et al. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. *J Med Screen*. 17(1): 25-30. 2010. https://dx.doi.org/10.1258/jms.2009.009094. **KQ1E3, KQ2E3, KQ3E3**

106. Duffy, SW, Tabar, L, et al. Beneficial Effect of Consecutive Screening Mammography Examinations on Mortality from Breast Cancer: A Prospective Study. *Radiology*. 299(3): 541-547. 2021. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2021203935. **KQ1E3, KQ2E3, KQ3E3**

107. Duffy, SW, Tabar, L, et al. The Swedish Two-County Trial of mammographic screening: cluster randomisation and end point evaluation. *Ann Oncol*. 14(8): 1196-8. 2003. https://dx.doi.org/10.1093/annonc/mdg322. **KQ1E3, KQ2E3, KQ3E3**

108. Duffy, SW, Vulkan, D, et al. Effect of mammographic screening from age 40 years on breast cancer mortality (UK Age trial): final results of a randomised, controlled trial. *Lancet Oncol*. 21(9): 1165-1172. 2020. https://dx.doi.org/https://dx.doi.org/10.1016/S1470-2045(20)30398-3. **KQ1E3, KQ2E4, KQ3E3**

109. Duijm, LEM, Broeders, MJM, et al. Effects of nonparticipation at previous

screening rounds on the characteristics of screen-detected breast cancers. *Eur J Radiol*. 154: 110391. 2022. https://doi.org/10.1016/j.ejrad.2022.110391. **KQ1E3, KQ2E3, KQ3E3**

110. Durand, MA, Friedewald, SM, et al. False-Negative Rates of Breast Cancer Screening with and without Digital Breast Tomosynthesis. *Radiology*. 298(2): 296-305. 2021. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2020202858. **KQ1E5, KQ2E5, KQ3E7c**

111. Durand, MA, Haas, BM, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology*. 274(1): 85-92. 2015. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.14131319. **KQ1E4, KQ2E5, KQ3E7j**

112. Ekpo, EU, Ujong, UP, et al. Assessment of Interradiologist Agreement Regarding Mammographic Breast Density Classification Using the Fifth Edition of the BI-RADS Atlas. *AJR Am J Roentgenol*. 206(5): 1119-23. 2016. https://doi.org/10.2214/ajr.15.15049. **KQ1E4, KQ2E4, KQ3E4**

113. El Bakry, Rehab Abdel Rahman. Breast tomosynthesis: A diagnostic addition to screening digital mammography. The Egyptian Journal of Radiology and Nuclear Medicine. 49(2): 529-535. 2018. https://dx.doi.org/https://doi.org/10.1016/j.ejrnm.2017.12.004. **KQ1E7d, KQ2E7d, KQ3E7d**

114. Elbakkoush, AA, Atique, S, et al. Screening Mammography Efficacy: A Comparison Between Screen-Film, Computed Radiography and Digital Mammography in Taiwan. *Stud Health Technol Inform.* 216: 914. 2015. **KQ1E2b, KQ2E2b, KQ3E2b**

115. Ellman, R, Angeli, N, et al. Psychiatric morbidity associated with screening for breast cancer. Br J Cancer. 60(5): 781-4. 1989. https://dx.doi.org/10.1038/bjc.1989.359. **KQ1E3, KQ2E3, KQ3E3**

116. Elmore, JG, Barton, MB, et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med*. 338(16): 1089-96. 1998. https://dx.doi.org/10.1056/nejm199804163381601. **KQ1E3a, KQ2E3a, KQ3E3a**

117. El-Zaemey, S, Liz, W, et al. Impact of the age expansion of breast screening on screening uptake and screening outcomes among older women in BreastScreen western. *Breast*. 56: 96-102. 2021. https://doi.org/10.1016/j.breast.2021.02.006. **KQ1E7c, KQ2E7c, KQ3E7c**

118. Endo, T, Morita, T, et al. Diagnostic performance of digital breast tomosynthesis and full-field digital mammography with new reconstruction and new processing for dose reduction. *Breast Cancer*. 25(2): 159-166. 2018. https://dx.doi.org/10.1007/s12282-017-0805-9. **KQ1E7d, KQ2E7d, KQ3E7d**

119. Espasa, R, Murta-Nascimento, C, et al. The psychological impact of a false-positive screening mammogram in Barcelona. *J Cancer Educ*. 27(4): 780-5. 2012. https://dx.doi.org/10.1007/s13187-012-0349-9. **KQ1E2b, KQ2E2b, KQ3E2b**

120. Etzioni, R, Gulati, R, et al. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. *Ann Intern Med*. 158(11): 831-8. 2013. https://dx.doi.org/10.7326/0003-4819-158-11-201306040-00008. **KQ1E7a, KQ2E7a, KQ3E7a**

121. Falk, RS, Hofvind, S, et al. Overdiagnosis among women attending a population-based mammography screening program. *Int J Cancer*. 133(3): 705-12. 2013. https://dx.doi.org/10.1002/ijc.28052. **KQ1E3, KQ2E3, KQ3E3**

122. Fann, JC, Chang, KJ, et al. Impact of Overdiagnosis on Long-Term Breast Cancer Survival. *Cancers (Basel)*. 11(3): 07. 2019. https://dx.doi.org/10.3390/cancers11030325. **KQ1E7h, KQ2E7h, KQ3E7h**

123. 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**

124. Fielder, HM, Warwick, J, et al. A case-control study to estimate the impact on breast cancer death of the breast screening programme in Wales. *J Med Screen*. 11(4): 194-8. 2004. https://dx.doi.org/10.1258/0969141042467304. **KQ1E7, KQ2E7, KQ3E2b**

125. Fitzpatrick, P, Fleming, P, et al. False-positive mammographic screening: factors influencing re-attendance over a decade of screening. *J Med Screen*. 18(1): 30-3. 2011. https://dx.doi.org/10.1258/jms.2010.010104. **KQ1E3, KQ2E3, KQ3E3**

126. Freer, PE, Riegert, J, et al. Clinical implementation of synthesized mammography with digital breast tomosynthesis in a routine clinical practice. *Breast Cancer Res Treat*. 166(2): 501-509. 2017. https://dx.doi.org/https://dx.doi.org/10.1007/s10549-017-4431-1. **KQ1E5, KQ2E5, KQ3E7j**

127. Friedewald, SM, Grimm, LJ. Digital Breast Tomosynthesis and Detection of Interval Invasive and Advanced Breast Cancers. *JAMA*. 327(22): 2198-2200. 2022. https://doi.org/10.1001/jama.2021.25018. **KQ1E7a, KQ2E7a, KQ3E7a**

128. Friedewald, SM, Rafferty, EA, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA*. 311(24): 2499-507. 2014. https://dx.doi.org/10.1001/*JAMA*.2014.6095. **KQ1E7c, KQ2E7c, KQ3E7c**

129. Fujii, MH, Herschorn, SD, et al. Detection Rates for Benign and Malignant Diagnoses on Breast Cancer Screening With Digital Breast Tomosynthesis in a Statewide Mammography Registry Study. *AJR Am J Roentgenol*. 212(3): 706-711. 2019. https://dx.doi.org/10.2214/AJR.18.20255. **KQ1E7, KQ2E7, KQ3E7**j

130. Gabe, R, Tryggvadóttir, L, et al. A case-control study to estimate the impact of the Icelandic population-based mammography screening program on breast cancer death. *Acta Radiol*. 48(9): 948-55. 2007. https://dx.doi.org/10.1080/02841850701501725. **KQ1E3, KQ2E3, KQ3E3**

131. Garayoa, J, Chevalier, M, et al. Diagnostic value of the stand-alone

synthetic image in digital breast tomosynthesis examinations. *Eur Radiol*. 28(2): 565-572. 2018. https://dx.doi.org/10.1007/s00330-017-4991-9. **KQ1E7d, KQ2E7d, KQ3E7d**

132. García Fernández, A, Chabrera, C, et al. Mortality and recurrence patterns of breast cancer patients diagnosed under a screening programme versus comparable non-screened breast cancer patients from the same population: analytical survey from 2002 to 2012. *Tumour Biol*. 35(3): 1945-53. 2014. https://dx.doi.org/10.1007/s13277-013-1260-7. **KQ1E3a, KQ2E3a, KQ3E3a**

133. Gard, CC, Aiello Bowles, EJ, et al. Misclassification of Breast Imaging Reporting and Data System (BI-RADS) Mammographic Density and Implications for Breast Density Reporting Legislation. *Breast J*. 21(5): 481-9. 2015. https://dx.doi.org/10.1111/tbj.12443. **KQ1E4, KQ2E4, KQ3E4**

134. Gastounioti, A, McCarthy, AM, et al. Effect of Mammographic Screening Modality on Breast Density Assessment: Digital Mammography versus Digital Breast Tomosynthesis. *Radiology*. 291(2): 320-327. 2019. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2019181740. **KQ1E7c, KQ2E7c, KQ3E7c**

135. Gatta, G, Cappabianca, S, et al. Second-Generation 3D Automated Breast Ultrasonography (Prone ABUS) for Dense Breast Cancer Screening Integrated to Mammography: Effectiveness, Performance and Detection Rates. *J Pers Med*. 11(9): 31. 2021. https://doi.org/10.3390/jpm11090875. **KQ1E7d, KQ2E7d, KQ3E7g**

136. Geiger-Gritsch, S, Daniaux, M, et al. Performance of 4 years of population-based mammography screening for breast cancer combined with ultrasound in Tyrol / Austria. *Wien Klin Wochenschr*. 130(3-4): 92-99. 2018. https://dx.doi.org/https://dx.doi.org/10.1007/s00508-017-1293-9. **KQ1E2b, KQ2E2b, KQ3E2b**

137. Gennaro, G, Hendrick, RE, et al. Performance comparison of single-view digital breast tomosynthesis plus single-view digital mammography with two-view digital mammography. *Eur Radiol*. 23(3): 664-72. 2013. https://dx.doi.org/10.1007/s00330-012-2649-1. **KQ1E1b, KQ2E1b, KQ3E1b**

138. Gibson, CJ, Weiss, J, et al. False-positive mammography and depressed mood in a screening population: findings from the New Hampshire Mammography Network. *J Public Health*. 31(4): 554-60. 2009. https://dx.doi.org/10.1093/pubmed/fdp064. **KQ1E3, KQ2E3, KQ3E3**

139. Giess, CS, Pourjabbar, S, et al. Comparing Diagnostic Performance of Digital Breast Tomosynthesis and Full-Field Digital Mammography in a Hybrid Screening Environment. *AJR Am J Roentgenol*. 209(4): 929-934. 2017. https://dx.doi.org/https://dx.doi.org/10.2214/AJR.17.17983. **KQ1E4, KQ2E5, KQ3E7j**

140. Giger, ML, Inciardi, MF, et al. Automated Breast Ultrasound in Breast Cancer Screening of Women With Dense Breasts: Reader Study of Mammography-Negative and Mammography-Positive Cancers. *AJR Am J Roentgenol*. 206(6): 1341-50. 2016.

https://dx.doi.org/10.2214/ajr.15.15367. **KQ1E7d, KQ2E7d, KQ3E7d**

141. Gilbert, FJ, Tucker, L, et al. Accuracy of Digital Breast Tomosynthesis for Depicting Breast Cancer Subgroups in a UK Retrospective Reading Study (TOMMY Trial). *Radiology*. 277(3): 697-706. 2015. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2015142566. **KQ1E1b, KQ2E1b, KQ3E1b**

142. Gilbert, FJ, Tucker, L, et al. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme--a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. *Health Technol Assess*. 19(4): i-xxv, 1-136. 2015. https://dx.doi.org/https://dx.doi.org/10.3310/hta19040. **KQ1E1b, KQ2E1b, KQ3E1b**

143. Giorgi Rossi, P, Djuric, O, et al. Validation of a new fully automated software for 2D digital mammographic breast density evaluation in predicting breast cancer risk. *Sci Rep*. 11(1): 19884. 2021. https://doi.org/10.1038/s41598-021-99433-3. **KQ1E3, KQ2E3, KQ3E3**

144. Girardi, V, Tonegutti, M, et al. Breast ultrasound in 22,131 asymptomatic women with negative mammography. *Breast*. 22(5): 806-9. 2013. https://dx.doi.org/10.1016/j.breast.2013.02.010. **KQ1E3, KQ2E3, KQ3E3**

145. Giuliano, V, Giuliano, C. Improved breast cancer detection in asymptomatic women using 3D-automated breast ultrasound in mammographically dense breasts. *Clin Imaging*. 37(3): 480-6. 2013. https://dx.doi.org/10.1016/j.clinimag.2012.09.018. **KQ1E7c, KQ2E7c, KQ3E7c**

146. Gong, AJ, Nguyen, DL, et al. Comparison of Outcomes for One-View Asymmetries Recalled From Digital Breast Tomosynthesis Versus Full-Field Digital Mammography Screening Examinations. *AJR Am J Roentgenol*. 15: 15. 2022. https://doi.org/10.2214/ajr.22.27820. **KQ1E4, KQ2E4, KQ3E7j**

147. Gorini, G, Zappa, M, et al. Breast cancer mortality trends in two areas of the province of Florence, Italy, where screening programmes started in the 1970s and 1990s. *Br J Cancer*. 90(9): 1780-3. 2004. https://dx.doi.org/10.1038/sj.bjc.6601744. **KQ1E2b, KQ2E2b, KQ3E2b**

148. Gøtzsche, PC, Jørgensen, KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*. 2013(6): Cd001877. 2013. https://dx.doi.org/10.1002/14651858.CD001877.pub5. **KQ1E7a, KQ2E7a, KQ3E7a**

149. Greenberg, JS, Javitt, MC, et al. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *AJR Am J Roentgenol*. 203(3): 687-93. 2014. https://dx.doi.org/https://dx.doi.org/10.2214/AJR.14.12642. **KQ1E5, KQ2E5, KQ3E7j**

150. Gunsoy, NB, Garcia-Closas, M, et al. Estimating breast cancer mortality

reduction and overdiagnosis due to screening for different strategies in the United Kingdom. *Br J Cancer*. 110(10): 2412-9. 2014. https://dx.doi.org/10.1038/bjc.2014.206. **KQ1E7h, KQ2E7h, KQ3E7h**

151. Gurando, AV, Babkina, TM, et al. Digital Breast Tomosynthesis and Full-Field Digital Mammography in Breast Cancer Detection Associated with Four Asymmetry Types. *Wiadomosci Lekarskie*. 74(4): 842-848. 2021. **KQ1E1b, KQ2E1b, KQ3E1b**

152. Haas, BM, Kalra, V, et al. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology*. 269(3): 694-700. 2013. https://dx.doi.org/10.1148/radiol.13130307. **KQ1E5, KQ2E5, KQ3E7j**

153. Haas, JS, Hill, DA, et al. Disparities in the use of screening magnetic resonance imaging of the breast in community practice by race, ethnicity, and socioeconomic status. *Cancer*. 122(4): 611-7. 2016. https://dx.doi.org/https://dx.doi.org/10.1002/cncr.29805. **KQ1E4, KQ2E4, KQ3E4**

154. Habbema, JD, van Oortmarssen, GJ, et al. Age-specific reduction in breast cancer mortality by screening: an analysis of the results of the Health Insurance Plan of Greater New York study. *J Natl Cancer Inst*. 77(2): 317-20. 1986. **KQ1E3, KQ2E3, KQ3E3**

155. Hafslund, B, Espehaug, B, et al. Effects of false-positive results in a breast screening program on anxiety, depression and health-related quality of life. *Cancer Nurs*. 35(5): E26-34. 2012. https://dx.doi.org/10.1097/NCC.0b013e3182341ddb. **KQ1E4, KQ2E4, KQ3E3**

156. Hafslund, B, Nortvedt, MW. Mammography screening from the perspective of quality of life: a review of the literature. *Scand J Caring Sci*. 23(3): 539-48. 2009. https://dx.doi.org/10.1111/j.1471-6712.2008.00634.x. **KQ1E7a, KQ2E7a, KQ3E7a**

157. Hakama, M, Pukkala, E, et al. Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *BMJ*. 314(7084): 864-7. 1997. https://dx.doi.org/10.1136/bmj.314.7084.864. **KQ1E3, KQ2E3, KQ3E3**

158. Hanley, JA, Hannigan, A, et al. Mortality reductions due to mammography screening: Contemporary population-based data. *PLoS ONE*. 12(12): e0188947. 2017. https://dx.doi.org/https://dx.doi.org/10.1371/journal.pone.0188947. **KQ1E3, KQ2E3, KQ3E3**

159. Harris, R, Yeatts, J, et al. Breast cancer screening for women ages 50 to 69 years a systematic review of observational evidence. *Prev Med*. 53(3): 108-14. 2011. https://dx.doi.org/10.1016/j.ypmed.2011.07.004. **KQ1E3, KQ2E3, KQ3E3**

160. Harvey, JA, Gard, CC, et al. Reported mammographic density: film-screen versus digital acquisition. *Radiology*. 266(3): 752-8. 2013. https://dx.doi.org/10.1148/radiol.12120221. **KQ1E4, KQ2E4, KQ3E4**

161. Haukka, J, Byrnes, G, et al. Trends in breast cancer mortality in Sweden before and after implementation of

mammography screening. *PLoS One*. 6(9): e22422. 2011. https://dx.doi.org/10.1371/journal.pone.0022422. **KQ1E7h, KQ2E7h, KQ3E7h**

162. Heleno, B, Siersma, VD, et al. Diagnostic invasiveness and psychosocial consequences of false-positive mammography. *Ann Fam Med*. 13(3): 242-9. 2015. https://dx.doi.org/https://dx.doi.org/10.1370/afm.1762. **KQ1E1b, KQ2E1b, KQ3E1b**

163. Hellquist, BN, Czene, K, et al. Effectiveness of population-based service screening with mammography for women ages 40 to 49 years with a high or low risk of breast cancer: socioeconomic status, parity, and age at birth of first child. *Cancer*. 121(2): 251-8. 2015. https://dx.doi.org/https://dx.doi.org/10.1002/cncr.29011. **KQ1E3, KQ2E3, KQ3E3**

164. Hellquist, BN, Duffy, SW, et al. Effectiveness of population-based service screening with mammography for women ages 40 to 49 years: evaluation of the Swedish Mammography Screening in Young Women (SCRY) cohort. *Cancer*. 117(4): 714-22. 2011. https://dx.doi.org/10.1002/cncr.25650. **KQ1E2b, KQ2E2b, KQ3E2b**

165. Hellquist, BN, Duffy, SW, et al. Overdiagnosis in the population-based service screening programme with mammography for women aged 40 to 49 years in Sweden. *J Med Screen*. 19(1): 14-9. 2012. https://dx.doi.org/10.1258/jms.2012.011104. **KQ1E2b, KQ2E2b, KQ3E2b**

166. Hendrick, RE. Radiation doses and cancer risks from breast imaging studies. *Radiology*. 257(1): 246-53. 2010. https://dx.doi.org/10.1148/radiol.10100570. **KQ1E7, KQ2E7, KQ3E7**

167. Heywang-Kobrunner, S, Jaensch, A, et al. Value of Digital Breast Tomosynthesis versus Additional Views for the Assessment of Screen-Detected Abnormalities - a First Analysis. *Breast Care*. 12(2): 92-97. 2017. https://dx.doi.org/https://dx.doi.org/10.1159/000456649. **KQ1E1b, KQ2E1b, KQ3E1b**

168. Hislop, TG, Harris, SR, et al. Satisfaction and anxiety for women during investigation of an abnormal screening mammogram. *Breast Cancer Res Treat*. 76(3): 245-54. 2002. https://dx.doi.org/10.1023/a:1020820103126. **KQ1E1b, KQ2E1b, KQ3E1b**

169. Hofvind, S, Skaane, P. Stage distribution of breast cancer diagnosed before and after implementation of population-based mammographic screening. *Rofo*. 184(5): 437-42. 2012. https://dx.doi.org/10.1055/s-0031-1299352. **KQ1E2b, KQ2E2b, KQ3E2b**

170. Hofvind, S, Ursin, G, et al. Breast cancer mortality in participants of the Norwegian Breast Cancer Screening Program. *Cancer*. 119(17): 3106-12. 2013. https://dx.doi.org/10.1002/cncr.28174. **KQ1E2b, KQ2E2b, KQ3E2b**

171. Hofvind, S, Wang, H, et al. Do the results of the process indicators in the Norwegian Breast Cancer Screening Program predict future mortality reduction from breast cancer? *Acta Oncol*. 43(5): 467-73. 2004.

https://dx.doi.org/10.1080/02841860410034315. **KQ1E3, KQ2E3, KQ3E3**

172. Hong, S, Song, SY, et al. Effect of Digital Mammography for Breast Cancer Screening: A Comparative Study of More than 8 Million Korean Women. *Radiology*. 294(2): 247-255. 2020. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2019190951. **KQ1E2b, KQ2E2b, KQ3E2b**

173. Honig, EL, Mullen, LA, et al. Factors Impacting False Positive Recall in Screening Mammography. *Acad Radiol*. 26(11): 1505-1512. 2019. https://dx.doi.org/https://dx.doi.org/10.1016/j.acra.2019.01.020. **KQ1E5, KQ2E5, KQ3E7j**

174. Hooley, RJ, Greenberg, KL, et al. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. *Radiology*. 265(1): 59-69. 2012. https://dx.doi.org/10.1148/radiol.12120621. **KQ1E3, KQ2E3, KQ3E3**

175. Houssami, N, Bernardi, D, et al. Breast cancer detection using single-reading of breast tomosynthesis (3D-mammography) compared to double-reading of 2D-mammography: Evidence from a population-based trial. *Cancer Epidemiol*. 47: 94-99. 2017. https://dx.doi.org/https://dx.doi.org/10.1016/j.canep.2017.01.008. **KQ1E7d, KQ2E7d, KQ3E7d**

176. 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**

177. Houssami, N, Hofvind, S, et al. Interval breast cancer rates for digital breast tomosynthesis versus digital mammography population screening: An individual participant data meta-analysis. *E Clinical Medicine*. 34: 100804. 2021. https://dx.doi.org/https://dx.doi.org/10.1016/j.eclinm.2021.100804. **KQ1E7, KQ2E7, KQ3E7**

178. Houssami, N, Lockie, D, et al. Pilot trial of digital breast tomosynthesis (3D mammography) for population-based screening in BreastScreen Victoria. *Med J Aust.* 211(8): 357-362. 2019. https://dx.doi.org/https://dx.doi.org/10.5694/mja2.50320. **KQ1E5, KQ2E5, KQ3E7j**

179. Houssami, N, Macaskill, P, et al. Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading--evidence to guide future screening strategies. *Eur J Cancer*. 50(10): 1799-1807. 2014. https://dx.doi.org/https://dx.doi.org/10.1016/j.ejca.2014.03.017. **KQ1E7d, KQ2E7d, KQ3E7d**

180. Houssami, N, Macaskill, P, et al. Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading--evidence to guide future screening strategies. *Eur J Cancer*. 50(10): 1799-1807. 2014. https://dx.doi.org/10.1016/j.ejca.2014.03.017. **KQ1E7d, KQ2E7d, KQ3E7d**

181. Hovda, T, Brandal, SHB, et al. Screening outcome for consecutive examinations with digital breast tomosynthesis versus standard digital mammography in a population-based

screening program. *Eur Radiol*. 29(12): 6991-6999. 2019. https://dx.doi.org/https://dx.doi.org/10.1007/s00330-019-06264-y. **KQ1E4, KQ2E7c, KQ3E7c**

182. Huang, C, Fann, C, et al. A Population-Based Cross-Over Randomized Controlled Trial of Breast Cancer Screening with Alternate Mammography and Ultrasound for Women Aged 40 to 49 Years in Taiwan. *Cancer Res*. 69(24 Supplement)**.** 2009. https://dx.doi.org/10.1158/0008-5472.SABCS-09-73. **KQ1E10, KQ2E10, KQ3E10**

183. 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**

184. Hunter, SA, Morris, C, et al. Digital Breast Tomosynthesis: Cost-Effectiveness of Using Private and Medicare Insurance in Community-Based Health Care Facilities. *AJR Am J Roentgenol*. 208(5): 1171-1175. 2017. https://doi.org/10.2214/ajr.16.16987. **KQ1E4, KQ2E4, KQ3E7j**

185. Huzarski, T, Gorecka-Szyld, B, et al. Screening with magnetic resonance imaging, mammography and ultrasound in women at average and intermediate risk of breast cancer. *Hered Cancer Clin Pract*. 15:4. 2017. https://dx.doi.org/https://dx.doi.org/10.1186/s13053-017-0064-y. **KQ1E7d, KQ2E7d, KQ3E7d**

186. Hwang, JY, Han, BK, et al. Screening Ultrasound in Women with Negative Mammography: Outcome Analysis. *Yonsei Med J*. 56(5): 1352-8. 2015. https://dx.doi.org/10.3349/ymj.2015.56.5.1352. **KQ1E3, KQ2E3, KQ3E3**

187. Jacklyn, G, McGeechan, K, et al. Trends in stage-specific breast cancer incidence in New South Wales, Australia: insights into the effects of 25 years of screening mammography. *Breast Cancer Res Treat*. 166(3): 843-854. 2017. https://dx.doi.org/10.1007/s10549-017-4443-x. **KQ1E2b, KQ2E2b, KQ3E2b**

188. Jensen, AR, Garne, JP, et al. Stage and survival in breast cancer patients in screened and non-screened Danish and Swedish populations. *Acta Oncol*. 42(7): 701-9. 2003. https://dx.doi.org/10.1080/02841860310010556. **KQ1E2b, KQ2E2b, KQ3E2b**

189. Jeter, LK, Morello, R, et al. Impact of Launching A High-Risk Breast Cancer Screening Program Using the Tyrer-Cuzick Model. *Am Surg*. 3134820956922. 2020. https://dx.doi.org/10.1177/0003134820956922. **KQ1E1a, KQ2E1a, KQ3E1a**

190. Jørgensen, KJ, Zahl, PH, et al. Breast cancer mortality in organised mammography screening in Denmark: comparative study. *BMJ*. 340: c1241. 2010. https://dx.doi.org/10.1136/bmj.c1241. **KQ1E3, KQ2E3, KQ3E3**

191. Jørgensen, KJ, Zahl, PH, et al. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Womens Health*. 9: 36. 2009. https://dx.doi.org/10.1186/1472-6874-9-36. **KQ1E3, KQ2E3, KQ3E3**

192. Kalager, M, Adami, HO, et al. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. *Ann Intern Med*. 156(7): 491-9. 2012. https://dx.doi.org/10.7326/0003-4819-156-7-201204030-00005. **KQ1E7h, KQ2E7h, KQ3E7h**

193. Kalager, M, Zelen, M, et al. Effect of screening mammography on breast-cancer mortality in Norway. *N Engl J Med*. 363(13): 1203-10. 2010. https://dx.doi.org/10.1056/NEJMoa1000727. **KQ1E2b, KQ2E2b, KQ3E2b**

194. Kamal, Rasha, Mansour, Sahar, et al. Detection and diagnosis of breast lesions: Performance evaluation of digital breast tomosynthesis and magnetic resonance mammography. *The Egyptian Journal of Radiology and Nuclear Medicine*. 47(3): 1159-1172. 2016. https://doi.org/10.1016/j.ejrnm.2016.06.008. **KQ1E7d, KQ2E7d, KQ3E7d**

195. Kang, E, Lee, EJ, et al. Reliability of Computer-Assisted Breast Density Estimation: Comparison of Interactive Thresholding, Semiautomated, and Fully Automated Methods. *AJR Am J Roentgenol*. 207(1): 126-34. 2016. https://doi.org/10.2214/ajr.15.15469. **KQ1E4, KQ2E4, KQ3E4**

196. Kaplan, SS. Clinical utility of bilateral whole-breast US in the evaluation of women with dense breast tissue. *Radiology*. 221(3): 641-9. 2001. https://dx.doi.org/10.1148/radiol.2213010364. **KQ1E7d, KQ2E7d, KQ3E7d**

197. Karzai, S, Port, E, et al. Impact of Screening Mammography on Treatment in Young Women Diagnosed with Breast Cancer. *Ann Surg Oncol*. 01: 01. 2022. https://doi.org/10.1245/s10434-022-11581-6. **KQ1E3, KQ2E3, KQ3E3**

198. Kelly, KM, Dean, J, et al. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *Eur Radiol*. 20(3): 734-42. 2010. https://dx.doi.org/10.1007/s00330-009-1588-y. **KQ1E7d, KQ2E7d, KQ3E7d**

199. 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**

200. Keyzer-Dekker, CM, De Vries, J, et al. Anxiety after an abnormal screening mammogram is a serious problem. *Breast*. 21(1): 83-8. 2012. https://dx.doi.org/10.1016/j.breast.2011.08.137. **KQ1E2b, KQ2E2b, KQ3E2b**

201. Kim, G, Mercaldo, S, et al. Impact of digital breast tomosynthesis (DBT) on finding types leading to true-positive and false-positive examinations. *Clin Imaging*. 71:155-159. 2021. https://dx.doi.org/https://dx.doi.org/10.1016/j.clinimag.2020.10.046. **KQ1E7c, KQ2E7c, KQ3E7c**

202. Kim, G, Mikhael, PG, et al. Ductal carcinoma in situ on digital mammography versus digital breast tomosynthesis: rates and predictors of pathologic upgrade. *Eur Radiol*. 30(11): 6089-6098. 2020. https://dx.doi.org/https://dx.doi.org/10.1007/s00330-020-07021-2. **KQ1E7c, KQ2E7c, KQ3E7c**

203. Kim, Wh, Chang, Jm, et al. Diagnostic performance of

tomosynthesis and breast ultrasonography in women with dense breasts: a prospective comparison study. *Breast Cancer Res Treat*. 162(1): 85-94. 2017. https://dx.doi.org/10.1007/s10549-017-4105-z. **KQ1E7d, KQ2E7d, KQ3E7d**

204. Kim, WH, Chang, JM, et al. Impact of prior mammograms on combined reading of digital mammography and digital breast tomosynthesis. *Acta Radiol*. 58(2): 148-155. 2017. https://dx.doi.org/10.1177/0284185116647211. **KQ1E7d, KQ2E7d, KQ3E7d**

205. Klevos, GA, Collado-Mesa, F, et al. Utility of supplemental screening with breast ultrasound in asymptomatic women with dense breast tissue who are not at high risk for breast cancer. *Indian Journal of Radiology & Imaging*. 27(1): 52-58. 2017. https://dx.doi.org/https://dx.doi.org/10.4103/0971-3026.202962. **KQ1E3, KQ2E3, KQ3E3**

206. Klompenhouwer, EG, Duijm, LE, et al. Re-attendance at biennial screening mammography following a repeated false positive recall. *Breast Cancer Res Treat*. 145(2): 429-37. 2014. https://dx.doi.org/10.1007/s10549-014-2959-x. **KQ1E2b, KQ2E2b, KQ3E2b**

207. Korpraphong, P, Limsuwarn, P, et al. Improving breast cancer detection using ultrasonography in asymptomatic women with non-fatty breast density. *Acta Radiol*. 55(8): 903-8. 2014. https://dx.doi.org/https://dx.doi.org/10.1177/0284185113507711. **KQ1E7d, KQ2E7d, KQ3E7d**

208. Kriege, M, Brekelmans, CT, et al. Factors affecting sensitivity and specificity of screening mammography and MRI in women with an inherited risk for breast cancer. *Breast Cancer Res Treat*. 100(1): 109-19. 2006. https://dx.doi.org/10.1007/s10549-006-9230-z. **KQ1E1a, KQ2E1a, KQ3E1a**

209. Kuhl, CK, Schrading, S, et al. Abbreviated breast magnetic resonance imaging (MRI): first postcontrast subtracted images and maximum-intensity projection-a novel approach to breast cancer screening with MRI. *J Clin Oncol*. 32(22): 2304-10. 2014. https://dx.doi.org/10.1200/jco.2013.52.5386. **KQ1E7d, KQ2E7d, KQ3E7d**

210. Kuhl, CK, Strobel, K, et al. Supplemental Breast MR Imaging Screening of Women with Average Risk of Breast Cancer. *Radiology*. 283(2): 361-370. 2017. https://dx.doi.org/10.1148/radiol.2016161444. **KQ1E7d, KQ2E7d, KQ3E7d**

211. Lai, YC, Ray, KM, et al. Microcalcifications Detected at Screening Mammography: Synthetic Mammography and Digital Breast Tomosynthesis versus Digital Mammography. *Radiology*. 289(3): 630-638. 2018. https://dx.doi.org/10.1148/radiol.2018181180. **KQ1E1b, KQ2E1b, KQ3E1b**

212. Lampic, C, Thurfjell, E, et al. The influence of a false-positive mammogram on a woman's subsequent behaviour for detecting breast cancer. *Eur J Cancer*. 39(12): 1730-7. 2003. https://dx.doi.org/10.1016/s0959-8049(02)00451-3. **KQ1E2b, KQ2E2b, KQ3E2b**

213. Lee, SH, Yi, A, et al. Supplemental Screening Breast US in Women with Negative Mammographic Findings: Effect of Routine Axillary Scanning.

Radiology. 286(3): 830-837. 2018. https://dx.doi.org/10.1148/radiol.2017171218. **KQ1E3, KQ2E3, KQ3E3**

214. Lehman, CD, Arao, RF, et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology*. 283(1): 49-58. 2017. https://dx.doi.org/10.1148/radiol.2016161174. **KQ1E3, KQ2E3, KQ3E3**

215. Leong, LC, Gogna, A, et al. Supplementary breast ultrasound screening in Asian women with negative but dense mammograms-a pilot study. *Ann Acad Med Singap*. 41(10): 432-9. 2012. **KQ1E7d, KQ2E7d, KQ3E7d**

216. Li, FY, Hollingsworth, A, et al. Feasibility of Breast MRI as the Primary Imaging Modality in a Large Asian Cohort. *Cureus*. 13(5): e15095. 2021. https://dx.doi.org/10.7759/cureus.15095. **KQ1E3, KQ2E3, KQ3E3**

217. Lo, G, Scaranelo, AM, et al. Evaluation of the Utility of Screening Mammography for High-Risk Women Undergoing Screening Breast MR Imaging. *Radiology*. 285(1): 36-43. 2017. https://dx.doi.org/10.1148/radiol.2017161103. **KQ1E7d, KQ2E7d, KQ3E7d**

218. Lourenco, AP, Barry-Brooks, M, et al. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology*. 274(2): 337-42. 2015. https://dx.doi.org/10.1148/radiol.14140317. **KQ1E7c, KQ2E7c, KQ3E7c**

219. Lowe, JB, Balanda, KP, et al. Psychologic distress in women with abnormal findings in mass mammography screening. *Cancer*. 85(5): 1114-8. 1999. https://dx.doi.org/10.1002/(sici)1097-0142(19990301)85:5<1114::aid-cncr15>3.0.co;2-y. **KQ1E4, KQ2E4, KQ3E3**

220. Lowry, KP, Coley, RY, et al. Screening Performance of Digital Breast Tomosynthesis vs Digital Mammography in Community Practice by Patient Age, Screening Round, and Breast Density. *JAMA Netw Open*. 3(7): e2011792. 2020. https://dx.doi.org/10.1001/*JAMA*networkopen.2020.11792. **KQ1E7, KQ2E5, KQ3E7j**

221. Malik, B, Iuanow, E, et al. An Exploratory Multi-reader, Multi-case Study Comparing Transmission Ultrasound to Mammography on Recall Rates and Detection Rates for Breast Cancer Lesions. *Acad Radiol*. 03: 03. 2020. https://dx.doi.org/https://dx.doi.org/10.1016/j.acra.2020.11.011. **KQ1E7e, KQ2E7e, KQ3E7e**

222. Mao, Z, Nystrom, L, et al. Breast cancer screening with mammography in women aged 40-49 years: Impact of length of screening interval on effectiveness of the program. *J Med Screen*. 28(2): 200-206. 2021. https://dx.doi.org/10.1177/0969141320918283. **KQ1E7h, KQ2E7h, KQ3E7h**

223. Mao, Z, Nystrom, L, et al. Effectiveness of Population-Based Service Screening with Mammography for Women Aged 70-74 Years in Sweden. *Cancer Epidemiol Biomarkers Prevent*. 29(11): 2149-2156. 2020. https://dx.doi.org/10.1158/1055-9965.EPI-20-0523. **KQ1E7c, KQ2E7c, KQ3E7c**

224. Mariscotti, G, Durando, M, et al. Digital breast tomosynthesis as an adjunct to digital mammography for detecting and characterising invasive lobular cancers: a multi-reader study. *Clin Radiol*. 71(9): 889-95. 2016. https://dx.doi.org/10.1016/j.crad.2016.04.004. **KQ1E1b, KQ2E1b, KQ3E1b**

225. Maxwell, AJ, Beattie, C, et al. The effect of false positive breast screening examinations on subsequent attendance: retrospective cohort study. *J Med Screen*. 20(2): 91-8. 2013. https://dx.doi.org/10.1177/0969141313499147. **KQ1E2b, KQ2E2b, KQ3E2b**

226. Maxwell, AJ, Michell, M, et al. A randomised trial of screening with digital breast tomosynthesis plus conventional digital 2D mammography versus 2D mammography alone in younger higher risk women. *Eur J Radiol*. 94: 133-139. 2017. https://dx.doi.org/10.1016/j.ejrad.2017.06.018. **KQ1E1, KQ2E1, KQ3E1**

227. McCann, J, Stockton, D, et al. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Res*. 4(5): R11. 2002. https://dx.doi.org/10.1186/bcr455. **KQ1E3, KQ2E3, KQ3E3**

228. McCarthy, AM, Kontos, D, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *J Natl Cancer Inst*. 106(11)**.** 2014. https://dx.doi.org/10.1093/jnci/dju316. **KQ1E7c, KQ2E7c, KQ3E7c**

229. McDonald, ES, McCarthy, AM, et al. Baseline Screening Mammography: Performance of Full-Field Digital Mammography Versus Digital Breast Tomosynthesis. *AJR Am J Roentgenol*. 205(5): 1143-8. 2015. https://dx.doi.org/https://dx.doi.org/10.2214/AJR.15.14406. **KQ1E7c, KQ2E7c, KQ3E7c**

230. McDonald, ES, McCarthy, AM, et al. BI-RADS Category 3 Comparison: Probably Benign Category after Recall from Screening before and after Implementation of Digital Breast Tomosynthesis. *Radiology*. 285(3): 778-787. 2017. https://dx.doi.org/10.1148/radiol.2017162837. **KQ1E7c, KQ2E7c, KQ3E7c**

231. McDonald, ES, Oustimov, A, et al. Effectiveness of Digital Breast Tomosynthesis Compared With Digital Mammography: Outcomes Analysis From 3 Years of Breast Cancer Screening. *JAMA Oncol*. 2(6): 737-43. 2016. https://dx.doi.org/10.1001/*JAMA*oncol.2015.5536. **KQ1E7c, KQ2E7c, KQ3E7c**

232. Meldrum, P, Turnbull, D, et al. Tailored written invitations for second round breast cancer screening: a randomised controlled trial. *J Med Screen*. 1(4): 245-8. 1994. https://dx.doi.org/10.1177/096914139400100412. **KQ1E2b, KQ2E2b, KQ3E2b**

233. Mesurolle, B, El Khoury, M, et al. Is there any added value to substitute the 2D digital MLO projection for a MLO tomosynthesis projection and its synthetic view when a 2D standard digital mammography is used in a one-stop-shop immediate reading mammography screening? *Eur Radiol*. 28: 28. 2021. https://dx.doi.org/https://dx.doi.org/10.1007/s00330-021-07999-3. **KQ1E7c, KQ2E7c, KQ3E7c**

234. Miller, AB, Wall, C, et al. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ*. 348: g366. 2014. https://dx.doi.org/10.1136/bmj.g366. **KQ1E3, KQ2E3, KQ3E3**

235. Miller, D, Livingstone, V, et al. Interventions for relieving the pain and discomfort of screening mammography. *Cochrane Database Syst Rev*. (1): Cd002942. 2008. https://dx.doi.org/10.1002/14651858.CD002942.pub2. **KQ1E7a, KQ2E7a, KQ3E7a**

236. Miltenburg, GA, Peeters, PH, et al. Seventeen-year evaluation of breast cancer screening: the DOM project, The Netherlands. Diagnostisch Onderzoek (investigation) Mammacarcinoom. *Br J Cancer*. 78(7): 962-5. 1998. https://doi.org/10.1038%2Fbjc.1998.609. **KQ1E3, KQ2E3, KQ3E3**

237. Moger, TA, Swanson, JO, et al. Cost differences between digital tomosynthesis and standard digital mammography in a breast cancer screening programme: results from the To-Be trial in Norway. *Eur J Health Econ*. 20(8): 1261-1269. 2019. https://dx.doi.org/https://dx.doi.org/10.1007/s10198-019-01094-7. **KQ1E4, KQ2E4, KQ3E4**

238. Mook, S, Van 't Veer, LJ, et al. Independent prognostic value of screen detection in invasive breast cancer. *J Natl Cancer Inst*. 103(7): 585-97. 2011. https://dx.doi.org/10.1093/jnci/djr043. **KQ1E3, KQ2E3, KQ3E3**

239. Moon, HJ, Jung, I, et al. Comparison of Cancer Yields and Diagnostic Performance of Screening Mammography vs. Supplemental Screening Ultrasound in 4394 Women with Average Risk for Breast Cancer. *Ultraschall Med*. 36(3): 255-63. 2015. https://dx.doi.org/10.1055/s-0034-1366288. **KQ1E7i, KQ2E7i, KQ3E7i**

240. 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**

241. Morrell, S, Barratt, A, et al. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. *Cancer Causes Control*. 21(2): 275-82. 2010. https://dx.doi.org/10.1007/s10552-009-9459-z. **KQ1E7h, KQ2E7h, KQ3E7h**

242. Moss, S, Thomas, I, et al. Randomised controlled trial of mammographic screening in women from age 40: results of screening in the first 10 years. *Br J Cancer*. 92(5): 949-54. 2005. https://dx.doi.org/10.1038/sj.bjc.6602396. **KQ1E3, KQ2E3, KQ3E3**

243. Moss, S. Overdiagnosis and overtreatment of breast cancer: overdiagnosis in randomised controlled trials of breast cancer screening. *Breast Cancer Res*. 7(5): 230-4. 2005. https://dx.doi.org/10.1186/bcr1314. **KQ1E7a, KQ2E7a, KQ3E7a**

244. Moss, SM, Cuckle, H, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet*. 368(9552):

2053-60. 2006. https://dx.doi.org/10.1016/s0140-6736(06)69834-6. **KQ1E4, KQ2E4, KQ3E4**

245. Moss, SM, Nyström, L, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies. *J Med Screen*. 19 Suppl 1: 26-32. 2012. https://dx.doi.org/10.1258/jms.2012.012079. **KQ1E7a, KQ2E7a, KQ3E7a**

246. Moss, SM, Wale, C, et al. 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. *Lancet Oncol*. 16(9): 1123-1132. 2015. https://dx.doi.org/10.1016/S1470-2045(15)00128-X. **KQ1E1a, KQ2E1a, KQ3E1a**

247. Narayan, AK, Visvanathan, K, et al. Comparative effectiveness of breast MRI and mammography in screening young women with elevated risk of developing breast cancer: a retrospective cohort study. *Breast Cancer Res Treat*. 158(3): 583-9. 2016. https://dx.doi.org/10.1007/s10549-016-3912-y. **KQ1E1a, KQ2E1a, KQ3E1a**

248. Narod, SA, Sun, P, et al. Impact of screening mammography on mortality from breast cancer before age 60 in women 40 to 49 years of age. *Current Oncology*. 21(5): 217-21. 2014. https://dx.doi.org/10.3747/co.21.2067. **KQ1E3, KQ2E3, KQ3E3**

249. Njor, S, Nyström, L, et al. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. *J Med Screen*. 19 Suppl 1: 33-41. 2012. https://dx.doi.org/10.1258/jms.2012.012080. **KQ1E7, KQ2E7, KQ3E7**

250. Njor, SH, Olsen, AH, et al. Overdiagnosis in screening mammography in Denmark: population based cohort study. *BMJ*. 346: f1064. 2013. https://dx.doi.org/10.1136/*BMJ*.f1064. **KQ1E3, KQ2E3, KQ3E3**

251. Noguchi, N, Marinovich, ML, et al. Evidence from a BreastScreen cohort does not support a longer inter-screen interval in women who have no conventional risk factors for breast cancer. *Breast*. 62: 16-21. 2022. https://doi.org/10.1016/j.breast.2022.01.015. **KQ1E3, KQ2E3, KQ3E3**

252. Nyström, L, Andersson, I, et al. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet*. 359(9310): 909-19. 2002. https://dx.doi.org/10.1016/s0140-6736(02)08020-0. **KQ1E3, KQ2E3, KQ3E3**

253. Ohashi, R, Nagao, M, et al. Improvement in diagnostic performance of breast cancer: comparison between conventional digital mammography alone and conventional mammography plus digital breast tomosynthesis. *Breast Cancer*. 25(5): 590-596. 2018. https://dx.doi.org/10.1007/s12282-018-0859-3. **KQ1E1b, KQ2E1b, KQ3E1b**

254. Olivotto, IA, Kan, L, et al. False positive rate of screening mammography. *N Engl J Med*. 339(8): 560; author reply 563. 1998. https://dx.doi.org/10.1056/nejm199808203390812. **KQ1E7a, KQ2E7a, KQ3E7a**

255. Olivotto, IA, Mates, D, et al. Prognosis, treatment, and recurrence of breast cancer for women attending or not attending the Screening Mammography Program of British Columbia. *Breast Cancer Res Treat*. 54(1): 73-81. 1999. https://dx.doi.org/10.1023/a:1006152918283. **KQ1E2b, KQ2E2b, KQ3E2b**

256. Olsen, AH, Agbaje, OF, et al. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast J*. 12(4): 338-42. 2006. https://dx.doi.org/10.1111/j.1075-122X.2006.00272.x. **KQ1E3, KQ2E3, KQ3E3**

257. Olsen, AH, Njor, SH, et al. Breast cancer mortality in Copenhagen after introduction of mammography screening: cohort study. *BMJ*. 330(7485): 220. 2005. https://dx.doi.org/10.1136/*BMJ*.38313.639236.82. **KQ1E2b, KQ2E2b, KQ3E2b**

258. Olsson, A, Garne, JP, et al. Overweight in relation to tumour size and axillary lymph node involvement in postmenopausal breast cancer patients-differences between women invited to vs. not invited to mammography in a randomized screening trial. *Cancer Epidemiol*. 33(1): 9-15. 2009. https://dx.doi.org/10.1016/j.canep.2009.04.008. **KQ1E3, KQ2E3, KQ3E3**

259. 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**

260. Ong, G, Austoker, J, et al. Breast screening: adverse psychological consequences one month after placing women on early recall because of a diagnostic uncertainty. A multicentre study. *J Med Screen*. 4(3): 158-68. 1997. https://dx.doi.org/10.1177/096914139700400309. **KQ1E3, KQ2E3, KQ3E3**

261. Orton, M, Fitzpatrick, R, et al. Factors affecting women's response to an invitation to attend for a second breast cancer screening examination. *Br J Gen Pract*. 41(349): 320-2. 1991. http://www.ncbi.nlm.nih.gov/pmc/articles/pmc1371753/. **KQ1E4, KQ2E4, KQ3E4**

262. Osteras, BH, Martinsen, ACT, et al. Digital Mammography versus Breast Tomosynthesis: Impact of Breast Density on Diagnostic Performance in Population-based Screening. *Radiology*. 293(1): 60-68. 2019. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2019190425. **KQ1E7d, KQ2E7d, KQ3E7d**

263. O'Sullivan, I, Sutton, S, et al. False positive results do not have a negative effect on reattendance for subsequent breast screening. *J Med Screen*. 8(3): 145-8. 2001. https://dx.doi.org/10.1136/jms.8.3.145. **KQ1E3, KQ2E3, KQ3E3**

264. Otto, SJ, Fracheboud, J, et al. Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. *Lancet*. 361(9367): 1411-7. 2003. https://dx.doi.org/10.1016/s0140-6736(03)13132-7. **KQ1E3, KQ2E3, KQ3E3**

265. Paci, E, Giorgi, D, et al. Assessment of the early impact of the population-based breast cancer screening programme in Florence (Italy) using mortality and surrogate measures. *Eur J Cancer*. 38(4): 568-73. 2002. https://dx.doi.org/10.1016/s0959-8049(01)00382-3. **KQ1E2b, KQ2E2b, KQ3E2b**

266. Paci, E, Mantellini, P, et al. Tailored Breast Screening Trial (TBST). *Epidemiol Prev*. 37(4-5): 317-327. 2013. **KQ1E8, KQ2E8, KQ3E8**

267. Paci, E, Miccinesi, G, et al. Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. *Breast Cancer Res*. 8(6): R68. 2006. https://dx.doi.org/10.1186/bcr1625. **KQ1E7h, KQ2E7h, KQ3E7h**

268. Paci, E, Warwick, J, et al. Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? *J Med Screen*. 11(1): 23-7. 2004. https://dx.doi.org/10.1177/09691413031100106. **KQ1E7h, KQ2E7h, KQ3E7h**

269. Padia, SA, Freyvogel, M, et al. False-positive Extra-Mammary Findings in Breast MRI: Another Cause for Concern. *Breast Journal*. 22(1): 90-5. 2016. https://doi.org/10.1111/tbj.12524. **KQ1E1b, KQ2E1b, KQ3E1b**

270. Pan, HB, Wong, KF, et al. Breast cancer screening with digital breast tomosynthesis - 4 year experience and comparison with national data. *JCMA*. 81(1): 70-80. 2018. https://dx.doi.org/10.1016/j.jcma.2017.05.013. **KQ1E4, KQ2E5, KQ3E7c**

271. Pang, JX, Newsome, J, et al. Impact of switching from digital mammography to tomosynthesis plus digital mammography on breast cancer screening in Alberta, Canada. *J Med Screen*. 9691413211032265. 2021. https://dx.doi.org/10.1177/09691413211032265. **KQ1E7c, KQ2E7c, KQ3E7c**

272. 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**

273. Parris, T, Wakefield, D, et al. Real world performance of screening breast ultrasound following enactment of Connecticut Bill 458. *Breast J*. 19(1): 64-70. 2013. https://dx.doi.org/10.1111/tbj.12053. **KQ1E7c, KQ2E7c, KQ3E7c**

274. Parvinen, I, Heinavaara, S, et al. Mammography screening in three Finnish residential areas: comprehensive population-based study of breast cancer incidence and incidence-based mortality 1976-2009. *Br J Cancer*. 112(5): 918-24. 2015. https://dx.doi.org/10.1038/bjc.2014.642. **KQ1E7c, KQ2E7c, KQ3E7c**

275. Philadelpho, F, Calas, MJG, et al. Comparison of Automated Breast Ultrasound and Hand-Held Breast Ultrasound in the Screening of Dense Breasts. *Rev Bras Ginecol Obstet*. 43(3): 190-199. 2021. https://dx.doi.org/10.1055/s-0040-1722156. **KQ1E7d, KQ2E7d, KQ3E7d**

276. Poplack, SP, Patel, AK, et al. The impact of adjunctive tomosynthesis on screening mammography outcomes in two widely diverse radiology practices.

Breast Journal. 27(1): 13-20. 2021. https://dx.doi.org/10.1111/tbj.14121. **KQ1E4, KQ2E5, KQ3E7j**

277. Porter, AJ, Evans, EB, et al. Full-field digital mammography: Is the apparent increased detection of microcalcification leading to over-investigation and over-diagnosis?. *J Med Imaging Radiat Oncol*. 61(4): 470-475. 2017. https://dx.doi.org/10.1111/1754-9485.12581. **KQ1E2b, KQ2E2b, KQ3E2b**

278. Powell, JL, Hawley, JR, et al. Impact of the Addition of Digital Breast Tomosynthesis (DBT) to Standard 2D Digital Screening Mammography on the Rates of Patient Recall, Cancer Detection, and Recommendations for Short-term Follow-up. *Acad Radiol*. 24(3): 302-307. 2017. https://dx.doi.org/10.1016/j.acra.2016.10.001, **KQ1E5, KQ2E5, KQ3E7j**

279. Procasco, M. Comparison of Digital Breast Tomosynthesis vs Full-Field Digital Mammography in Recall Rates and Cancer Detection Rates. *Radiol Technol*. 87(3): 349-51. 2016. **KQ1E5, KQ2E5, KQ3E1b**

280. Puliti, D, Duffy, SW, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen*. 19 Suppl 1: 42-56. 2012. https://dx.doi.org/10.1258/jms.2012.012082. **KQ1E7a, KQ2E7a, KQ3E7a**

281. Puliti, D, Miccinesi, G, et al. Effectiveness of service screening: a case-control study to assess breast cancer mortality reduction. *Br J Cancer*. 99(3): 423-7. 2008. https://dx.doi.org/10.1038/sj.bjc.6604532. **KQ1E2b, KQ2E2b, KQ3E2b**

282. Puliti, D, Zappa, M, et al. An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. *Eur J Cancer*. 45(18): 3166-71. 2009. https://dx.doi.org/10.1016/j.ejca.2009.06.014. **KQ1E2b, KQ2E2b, KQ3E2b**

283. Rafferty, EA, Durand, MA, et al. Breast Cancer Screening Using Tomosynthesis and Digital Mammography in Dense and Nondense Breasts. *JAMA*. 315(16): 1784-6. 2016. https://doi.org/10.1001/jama.2016.1708. **KQ1E4, KQ2E5, KQ3E7j**

284. Rafferty, EA, Durand, MA, et al. Breast Cancer Screening Using Tomosynthesis and Digital Mammography in Dense and Nondense Breasts. *JAMA*. 315(16): 1784-6. 2016. https://doi.org/10.1001/jama.2016.1708. **KQ1E5, KQ2E5, KQ3E7j**

285. Rafferty, EA, Park, JM, et al. Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. *Radiology*. 266(1): 104-13. 2013. https://dx.doi.org/10.1148/radiol.12120674. **KQ1E7d, KQ2E7d, KQ3E7d**

286. Rafferty, EA, Rose, SL, et al. Effect of age on breast cancer screening using tomosynthesis in combination with digital mammography. *Breast Cancer Res Treat*. 164(3): 659-666. 2017. https://dx.doi.org/10.1007/s10549-017-4299-0. **KQ1E7c, KQ2E7c, KQ3E7c**

287. 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

New South Wales, Australia. *Cancer Causes Control*. 20(5): 559-66. 2009. https://doi.org/10.1007/s10552-008-9264-0. **KQ1E4, KQ2E11, KQ3E4**

288. Rauscher, GH, Murphy, AM, et al. The "Sweet Spot" Revisited: Optimal Recall Rates for Cancer Detection With 2D and 3D Digital Screening Mammography in the Metro Chicago Breast Cancer Registry. *AJR Am J Roentgenol*. 216(4): 894-902. 2021. https://dx.doi.org/https://dx.doi.org/10.2214/AJR.19.22429. **KQ1E3, KQ2E3, KQ3E3**

289. Redondo, A, Comas, M, et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *Br J Radiol*. 85(1019): 1465-70. 2012. https://dx.doi.org/10.1259/bjr/21256379. **KQ1E7d, KQ2E7d, KQ3E7d**

290. Regen-Tuero, HC, Ram, S, et al. Community-Based Breast Cancer Screening Using Digital Breast Tomosynthesis Versus Digital Mammography: Comparison of Screening Performance and Tumor Characteristics. *AJR Am J Roentgenol*. 218(2): 249-257. 2022. https://doi.org/10.2214/ajr.21.26384. **KQ1E5, KQ2E5, KQ3E7j**

291. Romero Martin, S, Raya Povedano, JL, et al. Prospective study aiming to compare 2D mammography and tomosynthesis + synthesized mammography in terms of cancer detection and recall. From double reading of 2D mammography to single reading of tomosynthesis. *Eur Radiol*. 28(6): 2484-2491. 2018. https://dx.doi.org/10.1007/s00330-017-5219-8. **KQ1E7d, KQ2E7d, KQ3E7d**

292. Rose, SL, Shisler, JL. Tomosynthesis Impact on Breast Cancer Screening in Patients Younger Than 50 Years Old. *AJR Am J Roentgenol*. 210(6): 1401-1404. 2018. https://dx.doi.org/10.2214/AJR.17.18839. **KQ1E4, KQ2E5, KQ3E7j**

293. Rose, SL, Tidwell, AL, et al. A reader study comparing prospective tomosynthesis interpretations with retrospective readings of the corresponding FFDM examinations. *Acad Radiol*. 21(9): 1204-10. 2014. https://dx.doi.org/10.1016/j.acra.2014.04.008. **KQ1E7d, KQ2E7d, KQ3E7d**

294. Rose, SL, Tidwell, AL, et al. Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR Am J Roentgenol*. 200(6): 1401-8. 2013. https://dx.doi.org/10.2214/ajr.12.9672. **KQ1E7c, KQ2E7c, KQ3E7c**

295. Roth, RG, Maidment, AD, et al. Digital breast tomosynthesis: lessons learned from early clinical implementation. Radiographics. 34(4): E89-102. 2014. https://doi.org/10.1148/rg.344130087. **KQ1E4, KQ2E4, KQ3E7j**

296. Saadatmand, S, Geuzinge, HA, et al. MRI versus mammography for breast cancer screening in women with familial risk (FaMRIsc): a multicentre, randomised, controlled trial. *Lancet Oncol*. 20(8): 1136-1147. 2019. https://dx.doi.org/10.1016/S1470-2045(19)30275-X. **KQ1E1a, KQ2E1a, KQ3E1a**

297. Sanderson, M, Levine, RS, et al. Mammography Screening Among the Elderly: A Research Challenge. *Am J Med*. 128(12): 1362.e7-14. 2015.

https://dx.doi.org/10.1016/j.amjmed.2015.06.032. **KQ1E3, KQ2E4, KQ3E4**

298. Sarkeala, T, Heinävaara, S, et al. Breast cancer mortality with varying invitational policies in organised mammography. *Br J Cancer*. 98(3): 641-5. 2008. https://dx.doi.org/10.1038/sj.bjc.6604203. **KQ1E3, KQ2E3, KQ3E3**

299. Sarkeala, T, Luostarinen, T, et al. Breast carcinoma detection modes and death in a female population in relation to population-based mammography screening. Springerplus. 3: 348. 2014. https://dx.doi.org/https://dx.doi.org/10.1186/2193-1801-3-348. **KQ1E7h, KQ2E7h, KQ3E7h**

300. Savaridas, SL, Gierlinski, M, et al. Opting into breast screening over the age of 70 years: seeking evidence to support informed choice. *Clin Radiol*. 13: 13. 2022. https://doi.org/10.1016/j.crad.2022.01.057. **KQ1E3, KQ2E3, KQ3E3**

301. Schou Bredal, I, Kåresen, R, et al. Recall mammography and psychological distress. *Eur J Cancer*. 49(4): 805-11. 2013. https://dx.doi.org/10.1016/j.ejca.2012.09.001. **KQ1E4, KQ2E4, KQ3E3**

302. Scott, AM, Lashley, MG, et al. Comparison of Call-Back Rates between Digital Mammography and Digital Breast Tomosynthesis. *Am Surg*. 85(8): 855-857. 2019. **KQ1E4, KQ2E4, KQ3E7j**

303. Seely, JM, Peddle, SE, et al. Breast Density and Risk of Interval Cancers: The Effect of Annual Versus Biennial Screening Mammography Policies in Canada. *Can Assoc Radiol J*. 73(1): 90-100. 2022. https://doi.org/10.1177/08465371211027958. **KQ1E4, KQ2E4, KQ3E7i**

304. Seely, JM, Peddle, SE, et al. Breast Density and Risk of Interval Cancers: The Effect of Annual Versus Biennial Screening Mammography Policies in Canada. Can Assoc Radiol J. :8465371211027958. 2021. https://dx.doi.org/https://dx.doi.org/10.1177/08465371211027958. **KQ1E7, KQ2E7, KQ3E7**

305. Seigneurin, A, François, O, et al. Overdiagnosis from non-progressive cancer detected by screening mammography: stochastic simulation study with calibration to population based registry data. *BMJ*. 343: d7017. 2011. https://dx.doi.org/10.1136/*BMJ*.d7017. **KQ1E7h, KQ2E7h, KQ3E7h**

306. Sharma, N, McMahon, M, et al. The Potential Impact of Digital Breast Tomosynthesis on the Benign Biopsy Rate in Women Recalled within the UK Breast Screening Programme. *Radiology*. 291(2): 310-317. 2019. https://dx.doi.org/10.1148/radiol.2019180809. **KQ1E1b, KQ2E1b, KQ3E1b**

307. Sharpe, RE, Jr, Venkataraman, et al. Increased Cancer Detection Rate and Variations in the Recall Rate Resulting from Implementation of 3D Digital Breast Tomosynthesis into a Population-based Screening Program. *Radiology*. 278(3): 698-706. 2016. https://dx.doi.org/10.1148/radiol.2015142036. **KQ1E5, KQ2E5, KQ3E7j**

308. Shaughnessy, AF. Adding Ultrasonography to Mammography Increases False-Positive Findings Without an Increase in Cancer

Detection. *Am Fam Physician*. 101(1): 53-54. 2020. **KQ1E7a, KQ2E7a, KQ3E7a**

309. Shermis, RB, Wilson, KD, et al. Supplemental Breast Cancer Screening With Molecular Breast Imaging for Women With Dense Breast Tissue. *AJR Am J Roentgenol*. 207(2): 450-7. 2016. https://dx.doi.org/10.2214/AJR.15.15924. **KQ1E7d, KQ2E7d, KQ3E7d**

310. Shieh, Y, Eklund, M, et al. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *J Natl Cancer Inst*. 109(5): 01. 2017. https://dx.doi.org/10.1093/jnci/djw290. **KQ1E10, KQ2E10, KQ3E10**

311. Sia, J, Moodie, K, et al. A prospective study comparing digital breast tomosynthesis with digital mammography in surveillance after breast cancer treatment. *Eur J Cancer*. 61: 122-7. 2016. https://dx.doi.org/10.1016/j.ejca.2016.04.007. **KQ1E1b, KQ2E1b, KQ3E1b**

312. 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**

313. Skaane, P, Bandos, AI, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology*. 267(1): 47-56. 2013. https://dx.doi.org/10.1148/radiol.12121373. **KQ1E7d, KQ2E7d, KQ3E7d**

314. Skaane, P, Bandos, Ai, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a populationbased screening program. *Radiology*. 267(1): 47-56. 2013. https://dx.doi.org/10.1148/radiol.12121373. **KQ1E7d, KQ2E7d, KQ3E7d**

315. Skaane, P, Bandos, AI, et al. Digital Mammography versus Digital Mammography Plus Tomosynthesis in Breast Cancer Screening: The Oslo Tomosynthesis Screening Trial. *Radiology*. 291(1): 23-30. 2019. https://dx.doi.org/10.1148/radiol.2019182394. **KQ1E7d, KQ2E7d, KQ3E7d**

316. Skaane, P, Bandos, AI, et al. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. *Eur Radiol*. 23(8): 2061-71. 2013. https://dx.doi.org/10.1007/s00330-013-2820-3. **KQ1E7d, KQ2E7d, KQ3E7d**

317. Skaane, P, Bandos, AI, et al. Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology*. 271(3): 655-63. 2014. https://dx.doi.org/10.1148/radiol.13131391. **KQ1E7d, KQ2E7d, KQ3E7d**

318. Skaane, P, Sebuodegard, S, 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*. 169(3): 489-496. 2018.

https://dx.doi.org/10.1007/s10549-018-4705-2. **KQ1E7d, KQ2E7d, KQ3E7d**

319. Solbjor, M, Forsmo, S, et al. Psychosocial consequences among women with false-positive results after mammography screening in Norway. *Scand J Prim Health Care*. 36(4): 380-389. 2018. https://dx.doi.org/10.1080/02813432.2018.1523985. **KQ1E3, KQ2E3, KQ3E3**

320. Song, SE, Cho, N, et al. Undiagnosed Breast Cancer: Features at Supplemental Screening US. *Radiology*. 277(2): 372-80. 2015. https://dx.doi.org/10.1148/radiol.2015142960. **KQ1E3, KQ2E3, KQ3E3**

321. Spayne, MC, Gard, CC, et al. Reproducibility of BI-RADS breast density measures among community radiologists: a prospective cohort study. *Breast J*. 18(4): 326-33. 2012. https://dx.doi.org/10.1111/j.1524-4741.2012.01250.x. **KQ1E4, KQ2E4, KQ3E4**

322. Sprague, BL, Chen, S, et al. Cumulative 6-Year Risk of Screen-Detected Ductal Carcinoma In Situ by Screening Frequency. *JAMA Netw Open*. 6(2): e230166. 2023. https://dx.doi.org/10.1001/*JAMA*network open.2023.0166. **KQ1E7h, KQ2E7h, KQ3E7h**

323. Sprague, BL, Coley, RY, et al. Assessment of Radiologist Performance in Breast Cancer Screening Using Digital Breast Tomosynthesis vs Digital Mammography. *JAMA Network Open*. 3(3): e201759. 2020. https://dx.doi.org/10.1001/*JAMA*network open.2020.1759. **KQ1E4, KQ2E4, KQ3E3**

324. Starikov, A, Drotman, M, et al. 2D mammography, digital breast tomosynthesis, and ultrasound: which should be used for the different breast densities in breast cancer screening? *Clin Imaging*. 40(1): 68-71. 2016. https://dx.doi.org/10.1016/j.clinimag.2015.10.001. **KQ1E4, KQ2E5, KQ3E7j**

325. Stepanek, T, Constantinou, N, et al. Changes in the Utilization of the BI-RADS Category 3 Assessment in Recalled Patients Before and After the Implementation of Screening Digital Breast Tomosynthesis. *Acad Radiol*. 26(11): 1515-1525. 2019. https://dx.doi.org/10.1016/j.acra.2018.12.020. **KQ1E7c, KQ2E7c, KQ3E7c**

326. Strobel, K, Schrading, S, et al. Assessment of BI-RADS category 4 lesions detected with screening mammography and screening US: utility of MR imaging. *Radiology*. 274(2): 343-51. 2015. https://dx.doi.org/10.1148/radiol.14140645. **KQ1E7d, KQ2E7d, KQ3E7d**

327. Sumkin, JH, Ganott, MA, et al. Recall Rate Reduction with Tomosynthesis During Baseline Screening Examinations: An Assessment From a Prospective Trial. *Acad Radiol*. 22(12): 1477-82. 2015. https://dx.doi.org/10.1016/j.acra.2015.08.015. **KQ1E7d, KQ2E7d, KQ3E7d**

328. Svahn, TM, Chakraborty, DP, et al. Breast tomosynthesis and digital mammography: a comparison of diagnostic accuracy. *Br J Radiol*. 85(1019): e1074-82. 2012. https://dx.doi.org/10.1259/bjr/53282892. **KQ1E7d, KQ2E7d, KQ3E7d**

329. Tabár, L, Faberberg, G, et al. What is the optimum interval between

mammographic screening examinations? An analysis based on the latest results of the Swedish two-county breast cancer screening trial. *Br J Cancer*. 55(5): 547-51. 1987. https://doi.org/10.1038/bjc.1987.112. **KQ1E3, KQ2E3, KQ3E3**

330. Tabar, L, Fagerberg, G, et al. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. *Cancer*. 75(10): 2507-17. 1995. https://dx.doi.org/10.1002/1097-0142(19950515)75:10<2507::aid-cncr2820751017>3.0.co;2-h. **KQ1E3, KQ2E3, KQ3E3**

331. Tabar, L, Vitak, B, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology*. 260(3): 658-663. 2011. https://dx.doi.org/10.1148/radiol.11110469. **KQ1E2b, KQ2E2b, KQ3E2b**

332. Tagliafico, A, Astengo, D, et al. One-to-one comparison between digital spot compression view and digital breast tomosynthesis. *Eur Radiol*. 22(3): 539-44. 2012. https://dx.doi.org/10.1007/s00330-011-2305-1. **KQ1E1b, KQ2E1b, KQ3E1b**

333. Tagliafico, AS, Calabrese, M, et al. Adjunct Screening With Tomosynthesis or Ultrasound in Women With Mammography-Negative Dense Breasts: Interim Report of a Prospective Comparative Trial. *J Clin Oncol*. 34(16): 1882-1888. 2016. https://doi.org/10.1016/j.ejca.2018.08.029. **KQ1E7d, KQ2E7d, KQ3E7d**

334. Tagliafico, AS, Mariscotti, G, et al. A prospective comparative trial of adjunct screening with tomosynthesis or ultrasound in women with mammography-negative dense breasts (ASTOUND-2). *Eur J Cancer*. 104: 39-46. 2018. https://dx.doi.org/10.1016/j.ejca.2018.08.029. **KQ1E7d, KQ2E7d, KQ3E7d**

335. Taha Ali, TamerF, Magid, AsmaaMA, et al. Potential impact of tomosynthesis on the detection and diagnosis of breast lesions. *The Egyptian Journal of Radiology and Nuclear Medicine*. 47(1): 351-361. 2016. https://doi.org/10.1016/j.ejrnm.2015.10.006. **KQ1E7d, KQ2E7d, KQ3E7d**

336. Teertstra, HJ, Loo, CE, et al. Breast tomosynthesis in clinical practice: initial results. *Eur Radiol*. 20(1): 16-24. 2010. https://dx.doi.org/10.1007/s00330-009-1523-2. **KQ1E1b, KQ2E1b, KQ3E1b**

337. Teoh, KC, Manan, HA, et al. Comparison of Mean Glandular Dose between Full-Field Digital Mammography and Digital Breast Tomosynthesis. *Healthcare*. 9(12): 19. 2021. https://doi.org/10.3390/healthcare9121758. **KQ1E4, KQ2E4, KQ3E7j**

338. Thibault, F, Dromain, C, et al. Digital breast tomosynthesis versus mammography and breast ultrasound: a multireader performance study. *Eur Radiol*. 23(9): 2441-9. 2013. https://dx.doi.org/10.1007/s00330-013-2863-5. **KQ1E1b, KQ2E1b, KQ3E1b**

339. Thomassin-Naggara, I, Perrot, N, et al. Added value of one-view breast tomosynthesis combined with digital mammography according to reader experience. *Eur J Radiol*. 84(2): 235-41. 2015. https://dx.doi.org/10.1016/j.ejrad.2014.10.022. **KQ1E7d, KQ2E7d, KQ3E7d**

340. Timmermans, L, Bleyen, L, et al. Screen-detected versus interval cancers: Effect of imaging modality and breast density in the Flemish Breast Cancer Screening Programme. *Eur Radiol*. 27(9): 3810-3819. 2017. https://dx.doi.org/https://dx.doi.org/10.1007/s00330-017-4757-4. **KQ1E2b, KQ2E2b, KQ3E2b**

341. Tosteson, AN, Fryback, DG, et al. Consequences of false-positive screening mammograms. *JAMA Intern Med*. 174(6): 954-61. 2014. https://dx.doi.org/10.1001/*JAMA*internmed.2014.981. **KQ1E4, KQ2E4, KQ3E3**

342. Tsuruda, KM, Larsen, M, et al. Cumulative risk of a false-positive screening result: A retrospective cohort study using empirical data from 10 biennial screening rounds in BreastScreen Norway. *Cancer*. 128(7): 1373-1380. 2022. https://doi.org/10.1002/cncr.34078. **KQ1E3, KQ2E3, KQ3E3**

343. Uematsu, T. Sensitivity and specificity of screening mammography without clinical breast examination among Japanese women aged 40-49 years: analysis of data from the J-START results. Breast Cancer. 29(5): 928-931. 2022. https://doi.org/10.1007/s12282-022-01353-1. **KQ1E4, KQ2E4, KQ3E4**

344. Upneja, A, Long, JB, et al. Comparative Effectiveness of Digital Breast Tomosynthesis and Mammography in Older Women. *J Gen Intern Med*. 1-7. 2021. https://doi.org/10.1007/s11606-021-07132-6. **KQ1E5, KQ2E5, KQ3E7j**

345. Valencia, EM, Pruthi, S. Practice-Changing Opportunity to Reduce Disparities in Screening Mammography-Implementation of Risk Assessment in Primary Care. *JAMA Net Open*. 4(9): e2124535. 2021. https://doi.org/10.1001/jamanetworkopen.2021.24535. **KQ1E7a, KQ2E7a, KQ3E7a**

346. Van Dyck, Walter, Gassull, Daniel, et al. Unlocking the value of personalised healthcare in Europe—breast cancer stratification. *Health Policy and Technology*. 1(2): 63-68. 2012. https://doi.org/10.1016/j.hlpt.2012.04.006. **KQ1E7h, KQ2E7h, KQ3E7h**

347. van Schoor, G, Moss, SM, et al. Increasingly strong reduction in breast cancer mortality due to screening. *Br J Cancer*. 104(6): 910-4. 2011. https://dx.doi.org/10.1038/bjc.2011.44. **KQ1E2b, KQ2E2b, KQ3E2b**

348. Van, Ravesteyn Nt, Heijnsdijk, Eam, et al. Prediction of higher mortality reduction for the UK Breast Screening Frequency Trial: a model-based approach on screening intervals. *Br J Cancer*. 105(7): 1082-1088. 2011. https://dx.doi.org/10.1038/bjc.2011.300. **KQ1E7h, KQ2E7h, KQ3E7h**

349. Venturini, E, Losio, C, 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*. 268(2): 347-55. 2013. https://dx.doi.org/10.1148/radiol.13122278. **KQ1E3, KQ2E3, KQ3E3**

350. Wai, ES, D'Yachkova, Y, et al. Comparison of 1- and 2-year screening intervals for women undergoing screening mammography. *Br J Cancer*. 92(5): 961-6. 2005.

https://doi.org/10.1038/sj.bjc.6602393. **KQ1E7c, KQ2E7c, KQ3E7c**

351. Waldherr, C, Cerny, P, et al. Value of one-view breast tomosynthesis versus two-view mammography in diagnostic workup of women with clinical signs and symptoms and in women recalled from screening. *AJR Am J Roentgenol*. 200(1): 226-31. 2013. https://dx.doi.org/10.2214/ajr.11.8202. **KQ1E1b, KQ2E1b, KQ3E1b**

352. Weedon-Fekjaer, H, Romundstad, PR, et al. Modern mammography screening and breast cancer mortality: population study. *BMJ*. 348: g3701. 2014. https://dx.doi.org/https://dx.doi.org/10.1136/*BMJ*.g3701. **KQ1E7h, KQ2E7h, KQ3E7h**

353. Weigert, J, Steenbergen, S. The connecticut experiment: the role of ultrasound in the screening of women with dense breasts. *Breast J*. 18(6): 517-22. 2012. https://dx.doi.org/10.1111/tbj.12003. **KQ1E7, KQ2E7, KQ3E7**

354. Weigert, J, Steenbergen, S. The connecticut experiments second year: ultrasound in the screening of women with dense breasts. *Breast Journal*. 21(2): 175-80. 2015. https://dx.doi.org/https://dx.doi.org/10.1111/tbj.12386. **KQ1E3, KQ2E3, KQ3E3**

355. Welch, HG, Prorok, PC, et al. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. *N Engl J Med*. 375(15): 1438-1447. 2016. https://doi.org/10.1056/nejmoa1600249. **KQ1E2b, KQ2E2b, KQ3E2b**

356. Whelehan, P, Evans, A, et al. The effect of mammography pain on repeat participation in breast cancer screening: a systematic review. *Breast*. 22(4): 389-94. 2013. https://dx.doi.org/10.1016/j.breast.2013.03.003. **KQ1E7a, KQ2E7a, KQ3E7a**

357. Whitman, GJ, Scoggins, ME. Screening Breast Ultrasound Following Tomosynthesis. *Acad Radiol*. 29(3): 348-349. 2022. https://doi.org/10.1016/j.acra.2021.12.003. **KQ1E7a, KQ2E7a, KQ3E7a**

358. Wilczek, B, Wilczek, HE, et al. 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*. 85(9): 1554-63. 2016. https://dx.doi.org/10.1016/j.ejrad.2016.06.004. **KQ1E1b, KQ2E1b, KQ3E1b**

359. 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**

360. Winter, AM, Kazmi, S, et al. Comparison of interval breast cancers with 2D digital mammography versus 3D digital breast tomosynthesis in a large community-based practice. *Breast Journal*. 26(10): 1953-1959. 2020. https://dx.doi.org/10.1111/tbj.14047. **KQ1E7c, KQ2E7c, KQ3E7c**

361. Wu, Y, Alagoz, O, et al. Pursuing optimal thresholds to recommend breast biopsy by quantifying the value of

tomosynthesis. *Proc SPIE Int Soc Opt Eng*. 9037: 90370u. 2014. https://dx.doi.org/10.1117/12.2042905. **KQ1E7d, KQ2E7d, KQ3E7d**

362. Wu, YY, Yen, MF, et al. Individually tailored screening of breast cancer with genes, tumour phenotypes, clinical attributes, and conventional risk factors. *Br J Cancer*. 108(11): 2241-9. 2013. https://dx.doi.org/10.1038/bjc.2013.202. **KQ1E7h, KQ2E7h, KQ3E7h**

363. Yaffe, MJ, Mainprize, JG. Risk of radiation-induced breast cancer from mammographic screening. Radiology. 258(1): 98-105. 2011. https://dx.doi.org/10.1148/radiol.10100655. **KQ1E7h, KQ2E7h, KQ3E7h**

364. Yankaskas, BC, Taplin, SH, et al. Association between mammography timing and measures of screening performance in the United States. *Radiology*. 234(2): 363-73. 2005. https://dx.doi.org/10.1148/radiol.2342040048. **KQ1E4, KQ2E4, KQ3E7j**

365. Yen, AM, Duffy, SW, et al. Long-term incidence of breast cancer by trial arm in one county of the Swedish Two-County Trial of mammographic screening. *Cancer*. 118(23): 5728-32. 2012. https://dx.doi.org/10.1002/cncr.27580. **KQ1E3, KQ2E3, KQ3E3**

366. Yen, AM, Tsau, HS, et al. Population-Based Breast Cancer Screening With Risk-Based and Universal Mammography Screening Compared With Clinical Breast Examination: A Propensity Score Analysis of 1429890 Taiwanese Women. *JAMA Oncol*. 2(7): 915-21. 2016. https://dx.doi.org/10.1001/*JAMA*oncol.2016.0447. **KQ1E7c, KQ2E7c, KQ3E7c**

367. Yen, MF, Tabár, L, et al. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *Eur J Cancer*. 39(12): 1746-54. 2003. https://dx.doi.org/10.1016/s0959-8049(03)00260-0. **KQ1E7h, KQ2E7h, KQ3E7h**

368. Yi, A, Jang, MJ, et al. Addition of Screening Breast US to Digital Mammography and Digital Breast Tomosynthesis for Breast Cancer Screening in Women at Average Risk. *Radiology*. 298(3): 568-575. 2021. https://dx.doi.org/10.1148/radiol.2021203134. **KQ1E7d, KQ2E7d, KQ3E7d**

369. Youk, JH, Kim, EK, et al. Performance of hand-held whole-breast ultrasound based on BI-RADS in women with mammographically negative dense breast. *Eur Radiol*. 21(4): 667-75. 2011. https://dx.doi.org/10.1007/s00330-010-1955-8. **KQ1E1, KQ2E1, KQ3E1**

370. Zackrisson, S, Andersson, I, et al. Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ*. 332(7543): 689-92. 2006. https://dx.doi.org/10.1136/*BMJ*.38764.572569.7C. **KQ1E3, KQ2E3, KQ3E3**

371. Zahl, PH, Mæhlen, J. Overdiagnosis of breast cancer after 14 years of mammography screening. *Tidsskr Nor Laegeforen*. 132(4): 414-7. 2012. https://dx.doi.org/10.4045/tidsskr.11.0195. **KQ1E2b, KQ2E2b, KQ3E2b**

372. Zahl, PH. Overdiagnosis of breast cancer in Denmark. *Br J Cancer*. 90(8):

1686; author reply 1687. 2004. https://dx.doi.org/10.1038/sj.bjc.6601738. **KQ1E7a, KQ2E7a, KQ3E7a**

373. Zervoudis, S, Iatrakis, G, et al. Tomosynthesis improves breast cancer detection: our experience. Eur J Gynaecol Oncol. 35(6): 666-9. 2014. **KQ1E7d, KQ2E7d, KQ3E7d**

374. Zuckerman, SP, Conant, EF, et al. Implementation of Synthesized Two-dimensional Mammography in a Population-based Digital Breast Tomosynthesis Screening Program. *Radiology*. 281(3): 730-736. 2016. https://doi.org/10.1148/radiol.2016160366. **KQ1E7c, KQ2E7c, KQ3E7c**

375. Zuckerman, SP, Sprague, BL, et al. Multicenter Evaluation of Breast Cancer Screening with Digital Breast Tomosynthesis in Combination with Synthetic versus Digital Mammography. *Radiology*. 297(3): 545-553. 2020. https://dx.doi.org/https://dx.doi.org/10.1148/radiol.2020200240. **KQ1E7c, KQ2E7c, KQ3E7c**

Additional Details on Included NSRIs

Population-based Research to Optimize the Screening Process (PROSPR)

The fair-quality PROSPR NRSI[130] used data from three different academic US research centers and connected health care delivery systems that are members of the NCI-funded PROSPR consortium to compare the performance of DM screening with DBT and DM screening combined. The three study sites were the University of Pennsylvania (integrated health care delivery system), University of Vermont (statewide cancer surveillance system), and Geisel School of Medicine at Dartmouth College (in association with Brigham and Women's Hospital's primary care network). The study data included the findings from all bilateral screening mammography examinations provided at the study sites from approximately 2011 to 2014 (varied by study site) among women ages 40 to 74 years where screening was the coded indication. Those with a history of breast cancer or imaging conducted in the three months prior to screening were excluded. The study data were further limited to include only those exams provided by radiologists that had interpreted at least 50 DBT and 50 DM exams. Data from the same individuals could contribute multiple observations (e.g., two screening visits plus followup from the same individual could be eligible for inclusion). Screening visits were coded as first ever mammograms if there were no prior imaging exams available in PROSPR and no self-reported prior imaging.

The database included 103,401 individuals (55,998 DBT exams; 142,883 DM exams) and over a quarter of women (28.3%) contributed 3 or more exams to the analysis. The study reported interval cancers occurring in the year following DBT/DM or DM screening for the subset of screening visits that had at least one year of followup observation (e.g., remained in the consortium database) (25,268 DBT/DM exams; 113,061 DM exams) (70% of examinations eligible for the study). Effects were estimated with logistic regression adjusted for research center, age (categorized 40-49, 50-59, 60-74), breast density (4 BI-RADS density categories), and whether it was a first exam (prevalence screen). Additional analyses were conducted for some outcomes, including the use of generalized estimating equations (GEE) to account for effects of correlations within individuals contributing multiple screens, but the authors did not report these effects and stated that the results did not change based on the statistical model used.

Breast Cancer Surveillance Consortium (BCSC)

Ho et al. (2022)

The fair-quality Ho et al. BCSC NRSI[138] was conducted using data from 126 radiology facilities participating in the US Breast Cancer Surveillance Consortium (BCSC) to computed the cumulative probability of a false-positive result after 10 years of screening with DBT or DM during the years 2005 through 2018. The included observations were based on the screening visits of 903,495 women (444,704 DBT examinations, 2,524,351 DM examinations). The mean (SD) number of examinations per woman was 3.3 (2.5). First mammography examinations were excluded from the analysis, and had they been included the estimated cumulative false-positive rates would have been higher. Logistic regression was used to estimate the probability of false-positive recall, short-interval recall, and biopsy after a single round of screening as a function of age, breast density, screening interval, modality, and interactions among these variables. In addition, the interaction of screening round and modality was included in the model along with the total number of screening rounds for the individual. These round specific probabilities were

used to generate the cumulative probability of having at least one false-positive across 10 years of screening using discrete-time survival modeling to account for censoring. Estimates for annual compared with biennial screening and DBT compared with DM screening were presented and further stratified by age and breast density. Over the study period, the proportion of examinations conducted with DBT increased but the age and breast density distributions of those screened with DBT and DM were similar as were the proportion screened annually (nearly three quarters) versus biennially.

Kerlikowske et al. (2022)

The fair-quality BCSC NRSI by Kerlikowske et al.[145] was conducted using data from screening visits at 44 BCSC facilities to compare outcomes of screening with DBT or DM during the years 2011 through 2018. Additional followup for cancer diagnoses obtained from state and regional cancer registries continued through 2019. The cohort included 504,427 women ages 44 to 79 years that had at least one DBT or DM screening visit (based on radiologist indication in medical record). Individuals with a history of breast cancer or mastectomy were excluded. Screening visits were excluded from the analysis if they were: a first screening mammography (i.e., prevalence screening); unilateral screening; following a mammography within the previous 9 months; included an ultrasound within 3 months; or included an MRI within 12 months. Demographic and health history information was obtained from questionnaires or the electronic medical record (EMR) and breast cancer diagnoses were obtained from regional and state tumor registries or the SEER database (estimated 94.3% complete outcome reporting). The unit of analysis for the study was the screening examination, and among the individuals in the cohort 308,141 had only DM exams (mean 2.2 per person), 56,939 had only DBT (mean 1.6 per person), and 139,347 had both DBT and DM (mean 2.0 DBT and 2.3 DM per person). In total, the analysis included 1,377,902 screening examinations. To adjust for potential confounding related to the type of screening obtained, the statistical analysis used inverse-probability weighting (IPW) based on propensity scores. The propensity scores were obtained from a logistic model adjusted for age at examination, BCSC registry, facility type (academic versus not), calendar year, race and ethnicity, breast density, first-degree family history, time since last mammogram, and the most severe prior benign biopsy result. A generalized estimating equation was used to analyze the associations between the screening modality (with IPW) and cancer outcomes while accounting for clustering by BCSC facility and time since last mammogram. The study did not account for the patterns of screening over time when making comparisons between the two modalities, The potential remains in this observational study for unmeasured confounding and selection by indication related to access to health care, comorbidities, and ongoing screening patterns relative to studies that allocate participants to modality and observe outcomes across multiple screening rounds.

Lee et al (2019)

An NRSI by Lee et al.[150] reported results of an analysis using two BCSC registries (Vermont Breast Cancer Surveillance System and San Francisco Mammography Registry) to compare screening outcomes for individuals receiving ultrasonography on the same day as a screening mammogram (DM/US) compared with those that received only a mammogram (DM). Breast US examinations with a screening indication occurring from January 1, 2000 to December 31, 2013 were identified in the registries. Observations with the following characteristics were excluded: personal history of breast cancer, mastectomy, or known malignant neoplasm; unilateral breast

examination; and, self-reported breast symptoms (except pain). In one of the registries data from radiology reports were abstracted from 14% of included observations to confirm the screening indication for ultrasonography - since 96% were confirmed the remainder were included and assumed to be for screening indications. In the second registry, all reports were abstracted and the 78% confirmed to have screening indication for the US examination were included. Follow-up was for 12 months after the screening examination visit, or until the next examination. Outcomes were obtained from the cancer registries and their linkages to other data sources (E.g., SEER, state tumor registries, clinical data). Propensity scores were estimated using logistic regression. The probabilities for the screening type were calculated using the following variables: BCSC registry, age, year of examination, race/ethnicity, menopausal status, first-degree family history, time since last mammogram, breast density, and prior benign biopsy result. These probabilities were used to conduct 1:5 matching for mammography plus US (n = 6,081 examinations, 3,386 women) and mammography only (n=30,062 examinations, 113,293 women) from the same registries and without replacement. Comparisons between groups were tested using relative risk estimated from log binomial regression with adjustment for residual confounding using the propensity matching variables and a random effect for the matched sets to account for possible correlations. Before matching, those receiving ultrasonography were more likely to be younger, White non-Hispanic, have dense breasts, have a first-degree family history, and to have a BCSC risk score 2.50 percent or greater. Over a quarter (26%) of those having ultrasonography screening did not have dense breasts.

Miglioretti et al. (2015)

A fair-quality BCSC NRSI by Miglioretti et al.[152] used data on cancers detected in the BCSC registries from 1996 to 2012. The study compared the interval of screening relative to the characteristics of screen-detected and interval cancers. Individuals were included in the analysis if their cancer was preceded by at least two screening mammograms either 11 to 14 months apart (annual interval) or 23 to 26 months apart (biennial interval). Cancers were designated interval cancers if they followed a negative mammogram screening result and screen-detected interval cancers if they followed a positive screening result. The time of followup to assess interval cancers was 12 months for those with an annual interval before their cancer diagnosis and was 24 months for those with a biennial interval before their cancer diagnosis. Examinations were included as screening mammograms unless they were unilateral or if there was a mammography or ultrasonography visit within the prior 9 months. Cancers were defined as having 'less favorable prognostic characteristics' if they were at AJCC stage IIB or higher, size 15mm or greater, or had positive nodes. Comparisons of prognostic characteristics by screening interval were presented, with adjustments for race/ethnicity, first-degree family history, and BCSC registry using logistic binomial regression. Notably, the analysis of cancer prognostic characteristics grouped together screen-detected and interval cancers (23% of total cancer cases). The characteristics of women with cancers preceded by an annual screening interval (n = 12,070) and those preceded by a biennial interval (n = 3,370) differed on some reported factors; those with an annual interval preceding a cancer diagnosis were less likely to be ages 40 to 49 (14% versus 18%) or 70-85 (29% vs 27%), and more likely to have a first-degree family history of breast cancer (23% versus 18%). The groups did not differ in race/ethnicity composition, and over three-quarters of the study population was White, non-Hispanic (78%), with the remaining participants reported as Black (5%), Asian (5%), Hispanic (5%), American Indian or Alaska Native (<1%), and 7% reported as "other" or unknown. This study did not report overall effects

of the screening interval on the types of cancers diagnosed, but provides results stratified by age and menopausal status that are reported as KQ2a results below.

Malmö Breast Tomosynthesis Screening Trial (MBTST)
The fair-quality MBTST[79] is an NRSI using prospectively collected cohort data in Sweden comparing women screened with DBT/DM through their participation in a screening performance cohort study (n = 13,369) with a concurrently screening period (2010 to 2015) for an age-matched control cohort screened with DM (n = 26,738). Those screened with DBT/DM had two independent readings of their DBT image (read first), DM image from the visit, and any previous DM images. The age-matched controls were selected from the screening program registry records for women that did not participate in the DBT/DM study but were screened in the same setting which relied on the same radiologists as the DBT/DM study. A random selection of a single DM reading instance was drawn from the registry and used for 2:1 age-matching (+/- 1 year) and date of screen matching (+/- 1 year) with the DBT/DM trial participants. Matches were made for all but 1,479 DBT/DM screened individuals. Apart from the age- and date-matching, no descriptions of or additional adjustments for potential differences between the two groups were provided.

Blue Cross Blue Shield
A fair-quality NRSI by Richman et al.[160] conducted in the US used national medical claims registry data from the Blue Cross Blue Shield Axis which contains deidentified commercial insurance claims. The study comparing DBT/DM screening with DM screening among women ages 40 to 64 years included individuals receiving at least 1 screening mammogram between January 1, 2016 and December 31, 2017 who had been continuously enrolled for at least 2 years preceding the screen and for at least 1 year following the screen. Exclusions were any breast-cancer diagnosis in the 2 years preceding the screen and any insurance claims indicative of a genetic cancer syndrome or prophylactic mastectomy. To distinguish screening mammograms from diagnostic mammograms in the claims data, a previously validated algorithm was used. The analytic sample included 7,602,869 screening mammograms conducted among 4,580,698 women.

Know Your Risk: Assessment at Screening (KYRAS) study
The KYRAS study[151] calculated the cumulative risk of false positive screens over a median of 8.9 years. Eligible women were those screened at the Columbia University Medical Center (New York) during 2014 and 2015), for these women the study collected information on previous mammograms going back to 1989 based on their health record (N=2,019; median age 59 years). Women with a previous diagnosis of breast cancer or who did not speak English or Spanish were excluded from the study. Frequency of screening was determined in the EHR by calculating the median number of days between mammograms. If the median screening interval was between 274 days (9 months) and 548 days (18 months), then it was coded as yearly screening; a median interval between 548 days (18 months) and 913 days (30 months) was coded as biennial screening. Overall, women underwent a median of 7 mammograms during the study period.

The screening interval was categorized as annual if the previous examination was within 9-18 months and biennial if it was within 19 to 30 months. Intervals longer than 30 months were coded as triennial or longer (accounting for 11 percent of examinations), but were not reported

on in the study results. First mammography examinations were excluded from the analysis, and had they been included the estimated cumulative false-positive rates would have been higher. Logistic regression was used to estimate the probability of false-positive recall, short-interval recall, and biopsy after a single round of screening as a function of age, breast density, screening interval, modality, and interactions among these variables. In addition, the interaction of screening round and modality was included in the model along with the total number of screening rounds for the individual. These round specific probabilities were used to generate the cumulative probability of having at least one false-positive across 10 years of screening using discrete-time survival modeling to account for censoring. Estimates for annual compared with biennial screening and DBT compared with DM screening were presented and further stratified by age and breast density.

Appendix E Table 1. Additional Details on Included Trials and NRSI

Intervention Category	Study Design	Author, Year Study/Trial Name	Country	Comparison (IG vs. CG)	Number of readers	Reader experience/training	Type of reading	Consensus method	Case and mortality ascertainment method
Age to Stop	NRSI	Garcia-Albeniz, 2020	US	Continuing annual DM beyond 70 years of age vs. Stopping annual DM at 70 years of age	NR	NR	NR	NR	Medicare data and National Death Index
Screening Frequency	RCT	Blamey, 2002 UKCCCR	UK	Annual DM vs. Triennial DM	NR	NR	NR	NR	Local hospital data and regional cancer registries
	NRSI	Ho, 2022 BCSC	US	Annual DBT/SM or DM vs. Biennial DBT/SM or DM	NR	NR	NR	NR	BCSC registry data, pathology databases, stage/regional tumor registries, state death records
		McGuinness, 2018 KYRAS	US	Annual DM vs. Biennial DM	NR	NR	NR	NR	Electronic health record
		Miglioretti, 2015 BCSC	US	Annual DM vs. Biennial DM	NR	NR	NR	NR	BCSC registry data, pathology databases, stage/regional tumor registries
		Parvinen, 2011	Finland	Annual DM vs. Biennial DM	8	NR	Dual	NR	National cancer and mortality databases
Digital Breast Tomosynthesis	RCT	Armaroli, 2022 Proteus Donna Fair	Italy	DBT/DM (round 1), DM (round 2) vs. DM	38	Radiologists received basic training in integrated DM and DBT and pass a trial evaluation with the interpretation of 40 DBT cases. Readers met regional quality assurance of 5000+ mammograms per year	Dual independent	If either radiologist gave a score of 3 (probably benign) or higher the case was considered positive and recalled for	Population screening database, histology reports, hospital and population cancer registry

Appendix E Table 1. Additional Details on Included Trials and NRSI

Intervention Category	Study Design	Author, Year Study/Trial Name	Country	Comparison (IG vs. CG)	Number of readers	Reader experience/training	Type of reading	Consensus method	Case and mortality ascertainment method
						with periodic audits of performance.		investigation without consensus or arbitration	
		Heindel, 2022 TOSYMA Good	Germany	DBT/SM vs. DM	NR	Participated in all regular teaching courses for mammography screening program and having passed the yearly test of 50 screening case studies, a volume of at least 5,000 screening mammograms the year before participating in the study, readers regularly assessed with an emphasis on a comparable number of sets for DBT/SM and DM images	Dual independent	In case of any suspicious abnormality, reading results were clarified with an arbitrator to decide whether women had to be recalled for further diagnostic tests	Cancer registries
		Pattacini, 2022 RETomo Good	Italy	DBT/DM (round 1), DM (round 2) vs. DM	10	4 to 20 years. Regional quality assurance criterion of at least 5000 mammograms per year and period audits of individual performance indications and interval cancer imagining review.	Dual independent	Arbitration by third reviewer	Screening database and cancer registry
		Hofvind, 2021 To-Be Good	Norway	DBT/sDM vs. DM (round 1), DBT/SM (round 2)	8	Their experience in screen reading (screen film and digital mammography) before start-up of the trial varied from zero to approximately 110000 examinations	Dual independent	Consensus was done by pairs of radiologists, and a third radiologist was consulted if the pair could not agree	National cancer registry
	NRSI	Ho, 2022 BCSC-2022a	US	DBT vs. DM	699	NR	NR	NR	BCSC registry data, pathology databases,

Appendix E Table 1. Additional Details on Included Trials and NRSI

Intervention Category	Study Design	Author, Year Study/Trial Name	Country	Comparison (IG vs. CG)	Number of readers	Reader experience/training	Type of reading	Consensus method	Case and mortality ascertainment method
									stage/regional tumor registries, state death records
		Kerlikowske, 2022 BCSC-2022b	US	DBT vs. DM	NR	NR	NR	NR	BCSC registry data, pathology databases, regional/state tumor registries, SEER programs
		Johnson, 2021 MBTST	Sweden	DBT/DM vs. DM	7	2 to 41 years. Previous experience with DBT from clinical work or studies of previous DBTs	Dual independent	Examinations that scored as suspicious based on any modality were evaluated at a consensus meeting	National cancer registry
		Richman, 2021	US	DBT/DM vs. DM	NR	NR	NR	NR	Commercial insurance claims
		Hovda, 2020 OVVV	Sweden	DBT/SM (round 1), DM (round 2) vs. DM	NR	0 to 14 (using DM), 0 to 3 (using DBT)	Dual independent	Readings given a score of 1-5. If at least one radiologist gave score of 2 (probably benign) or greater a consensus meeting was help to determine recall. consensus with random pairs of radiologists	National cancer registry

Appendix E Table 1. Additional Details on Included Trials and NRSI

Intervention Category	Study Design	Author, Year Study/Trial Name	Country	Comparison (IG vs. CG)	Number of readers	Reader experience/training	Type of reading	Consensus method	Case and mortality ascertainment method
		Conant, 2016 PROSPR	US	DBT/DM vs. DM	NR	NR	NR	NR	Electronic health records, pathology databases, institutional and state cancer registries
Supplemental MRI	RCT	Veenhuizen, 2021 DENSE	Netherlands	DM plus MRI vs. DM	NR	5 to 23 years	Single reader	For those with a BI-RADS 3 score double reading was performed, consensus on level 3 lead to repeat MRI within 6 months.	National cancer registry
	NRSI	Ganguli, 2022	US	MRI vs. DM	NR	NR	NR	NR	Medical claims database
Supplemental Ultrasound	RCT	Ohuchi, 2016 J-START	Japan	DM plus US vs. DM	NR	<1 for US training. Ultrasonography is performed by qualified physicians, laboratory technologists, clinical radiological technologists or nurses having experienced with breast ultrasonography and completed the breast ultrasonography training program. The technologists and the physicians involved in this trial are asked to finish 2-day, 16-h education program for the standardization of US screening for breast cancer.	Dual independent	Results of ultrasound were reassessed by physicians at the study sites, including radiologists and breast surgeons.	Study database, postal survey, vital registry

Appendix E Table 1. Additional Details on Included Trials and NRSI

Intervention Category	Study Design	Author, Year Study/Trial Name	Country	Comparison (IG vs. CG)	Number of readers	Reader experience/training	Type of reading	Consensus method	Case and mortality ascertainment method
	NRSI	Lee, 2019 BCSC	US	DM plus US vs. DM	NR	NR	NR	NR	BCSC registry data, pathology databases, stage/regional tumor registries, state death records

Abbreviations: BCSC=Breast Cancer Surveillance Consortium; BI-RADS=Breast Imaging Reporting and Data System; **;** DBT=digital breast tomosynthesis; DENSE=Dense Tissue and Early Breast Neoplasm Screening; DM=digital mammography; J-START= Japan Strategic Anti-cancer Randomized Trial; MBTST=Malmo Breast Tomosynthesis Screening Trial; NR=not reported; NRSI=nonrandomized study of intervention; RCT=randomized controlled trial; RETomo=Reggio Emilia Tomosynthesis Trial; sDM=synthetic mammography; PROSPR=Population-based Research Optimizing Screening through Personalized Regimens; To-Be=Tomosynthesis Trial in Bergen; TOSYMA=TOmosynthesis plus SYnthesized MAmmography study; OVVV=Oslo-Vestfold-Vestre Viken; UKCCR=United Kingdom Coordinating Committee on Cancer Research trial; US=ultrasound

Appendix F Table 1. Breast Cancer Staging System

Description	
Primary Tumor (T)	T1=tumor size ≤20 mm T2=>20 mm but ≤50 mm T3=>50 mm T4=tumor of any size with direct extension to the chest wall and/or skin
Regional lymph nodes (N)	N0=no regional lymph node metastases N1mi=micrometastases N1=metastases to moveable ipsilateral axillary lymph nodes N2=metastases in ipsilateral axillary lymph nodes that are clinically fixed or matted N3=metastases that are more extensive
Distant metastasis (M)	M0=no evidence of distant metastases M1=distant detectable metastases as determined by clinical and radiographic means
Stage	**Anatomic Stage**
0	Tis, N0, M0
I	IA=T1, N0, M0 IB=T0, N1mi, M0 *or* T1, N1mi, M0
II	IIA=T0, N1, M0 *or* T1, N1, M0 *or* T2, N0, M0 IIB=T2, N1, M0 *or* T3, N0, M0
III	IIIA=T0, N2, M0 *or* T1, N2, M0 *or* T2, N2, M0 *or* T3, N1, M0 *or* T3, N2, M0 IIIB=T4, N0, M0 *or* T4, N1, M0 *or* T4, N2, M0 IIIC=any T, N3, M0
IV	Any T, any N, M1

Adapted from 2020 National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology.[106] The prognostic staging included in the most recent guidelines are not reflected here and are available on the NCCN website.

Appendix F Table 2. Imaging Technologies Approved for Primary and Supplemental Breast Cancer Screening

Mammography imaging	Description
Film mammography (FM)	• No longer widely used in United States • X-rays pass through compressed breast onto film to produce a grayscale image • Routine screening based on two views (craniocaudal and mediolateral oblique)
Digital mammography (DM)	• X-rays pass through compressed breast and converted to digital grayscale image • Routine screening based on two views (craniocaudal and mediolateral oblique)
Digital breast tomosynthesis (DBT)	• Modification of DM that obtains multiple images from many different angles from a brief x-ray scan • Sometimes referred to as 3D mammography • Synthetic mammogram refers to a two-view DBT image approximating the image obtained from two-view DM • May be conducted with less breast compression
Ultrasound (US)	• Sound wave images of the breast using a non-invasive, hand-held device (HHUS) • Whole breast ultrasound (ABUS) approved by FDA for supplemental screening among women with dense breasts • Not considered a primary breast cancer screening modality
Magnetic Resonance Imaging (MRI)	• Magnetic fields used to create image of the breast • Intravenous contrast agent given for the procedure • Not considered a primary breast cancer screening modality

Adapted from 2013 ACR BIRADS Atlas 5th Edition[70]

Appendix F Table 3. Cumulative False-Positive Followup Over Multiple Rounds of Screening in One Nonrandomized Study Comparing Digital Breast Tomosynthesis and Digital Mammography

Author, Year Study/ Trial Name	Comparison (IG vs. CG)	Population	Followup	Outcome Definition	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Rate difference per 1000 (95% CI)*
Ho, 2022 BCSC	Annual DBT/DM vs. Annual DM	Women ages 40 to 79 years	10 years of screening	Cumulative probability of at least one false-positive recall†	NR (496)	NR (563)	-67 (95% CI, -74 to -61)
				Cumulative probability of at least one false-positive short-interval followup recommendation‡	NR (166)	NR (178)	-11 (95% CI, -17 to -6)
				Cumulative probability of at least one false-positive biopsy recommendation§	NR (112)	NR (117)	-5 (95% CI, -10 to -1)
	Biennial DBT/DM vs. Biennial DM	Women ages 40 to 79 years	10 years of screening	Cumulative probability of at least one false-positive recall†	NR (357)	NR (381)	-24 (95% CI, -34 to -15)
				Cumulative probability of at least one false-positive short-interval followup recommendation‡	NR (103)	NR (105)	-1 (95% CI, -7 to 5)
				Cumulative probability of at least one false-positive biopsy recommendation§	NR (66)	NR (67)	-1 (95% CI, -5 to 4)

*Scale changed from study reported proportion for comparability across tables
†Recall was defined as a BI-RADS initial assessment of 0 (needs additional imaging evaluation), 3 (probably benign finding), 4 (suspicious abnormality), or 5 (highly suggestive of cancer). Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination.
‡Short-interval follow-up recommendation was defined as a BI-RADS final assessment of 3 after diagnostic imaging work up within 90 days of a recalled screening examination. Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination. Study imputed false-positive short-interval followup recommendations for examinations with a false positive recall but unresolved final assessment (n = 14171 [0.5%]) based on age, breast density, screening modality, and screening interval, imputing a single value because less than 1% of data was missing.
§Women were recommended for biopsy with a BI-RADS initial evaluation of 4 (suspicious abnormality) or 5 (highly suggestive of cancer). Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination. Study imputed false-positive biopsy recommendations for examinations with a false positive recall but unresolved final assessment (n=14171 [0.5%]) based on age, breast density, screening modality, and screening interval, imputing a single value because less than 1% of data was missing.

Abbreviations: CG=control group; CI=confidence interval; DBT=digital breast tomosynthesis; DM=Digital mammography; IG=intervention group

Appendix F Table 4. Cumulative False-Positive Followup Over Multiple Rounds of Screening in Nonrandomized Studies Comparing Breast Cancer Screening Frequencies

Author, Year Study/Trial Name	Comparison (IG vs. CG)	Population	Followup	Outcome Definition	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Effect (95% CI)*
Ho, 2022 BCSC	Annual DM vs. biennial DM	Women ages 40 to 79 years	10 years of screening	Cumulative probability of at least one false-positive recall†	NR (563)	NR (381)	Rate difference per 1000: -182 (95% CI, -186 to -177)
				Cumulative probability of at least one false-positive short-interval followup recommendation‡	NR (178)	NR (105)	Rate difference per 1000: -73 (95% CI, -77 to -69)
				Cumulative probability of at least one false-positive biopsy recommendation§	NR (117)	NR (67)	Rate difference per 1000: -50 (95% CI, -54 to -47)
	Annual DBT vs. biennial DBT	Women ages 40 to 79 years	10 years of screening	Cumulative probability of at least one false-positive recall†	NR (496)	NR (357)	Rate difference per 1000: -139 (95% CI, -149 to -128)
				Cumulative probability of at least one false-positive short-interval followup recommendation‡	NR (166)	NR (103)	Rate difference per 1000: -63 (95% CI, -70 to -56)
				Cumulative probability of at least one false-positive biopsy recommendation§	NR (112)	NR (66)	Rate difference per 1000: -46 (95% CI, -52 to -39)
McGuinness, 2018 KYRAS	Annual DM vs. biennial DM‖	Women 18 and older	Median of 8.9 years of screening	Cumulative rate of false positive. Defined as followup breast imaging or biopsy not resulting in a breast cancer diagnosis.	836/1399 (597.6)	139/335 (414.9)	OR: 2.18 (95% CI, 1.70 to 2.80) ¶

* Scale change from study reported proportion difference for consistency across tables

† Recall was defined as a BI-RADS initial assessment of 0 (needs additional imaging evaluation), 3 (probably benign finding), 4 (suspicious abnormality), or 5 (highly suggestive of cancer). Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination.

‡ Short-interval follow-up recommendation was defined as a BI-RADS final assessment of 3 after diagnostic imaging work up within 90 days of a recalled screening examination. Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination. Study imputed false-positive short-interval followup recommendations for examinations with a false positive recall but unresolved final assessment (n = 14171 [0.5%]) based on age, breast density, screening modality, and screening interval, imputing a single value because less than 1% of data was missing.

§ Women were recommended for biopsy with a BI-RADS initial evaluation of 4 (suspicious abnormality) or 5 (highly suggestive of cancer). Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination. Study imputed false-positive biopsy recommendations for examinations with a false positive recall but unresolved final assessment (n =14171 [0.5%]) based on age, breast density, screening modality, and screening interval, imputing a single value because less than 1% of data was missing.

‖ Annual screening classifications had a median screening interval between 274 days (9 months) and 548 days (18 months). Biennial screenings were had a median screening interval between 548 days (18 months) and 913 days (30 months). There was a median of 7 mammograms for entire sample (range, 1-27).

¶ Adjusted for: total years of follow up, age, race/ethnicity, BMI, breast density, and breast cancer risk status.

Abbreviations: BI-RADS=breast imaging-reporting data system; CG=control group; CI=confidence interval DM=Digital mammography; IG=intervention group; NRSI=nonrandomized study of intervention

Appendix F Table 5. False-Positive Followup Over Multiple Rounds of Screening in One Nonrandomized Study Comparing Digital Breast Tomosynthesis and Digital Mammography, by Population Subgroup

Author, Year Study/ Trial Name	Comparison (IG vs. CG)	Followup	Outcome Definition	Subgroup	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Rate difference per 1000 (95% CI)*
Ho, 2022 BCSC	Annual DBT/DM vs. Annual DM	10 years of screening	Cumulative probability of at least one false-positive recall†	40 to 49 years	NR (608)	NR (680)	-72 (95% CI, -87 to -58)
				50 to 59 years	NR (511)	NR (576)	-65 (95% CI, -74 to -56)
				60 to 69 years	NR (440)	NR (504)	-63 (95% CI, -73 to -54)
				70 to 79 years	NR (398)	NR (470)	-72 (95% CI, -86 to -57)
			Cumulative probability of at least one false-positive short-interval followup recommendation‡	40 to 49 years	NR (207)	NR (209)	3 (95% CI, -14 to 9)
				50 to 59 years	NR (170)	NR (185)	-15 (95% CI, -22 to -9)
				60 to 69 years	NR (147)	NR (162)	-15 (95% CI, -22 to -9)
				70 to 79 years	NR (133)	NR (142)	-9 (95% CI, -20 to 2)
			Cumulative probability of at least one false-positive biopsy recommendation§	40 to 49 years	NR (132)	NR (134)	-2 (95% CI, -12 to 8)
				50 to 59 years	NR (117)	NR (124)	-8 (95% CI, -14 to -2)
				60 to 69 years	NR (102)	NR (110)	-8 (95% CI, -13 to -2)
				70 to 79 years	NR (91)	NR (93)	-2 (95% CI, -11 to 8)
	Biennial DBT/DM vs. Biennial DM	10 years of screening	Cumulative probability of at least one false-positive recall†	40 to 49 years	NR (461)	NR (487)	-25 (95% CI, -47 to -3)
				50 to 59 years	NR (348)	NR (376)	-28 (95% CI, -39 to -16)
				60 to 69 years	NR (293)	NR (317)	-24 (95% CI, -36 to -12)
				70 to 79 years	NR (286)	NR (297)	-11 (95% CI, -32 to 11)
			Cumulative probability of at least one false-positive short-interval followup recommendation‡	40 to 49 years	NR (131)	NR (132)	-1 (95% CI, -14 to 13)
				50 to 59 years	NR (100)	NR (105)	-4 (95% CI, -11 to 3)
				60 to 69 years	NR (87)	NR (88)	-1 (95% CI, -8 to 7)
				70 to 79 years	NR (84)	NR (78)	6 (95% CI, -7 to 20)
			Cumulative probability of at least one false-positive biopsy recommendation§	40 to 49 years	NR (84)	NR (82)	2 (95% CI, -9 to 12)
				50 to 59 years	NR (67)	NR (68)	-1 (95% CI, -7 to 6)
				60 to 69 years	NR (55)	NR (58)	-3 (95% CI, -9 to 3)
				70 to 79 years	NR (51)	NR (51)	0 (95% CI, -12 to 12)

*Scale changed from study reported proportion for comparability across tables

† Recall was defined as a BI-RADS initial assessment of 0 (needs additional imaging evaluation), 3 (probably benign finding), 4 (suspicious abnormality), or 5 (highly suggestive of cancer). Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination.

‡ Short-interval follow-up recommendation was defined as a BI-RADS final assessment of 3 after diagnostic imaging work up within 90 days of a recalled screening examination. Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination. Study imputed false-positive short-interval followup recommendations for examinations with a false positive recall but unresolved final assessment (n = 14171 [0.5%]) based on age, breast density, screening modality, and screening interval, imputing a single value because less than 1% of data was missing.

§ Women were recommended for biopsy with a BI-RADS initial evaluation of 4 (suspicious abnormality) or 5 (highly suggestive of cancer). Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination. Study imputed false-positive biopsy recommendations for examinations with a false positive recall but unresolved final assessment (n=14171 [0.5%]) based on age, breast density, screening modality, and screening interval, imputing a single value because less than 1% of data was missing.

Abbreviations: CG=control group; CI=confidence interval; DBT=digital breast tomosynthesis; DM=Digital mammography; IG=intervention group

Appendix F Table 6. Cumulative False-Positive Followup Over Multiple Rounds of Screening in Nonrandomized Studies Comparing Breast Cancer Screening Frequencies, by Population Subgroup

Author, Year Study/ Trial Name	Comparison (IG vs. CG)	Followup	Outcome Definition	Subgroup	IG n/N (rate per 1000 screened)	CG n/N (rate per 1000 screened)	Rate difference per 1000 (95% CI)*
Ho, 2022	Annual DM vs. biennial DM	10 years of screening	Cumulative probability of at least one false-positive recall†	40 to 49 years	NR (680)	NR (487)	-194 (95% CI, -203 to -184)
				50 to 59 years	NR (576)	NR (376)	-200 (95% CI, -206 to -195)
				60 to 69 years	NR (504)	NR (317)	-186 (95% CI, -192 to -181)
				70 to 79 years	NR (470)	NR (297)	-173 (95% CI, -182 to -164)
			Cumulative probability of at least one false-positive short-interval followup recommendation‡	40 to 49 years	NR (209)	NR (132)	-77 (95% CI, -84 to -70)
				50 to 59 years	NR (185)	NR (105)	-81 (95% CI, -86 to -76)
				60 to 69 years	NR (162)	NR (88)	-75 (95% CI, -79 to -70)
				70 to 79 years	NR (142)	NR (78)	-64 (95% CI, -70 to -58)
			Cumulative probability of at least one false-positive biopsy recommendation§	40 to 49 years	NR (134)	NR (82)	-52 (95% CI, -58 to -46)
				50 to 59 years	NR (124)	NR (68)	-56 (95% CI, -61 to -52)
				60 to 69 years	NR (110)	NR (58)	-52 (95% CI, -56 to -48)
				70 to 79 years	NR (93)	NR (51)	-41 (95% CI, -47 to -36)
	Annual DBT vs. biennial DBT	10 years of screening	Cumulative probability of at least one false-positive recall†	40 to 49 years	NR (608)	NR (461)	-146 (95% CI, -171 to -122)
				50 to 59 years	NR (511)	NR (348)	-163 (95% CI, -177 to -149)
				60 to 69 years	NR (440)	NR (293)	-147 (95% CI, -161 to -134)
				70 to 79 years	NR (398)	NR (286)	-112 (95% CI, -136 to -87)
			Cumulative probability of at least one false-positive short-interval followup recommendation‡	40 to 49 years	NR (207)	NR (131)	-75 (95% CI, -92 to -59)
				50 to 59 years	NR (170)	NR (100)	-70 (95% CI, -79 to -61)
				60 to 69 years	NR (147)	NR (87)	-60 (95% CI, -69 to -51)
				70 to 79 years	NR (133)	NR (84)	-49 (95% CI, -65 to -33)
			Cumulative probability of at least one false-positive biopsy recommendation after§	40 to 49 years	NR (132)	NR (84)	-48 (95% CI, -61 to -35)
				50 to 59 years	NR (117)	NR (67)	-50 (95% CI, -58 to -41)
				60 to 69 years	NR (102)	NR (55)	-47 (95% CI, -55 to -39)
				70 to 79 years	NR (91)	NR (51)	-40 (95% CI, -54 to -26)

*Scale changed from study reported proportion for comparability across tables

† Recall was defined as a BI-RADS initial assessment of 0 (needs additional imaging evaluation), 3 (probably benign finding), 4 (suspicious abnormality), or 5 (highly suggestive of cancer). Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination.

‡Short-interval follow-up recommendation was defined as a BI-RADS final assessment of 3 after diagnostic imaging work up within 90 days of a recalled screening examination. Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination. Study imputed false-positive short-interval followup recommendations for examinations with a false positive recall but unresolved final assessment (n = 14171 [0.5%]) based on age, breast density, screening modality, and screening interval, imputing a single value because less than 1% of data was missing.

§ Women were recommended for biopsy with a BI-RADS initial evaluation of 4 (suspicious abnormality) or 5 (highly suggestive of cancer). Results considered false-positive if no diagnosis of invasive carcinoma or ductal carcinoma in situ occurred within 1 year of the screening examination and before the next screening examination. Study imputed false-positive biopsy recommendations for examinations with a false positive recall but unresolved final assessment (n=14171 [0.5%]) based on age, breast density, screening modality, and screening interval, imputing a single value because less than 1% of data was missing.

Abbreviations: BI-RADS=breast imaging-reporting data system; CG=control group; CI=confidence interval DM=Digital mammography; IG=intervention group; NRSI=nonrandomized study of intervention

Appendix F Table 7. Summary of Evidence, by Population Subgroup

Key Question Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
KQ1a. All comparisons*	k = 0	NA	**NA**	NA	**Insufficient**	NA
KQ2a. Age to start or stop screening	k = 0	NA	**NA**	NA	**Insufficient**	NA
KQ2a. Screening interval: Biennial versus Annual Subgroups addressed: **Age:** (40 to 49, 50 to 59, 60 to 69, 70 to 85) **Hormonal status**	k = 1 (n = 15,440) NRSI Fair-quality	Age: No difference in risk of stage IIB+ or other cancers with less favorable prognosis diagnosed after a biennial compared with annual interval for any age group Hormonal status: More stage IIB+ and 'less favorable prognosis' cancers for premenopausal persons after biennial compared with annual interval No difference in stage IIB+ cancers for postmenopausal persons, trend toward more 'less favorable prognosis' cancers after biennial interval for postmenopausal persons using hormone therapy.	**NA (for consistency), Imprecise**	Risk of bias due to limited adjustment for confounding and potential unmeasured confounding and selection into study groups. No adjustment for multiple comparisons, increased risk that significant findings could be due to chance. Stratified analysis without tests for interaction.	**Insufficient**	Conducted using BCSC data linked with US SEER and other tumor registry sources; Ages 40 to 85; >77% population White non-Hispanic
KQ2a. Modality: DBT versus DM Subgroup addressed: **Age** and **density** subgroups	k = 2 (n = 55,119) RCT Good-quality	Age/Density: One RCT reported invasive cancer detection analyses stratified by breast density and age. Similar to the overall results, increased invasive cancer detection at round 1 was observed for women ages 50 to 69 and for those with nondense breasts. Women ages 45 to 49 and those with dense breasts did not have increased detection at round 1.	**Inconsistent Imprecise**	The studies did not power the study for subgroup comparisons and did not test for interactions. Information on the tumor characteristics was not stratified by density or age.	**Insufficient**	Two trials conducted organized screening programs in Europe (Norway, Italy) that use independent dual mammography reading and consensus.

Appendix F Table 7. Summary of Evidence, by Population Subgroup

Key Question Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
		Density: One RCT stratified results by breast density categories and found no difference in invasive cancer detection with DBT at either screening round.				
KQ2a. Supplemental screening with MRI	k = 0	NA	NA	NA	**Insufficient**	NA
KQ2a. Supplemental screening with ultrasound	k = 0	NA	NA	NA	**Insufficient**	NA
KQ3a. Age to start or stop screening	k = 0	NA	**NA**	NA	**Insufficient**	NA
KQ3a. Screening interval: Biennial versus Annual Subgroups addressed: Age X density	k = 1 (n = 903,495) NRSI Fair-quality	Age by density: People in higher dense breast category and younger age groups had highest cumulative FP recall and FP biopsy rates, regardless of screening modality or interval. Annual screening associated with higher FP recall and biopsy across all age and density categories	**NA (for consistency), Precise**	Risk of bias from potential selection and confounding bias, including time varying factors Study did not include prevalence screen, may underestimate FP starting from start of screening.	**Low** for greater 10-year probability of false-positive results with annual screening regardless of age or density category	BCSC NRSI includes population representative of US population undergoing screening; population distribution reflects US demographics; subgroup comparisons by race/ethnicity not reported

Appendix F Table 7. Summary of Evidence, by Population Subgroup

Key Question Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
KQ3a. Modality: DBT versus DM Subgroups addressed: Age and breast density	k=2 (n = 56,330) RCT Good-quality k = 4 (n = 6,028,727) NRSI Fair-quality	Age: One trial presented age group stratified analyses (45-49 v 50-69) and reported no difference in recall rates but a potential increased risk of biopsy or surgery at the first screening round for both age groups followed by lower risk for ages 45-49 at round 2. One RCT and two NRSI found no difference in the interval cancer rate by age group by screening modality. Density: Individuals with extremely dense breasts did not have different 10-year probabilities of false positive recall or biopsy when screened with DBT versus DM in a BCSC NRSI, but those ages 50-59 with extremely dense breasts were at increased risk of FP recall with a biennial screening interval.	**NA (for consistency), Imprecise**	Apart from one NRSI, studies did not report interaction tests and were not powered to test subgroup differences. NRSI had substantial risk of bias, limited adjustment for potential confounding and selection. Only one RCT used the same screening modality at rounds 1 & 2. Low event rates limit power for subgroup comparisons.	**Insufficient**	Two European trials conducted in organized screening programs using dual independent mammogram reading with consensus may be less applicable to US settings BSCS and private insured population NRSIs more applicable to US populations, but include mostly White, non-Hispanic participants and lack subgroup analysis by race/ethnicity.
KQ3a. Supplemental screening with MRI	k = 0	NA	NA	NA	**Insufficient**	NA
KQ3a. Supplemental screening with ultrasound Subgroup addressed: Density	k = 1 (n = 72,717) RCT Fair-quality	Density: The interval cancer rate was the same in the US arm regardless of breast density group (0.5 per 1,000) and the difference between study groups in the interval cancer rate was not statistically different for either density subgroup Additional recall due to US was higher among those with dense	**NA (for consistency), Imprecise**	Interval cancers rare in young women enrolled in RCT, limited power to detect differences Population-averaged GEE effect estimate for interval cancer reported in RCT	**Insufficient**	RCT conducted in Japan included people ages 40 to 49; 23% of study population prevalence screened; 58% reported to have dense breasts, distribution not reported; US and DM

Appendix F Table 7. Summary of Evidence, by Population Subgroup

Key Question Intervention	Studies (k), Study Design, Observations (n) Quality	Summary of Findings	Consistency and Precision	Other Limitations	Strength of Evidence	Applicability
		breasts (69.7 per 1,000) compared with those with nondense breasts (39.4 per 1,000)		including DCIS lesions was statistically significant between study groups for both density categories; no interaction test reported RCT has not reported second round results despite trial completion		results interpreted independently; performance could differ if considered together

*Includes Age to start/stop, screening interval, DBT vs. DM, Supplemental screening with MRI, Supplemental screening with ultrasound

Detailed Results for KQ3. What Are the Comparative Harms of Different Breast Mammography-Based Cancer Screening Strategies?

Screening Age to Start or Stop

Study and Population Characteristics

One fair-quality NRSI (n = 1,058,013)[134] analyzed data to emulate a trial of discontinuation of mammography screening at age 70 compared with continued annual screening beyond this age (described in detail in KQ1 in this report) (**Table 4**). The authors used statistical techniques to account for factors that could influence decisions to stop or continue screening using observed data from the US Medicare population (ages 70-84 years) among individuals that received a screening mammogram, had a predicted life expectancy of at least 10 years, and no previous breast cancer diagnosis. Results were presented separately for people ages 70 to 74 years and people ages 75 to 84 years. Individuals represented in this study were primarily White (92%) with an additional 5 percent described as Black and 3 percent as "other" (with no additional details) (**Table 5**).

Outcomes

Overdiagnosis

Based on the natural history of breast cancer, the additional cancers observed with continued annual screening after age 70 are a combination of aggressive asymptomatic cancers that could be treated once detected plus cancers that would never have become clinically apparent before the end of life (overdiagnosis). Overall, the 8-year cumulative risk of a breast cancer diagnosis was higher for the continued annual screening strategy after age 70 (5.5%) (5.3% ages 70-74, 5.8% ages 75-84) compared with the stop screening strategy (3.9%) (same proportion for both age groups) (**Table 10**). Paired with the mortality outcomes presented in this study, showing no benefit of continued screening for those ages 75 to 84 and only a trend toward a small mortality reduction for those ages 70 to 74, the detection of breast cancer is predominantly attributable to overdiagnosis. Additionally, the difference in diagnoses between the strategies reflects the proportion exposed to treatment harms that did not benefit, especially among those ages 75 to 84 (5.8% minus 3.9% = 1.9%).

Overtreatment

As noted above, there were more people diagnosed with breast cancer in the continue screening strategy beyond age 70 years, compared with the stop screening strategy. The specific cancer treatments received by those with a diagnosis were presented by study group (standardized to age distribution, comorbidity score, chronic conditions, and long-term care institutionalization). Lumpectomy and radiotherapy were more common for cancers diagnosed among individuals in the continued annual screening strategy compared with those that stopped screening after age 70, whereas radical mastectomy and chemotherapy were more common for cancers diagnosed in

those that discontinued screening after age 70 **(Table 10)**. Overall, because fewer individuals were diagnosed for the stop screening strategy (ages 70 to 84), there was a lower risk of undergoing follow-up and treatment (1.4% lower 8-year cumulative risk of a diagnosis).

KQ3a. Does Comparative Effectiveness Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

No studies of ages to start or stop screening presented data that would allow for testing of effect differences or stratification of results by different population characteristics or risk markers.

Screening Interval

Study and Population Characteristics

Three of the studies included to address potential harms of different screening intervals have been described elsewhere in this report (**Table 4**). The United Kingdom Co-ordinating Committee on Cancer Research (UKCCCR) is a RCT that was conducted within the UK National Breast Screening Programme (described above in KQ2).[124] The trial randomized 76,022 women to screening with DM at 50 to 62 years of age to receive an invitation for annual screening each year for the next 3 years or to one screening visit 3 years later. A fair-quality NRSI study by Parvinen et al.[157] was conducted in one city (Turku) as part of the Finnish national screening program, and used quasi-randomization to assign women ages 40 to 49 years to screening very year or every 3 years depending on whether their birth year was odd or even (see detailed description in KQ1 results). A fair-quality BCSC NRSI by Miglioretti et al.[152] compared cancer outcomes for women with at least two screening visits prior to a cancer diagnosis and a followup period (detailed description above for KQ2). The aim was to test whether the incidence and characteristics of cancer differed by the interval between the two visits (annual defined as 11 to 14 months versus biennial defined as 23 to 26 months) among 15,440 cancers identified in the BCSC screening program.

Two additional studies were identified to address potential harms of screening intervals by examining the potential cumulative harms across multiple rounds of screening (**Table 4**). A fair-quality NRSI by Ho et al.[138] was conducted using BCSC data to compute the cumulative probability of a false-positive result after 10 years of screening on an annual or biennial basis with either DM or DBT during the years 2005 through 2018. The included observations were based on the screening visits of 903,495 women (444,704 DBT examinations and 2,524,351 DM examinations). The mean (SD) number of examinations per woman was 3.3 (2.5). The second additional study was the Know Your Risk: Assessment at Screening (KYRAS) study[151] that calculated the cumulative risk of false positive screens over a median of 8.9 years. Eligible women were those screened at the Columbia University Medical Center (New York) during 2014 and 2015); for these women the study collected information on previous mammograms going back to 1989 based on their health record (N=2,019; median age 59 years). Overall, women underwent a median of seven mammograms during the study period.

Demographic characteristics were not commonly reported in the studies of screening interval (**Table 5**). As described in KQ2, the Miglioretti et al. BCSC[152] study population was primarily White (78%) with the remaining participants reported as Black (5%), Asian (5%), Hispanic

(5%), American Indian or Alaska Native (<1%), and 7% reported as "other" or unknown. In the KYRAS study[151] the population was majority Hispanic (76%) with the remaining reported as White (10%), Black (10%), or other (4%) including Asian, Pacific Islander, Native American, or Alaskan Native. Twenty-four percent of the non-Hispanic White women were of Ashkenazi Jewish descent.

Outcomes

Interval Cancers

Three studies presented data on interval cancers by participant screening interval (**Table 11**). The UKCCCR RCT[124] reported interval invasive cancers for women invited to annual screening (N= 37,530) compared with triennial screening following an initial prevalence screening visit (N= 38,492) during a three-year followup period. In the triennial screening group, these were cancers detected prior to rescreening after 3 years. In the annual screening group, these were cancers detected in the three intervals between screening visits during the 3 years of followup. Overall, the rate of interval cancers was significantly lower in the annual invitation group (1.84 per 1,000 women initially screened) than in the triennial invitation group (2.70 per 1,000 women initially screened) (RR 0.68, 95% CI, 0.50 to 0.92).

The Parvinen et al.[157] quasi-randomized study comparing annual with triennial screening from ages 40 to 49 in Finland reported mortality and the number of invasive interval cancers that occurred from 1987 to the end of 1994. Interval cancers were defined as those occurring after a negative mammogram and between two subsequent screening visits. Similar numbers of cases were reported in the annual screening and triennial screening groups and a statistical test for the difference was null (p=.22).

The Miglioretti et al. BCSC NRSI[152] presented analyses comparing interval cancers occurring one year after annual screening or two years after biennial screening. In the group with an annual screening interval preceding their cancer diagnosis, 22.2 percent (2,680/12,070, 95% CI, 21.5 percent to 23.0 percent) of all cancers diagnosed were interval cancers and this rate was lower than in the group with a biennial screening interval preceding their cancer diagnosis, where 27.2 percent (917/3,370, 95% CI 25.7 percent to 28.8 percent) were interval cancers. The study considered interval cancers to be those occurring within 12 months following an annual screening interval and within 24 months following a biennial screening interval. This was based on the assumption that the screening interval before a diagnosis was received would have continued. The study did not provide adjusted comparisons, limiting the ability to draw inferences about differences in the interval cancer rate associated with biennial and annual screening.

False-Positive Recall

The comparative NRSI from Ho et al.[138] used BCSC data to estimate the cumulative probability of having at least one false-positive recall over 10 years of screening with DBT or DM on an annual or biennial basis (**Figure 8, Appendix F Tables 3, 4**). Most individuals in the cohort were screened on an annual basis (73% DBT, 72% DM) versus biennial (15% DBT, 17% DM)

or triennial (12% DBT, 11% DM). Exams were defined as a false positive if no diagnosis of invasive cancer or DCIS occurred within one year of screening. For individuals screened with DBT the estimated cumulative probability of at least one false positive recall was 49.6 percent for those screened annually and 35.7 percent for those screened biennially (proportion difference: -13.9%, 95% CI, -14.9% to -12.8%). For individuals screened with DM the estimated cumulative probability of at least one false positive recall was 56 percent for those screened annually and 38 percent for those screened biennially (proportion difference: -18.2%, 95% CI -18.6% to -17.7%). The difference in cumulative false positive recalls between annual and biennial screening was larger for DM (-18.2, 95% CI, –18.6 to –17.7) than for DBT screening (-13.9, 95% CI, –14.9 to –12.8). The study also reported cumulative probabilities of false-positive short-interval followup recommendations (return for diagnostic imaging after 6 months). Approximately 17 percent of screened individuals undergoing annual screening expected to experience at least one short-interval follow-up recommendation compared with 10 percent of those undergoing biennial screening. The probability of short interval follow-up was slightly lower with DBT than DM.

In the KYRAS study[151] individuals screened with DM annually had 2.18 times the odds of having a false positive result compared with those who screened biennially (OR, 2.18; 95% CI, 1.70 to 2.80) after controlling for total years of follow-up, age, race/ethnicity, BMI, breast density, and breast cancer risk status (**Appendix F Table 4**).

False-Positive Biopsy

The comparative NRSI from Ho et al.[138] used data from the BCSC to estimate the cumulative probability of having at least one false-positive biopsy over 10 years of screening with DBT or DM on an annual or biennial basis (**Figure 9, Appendix F Tables 3, 4**). Most individuals in the cohort were screened on an annual basis. A false-positive biopsy was defined as a false positive if no diagnosis of invasive cancer or DCIS occurred within one year of screening. Biennial screening compared with annual screening led to a 5 percent lower cumulative false positive biopsy rate whether the screening was conducted with DBT or DM. For individuals screened with DBT, the estimated cumulative probability of at least one false-positive biopsy recommendation was 11.2% for those screened annually and 6.6% for those screened biennially (proportion difference: -4.6%, 95% CI, -5.2 % to -3.9%). For individuals screened with DM, the estimated cumulative probability of at least one false positive biopsy was 11.7% for those screened annually and 6.7% for those screened biennially (proportion difference: -5.0%; 95% CI, -5.4% to -4.7%) among those screened with DM.

KQ3a. Does Comparative Effectiveness Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

The Ho et al. NRSI[138] reported 10-year cumulative false-positive and biopsy rates by age and breast density category using BCSC data. Annual screening was associated with higher cumulative FP recall and biopsy for most age and density groups, though this difference in recall by screening interval was not seen among those with the lowest density (almost entirely fatty tissue) (**Figures 8, 9**). The cumulative risk of experiencing a false positive recall was positively associated with both younger age and higher breast density categories, except among those with

the lowest breast density category, where age differences were less pronounced (**Figure 8**). A similar pattern was seen for the cumulative risk of false positive biopsy (**Figure 9**). There was not a strong association between and age and cumulative false positive biopsy regardless of the screening interval among those with the lowest breast density (**Figure 8, 9**, **Appendix F Tables 5, 6**).

Digital Breast Tomosynthesis

Study and Population Characteristics

We identified 10 eligible studies, 4 RCTs (3 good-quality, 1 fair-quality)[127, 137, 141, 158] and six fair-quality NRSIs,[79, 130, 138, 142, 145, 160] that reported on potential harms of screening associated with the use of DBT (plus DM or sDM) compared to DM only screening (**Tables 4 and 5**).

Four large trials were conducted with individuals participating in organized screening programs in Germany,[137] Italy,[127, 158] and Norway.[141] Three of these trials were previously discussed in KQ2. The population characteristics were not well reported for the trials of DBT compared with DM. No trial reported race or ethnicity characteristics of the population.

The fair-quality Proteus Donna (N= 73,866)[127] and good-quality RETomo (N= 26,877) [30] RCTs were both conducted in Italy (detailed description above for KQ2) and reported screening results from two rounds of screening with randomization to DBT/DM or DM for the first round of screening and DM screening for all participants at the second round of screening (annual screening for ages 45 to 49, biennial screening for ages 50 to 69). The To-Be study[141] is a good-quality RCT conducted in Norway (described in detail for KQ2) that randomized participants to DBT/sDM screening (n = 14,380) or DM screening (n = 14,369) and followed them for two years, or until the next screening episode. The second screening round occurred two years later and consisted of DBT for all participants. Therefore, the study compares the findings from two rounds of screening with DBT (n = 11,201) compared with one round of DM screening followed by one round of DBT screening (n = 11,105).

One additional RCT was identified that addresses the potential harms of screening with DBT compared with DM. The TOmosynthesis plus SYnthesised MAmmography Study (TOSYMA)[137] is a good-quality RCT conducted in Germany that assigned 99,634 woman ages 50 to 69 to DBT/sDM (DBT with synthetic two-view imaging) versus DM alone between July 5, 2018 and December 30, 2020. Available results from the trial report on performance at a single round of screening and for this review was included only for rare or uncommonly reported harms (adverse events, radiation exposure). Future planned publications from the trial will report on interval cancers and cancer incidence at a second round of screening (see Discussion).

The six NRSI included for KQ3 were conducted using data from populations screened with DBT and DM in the US,[130, 138, 145, 160] Sweden[79] and Norway.[142] Additional details of the analysis for each of the NRSI are included in **Appendix E Table 1**.

The fair-quality PROSPR NRSI[130] used data from three US academic research centers and connected health care delivery systems that are members of the NCI-funded PROSPR

consortium to compare the performance of DM screening with DBT and DM screening combined. The study data included the findings from all bilateral screening mammography examinations provided at the study sites from approximately 2011 to 2014 (varied by study site) among women ages 40 to 74 years. Those with a history of breast cancer or imaging conducted in the 3 months prior to screening were excluded. The database included 103,401 individuals (55,998 DBT exams; 142,883 DM exams) and more than a quarter of women (28.3%) contributed three or more exams to the analysis.

The fair-quality Ho et al. BCSC NRSI study[138] (described above in studies of screening interval) provided estimates of the cumulative probability of a false-positive result after 10 years of screening with DBT or DM during the years 2005 through 2018. The included observations were based on the screening visits of 903,495 women (444,704 DBT examinations; 2,524,351 DM examinations). The mean (SD) number of examinations per woman was 3.3 (2.5). These round-specific probabilities were used to generate the cumulative probability of having at least one false-positive across 10 years of screening using discrete-time survival modeling to account for censoring. Over the study period, the proportion of examinations conducted with DBT increased but the age and breast density distributions of those screened with DBT and DM were similar as were the proportion screened annually (nearly three-quarters) versus biennially.

A second study utilizing BCSC data was the fair-quality NRSI by Kerlikowske et al.[145] conducted using data from screening visits at 44 BCSC facilities to compare outcomes of screening with DBT or DM during the years 2011 through 2018. Additional followup for cancer diagnoses obtained from state and regional cancer registries continued through 2019. The cohort included 504,427 women ages 40 to 79 years that had at least one DBT or DM screening visit (based on radiologist indication in medical record). Individuals with a history of breast cancer or mastectomy were excluded. The unit of analysis for the study was the screening examination and the analysis included 1,377,902 screening examinations.

The fair-quality Malmö Breast Tomosynthesis Screening Trial (MBTST)[79] is an NRSI using prospectively collected cohort data in Sweden comparing women screened with DBT/DM through their participation in a screening performance cohort study (n = 13,369) with a concurrently screening period (2010 to 2015) for an age-matched control cohort screened with DM (n = 26,738). The age-matched controls were selected from the screening program registry records for women that did not participate in the DBT/DM study but were screened in the same setting which relied on the same radiologists as the DBT/DM study.

A fair-quality NRSI by Richman et al.[160] conducted in the US used national medical claims registry data from the Blue Cross Blue Shield Axis data resource which contains deidentified commercial insurance claims. The study comparing DBT/DM screening with DM screening among women ages 40 to 64 years included individuals receiving at least 1 screening mammogram between January 1, 2016 and December 31, 2017 who had been continuously enrolled for at least 2 years preceding the screen and for at least 1 year following the screen. Individuals with any breast-cancer diagnosis in the 2 years preceding the screen and any insurance claims indicative of a genetic cancer syndrome or prophylactic mastectomy were excluded. The analytic sample included 7,602,869 screening mammograms conducted among 4,580,698 women.

The fair-quality OVVV study[142] is an NRSI using a geographical comparison cohort design that was conducted within the BreastScreen national screening program in Norway (detailed description for KQ2 above). The cohort study reported cancer screening outcomes from one round of screening with DBT/sDM (n = 34,641) or DM (n = 57,763) followed 2 years later with DM for those attending a second round of screening (n = 72,017).

Outcomes

Interval Cancers

Three trials reported interval cancers following screening with DBT or DM (**Table 12)**.[127, 141, 158] The three RCTs did not show statistically significant differences in the risk of interval cancer following screening with DBT or DM (pooled RR 0.87 [95% CI 0.64 to 1.17], I^2 0%, k = 3, n = 130,196) (**Figure 10**). Both Proteus Donna and RETomo used 12-month followup for those ages 45 to 49 years and 24-month followup for those ages 50 to 69 years. The relative risks for invasive interval cancers in these trials were RR: 0.92 (95% CI 0.60 to 1.42) and RR: 0.96 (95% CI 0.5 to 1.8), respectively. The To-Be RCT reported interval cancers from up to 24 months of followup for all participants (only individuals ages 50 to 69) and reported a relative risk of 0.71 (95% CI 0.40 to 1.27).

Five observational studies used data from medical systems, registries, and cancer screening and surveillance programs to compare interval cancers occurring after screening with DBT or DM obtained through recommended mammography screening visits or screening programs **(Table 12)**. These studies differed in the timeline of follow up and method of identifying interval cancers (**Appendix E Table 1**) highlighting the variability in interval cancer definitions and data used to assess the outcome across the included NRSI, and the need for more standardization of definitions and study protocols.

Three of the NRSIs found no significant difference in the rate of interval cancers diagnosed following screening with DBT or DM (including data from the BCSC,[145] PROSPR consortium,[130] and the OVVV comparative cohort study[142]). The Richman et al.[160] NRSI analysis of commercial insurance claims included an exploratory analysis of rates of cancers occurring between 5 and 12 months following the index screening with DBT or DM. These cancers were presumed to be identified clinically before the next scheduled mammogram. The study did not report invasive cancers separately from DCIS. Results from an adjusted multilevel model suggested small but statistically significantly higher incidence of interval cancer in the DBT/DM arm (0.52 per 1,000, 99% CI, 0.47 to 0.56) compared with the DM arm (0.45 per 1,000, 99% CI 0.43 to 0.48) with an adjusted difference of 0.07 interval cancers per 1,000 screens (99% CI, 0.01 to 0.12). The NRSI comparing the MBTST single-arm trial with an age-matched population-based cohort[79] examined rates of interval cancers 18 to 24 months after screening (depending on age). The rate of invasive interval cancers for DBT/DM was 1.4 per 1,000 women screened and for DM was 2.7 per 1,000 women screened (unadjusted RR 0.53, 95% CI, 0.32 to 0.87). The study did not report or adjust for characteristics of the MBTST NRSI participants and the comparison cohort. Both groups were drawn from a population-based screening program, but participants and outcome ascertainment in the MBTST NRSI could have differed from those not participating in a study.

Recall

The same three RCTs[127, 141, 158] and one NRSI[142] included for KQ2 reporting data across multiple rounds of screening were also included to assess screening recall rates **(Table 13)**. Recall was defined similarly across the studies, with any positive or suspicious results after double-reading and arbitration leading to recall for additional followup which could include or lead to additional imaging, percutaneous biopsy, open biopsy, or surgery. Results regarding recall rates were mixed across the first round of screening. In the RETomo RCT,[158] recall was similar in both study arms (RR: 0.99 [95% CI, 0.88 to 1.10]). The Proteus Donna RCT[127] reported a higher recall rate in the DBT/sDM study group (63.4 per 1,000) compared with the DM (50.9 per 1,000) group at round one (RR 1.24, 95% CI, 1.17 to 1.32). In the To-Be trial,[141] a lower proportion were recalled at round one in the DBT/sDM group (30.9 per 1,000) compared with the DM group (39.7 per 1,000) (RR 0.78, 95% CI 0.69 to 0.88). Inconsistency in the recall rates across trials at round one resulted in high statistical heterogeneity, so a pooled effect is not presented. The studies varied in their approaches to screening at round two: two RCTs used DM screening for both study groups (Proteus Donna, RETomo) and one used DBT for both study groups (To-Be) at round two. Recall rates at round two were more consistent and did not suggest a difference in recall between study groups when combined using meta-analysis (pooled RR, 0.97 [95% CI, 0.91 to 1.03] I^2 0%, k = 3, n = 105,244) (**Figure 11**).

Among those recalled, further evaluation was conducted and those without a DCIS or cancer diagnosis were reported as false positives (**Table 14**). Results were inconsistent across the four studies included for this outcome. In the Proteus Donna RCT the risk of a false-positive result was higher in the DBT/sDM group (55.1 per 1,000 screened) compared with the DM group at the first screening round (45.2 per 1,000 screened) (RR 1.22, 95% CI, 1.14 to 1.30). In the To-Be RCT, the first round of screening with DBT/sDM resulted in fewer false-positive recalls compared with DM (24.3 versus 33.7 per 1,000, RR 0.72, 95% CI, 0.63 to 0.83). There was not a statistically significant difference in the false-positive recall rates in the RETomo RCT[158] (RR: 0.90 [95% CI, 0.79 to 1.00]). Again, in round one inconsistency in effects for the two Italian trials compared with the To-Be trial resulted in high statistical heterogeneity, thus a pooled effect is not presented for the effect of DBT on false-positive recall. In all three trials, the relative risks of false-positive recall were near 1.0 at round two and the effects were sufficiently homogeneous to combine (pooled RR 0.99 [95% CI, 0.92 to 1.05] I^2 0%, k = 3, n = 105,244) (**Figure 12**).

The included NRSI OVVV study that used a concurrent regional comparison did not report a statistically significant difference in recall rates between the DBT and DM arms at round one (RR: 1.02 [95% CI, 0.95 to 1.09]) and reported slightly lower false-positive recall rates for those screened with DBT at the first screening round (RR 0.91, 95% CI, 0.84 to 0.98). At round two, both groups received DM and the recall and false positive rates were higher (~25%) in the DM compared with the DBT group (31 versus 24 per 1,000) (**Tables 13 and 14**).

Biopsy

Two of the included RCTs reported on the rate of biopsy following screening (**Table 13**). The RETomo study specified these biopsies were percutaneous needle biopsies, while the type of biopsy was not defined in the ToBe study. The RETomo RCT[158] reported more people had a

biopsy in the group randomized to DBT/DM (11.9 per 1,000) compared to those randomized to DM (8.1 per 1,000); the relative risk was 1.50 (95% CI, 1.10 to 1.90). At the second round of screening when all participants underwent screening with DM, there were fewer percutaneous biopsy referrals in the study arm originally screened with DBT/DM (6.1 per 1,000) compared with the DM study arm (8.1 per 1,000), although the relative risk was on the margin of statistical significance (RR 0.76, 95% CI, 0.57 to 1.00). Thus, cumulatively over the two rounds of screening a similar proportion of study participants were referred for percutaneous biopsy regardless of the screening modality used at the first screening round. Since both arms received DM screening at round two it is unclear whether an additional round of screening with DBT/DM would have resulted in higher biopsy rates for the intervention arm.

In the ToBe study, there was no evidence that rates of biopsy reported differed between the DBT/sDM arm (17.5 per 1,000) and the DM arm (18.9 per 1,0000), relative risk 0.93 (95% CI, 0.78 to 1.10). Similarly, no difference in the rates of biopsy were reported at round two when all participants were screened with DBT/sDM (RR: 0.95 (95% CI, 0.80 to 1.13). Only the To-Be RCT reported the necessary data to compute false-positive biopsy rates (**Table 14**). There was not a statistically significant difference between study arms in this outcome at either round of screening. At the first round false-positive biopsies occurred for 10.9 per 1,000 screened in the intervention group and 12.8 per 1,000 in the DM control group (RR 0.85, 95% CI, 0.69 to 1.05). The false-positive biopsy rate was approximately 14 per 1,000 in both study groups at the second screening round (RR 0.99, 95% CI, 0.80 to 1.24).

The RETomo RCT[158] also reported referrals to surgical followup on screening results, including open biopsy (**Table 13**). Similar to the percutaneous biopsy findings, surgical followup was higher in the DBT/DM group at round one (8.7 versus 5.0 per 1,000, RR 1.70, 95% CI, 1.3 to 2.30) but was not statistically different at round two when both arms received screening with DM (5.3 versus 6.4 per 1,000, RR 0.83, 95% CI 0.60, to 1.10). In the Proteus Donna RCT there were more surgery referrals in the DBT/sDM group at round one (9.9 versus 6.4 per 1,000, RR 1.54, 95% CI, 1.31 to 1.82) and fewer at round two in the intervention group (4.3 versus 5.7 per 1,000, RR 0.76, 95% CI, 0.59 to 0.97) after all participants were screened with DM.

Cumulative False-Positive Recall

The comparative BCSC NRSI from Ho et al. reported the estimated cumulative probability of having at least one false-positive recall over 10 years of screening with DBT or DM on an annual or biennial basis (**Figure 13, Appendix F Tables 3, 4**). Probabilities were lower with DBT screening compared with DM screening, regardless of the screening interval, but the difference was greater with annual screening. With annual screening, the 10-year cumulative probability of a false-positive recall was 49.6% with DBT and 56.3% with DM (Difference -6.7%, 95% CI, -7.4 to -6.1). With biennial screening, the 10-year cumulative probability was 35.7% for DBT and 38.1% for DM (Diff -2.4%, 95% CI, -3.4 to -1.5). The study also reported cumulative probabilities of receiving a false-positive short-interval followup recommendation, which was 16.6% for DBT and 17.8% for DM annual screening (Difference -1.1, 95% CI, -1.7 to -0.6) and ~10% for both modalities with biennial screening (Difference -0.1, 95% CI, -0.7 to 0.5).

Cumulative False-Positive Biopsy

The comparative BCSC NRSI from Ho et al. reported the estimated cumulative probability of having at least one false-positive biopsy over 10 years of screening with DBT or DM on an annual or biennial basis (**Figure 14, Appendix F Tables 3, 4**). Differences were small, and not statistically significant when comparing biennial DBT and DM. With annual screening, the 10-year cumulative probability of a false-positive biopsy was 11.2% with DBT and 11.7% with DM (Difference -0.5, 95% CI, -1.0 to -0.1). With biennial screening, the 10-year cumulative probability was 6.6% for DBT and 6.7% for DM (Diff –0.1%, 95% CI, –0.5 to 0.4).

Overdetection and Overtreatment

In the three RCTS (To-Be, RETomo,[158] and Proteus Donna) rates of DCIS detected at each screening round and between study arms were similar, ranging from 0.7 to 1.3 per 1,000 screened at the first screening round and from 0.6 to 1.3 per 1,000 screened at the second screening round, with no statistical differences between the DBT and DM screened groups (**Table 15**). Meta-analysis was used to generate combined estimates that also did not show statistically significant differences at round one (pooled RR 1.33, 95% CI 0.92 to 1.93, k = 3 RCT, n = 130,196, I^2 = 0%) or round two (pooled RR 0.75, 95% CI 0.49 to 1.14, k = 3 RCT, n = 130,196, I^2 = 0%) (**Figure 15)**. The OVVV NRSI reported higher DCIS detection at the first screening round in the DBT group compared with the DM group (1.8 versus 0.8 per 1,000 screened; RR 2.16, 95% CI, 1.49 to 3.12).

Adverse Events

The TOSYMA RCT reported on adverse events from a single round of screening using DBT/sDM compared with DM only. The study randomized 49,804 individuals to DBT/sDM and 49,830 to DM. Six adverse events were reported in each study arm and none were serious (e.g., fainting, circulatory collapse/problem, allergic skin reaction). Device deficiencies occurred in 0.05 percent (23/49,179) of examinations in the DBT/sDM group and 0.01 percent (5/50455) of DM examinations. None were determined to cause any serious adverse events.

Radiation Exposure

Five studies (4 RCTs, 1 NRSI) reported the median, mean, or relative radiation dose by study arms from a single screening round (**Table 16**). In three of these studies participants underwent a DBT and DM screening (in one or two compressions) and in two studies participants underwent DBT with a synthetic reconstruction of a 2D DM image.

Studies using DBT/DM screening reported radiation exposure approximately two times higher in the intervention group compared with the DM only control group. The RETomo RCT[158] reported higher median dosages, with the DBT/DM group having 6.40 milligray (mGy) (IQR 5.68 to 7.36) and the DM group 4.84 mGy (IQR 4.24 to 5.72), with the dose in the DBT/DM arm reported to be 2.3 times higher than in the DM study arm. The Proteus Donna RCT reported that the DBT/DM study group received radiation doses approximately 2.5 times higher than the DM only study group, but did not report details. The MBTST NRSI reported mean (SD) radiation

doses with DBT (mean 2.3 mGy, SD 0.7) and DM (2.7 mGy, SD 0.8) screening in the study population that comprised the intervention group for the larger analysis presented in this NRSI. Assuming the radiation dose with DM in the comparison population was similar, the DBT/DM group was exposed to nearly two times more radiation than the DM only group.

Differences between study groups in radiation exposure were smaller in studies using DBT/SM. The TOSYMA RCT reported median glandular radiation dose in the DBT/sDM group was 1.86 mGy (IQR 1.48 to 2.45) and in the DM group was 1.36 mGy (IQR 1.02 to 1.85). In the To-Be RCT which also used DBT/SM, the mean radiation dose was 2.96 mGy compared with 2.95 mGy in the DM group.

KQ3a. Does Comparative Effectiveness Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

None of the included studies were designed to enroll populations to support comparisons in the screening outcomes of DBT and DM by race or ethnicity, or family history.

Two RCTs and four NRSIs that compared DBT-based screening strategies with DM only screening strategies presented results stratified by age and/or breast density.

Age

The RETomo RCT[158] reported the effects of DBT/sDM versus DM on recall, biopsy, and surgical procedures stratified by age category (45-49 versus 50-69) (**Table 17**). Recall rates at the first round of screening were similar for both age groups regardless of study intervention (~40 per 1,000 screened). At the second screening (DM for both groups), those ages 45 to 49 in the DBT group had lower recall (34 vs 40 per 1,000 screened; RR 0.84, 95% CI, 0.69 to 1.00) and lower rates of percutaneous biopsy (3 vs 6 per 1,000 screened; RR 0.50, 95% CI, 0.27 to 0.94).

Conversely, those ages 50 to 69 screened with DBT experienced no difference in recall at either round, but more percutaneous biopsy (14 vs 9 per 1,000 screened, RR 1.40, 95% CI, 1.10 to 1.90) at the first screening round and no difference at round two. False positive recall was not statistically different for either age group at either screening round, although a trend toward lower FP recall at the first round was seen for those ages 50 to 69 screened with DBT (**Table 18**). After the first screening round both age groups were more likely to have surgical procedures, including open biopsy, in the DBT screened intervention group. At the second round, participants ages 45 to 49 were less likely to have surgical procedures, but the effect was on the margin of statistical significance (RR 0.50, 95% CI, 0.24 to 1.00). There was no difference in surgical procedures for those ages 50 to 69 at round two.

RETomo[158] also presented stratified analyses comparing interval cancer incidence by age groups; the event rates were low and all confidence intervals contained null (**Table 19**). The study did not report interaction tests, and was not designed to test for subgroup differences, making it difficult to draw conclusions about differences by age. Overall, these stratified results suggest some risk of increased biopsy or surgery with DBT screening at the first round for all, followed

by lower rates at the next round for those ages 45 to 49. Analytic and study design limitations preclude firm subgroup conclusions.

The Richman et al. NRSI using administrative data from a large US private insurer compared invasive interval cancers occurring 5-12 months after negative DBT or DM screening among individuals ages 40 to 64 years and conducted a test for interaction across 5-year age group categories (**Table 19**). Although there was a main effect of DBT screening (increased odds of invasive interval cancer – reported above), the age group by intervention interaction term was not statistically significant (p=0.54, α 0.01).

The MBTST NRSI reported interval cancers stratified by two age groups: 40 to 54 versus 55 to 74 years (**Table 19**). The adjusted OR of being diagnosed with an interval cancer among women less than 55 years of age was 0.5 (95% CI, 0.2 to 1.1) and 0.6 (95% CI, 0.3 to 1.1) among women 55 years and older, similar to the overall effect for both groups combined which was statistically significant. The study did not report interaction tests and was not designed to test these subgroup comparisons, making it difficult to draw conclusions about differences by age group.

Breast Density

The RETomo RCT[158] presented stratified analyses comparing interval cancer incidence by density groups; event rates were low, and all confidence intervals contained null (**Table 19**). The study did not report interaction tests, and was not designed to test for subgroup differences, making it difficult to draw conclusions about differences by age.

The To-Be trial reported recall and biopsy stratified by Volpara density grade categories (VDG1-VDG4). There was lower recall at the first screening round for those screened with DBT that had lower density breasts (VDG1 and VDG2) but not for those with higher density breasts (VDG3 and VDG4), and no statistical difference in recall at the second screening (**Table 17**). No statistical difference in biopsy at round one or two was reported for any of the breast density categories. The risk of a false positive recall in the DBT/SM group was, however, lower at round one for those with less dense breasts (VDG1 RR 0.58, 95% CI, 0.43 to 0.80; VDG2 RR 0.66, 95% CI, 0.54 to 0.81) and this trend was also seen for false positive biopsy (VDG1 RR 0.57 95% CI, 0.33 to 1.00; VDG2 RR 0.64, 95% CI, 0.46 to 0.89). Those in density category VDG3 had a higher risk of false positive biopsy (RR 1.79, 95% CI 1,.23 to 2.61) At the second round of screening there was no evidence of a difference in false positive recall for any of the density categories, but in the highest density category false positive biopsy risk neared statistical significance (RR 2.08, 95% CI, 0.98 to 4.41). (**Table 18**) and was not statistically different for those in the VDG3 or VDG4 breast density categories. In density stratified comparisons of invasive interval cancer, event rates were small, and the confidence intervals contained null (**Table 19**). The study did not report interaction tests and was not designed to test these subgroup comparisons making it difficult to draw conclusions about differences by breast density.

The Kerlikowske et al. BCSC NRSI presented comparisons of interval cancer incidence following DBT and DM screening examinations stratified by breast density category and additionally stratified within density categories by BCSC risk score (<1.67% versus ≥ 1.67). No

statistically significant differences in the incidence of interval cancer were reported for the breast density stratified comparisons (**Table 19**) or the density and risk stratified comparisons.

The To-Be RCT reported mean radiation doses for the study groups, stratified by breast density in a figure. The study reported that there were not statistically significant differences in radiation dose for DBT/SM compared with DM for any of the density categories.

Age and Breast Density Subgroups

The Ho et al. BCSC NRSI presented 10-year cumulative false positive recall and biopsy probabilities stratified by breast density and age and comparing DBT to DM screening. Overall, the study reported lower false positive recall with DBT screening. In stratified analyses, however, there was not a statistical difference in cumulative false positive recall among those with extremely dense breasts in any age group. Among individuals ages 50 to 59 with extremely dense breasts screened on a biennial basis the risk of false positive recall was higher with DBT compared with DM screening (**Figure 16**). Cumulative false-positive biopsy was also no different with DBT versus DM screening among individuals with extremely dense breasts.

Magnetic Resonance Imaging

Study and Population Characteristics

The Dense Tissue and Early Breast Neoplasm Screening (DENSE) trial is a good-quality RCT conducted in the Netherlands that enrolled participants from December 2011 to November 2015 (N = 40,373) (**Table 4**). The aim of the study was to determine whether an invitation to supplemental MRI screening after a negative mammogram for those with extremely dense breast tissue would reduce the incidence of interval cancer.

Women ages 50 to 75 years old participating in the national digital mammography program that were assessed to have extremely dense breasts and had a negative screening result (BI-RADS 1 or 2) were randomized (1:4) to an invitation to supplemental MRI screening (n = 8,061) or usual care screening invitation (mammography in 2 years) (n = 32,312). The baseline characteristics of the study groups were balanced on the reported characteristics (**Table 5**). Among those invited to MRI screening 59 percent underwent the MRI examination (n = 4,783). The characteristics of MRI completers and noncompleters were reported to be similar based on the limited set of variables, but the groups could differ in their motivation to be screened, access to health care, and risk factors for invasive breast cancer, such as family history and other characteristics. Therefore, study outcomes reported only for MRI screened participants (detection, recall, false-positives) are subject to higher risk of bias than outcomes based on all randomized study participants (interval cancer). Single-read MRI examinations were conducted by breast radiologists with findings of BI-RADS 3 referred for follow-up MRI imaging in 6 months and BI-RADS 4 or 5 recalled for diagnostic evaluation. Cancer diagnoses and tumor characteristics were obtained from the Netherlands Cancer Registry. While this study included two rounds of screening with MRI, findings from the second round of screening in the mammography only arm have not been published. Therefore, this study was not eligible for inclusion in KQ2, but it is included for interval cancers and potential harms of supplemental MRI imaging.

A fair-quality NRSI compared commercially insured women ages 40 to 64 years of age in the MarketScan database who had received at least one bilateral screening breast MRI (n = 9208) or mammogram (n = 9,208) between January 2017 and June 2018 (**Tables 4 and 5**). Propensity score matching was used to compare cascade events (mammary and extramammary) in the 6 months following the MRI or mammogram that were potentially attributable to having a breast MRI (see **Appendix E** for Detailed Methods).

Interval Cancers

In the DENSE RCT the ITT analysis based on invitation to MRI screening found a rate of invasive interval cancers for the DM+MRI of 2.2 per 1,000 invited to screening compared with 4.7 per 1,000 screened for the DM only control group (RR 0.47, 95% CI, 0.29 to 0.77) (**Table 12**).

Adverse Events

In the DENSE RCT, eight adverse events (including 5 classified as serious adverse events) occurred during or immediately after the MRI screening. Adverse events included two vasovagal reactions and three allergic reactions to the contrast agent (serious adverse events) as well as two reports of extravasation (leaking) of the contrast agents and one shoulder subluxation. Twenty-seven individuals (0.6% of MRI arm) reported as serious adverse event within 30 days of the MRI; however, the authors did not determine whether any of these were attributable to the MRI. Two of these serious adverse events were unspecified complications during or after a biopsy that took place after the initial MRI screening exam.

Downstream Consequences of Supplemental Imaging Including Incidental Findings

Because women were selected for the DENSE trial based on a negative mammography, all of the additional imaging (including additional exposure to radiation and contrast agents) and biopsy procedures would not have occurred in the absence of MRI screening. Among those who underwent MRI in the first round of the DENSE trial the rate of recall for additional imaging following MRI was 94.9 per 1000 screened and the false positive rate was 79.8 per 1,000 screened (**Table 20**). The rate of biopsy for those undergoing supplemental MRI was 62.7 per 1000 screened. Among the cancers diagnosed by MRI over 90 percent were classified as DCIS (stage 0) or stage 1 cancer. Without information for two rounds of screening from both arms of the study there is not sufficient information to weigh the relative benefit versus harms of these diagnoses and downstream imaging consequences. The DENSE trial did not report on incidental findings from MRI imaging.

In the US insurance claims NRSI, individuals that had an MRI compared to those receiving only a mammogram were more likely in the subsequent 6 months to have additional cascade events, most related to breast conditions. Since individuals could contribute multiple events, rates were per 100 screened and could exceed that number. Events unrelated to breast diagnostic codes were higher in the MRI group (304.5 per 100) than in the mammography group (284.8 per 100), and the adjusted difference between groups (19.6 per 100, 95% CI, 8.6 to 30.7) was mostly comprised of additional health care visits. There were no statistically significant differences in

laboratory tests, imaging tests, procedures, hospitalizations, or new diagnoses (unrelated to breast conditions) (**Table 20**).

KQ3a. Does Comparative Effectiveness Differ By Population Characteristics And Risk Markers (e.g., age, Breast Density, Race/Ethnicity, Family History)?

No studies of supplemental MRI screening presented data that would allow for testing of effect differences or stratification of results by different population characteristics or risk markers.

Ultrasound

Study and Population Characteristics

The Japan Strategic Anti-cancer Randomized Trial (J-START) is a fair-quality RCT that randomly assigned asymptomatic women ages 40 to 49 years of age in 23 prefectures in Japan to breast cancer screening with mammography plus handheld ultrasound ((DM/US) (n = 36,859) or mammography only (DM) (n = 36,139) over two rounds of annual screening from 2007 to 2011 (**Table 4**). Those with a personal history of breast cancer or in situ lesions, any other cancers in the previous 5 years, or a life expectancy of 5 or fewer years were not eligible for the study. The two study groups were balanced across a range of characteristics, and for nearly one-quarter (23.2%) the first round of screening was their first breast cancer screen (**Table 5**). The authors note that 58 percent of women were classified as having dense breasts; however, the distribution of breast density across study arms was not reported. The findings of the DM, clinical exam (when performed), and ultrasound exams were considered independently. An ITT analysis was published in 2016, reporting on the first screening round, but there have been no further publications from the main trial.[286] The absence of second round screening results limits conclusions that can be drawn with regard to the effectiveness of supplemental ultrasound screening. Therefore, this study was not eligible for inclusion in KQ2, but it is discussed here for interval cancers and potential harms related to supplemental ultrasound imaging.

An NRSI by Lee et al. reported results of an analysis using data from two BCSC registries to compare screening outcomes for individuals receiving ultrasonography on the same day as a screening mammogram (DM/US) (n = 3,386, contributing 6081 screens) compared with those that received only a mammogram (DM) (n = 15,176, contributing 30,062 screens) (**Tables 4 and 5**, see **Appendix E** for Detailed Methods). Screening exams occurred between 2000 and 2013 and follow-up was for 12 months after the screening examination visit, or until the next examination. The majority of individuals were White (accounting for 80% of the screening examinations) with 11 percent reported as Asian/PI, 7 percent as Hispanic, less than 1 percent as Black, and 2 percent as “mixed/other.” Sixty-five percent of exams were performed in individuals classified as having heterogeneously or extremely dense breasts. This study enrolled a higher risk population with 31 percent of exams were among those with a known first-degree family history of breast cancer and 35 percent of women classified with at least an intermediate 5-year risk of breast cancer. Forty percent of exams were among those with a previous breast biopsy. Propensity score matching was used to adjust for confounding on using variables available in survey, EMR, and registry data.

Outcomes

Interval Cancers

In the J-START RCT there were 16 interval cancers reported (0.4 per 1,000) in the DM/US group and 27 (0.8 per 1,000) in the DM group after the round one exam (**Table 12**). The relative risk was not statistically different (RR: 0.58, 95% CI 0.31 to 1.08) when calculated based on the study reported event rates. The study reported effect of the intervention was estimated with a generalized estimating equation (GEE), aimed at accounting for the mix of cluster and individual randomization. The analysis produced a population-averaged effect that was on the margin of statistical significance for a difference in the risk for invasive interval cancer by study arm (proportion difference -0.05, 95% CI, - 0.09 to 0). We calculated the individual-level relative risk to support comparability across the studies in this review. Not accounting for clustering in analysis usually results in narrower confidence intervals, so the naïve estimate we calculated is more likely to be biased toward a statistically significant effect (type II error).

The Lee et al. NRSI using BCSC found no statistical difference in the interval cancer rate, with 9 interval cancers (invasive and DCIS) following examinations with DM/US (1.5 per 1,000 screens) and 56 interval cancers following examinations with DM only (1.9 per 1,000 screens) in the propensity matched comparison groups (aRR 0.67, 95% CI 0.33 to 1.37) (**Table 12**).

Downstream Consequences of Supplemental Imaging

The findings of each modality were considered separately in the J-START trial, allowing estimation of additional followup (including imaging and biopsy) attributable to supplemental ultrasound screening. The rate of recall based only on ultrasound was 49.7 per 1,000 in the ultrasound arm and 48.0 per 1,000 had a false positive recall (**Table 20**). Of those cancers identified only by ultrasound 76.2 percent were classified as stage 0 or 1 cancer. Without information on cancers detected over two rounds of screening from both arms of the study, or health outcomes, there is not sufficient information to weigh the relative benefit versus harms of these diagnoses and downstream imaging consequences. The J-START trial did not report on any incidental findings from ultrasound imaging.

The NRSI by Lee did not report the findings of ultrasound and mammography separately; therefore, we are unable to account for how much followup is attributable to the use of ultrasound alone (**Table 20**). Referral to biopsy and false positive biopsy recommendations were twice as high and short interval followup three times as high for the group screened with ultrasound, despite there being no statistical difference in cancer detection. Despite the use of propensity scoring to adjust analyses, unmeasured differences in the groups screened with ultrasound and those not screened may confound the results and bias estimates.

KQ3a. Does Comparative Effectiveness Differ by Population Characteristics and Risk Markers (e.g., Age, Breast Density, Race/Ethnicity, Family History)?

A secondary analysis of J-START reported results for trial participants from a single screening center in one Japanese prefecture (Miyagi) to compare interval cancer rates for DM/US and DM screening among women ages 40 to 49. The interval cancer rates (invasive and DCIS) were

reported stratified by two breast density groups: nondense (BI-RADS A, B) and dense (BI-RADS C, D). Among those with non-dense breasts the rate of invasive interval cancers among the DM/US arm was 0.5 per 1,000 compared with 2.3 per 1,000 for those undergoing DM only (RR 0.229, 95% CI, 0.05 to 1.03). For those with dense breasts the invasive interval cancer rate was 0.5 per 1,000 among those randomized to DM/US compared with 1.8 per 1,000 for DM only (RR 0.29, 95% CI, 0.08 to 1.05). (**Table 19**). The study used GEE to estimate population-averaged effects accounting for the mix of cluster and individual randomization per the statistical analysis protocol. As noted above, we present raw calculated relative risks for consistency across studies. By not accounting for the clustered data, the relative risks we calculated based on event rates would be more likely to underestimate correlations in the error terms, leading to narrower confidence intervals than analyses that use robust standard errors to account for clustering.

The rates of recall based only on ultrasound were 69.7 per 1000 (95% CI 63.3 to 76.6) among those with dense breast and 39.4 per 1000 (95% CI 33.5 to 46.0) among those with nondense breasts. Of cancers identified only by ultrasound 76.5 percent were classified as stage 0 or 1 cancer among those with dense breasts, and 86.7 were classified as stage 0 or 1 among those with non-dense breasts. Without data for two rounds of screening from both arms of the study there is not sufficient information to weigh the relative benefit versus harms of these diagnoses and downstream imaging consequences.

Personalized Screening Programs Using Risk Assessment

No eligible studies were identified that reported on the potential harms of screening comparing usual care mammography with personalized screening programs using risk assessment.

Appendix H Table 1. Ongoing Trials of Mammography Screening Strategies and Modalities

Intervention	Study Name Trial Identifier Location	Planned Study Population	Intervention	Relevant Outcomes	Status*
Age to Start/Stop					
Screening Interval	MISS (NCT04590560) Italy	60,000 women ages 45-49 years	Annual vs. biennial screening with DBT and sDM (based on breast density)	Cancer detection Interval cancer rate Recall rate	Ongoing Expected completion date, February 2026
	TBST (NCT02619123) Italy	33,200 women ages 44-45 years	Annual screening vs. tailored screening *(women in the intervention group with BIRADs 3 or 4 density will be invited to screen again after 1 year, while women with BIRADS 1 or 2 density will be invited after 2 years. After age 50, all women will be screened according to usual care)*	Cancer detection Interval cancer rate False positive rates	On Hold Results will be part of a pooled analysis with a recently funded study looking at screening intervals
Modality	PROSPECTS (NCT03833106) United Kingdom	100,000 women ages 49-71 years	DBT + DM vs. DM	Cancer detection Interval cancer rate Recall rate	Ongoing Estimated completion date, July 2024
	TOSYMA (NCT03377036) Germany	80,000 women ages 50-69 years	DBT vs. DM	Cancer detection Interval cancer rate Recall rate	Ongoing Estimated completion date, March 2025
	MAITA (NCT04461808) Italy	8,000 women ages 45-65 years	DBT vs. DM	Cancer detection Interval cancer rate Recall rates Biopsy rates	Ongoing Expected completion date, June 2026 Interim findings (first-round followup) will be published in 2024
	TMIST **(**NCT03233191) United States, Canada	128,905 women ages 45-74 years with BIRADS density C or D†	DBT vs. DM	Breast cancer-specific mortality Cancer detection Interval cancer rate Recall rates Biopsy rates	Ongoing Expected completion date, December 2030
	IMPETO (NCT03587259) Italy	6,000 women ages 45-46 years	DBT vs. DM	Cancer detection Recall rate Biopsy rate	Ongoing

Appendix H Table 1. Ongoing Trials of Mammography Screening Strategies and Modalities

Intervention	Study Name Trial Identifier Location	Planned Study Population	Intervention	Relevant Outcomes	Status*
					Enrollment postponed due to COVID-19
Personalization	WISDOM (NCT02620852) United States	100,000 women ages 40-74 years	Annual screening vs. tailored screening *(annual screening with an individualized, risk-based screening schedule)*	Cancer detection Interval cancer rate Recall rate Biopsy rate	Ongoing Estimated completion date, March 2025 Interim findings (patient centered outcomes, mutation carrier characteristics, polygenic risk score analysis) will be published in late 2022/early 2023
	MyPeBS (NCT03672331) Europe, Israel	85000 women aged 40-70 years	Standard screening based on national/regional guidelines vs. risk-based screening *(screening interval based on estimated 5-year risk of developing breast cancer, via DM and/or DBT every 1 to 4 years with or without US depending on breast density)*	Breast cancer-specific survival (including combined analysis with WISDOM trial) Cancer detection Interval cancers False positive imaging and benign breast biopsies	Ongoing Estimated completion date, December 2025

*Status is based on published results, information provided by investigators or expected completion date as reported in clinical trials.gov