Mammographic Breast Density as a Predictive Tool in Average-risk Women aged 40 to 59: a systematic review

By

Amber Kernodle

A Master’s Paper submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Public Health in the Public Health Leadership Program.

Chapel Hill

2014

Russell P. Harris, MD, MPH

July 14, 2014

Cherie Kuzmiak, DO

July 15, 2014
Table of Contents:

Abstract .............................................................................................................................................. 1

Introduction ......................................................................................................................................... 3
  Breast cancer facts .......................................................................................................................... 3
  Breast cancer risk factors .............................................................................................................. 3
    Table 1. Breast cancer risk according to risk factors ................................................................. 3
  Breast density and legislation ........................................................................................................ 3
  Breast cancer screening recommendations .................................................................................. 3
  Screening strategies ...................................................................................................................... 3
    Figure 1: Prevention strategy to relieve the “burden of suffering” ........................................... 3
  Potential for breast density .......................................................................................................... 3

Background ........................................................................................................................................ 13
  History of mammographic breast density ................................................................................. 13
  Mammography Quality Standards Act and Program (MQSA) ...................................................... 13
    Table 2. Breast tissue composition and frequency distribution ............................................. 13
  Risk Factors as Predictive Tools .................................................................................................. 13
    Table 3. Criteria for evaluating predictive tools ........................................................................ 13

Methods ............................................................................................................................................. 18
  Key Questions ............................................................................................................................... 18
    Table 4: PICOTSS ....................................................................................................................... 18
  Data Sources and Search Strategy .............................................................................................. 18
  Study Selection ............................................................................................................................. 18
  Data Collection and Quality Assessment .................................................................................... 18
  Data Synthesis and Analysis ........................................................................................................ 18

Results ................................................................................................................................................. 24

Discussion .......................................................................................................................................... 35
  Significance of main findings ....................................................................................................... 35
  Limitations of review .................................................................................................................... 35
  Limitations of evidence ................................................................................................................ 35
  Implication for research and practice .......................................................................................... 35

Conclusion .......................................................................................................................................... 40

Appendices .......................................................................................................................................... 47
  Detailed search terms .................................................................................................................... 47
  Flow diagram for search strategy ................................................................................................. 47
  Data extraction and critical Appraisal .......................................................................................... 47
Abstract

Background: In 1976 Dr. John Wolfe introduced the relationship between breast tissue composition and breast cancer risk. Breast tissue composition is one of several well-established risk factors for breast cancer, but is not currently incorporated into commonly used risk prediction models or routinely used in clinical practice to recommend for or against routine screening. Knowledge of risk factors and their predictive abilities is important for creating appropriate screening strategies. It would be of value if breast tissue composition, alone or in combination with other risk factors, would allow for creating personalized screening plans for every woman based on her individual risk as screening can be harmful and costly.

Objective: To assess whether the American College of Radiology (ACR) Breast Imaging-Reporting and Data System (BI-RADS) breast tissue composition categorization has a sufficient degree of incremental value beyond age to predict incidence and/or breast cancer-related mortality in average-risk women aged 40 to 59. Evaluation is based on the following criteria: magnitude of risk estimates, discriminatory ability, net reclassification index, clinical utility and cost-effectiveness.

Search methods: I searched PubMed and EMBASE from January 1, 2003 to April 19, 2014. Radiologist assessment of breast tissue composition using mammography was the intervention of interest. I focused on outcomes assessed in the short-term, within 10 years of intervention. I excluded lifetime estimates. Exclusion criteria for the population included risk factors, such as personal history of breast cancer and BRCA 1 or BRCA 2 mutation carrier status.
**Main results:** Eleven studies met the inclusion criteria. While exposure and outcome measurement were relatively consistent across studies, there was heterogeneity in study design, population age range, follow-up duration and outcomes assessed. Despite this, it is clear that there is a dose-response relationship between breast tissue composition and risk of in situ and invasive breast cancer. Across all of the studies, effect sizes were moderate. There was not a meaningful relationship with short-term mortality, though few studies in this review evaluated this outcome. The risk factor’s ability to discriminate, separate cases from non-cases, was close to chance and not clinically proven to be superior to existing Breast Cancer Risk Assessment Tool (Gail Model). There were no studies that calculated the Net Reclassification Index (NRI), or evaluated clinical utility or cost-effectiveness.

**Limitations of the evidence:** The BI-RADS lexicon used to describe breast tissue composition is subjective; creating measurement bias that can lead to inaccurate risk estimates. Outcome data (incidence and vital status) were obtained from population-based registries. Equal, valid and reliable measurement of the outcome rests on the completeness, validity and reporting timeliness of utilized databases. This information was variably reported and not easily obtainable. Lastly, breast density is not stable over time. Crossovers, movement from one category to another, can diminish risk estimates.

**Conclusion:** Based on the available evidence, breast tissue composition, as currently assessed in routine clinical practice, does not fulfill the criteria to be valuable as a predictive tool. Authors of breast density notification laws should clarify intentions, critically examine the evidence and evaluate outcomes. Given subjectivity, focus could be shifted to objective measurements and other enterprising technologies, such as breast magnetic resonance imaging or 3D digital breast tomosynthesis.
Introduction

In his article “Sick individuals and sick populations”, well-known epidemiologist Geoffrey Rose writes that he encourages his medical students when thinking about their patients to consider – Why did this patient get this disease at this time?¹ For many diseases, the answers to this multi-step question are not easy. Other pertinent questions for population level interventions include – What are the harms and benefits to intervening on this disease process? How much will it cost the patient and society?

Currently, many organizations recommend routine screening mammography to women for early detection of breast cancer. Mammography has harmful downstream consequences for some women and is costing the United States progressively more money over time.² ³ So, to decrease the undesirable outcomes of this screening intervention, we must go back to part of Rose’s question – Why this patient?

Breast Cancer Facts

According to the Surveillance, Epidemiology and End Results (SEER) annual Cancer Statistics Review, a report from a population-based registry that covers 28% of the general US population, breast cancer has the highest incidence of any cancer, about 125 cases per 100,000 women. It is estimated that in 2014, 40,000 women will die of breast cancer. The highest proportion of breast cancer deaths occur in women aged 55 to 64. About 5% of deaths occur in women aged 35 – 44, and 16% occur in women greater than 84-years-old.⁴ There is persistent controversy around screening recommendations for all women – young and old, “higher risk” and average risk, and with and without pertinent medical findings. While the conversation around breast cancer screening is fluid and polemic, participating stakeholders have the same interest in mind – to relieve the burden of breast cancer and to preserve the health and quality of life of women. The most reasonable screening recommendation for breast cancer is the result
of interpreting statistics, like those from the SEER database, and integrating these facts with population health concepts.

Health statistics include disease outcome and frequency data. Incidence, a frequency measure, refers to the number of new breast cancer cases among an initially disease free population over a specified time. Screening is the detection of disease in clinically asymptomatic individuals. The desired effect of screening mammography is to reduce breast cancer-related morbidity and mortality. Due to a number of factors, such as disease course and disease detection capabilities, early detection by screening can increase documented incidence.

Ductal carcinoma in situ (DCIS) is classified as a breast cancer and thought by some to be a precursor to invasive breast cancer.\(^5,6\) DCIS is a biologically heterogeneous disease, each variant having different prognoses. Unfortunately, at this time, we are not able to predict with certainty the DCIS lesions that will progress and those that will not. The incidence of DCIS has increased with the increased utilization of screening mammography.\(^7\) Low-grade DCIS is thought to be a slowly progressive lesion, thus introducing length-time bias and contributing to over-diagnosis. The early detection of DCIS is giving rise to the increased incidence of breast cancer. However, detecting these lesions that will not cause relevant outcomes by screening is not beneficial.

Mortality, a prognostic health outcome, is a useful statistic to help balance this information. For example, a high incidence rate for a clinically benign condition, with a very low mortality rate is not distressing. Short-term mortality, at 5 or 10 years, is preferable to lifetime estimates because it reduces the risk of influence by competing comorbidities. To determine the effect of screening mammography, the incidence of both in situ and invasive breast cancer, and pertinent health outcomes, such as near-term mortality are helpful. These pieces of information, integrated with other concepts, can be used to determine at what age and how often a women should be screened.
Breast Cancer Risk Factors

A risk factor is a characteristic, condition or behavior that makes an individual more likely to experience an outcome. Thus, risk factors are often used when predicting outcomes, such as invasive breast cancer diagnosis and breast cancer-related death. Risk factors can be causally related or statistically correlated to the outcome of interest. Support for causality can come from evidence of biologic mechanism or temporal sequence of events for example. Both causal and statistical risk factors may be useful in predicting breast cancer incidence and mortality.

Risk factors for breast cancer include age, number of first-degree relatives with a history of breast cancer, having a first-degree relative with breast cancer before the age of 40 and having "extremely dense" breast tissue composition. In the United States, breast tissue composition is most commonly reported using standard language outlined by the American College of Radiology (ACR) in the breast imaging-reporting and data system (BI-RADS) Atlas®. Other commonly associated risk factors for primary breast cancer in all women include benign proliferative breast lesions, inherited genetic mutations, reproductive factors, such as age at menarche and nulliparity; and some lifestyle factors, such as alcohol use, smoking, decreased physical activity, and weight gain after menopause. Table 1 provides the risk ratios associated with each risk factor.
Table 1. Breast cancer risk according to risk factors

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Breast Cancer Risk Ratio (95% CI)</th>
<th>Risk Factor</th>
<th>Breast Cancer Risk Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two first-degree relatives with breast cancer*</td>
<td>3.84 (2.37 – 6.22)</td>
<td>Menarche ≤ 12 years of age*</td>
<td>1.10 (0.98 – 1.23)</td>
</tr>
<tr>
<td>First-degree relative diagnosed at &lt;40 years of age*</td>
<td>3.0 (1.8 – 4.9)</td>
<td>Nulliparity*</td>
<td>1.16 (1.04 – 1.26)</td>
</tr>
<tr>
<td>Extremely dense breast tissue*</td>
<td>2.04 (1.84 – 2.26)</td>
<td>Alcohol use ≥ 14 drinks per week*</td>
<td>1.24 (0.87 – 1.78)</td>
</tr>
<tr>
<td>Benign proliferative lesions</td>
<td></td>
<td>Smoking*</td>
<td>1.05 (0.98 – 1.13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weight gain of 10.0 kg after menopause***</td>
<td>1.18 (1.03 – 1.35)</td>
</tr>
</tbody>
</table>

*In women aged 40 to 49
** Women less than 70 years of age
***Post-menopausal women aged 30 to 55

Breast Density and Legislation

In 2008, a breast cancer survivor founded Are You Dense, Inc. and Are You Dense Advocacy, Incorporated, organizations whose mission is to raise breast density awareness. Her experience with self-reported, non-screen detected, advanced stage breast cancer was the impetus for organizational establishment. The advocates describe breast density as the “best kept secret”, referencing the inaction of the medical community. Although breast density has long been recognized as one of the multiple risk factors for breast cancer and included in reports to physicians, its integration into clinical decision-making has been difficult and problematic. Despite some hesitance of health care professionals, patient advocates have helped to foster a legislative movement, encouraging action on both state and federal levels.

As of July 2014, nineteen states have enacted breast density notification laws requiring radiologists to report breast density in routine patient summaries. The specific requirements outlined by the laws vary by state. Some require that lay patient
summaries inform all women about their breast composition while others require that only a subset be notified. In North Carolina, women who are categorized as “heterogeneously dense” or “extremely dense” must receive the following statement by law.

“We are providing this information to raise your awareness of this important factor and to encourage you to talk with your physician about this and other breast cancer risk factors. Together, you can decide which screening options are right for you. A report of your results was sent to your physician.”

Breast density notification regulations have also reached the federal arena. Congresswoman Rosa DeLauro of Connecticut introduced the Breast Density and Mammography Reporting Act of 2013 (H.R. 3404). This federal law would require reports to include whether a woman is at, above or below average breast cancer risk according to parenchymal pattern. In addition, the report would state that women with extremely dense breast tissue composition, about 10 percent of the general population, might benefit from supplemental screening. The most recent action was referral to the Subcommittee on Health.

Breast Cancer Screening Recommendations

Organizations, such as the US Preventive Services Task Force (USPSTF) and American Congress of Obstetricians and Gynecologist (ACOG), have taken differing courses of action in response to including breast density in guidance statements. In
In 2009, the USPSTF recommended regular, biennial screening mammography for women aged 50 to 74.\textsuperscript{15} They gave screening women aged 40 to 49 a Grade C recommendation, which means that it can be selectively offered based on professional judgments and patient preferences. In regards to breast density, they stated that it was unclear whether additional screening of women with dense breast tissue would reduce mortality. A need for better understanding of breast tumor biology that predisposes women to fast growing cancers and mortality was also noted.

While the ACOG recommendation deviates from the USPSTF in that its screening recommendation starts at age 40, their approach to breast density is similar. In a recent Committee Opinion, crafted specifically to address management of women with dense breast tissue, ACOG recommended against any alternative or adjunctive tests, such as breast ultrasound, to screening mammography. Research needs are also highlighted since there is very limited data for supplementary screening.\textsuperscript{16} The American College of Radiology Imaging Network (ACRIN) sponsored a clinical trial investigating the diagnostic yield and diagnostic accuracy of combined mammography and ultrasound screening versus mammography alone in women with heterogeneously dense breast tissue in at least one quadrant. Investigators found that the addition of ultrasound to mammography would yield 1 to 7 more breast cancer diagnoses per 1000 women, but it would also substantially increase the rate of false positives. The false positive rate was about 10\% percent for the combined approach and only 4\% for mammography alone.\textsuperscript{17} In summary, research gaps around breast density exist and efforts to add supplementary screening have demonstrated to be of poor overall benefit.

The American College of Radiology (ACR) released a report in 2012 cautioning breast density description in lay patient summaries; however, they support provider notification. They highlight several important considerations. One, the assessment of breast density is not reliably reproducible among interpreting radiologists, which may
lead to confusion and the perception that mammography is an unreliable imaging technique. Second, it would lead to increased demand for supplemental imaging techniques, such as ultrasound (US) and magnetic resonance imaging (MRI). The utilization of these techniques may lead to more false positive test results and subsequent benign breast biopsies. Non-mammographic screening also has not been demonstrated to reduce mortality. Lastly, informing patients might create undue anxiety and worry about breast cancer risk.

In this statement, the common tradeoffs associated with cancer screening are highlighted. On the benefit side, cancer screening reduces morbidity and mortality by detecting localized disease. On the harms side, screening is accompanied by false positive test results, over-diagnosis, over-treatment and associated psychological harms. While some advocacy groups and legislative bodies are ready to move forward with requiring actionable steps on the grounds of breast density, caution must be exercised as we explore what prevention strategy should be used based on epidemiology and scientific knowledge as screening is not without harm. The benefits of considering breast tissue composition in screening conditions are not clear. Its careless integration could increase the harms already noted.

**Screening Strategies**

Screening for breast cancer using mammography is an example of secondary prevention, detecting an asymptomatic “disease” state. Prevention can be further characterized as selective or population-based. In a selective prevention strategy, people at “higher risk”, on a continuum, than what is deemed acceptable are identified and targeted for intervention. In a population-based strategy, interventions are created to mitigate the determinants of incident disease in the general population and not just for certain individuals. High-risk screening strategies can be viewed on a spectrum from moderately selective to very selective. The recommendations from the USPSTF and
ACOG, for example are all high-risk screening strategies because they target women of a certain age. However, including risk factors other than age into intent-to-screen decision makes the strategy even more selective.

Both risk-based and population-based strategies have advantages and disadvantages. Advantages of targeting prevention to higher risk individuals are that there is a more favorable benefit-to-risk ratio in comparison to population-based strategies; it effectively and efficiently allocates limited resources providing cost-effective care and it motivates both patients and providers to participate. While risk-based strategies have their advantages, their applicability is limited by the ability to appropriately identify and demarcate the highly susceptible individuals from those at lower risk.

Challenging aspects of risk-based prevention strategies are that some risk factors, such as breast density and age, are not stable over a lifetime. In addition, established cutoff values may exclude borderline patients who will actually go on to develop cancer. Lastly, identifying higher risk individuals alone, on the basis of breast density, is palliative and temporary in that it identifies those that may go on to develop breast cancer. There are strategies that go further to modify risk factors after identification. In this case, if breast density was found to be predictive of breast cancer outcomes, chemoprevention could be used to modify breast tissue, decreasing a woman's susceptibility to the outcome.

Health care professionals and policymakers are tasked with selecting an appropriate strategy for breast cancer prevention. When selecting an approach, one should consider the "burden of suffering" and the "preventable burden". The burden of suffering includes the population attributable risk proportion, the amount of disease attributable to a risk factor within a population, and heuristics and biases that influence patients' risk perceptions. In this case, the population attributable risk proportion would
be the proportion of incident breast cancer cases or breast cancer-related mortality due
to dense breasts. If this proportion were high, then a strategy that selects for this
population would be helpful. A risk-based strategy is preferred over a population-based
strategy when the “cases” are concentrated in a group of individuals with an identifiable
risk factor. This concept is further demonstrated in Figure 1.

**Figure 1. Prevention strategy to relieve “burden of suffering”**

In Figure 1b, all of the “cases” are concentrated in one sub-group D, while in Figure 1a;
the “cases” are evenly distributed across the entire population. A strategy that targets
group D in Figure 1b. would be highly effective if the majority of “cases” were
concentrated in that subgroup.

The concept of preventable burden incorporates the screening strategy
effectiveness, in this case mammography. While screening mammography has reduced
breast cancer mortality over the years, it is not without limitations.\(^{15, 19, 20}\) Welch and
Passow sought to quantify the benefits and harms of mammography screening to help
women make decisions about whether to be screened or not screened.\(^{21}\) In their study,
they focused on women in three different age groups: 40 to 49, 50 to 59 and 60 to 69
years of age. They quantified the following outcomes: number of breast cancer-related
deaths, number of false-positive recalls, number of false positive biopsy
recommendations and number of over-diagnosed breast cancers. Annual screening was
used in the model because of common clinical practice though conflicting with USPSTF recommended screening strategy.\textsuperscript{15,21} They found that as age increases more women will avoid dying from breast cancer and fewer will undergo biopsies because of a false positive recall. Young women are more often faced with screening harms. One could infer from this study that using an age-based risk strategy in a certain age subgroup to target mammography would relieve the burden of suffering – fewer women would die and fewer would undergo burdensome biopsies. Using age alone to target screening is not perfect. Some women younger than the USPSTF recommended screening age of 50 will develop life-threatening advanced stage breast cancer. One of the challenges for prevention is uncovering the ideal combination of risk factors that can be used to target interventions.

In addition to harms, the cost of screening mammography creates a moment of pause. In the US, actual screening practice of women aged 40 to 85 years old costs approximately 7.8 billion dollars per year.\textsuperscript{22} O’Donaghue et al. developed models of various screening strategies to estimate the actual cost of screening in 2010 and to compare it to existing screening recommendations. The USPSTF screening strategy, which included biennial screening for women 50 to 74 years of age, personalized screening of high-risk women aged 40 to 49 and screening for women 75 to 85 years old with few comorbidities; costs less than actual practice. The estimated cost was 3.5 billions dollars per year. The authors highlight that the Patient Protection and Affordable Care Act of 2010 (ACA) is likely to increase screening mammography rates, which further emphasizes the need to curtail spending and be more cost-conscious, spending finite financial resources on necessary and beneficial health care.

**Potential for Breast Density**

A high-risk screening strategy would be a desirable approach to breast cancer screening because it is cost-effective, efficient and would relieve the burden of suffering
while minimizing harm to the low-risk, general population. It would be valuable if breast density would allow for the narrowing of our focus on a smaller population to screen at regular intervals or a personalized plan for every woman based on the risk she carries. There are several relevant questions – Would targeted screening of individuals with dense breast tissue relieve the burden of suffering? At what interval and with what imaging modality, should a 42-year-old woman with ‘extremely dense’ breast tissue be screened? Should a 52-year-old woman with less dense, ‘almost entirely fatty’ breast tissue composition be screened at all? These are complex questions, but the first step is discerning the ability of breast density to discriminate between women who will and will not be diagnosed with breast cancer. This review compares the predictive ability of breast tissue composition to age to evaluate whether it adds incremental value to cancer risk prediction and if the evidence permits personalizing or targeting screening strategies.

**Background**

In this section, I will provide the background information necessary to interpret the findings and significance of this systematic review. If breast tissue composition is used as a method to target higher risk individuals for screening, we must understand how breast density is determined and how to evaluate risk factors as predictive tools.

**History of Mammographic Breast Density**

A mammogram is an x-ray image of the breast. Breast connective and epithelial tissues attenuate x-rays creating a radio-opaque image. This is in contrast to adipose tissue, which appears radiolucent on mammogram. Mammographic breast density refers to the characteristics the opaque areas. Dr. John Wolfe was the first to introduce the relationship between breast parenchymal patterns and subsequent incident cancer. In a prospective cohort of 5284 women 30 years or older with a median follow-up time of
2.5 years, he found that there was a positive linear relationship between mammographic breast density and breast cancer risk. He divided breast parenchymal patterns into four categories: N1, P1, P2 and DY. The N1 pattern was considered “normal”. He did, however, note that normal varied with age. He stated that in younger women, the mammogram would appear trabeculated because connective tissue transverses the fatty breast. In older women, normal would be almost completely fat. DY denotes the most dense parenchymal pattern. Women classified as N1 had a breast cancer incidence lower than those classified as DY.

Since this study, several investigators have published findings demonstrating a similar relationship.\textsuperscript{24} However, the biologic mechanism linking mammographic breast density to increased breast cancer risk isn’t completely resolved. Masking bias, the concealing of non-calcified lesions by dense breast tissue, is part of the story. Boyd et al in three nested case-control studies, found that the mammographic breast density and risk for breast cancer persisted regardless of whether the cancer was screen detected or identified by other means.\textsuperscript{25} This demonstrates that there are likely other pathways other than screening failure to explain the relationship.

**Mammography Quality Standards Act and Program**

The Mammography Quality Standards Act of 1992 authorizes the Food and Drug Administration (FDA), under its regulation of radiation-emitting products, to ensure that facilities perform high-quality mammograms. The MQSA recognizes the ACR as an accrediting body and thus supports their standard mammography reporting recommendations.\textsuperscript{26} The 5\textsuperscript{th} edition of the BI-RADS Atlas provides standards for reporting results of breast mammography, magnetic resonance imaging (MRI) and ultrasound (US). The BI-RADS lexicon was created after widespread use of screening mammography accentuated marked variability in radiation dose, image quality and processing speed.\textsuperscript{27} Also of issue were diagnostic inconsistencies and reporting.\textsuperscript{19} The
resource establishes terminology, report organization, classification systems and 
assessment and management recommendations. The specifications are important for 
patient satisfaction, clinical decision-making, outcome monitoring and quality assurance.

A mammography report must include the indication, description of breast 
composition, important findings, statement of comparison to previous examinations, final 
assessment and management recommendation. There are seven final assessment 
categories, ranging from “Category 0: Incomplete” to “Category 6: Know Biopsy-Proven 
Malignancy”. Each assessment category is linked to a management recommendation 
and a statement of likelihood of cancer. For example, a “Category 4A: Low suspicion for 
malignancy” has a 2 to 10% likelihood of malignancy and tissue diagnosis is 
recommended. While the BI-RADS lexicon has been in practice for over 20 years, the 
breast composition categorization was added in 2002.

There are four breast composition categories, which are listed in Table 2. The 
most recent, fifth, edition of the BI-RADS Atlas published in 2013 deviates from previous 
editions in that the organization abandoned quartile ranges for percentage dense tissue. 
The Committee on BI-RADS wanted to emphasize the text descriptions of density, as 
they are clinically more important than percentage dense area when considering the 
sensitivity of mammography. Dense breast tissue that masks non-calcified lesions is 
more meaningful than a relatively less dense tissue that occupies a larger area. There 
are other qualitative and quantitative assessments of breast density. Other qualitative 
assessments include Wolfe’s parenchymal patterns, and Tabar and Boyd classifications. 
The qualitative assessments are subjective, relying on the interpreting radiologist. 
Quantitative, computer-assisted measurements include percentage mammographic 
density based on digitized images, and volumetric measurements. While quantitative 
assessment software is more objective, it is expensive technology with few models that 
are readily, commercially available.
While the BI-RADS lexicon is useful in standardizing reporting, it does have limitations. Given that the breast composition categories are qualitatively assessed and assigned independently by radiologist, there is limited inter-observer agreement.\textsuperscript{28} There is substantial agreement between repeat breast density assessments by the same radiologist using the BI-RADS lexicon ($\kappa$: 0.72 95% CI: 0.66 – 0.78).\textsuperscript{29,30} Moderate to low agreement is noted between radiologists interpreting screening mammography images. Agreement at the extremes of breast composition, “almost entirely fat” and “extremely dense” is also higher. This is unfortunate given that the middle categories, “scattered fibroglandular densities” and “heterogeneously dense” make up the highest proportion of breast composition in the general population, Table 1.

Table 2. Breast tissue composition categories and frequency distribution\textsuperscript{31}

<table>
<thead>
<tr>
<th>Breast Composition Categories</th>
<th>Percent of General Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. The breasts are almost entirely fatty</td>
<td>10%</td>
</tr>
<tr>
<td>b. There are scattered areas of fibroglandular density</td>
<td>40%</td>
</tr>
<tr>
<td>c. The breasts are heterogeneously dense, which may obscure small masses</td>
<td>40%</td>
</tr>
<tr>
<td>d. The breasts are extremely dense, which lowers the sensitivity of mammography</td>
<td>10%</td>
</tr>
</tbody>
</table>

Risk Factors as Predictive Tools

In medicine, clinicians gather evidence to estimate risk, disease likelihood, and prognosis. Evidence comes from sources, such as patient anecdote, medical history, physical exam and diagnostic test results. Predictive tools in practice integrate the truth; observed disease outcomes, with the biological and environmental factors that give rise to disease into statistical models. Prediction is important to both prevention and treatment, and allows for intervention targeting to higher risk individuals and avoiding harm to those likely to receive little benefit.
Cancer risk prediction models in breast cancer include known risk factors for the outcome. Risk factors that are most suitable for prediction models are those that have a strong association and exhibit a dose-response relationship with the outcome.\textsuperscript{32} Risk factors that reflect determinism based on pathophysiology rather than the factor preceding the outcome, or temporal determinism, are preferred.\textsuperscript{33} Chosen risk factors are strongly associated and often exhibit a dose-response relationship with the outcome.

Existing breast cancer prediction models have incorporated some of these risk factors. The Breast Cancer Risk Assessment Tool, also known as the Gail Model, incorporates personal history of breast cancer or radiation to the chest, genetic history, age, age at menarche, age at first live birth of child, number of first-degree relatives that have had breast cancer, personal history of breast biopsies and notable results, and race/ethnicity. Studies have also explored adding breast density to the prediction model, more specifically, assessing whether its addition adds incremental value.

Predictive ability moves far beyond the strength of association and relationship characterization. A good prediction model is internally valid, produces the right answers for the source population, provides clinically important discrimination and is generalizable. The key aspects used to evaluate predictive tools are the magnitude and certainty of risk score, discrimination, calibration, and net reclassification. The criteria and their definitions are outlined in Table 2.\textsuperscript{33-36} The actual use of a prediction tool in everyday practice is contingent on its clinical utility and cost-effectiveness, also defined in Table 2. The incremental value that a risk factor adds to an existing model is based on all of these factors.

Evaluating all facets of predictive ability is necessary to minimize harm and to make output data easily interpretable. In addition, one must be mindful that risk factors have their limitations as prognostic tools. For example, to see meaningful, incremental change in discrimination, very large measures of association are required. To easily
create cutoffs, the distribution of breast cancer in those with and without dense breast tissue needs to be sufficiently separated. The deleterious BRCA1 and BRCA2 mutations that cause hereditary breast and ovarian cancer are examples of risk factors associated with substantial risk.\textsuperscript{15,37} It would be valuable to identify risk factors or a combination of risk factors that would amount to a very strong association with breast cancer incidence and/or mortality.

**Table 3. Criteria for evaluating prediction tools**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td>The ability of a tool to accurately predict the observed risk level, whether the predicted outcome is equal to the observed outcome. The statistical test that is used to assess for calibration is the Hosmer-Lemeshow ( \chi^2 ) test.</td>
</tr>
<tr>
<td>Clinical utility</td>
<td>In a setting of efficacious uptake and implementation, clinical utility refers to the ability of a screening test to prevent adverse health outcomes. This concept integrates implementation science.</td>
</tr>
<tr>
<td>Discrimination</td>
<td>The probability that a randomly selected patient who develops breast cancer (“a case”) will have a higher risk score than a randomly selected non-case. It is described using the concordance (c) statistic and ranges from 0.5 to 1.0. In a model that discriminates well, there is little overlap between cases and non-cases. A c statistic of 1.0 indicates perfect discrimination.</td>
</tr>
<tr>
<td>Incremental cost-effectiveness</td>
<td>The incremental difference in costs and health outcomes between an intervention and an alternative that competes for the same resources. The cost is multi-factorial and can include program and productivity costs for example.</td>
</tr>
<tr>
<td>Net Reclassification Improvement</td>
<td>Reclassification refers to a change in risk category that leads to a new action or end to an old action. Net reclassification improvement would be the appropriate movement of one risk category to another. The movement of a “case” to a higher risk stratum would be appropriate, while movement down an inappropriate or poor reclassification. Net Reclassification Improvement (NRI) is the sum of the difference in proportions of cases moving up minus those moving down, and the proportion of non-cases moving down minus those moving up.</td>
</tr>
<tr>
<td>Risk Score</td>
<td>The absolute risk of disease over a specified time interval.</td>
</tr>
</tbody>
</table>

**Methods:**

**Key Questions**
In this review, I assessed whether the ACR BI-RADS breast composition categorization had ability as a predictive tool. Specifically, whether it had superior ability over age to predict incidence and/or breast cancer-related mortality in average-risk women aged 40 to 59. The following key questions guided the review. Table 3 contains a PICOTSSS table with inclusion and exclusion criteria.

(1) What are the age-adjusted relative risk and/or odds ratio of in situ and/or invasive breast cancer according to breast tissue composition category? How does that risk compare to age alone?

(2) How many incident cancers per 100,000 women ages 40 to 49 are diagnosed within 5 or 10 years of indexed mammogram?

(3) How many incident cancers per 100,000 women ages 50 to 59 are diagnosed within 5 or 10 years of indexed mammogram?

(4) Does breast composition categorization offer clear advantage in discriminatory power over age alone?

(5) Using the risk estimate for an average-risk 50-year-old woman as the threshold, does age-adjusted or age-specific breast composition lead to net reclassification improvement?

(6) What is the clinical utility of BI-RADS breast composition categories as a breast cancer predictive tool?

(7) Is it cost-effective to screen women ages 40 to 49 with “heterogeneously dense” or “extremely dense” breasts as classified by BI-RADS, compared with no screening?
Table 4. PICOTSS description for systematic review

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population</strong></td>
<td>Average-risk women aged 40 to 59.</td>
</tr>
<tr>
<td></td>
<td>Men; women with a personal history of in situ or invasive breast cancer, chest irradiation, benign proliferative changes, atypia or breast augmentation; first-degree family history of breast cancer; BRCA 1 and BRCA 2 mutation carriers; women less than 40 and greater than 60 years of age.</td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td>Breast tissue composition as assessed by a radiologist using film or digital mammography and described using ACR BI-RADS breast tissue composition categories.</td>
</tr>
<tr>
<td></td>
<td>Breast density as assessed by other imaging technologies, such as MRI or ultrasound, and automated or computer-assisted methodologies; quantitative measurements of breast density, such as percentage dense area and percentage non-dense area.</td>
</tr>
<tr>
<td><strong>Comparison</strong></td>
<td>Age as a predictive tool.</td>
</tr>
<tr>
<td></td>
<td>Risk predictions based on the following risk factors alone or in combination: first-degree family history, reproductive factors: parity, age at menarche, menopausal status and age at menopause; BMI, benign breast pathology, bone mineral density and lifestyle factors: alcohol consumption and smoking history.</td>
</tr>
<tr>
<td><strong>Outcome</strong></td>
<td>Measures of effect: absolute risk, risk ratio or odds ratio.</td>
</tr>
<tr>
<td></td>
<td>Number of incident in situ and/or invasive breast cancer cases within 5 years and/or 10 years per 100,000 women.</td>
</tr>
<tr>
<td></td>
<td>Breast-cancer related mortality within 10 years of indexed mammogram.</td>
</tr>
<tr>
<td></td>
<td>Lifetime risk.</td>
</tr>
<tr>
<td></td>
<td>Incidence 10 years after indexed mammogram.</td>
</tr>
<tr>
<td></td>
<td>Mortality 10 years after indexed mammogram and/or mortality due to causes other than breast cancer.</td>
</tr>
<tr>
<td><strong>Time for outcome</strong></td>
<td>10 years</td>
</tr>
<tr>
<td></td>
<td>After 10 years</td>
</tr>
<tr>
<td><strong>Literature Search</strong></td>
<td>2003 to present</td>
</tr>
<tr>
<td></td>
<td>Before 2003</td>
</tr>
<tr>
<td><strong>Study Designs</strong></td>
<td>Cohort studies.</td>
</tr>
<tr>
<td></td>
<td>Case-control studies.</td>
</tr>
<tr>
<td></td>
<td>Modeling studies.</td>
</tr>
<tr>
<td></td>
<td>Meta-analysis/Systematic Reviews.</td>
</tr>
<tr>
<td></td>
<td>Randomized controlled trials.</td>
</tr>
<tr>
<td></td>
<td>Cross-sectional studies.</td>
</tr>
<tr>
<td></td>
<td>Case reports.</td>
</tr>
<tr>
<td></td>
<td>Case series.</td>
</tr>
</tbody>
</table>
The target population for this systematic review was women aged 40 to 59 given the proximity to screening age and heterogeneity in breast tissue composition. Men were excluded from this review, as well as, other high-risk groups, such as those with a personal history of breast cancer, a history of chest irradiation, a significant first-degree family history of breast cancer and BRCA 1 and BRCA 2 mutation carriers. These high-risk populations were excluded from the review because prevention strategies will be different than those implemented for the general population.

There are several methods of assessing breast tissue composition to include subjective assessments by radiologist, objective computerized assessments that estimate the percentage of dense area of the total area, volumetric measurements and other computer-assisted methodologies. Currently, automated methods are an active area of research given the need for equal, valid and reliable measurements that remove the subjectivity of the radiologist. For this review, I focused on the BI-RADS reporting system given its widespread and routine use in the US. I excluded computerized and automated methods of breast density measurement because they are in limited use at the population level currently.

The outcomes of interest included effect measures, such as the risk ratio (RR) and odds ratio (OR), absolute risk increase above age-specific risk and breast-cancer related mortality. The number of incident in situ and/or invasive breast cancers with 5 or 10 years of indexed mammogram was another desired outcome. I wanted to quantify the number of cancers per 100,00 women ages 40 to 49, before they reach the USPSTF recommended screening age of 50 years old. In those that older than 50, I wanted to investigate how their risk changed after taking into account their breast tissue composition. I excluded lifetime risk, and incidence and mortality after 10 years to focus on 10-year outcomes that might influence current guidelines and screening decision-making. In comparison to cumulative risk estimates calculated in the short-term,
competing risk can confound the relationship between exposure and lifetime risk estimate.

The literature search begins in 2003 when the ACR published its fourth edition of BI-RADS Atlas. Prior editions used quartile ranges for subjective amount of dense breast tissue that was observed on mammogram. This edition included text descriptors, such as ‘almost entirely fat’ or ‘extremely dense’. The Committee on BI-RADS made this decision because they acknowledged that the breast tissue composition’s masking ability, causing a reduction in the sensitivity of mammography, was clinically more important than the amount of area covered.

The study designs of interest include cohort studies, case-control studies, modeling studies, systematic reviews, randomized controlled trials and meta-analysis. Cross-sectional were not included in this review given that by design they lack longitudinal data. Case reports and case series have been excluded because they are low in the hierarchy of research design.

**Data Sources and Search Strategy**

I searched PubMed and EMBASE from January 1, 2003 to April 19, 2014 to identify potentially relevant articles to address the key questions. For PubMed, I used a combination of MeSH terms and keywords to capture relevant literature. I did not apply limits to the search so that recent, non-indexed articles would not be excluded. I adapted the search strategy for PubMed for utilization in EMBASE. For all of the articles that met eligibility criteria for full-text review, I hand searched reference lists to identify missed articles and to corroborate the initial search strategy. I consulted two research librarians to formulate the most appropriate search strategy. Appendix A contained the detailed search terms used for the various online databases.

**Study Selection**
I reviewed titles and abstracts that were retrieved from the initial online search to determine whether an article was admissible based on inclusion and exclusion criteria. I then reviewed the full-text of the articles to determine final eligibility for inclusion in systematic review. For this paper, I alone reviewed articles to determine eligibility.

**Data Collection and Quality Assessment**

For all of the articles that met the eligibility criteria, I extracted pertinent data and assessed each study’s quality. Dual review was not utilized for this paper given time and resource limitations. I used a common template to extract data from articles. Data of interest included study design, registry source, sample size, study population age range and effect measures. Effect measures included odds ratios, relative risks and hazard ratios.

To evaluate individual studies’ internal validity, I used a list of quality criteria, adapted from the USPSTF Procedure Manual for each type of included study design. For the cohort studies, internal validity was assessed based on methodology of selecting participants, the comparability of selected groups, whether there was important differential loss to follow-up, their explanation and measurement of both the exposure and outcome. For case-control studies, internal validity was based on the methodology of identifying cases and selecting controls and whether investigators paid appropriate attention to potential confounding variables. For systematic reviews and meta-analyses, I paid attention to whether the search strategy and study selection was comprehensive and systematic, whether appraisal of the individual studies included a quality assessment and, lastly, if in their conclusion, they elaborate on both the certainty and magnitude of the conclusion after evidence synthesis. For each study, I made a subjective, overall assessment of the risk of bias taking into account all of the aforementioned factors and assigned a rating of good, fair or poor for internal and external validity.
Data Synthesis and Analysis

After assessing the quality of individual studies, I qualitatively synthesized the results. I did not use quantitative statistical methods because of time and resource limitations. The USPSTF Procedure Manual was referenced again to guide the assessment of strength of the overall body of evidence. To evaluate the evidence, important considerations were the hierarchy of research design, the quantity of good quality studies, generalizability to the U.S. primary care setting, number of study subjects’ and consistency of results. Based on these factors, I assigned a subjective rating of convincing, adequate, or inadequate to the entire body of evidence. Biases that affect the cumulative evidence, such as publication bias and selective reporting, were also considered.

Results

The search strategy identified 544 unduplicated references. I excluded articles after reviewing titles and abstracts if they did not meet the eligibility criteria. Common reasons for exclusion included not including target population, using quantitative measurements, such as percentage dense area, to measure breast density, using computerized and/or automated tools to measure breast density, using imaging technologies other than film or digital mammography and simply not being relevant to key questions. I selected 25 articles for full-text review and excluded conference abstracts and studies using quantitative measurements. Eleven articles remained for inclusion in this systematic review. Appendix B contains a flow chart documenting the process from initial search strategy to inclusion in the systematic review.

The study designs included prospective cohort studies, case-control studies, a nested case-control study, two modeling studies and a meta-analysis. The age range captured by the data was 25 years and older. A minority of studies used an upper age
limit for exclusion. Some studies used subgroup analysis, calculating outcomes for specific age ranges within overall study population. Others used menopausal status to stratify results. The frequency distribution of breast tissue composition in all of the studies mirrored that of the general population.

The measurement of the exposure, BI-RADS breast tissue composition, was similar across the majority of the studies. The majority of the studies used breast density assessments that were performed in routine clinical practice across the various radiology facilities using the ACR BI-RADS lexicon despite the recognized problems with observer agreement. One study utilized one observer and another study utilized double blind readings. The application of the BI-RADS lexicon is inherently not equally applied or reliable because it is subjective. There were some differences across the studies in which mammography report was used if more than one mammography report was available. Some studies included the first mammogram documented in the study period, while others used the mammogram closest to the diagnosis. The majority of studies included only breast density assessments associated with screening mammograms, while a minority included assessments from diagnostic mammograms. Investigator obtained information related to patient’s medial history by self-report using questionnaires. The relevant health history was similar across all studies and included age at birth of first child, age at menarche, first-degree family history of breast cancer, menopausal status, use of hormone replacement and personal history of breast cancer, biopsy or augmentation.

The majority of studies included DCIS and invasive breast cancers together as the outcome, while a minority focused on invasive cancer alone (N=3) or DCIS (N=1) alone. There was also variation in whether investigators used only screen-detected cancers or a composite of screen-detected and interval-cancers. The duration of follow-up after indexed mammogram ranged from 1 to 8 years.
A common feature of all included studies was cancer registry utilization to obtain outcome data. The cancer registries included hospital-based registries and purportedly population-based registries. Only one study, the Mayo Mammography Health Study, used a hospital-based registry. The population-based registries included state tumor registries, SEER and some or all of the registry sites within the Breast Cancer Surveillance Consortium (BCSC). The BCSC, a program of the National Cancer Institutes Division of Cancer Control and Population Sciences, was established in the early 1990s. There are seven BCSC sites – Carolina Mammography Registry, New Hampshire Mammography Network, Vermont Breast Cancer Surveillance System, Colorado Mammography Project, Group Health, San Francisco Mammography Registry and New Mexico Mammography Project. The BCSC registry sites collect mammography examination data from radiology facilities, medical history from patient questionnaires and cancer outcomes data from state tumor registries, SEER and pathology facilities.

I was not able to easily obtain information about the geographic catchment areas and baseline characteristics for each site, however, information about the BCSC overall was available. Between 1994 and 2009, the BCSC collected data from 296 radiology facilities across their seven sites. More than 90% of these facilities were associated with academic medical centers. Using 2000 census data to compare BCSC counties to all other US counties, women living in BCSC catchment counties had a higher median family income, had a higher proportion of individuals with high school degrees and had a lower proportion of Hispanics and Blacks. The number of participating radiology facilities and their catchment area varies in each state. Few of the studies included in this systematic review provided details about the proportion of the state population captured by the registry, baseline characteristics of captured population in comparison to others in the geographic area and the registries quality; quality being evaluated by the
completeness of the data, the timeliness of outcome reporting and the accuracy of recorded data.

The majority of studies also used a passive methodology for obtaining follow-up, using only databases to obtain cancer incidence and mortality data. Only one study used active follow-up, attempting to contact patients by mail or telephone, to confirm cancer or cancer-free status, vital status or movement outside of registry catchment area.

The overall body of evidence was convincing, but many of the studies had fair internal and external validity, and lack of precision of estimates in women with ‘extremely dense’ breast tissue composition. Risk of bias was present due to the the use of a subjective measure for exposure, passive follow-up methods for outcome assessment and use of registry sites whose individual quality cannot be easily ascertained. There was substantial heterogeneity in the age of women included and whether subgroup analysis. Therefore, it is difficult to make conclusions about the estimates for women aged 40 to 59. Some studies include age subgroups, demarcated at different intervals, and others divide population according to menopausal status.

The key questions addressed by the literature were those related to age-adjusted risk of breast cancer according to breast density and discriminatory ability. There were no studies that addressed net reclassification ability, clinical utility and cost-effectiveness of screening on the grounds of BI-RADS breast tissue composition. The risk estimates were small-to-moderate in magnitude and exhibited a consistent dose-response gradient from ‘almost entirely fat’ to ‘extremely dense’. In all of the studies, women with ‘extremely dense’ breast tissue composition represented one of the smallest samples in the overall study population creating less precise estimates, which was indicated by wider confidence intervals. There was also heterogeneity in how studies grouped exposure status, some combining ‘heterogeneously dense’ with ‘extremely dense’ for
example leading to over-estimation or under-estimation of effect for women in different breast tissue composition categories.

In the studies that calculated a c-statistic, the value was around 0.6, slightly above the value of chance, 0.5. One study reported that this value was statistically significant and greater than that of the Breast Cancer Risk Assessment Tool. They did not prove clinical signification, however. Appendix C, Tables 1 – 22 contains data extraction and appraisal of each individual study.

Gierach et al. 2012

The study aim was to evaluate the relationship between mammographic breast density and risk of breast cancer-related death. The study participants were from five of the seven BCSC sites, chosen because they routinely collecting body mass index (BMI) data. The study only included women who were diagnosed with primary invasive cancer. The majority of the population was non-Hispanic, White and post-menopausal (74.6%). The average age was 59. The mean-follow-up time from index mammogram to diagnosis was about 7 years. In this study, high mammographic breast density was not associated with breast cancer-related death. In their fully adjusted model, the hazard ratio of breast cancer-related death in those with ‘extremely dense’ breast tissue composition was 0.92 (95% CI: 0.71 – 1.19) when compared those with ‘scattered fibroglandular densities’. Women with ‘almost entirely fat’ breast tissue composition actually had a higher risk of death; the hazard ratio was 1.36 (95% CI: 1.04 – 1.77). BMI modified the relationship between breast density and death. The authors hypothesize that risk factors for breast cancer-related death may not be the same as those influencing the development of breast cancer.

The internal validity of this study is threatened by small sample size, measurement bias and inappropriate attention to important confounders. State vital records and cancer registries were used to measure vital status. The use of death
certificates is subject to ascertainment bias as physicians commonly confuse underlying case of death with the mechanism of death.\textsuperscript{40} There is no detail about how the cause of death is documented in these sources, and the accuracy and timeliness of the reporting. In addition, participants were assumed alive if there was no documentation of death in these sources. There could have been more deaths than what was documented. Other potential confounders that should have been measured include comorbidity and other socioeconomic variables, such as race, education and health insurance status. This study received a poor rating for internal and external validity.

\textit{Kerlikowske et al. 2007}\textsuperscript{41}

The study objective was to determine whether a change in mammographic breast density over time was associated with breast cancer risk. The stated rationale for this study was based on knowledge that breast density is modifiable and changes over time. Specifically, breast density changes with age, body mass index (BMI), menopause and other reproductive factors. Given that exposure of interest was change in density over time, study participants had to have at least two screening mammograms at least 9 months apart. The median time between screening examinations was 3.2 years. For this systematic review, only breast density measurements and outcomes associated with first screening mammography were extracted. In addition, in the stratified analysis by age, only outcomes associated with stable breast density measurements were extracted so that systematic review results would be consistent.

Cancer outcomes included both invasive and in situ breast cancers, but the majority of cancers (80%) were invasive. In women with ‘extremely dense’ breast tissue composition documented on the first screening mammogram, the odds ratio of in situ invasive breast cancer was 1.75 (95\% CI: 1.52 – 2.00) as compared to those with “scattered fibroglandular densities’. For women with “almost entirely fat” composition the odds ratio was 0.63 (95\% CI: 0.54 – 0.74). An increase in breast density over time was
associated with a higher breast cancer risk in women with “almost entirely fat”, ‘scattered fibroglandular densities’ and ‘heterogeneously dense’ breast density as compared to those with assessments that remained unchanged. For example, a woman with ‘almost entirely fat’ on first screen and ‘scattered fibroglandular densities’ on last screen had an adjusted rate of breast cancer per 1000 women of 9.9 (95% CI: 6.4 – 15.5).

This study was rated as having good internal validity and fair external validity. An aspect of their methods that is unique is adjustment for misclassification of breast tissue composition using the Reade-Christopher and Kupper method. Statistical analysis was performed with and without adjustment for misclassification. Given that results were unchanged, final reported values were actually unadjusted. The external validity was limited for reasons similar to others in this study, demographics and inclusion of radiology facilities affiliated with academic medical centers.

Kerlikowske et al. 2010

Kerlikowske et al. aimed to report whether the association between breast tissue composition and breast cancer risk and severity differed according to menopausal status and post-menopausal hormone use. The average age of the study population was 56 years old. They found that the magnitude of the association between ‘extremely dense’ breast tissue and breast cancer risk was stronger in pre-menopausal women. For women classified as having ‘extremely dense’ breasts, the hazard ratio (HR) of breast cancer was 2.04 (95% CI: 1.84 – 2.25) using ‘scattered fibroglandular densities’ as the referent category. The 5-year risk of breast cancer for a pre-menopausal woman between the ages of 45 and 59 was estimated to be 2.1 (95% CI: 1.9 to 2.2). There was not a meaningful change in magnitude of association according to menopausal status for women with almost entirely fat breast tissue composition. Breast cancer risk in post-menopausal women with almost entirely fatty breasts was less than the referent category (HR = 0.57, 95% CI: 0.53, 0.62). The authors conclude that women with low
breast density have a low breast cancer risk regardless of age, menopausal status or post-menopausal hormone use. The management of the covariate BMI threatens the validity of the study. In their statistical analysis, the average BMI for the entire population was incorporated. The estimates may be inaccurate for women of different weight classes. One of the study’s strengths is the large sample size, as they pooled data from all of the existing BCSC sites.

*Mackenzie et al. 2007*⁴³

In this study, investigators examined the association between breast tissue compositions and risk of ductal carcinoma in situ (DCIS). The outcome DCIS is a heterogeneous disease, but was not further characterized as low-grade or high-grade for example. The rationale for this study was that there was little prior evidence examining this relationship, the majority of prior literature focusing solely on invasive breast cancer. In this prospective cohort study, following about 155,000 women in New Hampshire and Vermont over an average follow-up time of about 4 years, the relative risk (RR) of DCIS in pre-menopausal women with ‘extremely dense’ breast tissue was 2.4 (95% CI, 1.47 – 3.91) relative to women with ‘scattered fibroglandular density’ breast tissue composition. The relationship was not as strong in post-menopausal women (RR = 1.49, 95% CI: 0.93 – 2.37).

*Olson et al. 2012*⁴⁴

The Mayo Mammography Health Study (MMHS) was a study whose goal was to identify mammographic features that can be used to identify women at higher risk for breast cancer. The MMHS was a prospective cohort enrolling women from three states – Minnesota, Iowa and Wisconsin. In a nested case-control study, investigators sought to determine whether acquisition parameters were associated with mammographic density and whether the parameters confounded the relationship between breast density and incident breast cancer. The mammogram acquisition technique parameters included x-
ray tube voltage peak, milliamperes-seconds, thickness and compression force. The rationale for studying mammogram acquisition parameters was that the majority of prior studies of breast density and breast cancer risk have been conducted across multiple institutions using different image acquisition techniques. The relationship between mammographic breast density and breast cancer was confirmed. Women with extremely dense breast were about 3-times as likely to develop breast cancer than those with almost entirely fat (HR 2.96, 95% CI: 1.73 – 5.07). The studied acquisition parameters alone and in combination did not confound this relationship, as there was little to no change in calculated hazard ratios.

Tesic et al. 45

Tesic et al highlight that breast density varies by ethnicity and geography and use this as the justification for conducting this study in Croatia. The study aim was to determine whether the relationship between breast density and breast cancer risk held in a non-Western population. The Zagreb Breast Cancer Surveillance population-based registry in Croatia captures 22% of the population registry. The participating facilities include university hospitals, outpatient clinics and mobile mammography units. The study population consisted of post-menopausal women aged 50 to 69 with no personal history of breast cancer or breast augmentation. The majority of the study population (87%) had ‘almost entirely fat’ or ‘scattered fibroglandular densities’ breast tissue composition. Unique to this study, the exposure, breast tissue composition, was assessed by radiologist in routine practice using double blind readings. There was substantial intra-observer agreement and moderate inter-observer agreement. The absolute number of breast cancers diagnosed was highest in women with ‘scattered fibroglandular densities’. For analysis, investigators combined women with ‘heterogeneously dense’ and ‘extremely dense’ breast composition. Women in this group in situ and invasive breast cancer risk estimates higher than women classified as having
‘scattered fibroglandular densities’ (Odds ratio 1.9 (95% CI: 1.3 – 2.7). Per report, the risk estimate did not change substantially after adjusting for relevant covariates.

*Tice et al. 2005*[^46]

In this study, Tice et al. sought to determine if the addition of breast density to the Gail model had better predictive accuracy than the Gail model alone. The study population was small relative to other studies in this review (N = 81,777). The investigators only included women from the San Francisco Mammography Registry. They found that the addition of breast density to the model significantly increased model fit. There wasn’t much difference in the c-statistics however. The new model slightly increased the c-statistic, calculated to be 0.68 (95% CI: 0.66 – 0.7). The value for the existing Gail model is 0.67. A study strength that stands out in comparison to the other studies included in the review is the composition of the cohort. The cohort was racially and ethnically diverse. This makes the results more generalizable. Another factor that increases its external validity is the variety among the radiology facilities that participate in this registry site. The facilities include hospital-based clinic, private practices and a mobile mammography unit. Only one of the facilities is affiliated with an academic medical center, this stands out as overall most of the BCSC sites are affiliated with academic institutions.

*Tice et al. 2008*[^47]

This study expanded on the previous work by Tice et al. in 2005 that tested the accuracy of a new prediction model. In this study, using a larger population of women, they developed and validated a risk assessment tool that incorporates breast density. The two other risk factors that were incorporated as yes or no categorical variables were history of breast cancer in first-degree relative and history of breast biopsy. The model was well calibrated. The expected rate of invasive breast cancer according to the model
was 1.41%. The observed rate was 1.38%. Discrimination was statistically significant slightly better than that of the Gail model (c-statistic 0.660, 95% CI: 0.651 – 0.669).

*Vacek et al. 2004*48

The study aim was simple, to ascertain the breast cancer risk among women in different breast tissue composition categories. The study included only one of the registry sites in the BCSC, the Vermont Breast Cancer Surveillance System. This registry site collects outcome information from Vermont and New Hampshire to account for the fact that some VT residents actually receive health care in NH. This method reduces selection bias, increasing the proportion of individuals for which follow-up information is available. Unique to this study only, in comparison to others included in this review, is confirmation of breast cancer-free status. They did not assume that the absence of cancer documentation indicated that the patient did not have breast cancer. They reviewed results of exit mammograms and biopsies to check for disease. If neither of these resources were available, participants were excluded. The frequency distribution of breast tissue composition of the excluded women did not differ greatly from those included in the study. Relative to ‘almost entirely fat’ breast tissue, pre-menopausal women with ‘extremely dense’ breast tissue had an increased risk of invasive breast cancer (Risk ratio (RR) 4.6 (95% CI: 1.7 – 12.6).

*Zhang et al. 2012*49

This study was conducted using data from the Carolina Mammography Registry. The objective was to examine the relationship between breast tissue composition and breast cancer-related mortality according to age, race and cancer stage. Unique to this study design was the combination of breast tissue composition categories into smaller groups. Women with ‘almost entirely fat’ and ‘scattered fibroglandular densities’ breast tissue were combined into one group and the remaining in another. They found that there was no statistical difference in mortality for either group after adjusting for pertinent
covariates. The internal validity is threatened by loss to follow-up. The study is also not applicable to the general population as it was restricted to women of certain racial and ethnic groups.

There was one meta-analysis that was identified in the search and was a commonly found in the reference lists of included studies. This article was ultimately excluded from the review because it included studies utilizing quantitative measures and also captured some of the literature already included in this review. This meta-analysis incorporated studies that utilized both quantitative and subjective breast tissue composition assessments. The qualitative measures included the Wolfe classification, BI-RADS lexicon and Tabar classification. The quantitative measure included percentage breast density. The association between breast density and breast cancer risk was stronger using the quantitative measurement of breast density. Only 3 studies included in this analysis used the BI-RADS lexicon. In two of the studies, ‘scattered fibroglandular densities’ breast tissue composition was used as the referent category. The combined risk ratio of breast cancer was 4.08 for women with ‘extremely dense’ breast tissue. The meta-analysis only included literature published up to November 2005, threatening the validity of the conclusions for present day practice. The authors conduct a comprehensive literature search, but do not use organized, standard criteria for appraising included studies.

Discussion

The objective of this systematic review was to determine whether the ACR BI-RADS breast tissue classification system could be used as a predictive tool to identify women between the ages of 40 and 59 who are at a higher risk for early primary incident breast cancer and/or mortality. The value and ability as a predictive tool is based on fulfilling the following criteria: magnitude of risk estimates, discriminatory ability, net
reclassification index, clinical utility and cost-effectiveness. The identification of strong risk factors that can be used in risk prediction is an important goal because screening interventions can be targeted to populations more likely to develop the majority of breast cancer cases, thus minimizing harm to lower risk women and controlling health care cost.

**Significance of main findings**

After the search strategy, one finding that was notable was that most of the studies evaluating breast density and its relationship to breast cancer risk use quantitative measures, such as percentage dense area, rather than the routinely used BI-RADS lexicon to assess breast tissue composition. In research conceptual frameworks, effectiveness research concerns the results that are achieved in actual practice while efficacy research is performed under ideal conditions. The use of computerized or automated quantitative measures are objective and reproducible, an ideal condition. Unfortunately, radiologists in clinically practice do not use these measurements routinely; instead, they are implemented under research environments. Given that the BI-RADS lexicon is used in actual practice, studies that utilize this measurement are more important in determining the effectiveness of this practice in day-to-day operations. While quantitative measures are scientifically sound and may offer stronger effect measures, its utility is limited if not routinely used. If more research is done evaluating this relationship, methods used in routine practice should be studied or efforts to make affordable, objective measurement tools available to the general population should be initiated.

Another finding was that the literature only addressed a few of the criteria used to evaluate risk factors as predictive tools. Specifically, the majority of studies calculated risk and a minority calculated the c-statistic for discriminatory ability. There were no studies that calculated the Net Reclassification Index (NRI). Due to these failures,
clinical utility and cost-effectiveness of a personalized screening intervention based on breast tissue composition could not be assessed. The only conclusion that can be stated with certainty about the data retrieved for this review is that there is a dose-response relationship between breast tissue composition and incidence of both in situ and invasive breast cancers. Discrimination was close to chance and almost equivalent to what the existing Gail model provides. This value is likely created because breast density, even the ‘extremely dense’ category, is not a strong enough risk factor. In addition, the prevalence of ‘scattered fibroglandular densities’ and ‘heterogeneously dense’ breast tissue compositions are prevalent in the population, capturing the majority of women. There is no evidence of a significant effect on short-term mortality, though few studies in this review evaluated mortality as an outcome and their internal validity was poor.

While the measurement of exposure and outcome were relatively consistent across studies, there is significant heterogeneity in age group studied and duration of follow-up time. Therefore, the magnitude of the effect and at what time point it becomes strong or clinically significant cannot be concluded with certainty for the population of interest.

**Limitations of review**

This systematic review is limited because I was the sole reviewer who assessed eligibility for inclusion. Dual-review is the preferred methodology. I also relied completely on published literature. There were conference abstracts without subsequent published manuscripts that I identified in my search. I did not attempt to contact study investigators due to time and resource limitations. Lastly, I qualitatively synthesized the results instead of pooling data in a meta-analysis. Meta-analyses are useful in increasing power and precision of estimates.

**Limitations of evidence**
Significant limitations of the evidence are the inherent bias in measurement of the exposure and the outcome, small body of literature using BI-RADS lexicon, few quality studies and heterogeneity of study design. The breast tissue composition is assessed using a subjective ordinal scale. The radiologist categorizes a woman’s breast tissue composition based on how it appears on imaging using narrative guidance outlines in BI-RADS atlas.

Using the BI-RADS lexicon, exposure can be misclassified. Observer agreement is less when considering women who fall into the two intermediate breast tissue composition categories, ‘scattered fibroglandular densities’ and ‘heterogeneously dense’, rather than at the extremes. For the two intermediate categories in comparison to the extremes, misclassification is differential, occurring more frequently in the intermediate categories. This misclassification has the potential to over or under-estimate the effect. In the majority of studies included in this review, ‘scattered fibroglandular densities’ was used as the referent category. Depending on the amount and direction of the misclassification, the effect of density on risk in women with ‘extremely dense’ breast could be different than what was measured. Using this subjective measure moving forward, the only way to reduce misclassification would be training radiologist to improve observer agreement.

Breast tissue composition is also not stable over time. The included studies had different durations of follow-up and most used exposure measurement obtained with the index mammogram. Migration bias, when study subjects more from one exposure category to another, can be inherently produced. Crossovers can diminish risk estimates. Using a one-time measurement to predict what happens 5 or 10 years out, after the potential influence of weight, age and hormones on breast density, may not be producing the most accurate results.
Issues with measurement bias are also apparent in outcome assessment. The quality of cancer registries is assessed by its completeness, validity, timeliness and resemblance to the general population. The validity of outcome measurement rest on the accuracy of the population-based registries used. For this paper, I used web-available information about the BCSC and their respective sites. I did not contact the organization for more details, which may have been helpful to answering pertinent questions about the quality of registry data. I was not able to find important details, such as whether they used active or passive follow-up to contact physicians or patients, completeness of data and timeliness of outcome reporting.

While the BI-RADS classification scheme and cancer registries represent the most feasible resources for research, there are improvements that can be made to improve the quality of the overall body of evidence. There is so much heterogeneity in study design, specifically, in the baseline characteristics of study subjects, follow-up time and outcomes of interest, that meta-analysis is likely not even possible.

Implications for practice and research

Breast density does not provide the information needed to determine for what population of women health care professionals and policymakers should recommended or dissuade from undergoing screening mammography. Given the problems associated with measurement, the focus could be shifted away from the BI-RADS lexicon to objective breast density measurements or away from mammography to other enterprising imaging technologies such as breast magnetic resonance imaging (MRI) or 3D digital breast tomosynthesis. The findings of this review also suggest that we have reached the limitations of screening for breast cancer with mammography and may need to shift our focus to lifestyle modifications, such as diet and exercise.50

Another option for future research is to focus on the breast microenvironment, developing a better understanding of what leads to the differences seen on imaging. Sun
et al., in a population-based Polish Women’s Breast Cancer Study, studied the gene expression profile of high mammographic breast density. They found that mammographic density reflects transcriptional and signaling changes at the molecular level. Using data from the Carolina Breast Cancer Study (CBCS) and Carolina Mammography Registry (CMR), Razzaghi and Troester et al. have done research evaluating molecular features, or micro-environment, under different circumstances, following surgical wounds, with variable mammographic density patterns, and in women of different ages. The breast microenvironment may offer more valuable and predictive information than subjective imaging assessments. One priority would be to determine whether mammographic breast density reflects differences in the microenvironment. A proposed research agenda would select participants with normal breast tissue, but classified differently according to breast tissue composition. It would more ethically sound to select participants who have undergone breast biopsy with benign findings. The intervention would be to profile gene expression and with the outcome of describing features, looking for patterns. This may reveal a new biomarker with more predictive promise.

**Conclusion**

Rose encourages considering – Why did *this* patient get *this* disease at *this* time? Why do average-risk women in between the ages of 40 and 59 get breast cancer at any given time? Unfortunately, there is no clear answer to this question because there is substantial uncertainty given scientific limitations. Strong risk factors, however, can have predictive value. So even if we do not have the answer to Rose’s question, we can use risk factors that are casually or statistically related to provide estimates of risk. These estimates of risk are incorporated into the weighing of benefits and harms when making population level decisions. While breast tissue composition has been a well-established
risk factor for some time, the evidence does not suggest that the ACR BI-RADS lexicon has predictive ability. The classification scheme does not discriminate well, which means that it is not clear at which breast tissue composition one is very likely to become a “case”. Without meeting this basic requirement of predictive ability, clinical utility and cost-effectiveness have no place. There is a clear, reproducible dose-response relationship between breast density and breast cancer, but more research is needed. In addition, continuous development of population-based registries held to high-quality standards should be encouraged. Policymakers and health care professionals should approach decisions around breast tissue composition with caution, as the evidence is not yet powerful and meaningful enough.
Acknowledgements

I would like to first thank Dr. Russell Harris and Dr. Cherie Kuzmiak for their wonderful guidance in writing this paper. Dr. Harris is an outstanding educator, easily one of the best that I’ve had the opportunity to learn from while in medical school. Upon meeting Dr. Kuzmiak, her passion for radiology and breast imaging was readily apparent. I was very appreciative of her warm welcome and attention to my work. I would also like to thank my family and my two puppies Bo and Bear for providing a loving environment for my work. Lastly, I would like to thank my colleagues Sarah Smiley and Ali Annaim for their edits and input along the way.
References:

50. Harris R. Screening is only part of the answer to breast cancer. Annals of internal medicine. 2014;160(12):861-3.
of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2012;21(10):1735-44.


Appendix A: Detailed search terms

PubMed:
(((((breast OR breast cancers OR breast cancer OR breast tumor OR breast tumors OR breast carcinoma OR breast neoplasm OR breast neoplasms) AND (mortality OR survival OR incidence OR prevalence OR risk OR risk factor OR risk assessment) AND (age OR breast density OR mammographic breast density OR mammographic density) AND (mammography OR mammogram) AND (mass screening OR screen OR screening OR screened) NOT (BRCA*) AND (BIRAD OR bi-rad OR bi rad OR BIRADs OR breast imaging-reporting and data system OR dense OR "extremely dense breast" OR "heterogeneously dense" OR "fatty" OR "almost entirely fat" OR "scattered" OR "fibroglandular")))

EMBASE:
'brest' OR 'breast'/exp OR breast OR 'breast cancer'/exp OR 'breast cancer' OR 'breast cancers' OR 'breast tumor'/exp OR 'breast tumor' OR 'breast tumors' OR 'breast carcinoma'/exp OR 'breast carcinoma' OR 'breast neoplasm' OR 'breast neoplasms'/exp OR 'breast neoplasms' AND ('mortality' OR 'mortality'/exp OR mortality OR 'survival' OR 'survival'/exp OR survival OR 'incidence' OR 'incidence'/exp OR incidence OR 'prevalence' OR 'prevalence'/exp OR prevalence OR 'risk' OR 'risk'/exp OR risk OR 'risk factor'/exp OR 'risk factor' OR 'risk factors'/exp OR 'risk factors' OR 'risk assessment'/exp OR 'risk assessment') AND ('age' OR 'age'/exp OR age OR 'density' OR 'density'/exp OR density OR 'breast density' OR 'mammographic density' OR 'mammographic breast density') AND (mammogram OR 'mammography' OR 'mammography'/exp OR mammography) AND (birad OR 'bi rad' OR 'breast imaging reporting and data system' OR 'extremely dense breast' OR 'heterogenously dense' OR 'fatty' OR 'almost entirely fat' OR 'scattered' OR 'fibroglandular' OR 'dense') AND [female]/lim AND [english]/lim AND [2000-2014]/py
Appendix B: Flow diagram for search strategy

Identification

- 362 records identified through PubMed searching
- 450 records identified through EMBASE searching

Screening

- 268 duplicates removed
- 544 titles and abstracts screened for eligibility
- 519 records excluded
  - Reasons for exclusion included lack of focus on target population, use of quantitative breast density tools (ex. percent dense area or absolute dense area), use of automated and/or computerized tools and use of technology other than mammogram (ex. US or MRI)

Eligibility

- 25 full-text articles assessed for eligibility
- 14 records excluded
  - Reasons for exclusion included lack of published manuscript for abstract, outcome not of interest and using quantitative measurements.

Included

- 11 articles included in qualitative synthesis
### Appendix C. Data extraction and critical appraisal of eligible studies

#### Table 1. Data Extraction Table for Gierach et al. 2012

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Sample size for analysis</th>
<th>Age range</th>
<th>Average follow-up time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prospective cohort</td>
<td>9232</td>
<td>30 years or older</td>
<td>6.6 years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Registry</th>
<th>Age range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Five Breast Cancer Surveillance Consortium mammography registries: Group Health Cooperative in Washington State, the New Hampshire Mammography Network, the New Mexico Mammography Project, the San Francisco Mammography Registry, and the Vermont Breast Cancer Surveillance System</td>
<td>30 years or older</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dates of Recruitment</th>
<th>Average follow-up time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996 – 2005</td>
<td>6.6 years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Breast Tissue Composition</th>
<th>Simple adjusted hazard ratio of invasive breast cancer death (95% CI) Adjusted for registry site, age at diagnosis, year of diagnosis and BMI</th>
<th>Fully adjusted hazard ratio of invasive breast cancer death (95% CI) Adjusted for registry site, age at diagnosis, year of diagnosis and BMI, mode of detection, treatment and median family income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost entirely fat</td>
<td>1.24 (1.03 – 1.74)</td>
<td>1.36 (1.04 – 1.77)</td>
</tr>
<tr>
<td>Scattered fibroglandular densities</td>
<td>Referent</td>
<td>Referent</td>
</tr>
<tr>
<td>Heterogeneously dense</td>
<td>1.01 (0.87 – 1.18)</td>
<td>1.07 (0.92 – 1.25)</td>
</tr>
<tr>
<td>Extremely dense</td>
<td>0.91 (0.71 – 1.18)</td>
<td>0.92 (0.71 – 1.19)</td>
</tr>
</tbody>
</table>
Table 2. Critical Appraisal of Gierach et al. 2012\(^{39}\)

<table>
<thead>
<tr>
<th>Selection Bias</th>
<th>Selection: The authors identified all women 30 years and older that had been diagnosed with cancer (N = 26,571). They then selected subjects with breast density assessments and at least three years of follow-up. Subjects were excluded for low BMI and missing data. Within the text, there are no details about selection of radiology facilities and their resemblance to respective catchment area. Comparability of groups: Women with dense breast were younger, had lower BMI’s, etc. Follow-up: If state vital records were complete and no breast cancer related death information was available subjects were presumed to be alive. Given this methodology, follow-up vital status was available for all study participants creating no lost to follow-up.</th>
<th>Medium potential for selection bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement Bias</td>
<td>Exposure: One of the exposures was the recorded ACR BI-RADS breast tissue composition assessment routinely used by radiologists in clinical practice. The ordinal scale is inherently unequally applied and unreliable due to its subjectivity. Index mammogram was defined as an exam within 5 years of diagnosis. If there were multiple exams, they used exam closest to diagnosis. If no screening mammogram was available, they used diagnostic mammogram results. Mammogram reports within 30 days after diagnosis were used in neither of the prior were available. Breast cancer data was retrieved from SEER, state tumor registries and/or pathology databases. Outcome: Vital status (alive or dead) and cause of death was obtained from cancer registries and state vital records. Completeness of records varied by database. A participant was assumed to be alive if there was no documentation of death.</td>
<td>Medium potential for selection bias</td>
</tr>
<tr>
<td>Confounding</td>
<td>Variables potentially confounding the relationship between breast density and breast cancer related death include age, overall health status, stage of diagnosis, etc. There was no attempt to assess for competing comorbidities and general medical condition.</td>
<td>High potential for confounding</td>
</tr>
<tr>
<td>Analysis</td>
<td>The fully adjusted survival models were included pertinent covariates. Notably, authors did not adjust for race and year of diagnosis.</td>
<td>Less than appropriate</td>
</tr>
<tr>
<td>Internal Validity</td>
<td>There are inherent issues with passive follow-up methodologies for obtaining vital status information. In regards to vital status, all appropriate confounders were also not considered.</td>
<td>Poor</td>
</tr>
<tr>
<td>External Validity</td>
<td>The majority of the study population was White, non-Hispanic.</td>
<td>Poor</td>
</tr>
</tbody>
</table>
Table 3. Data Extraction Table for Kerlikowske et al. 2007\textsuperscript{55}

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Sample size for analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prospective cohort</td>
<td>301,955</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Registry</th>
<th>Age range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seven BCSC mammography registries, which included the San Francisco Mammography Registry, Group Health’s Breast Cancer Surveillance, Colorado Mammography Advocacy Project, Vermont Breast Cancer Surveillance System, New Hampshire Mammography Network, Carolina Mammography Registry, and New Mexico Mammography Registry</td>
<td>30 years or older</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dates of Recruitment</th>
<th>Average follow-up time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993 – 2003</td>
<td>3 year</td>
</tr>
</tbody>
</table>

Results for the entire study population

<table>
<thead>
<tr>
<th>Breast Tissue Composition at first screen</th>
<th>Number of invasive and in situ breast cancers per 1000 women</th>
<th>OR for breast cancer (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Adjusted for age at last screening mammogram</td>
</tr>
<tr>
<td>Almost Entirely Fat</td>
<td>5.0 (4.0 – 5.4)</td>
<td>0.63 (0.54 – 0.74)</td>
</tr>
<tr>
<td>Scattered fibroglandular densities</td>
<td>7.0 (6.8 – 7.8)</td>
<td>Referent</td>
</tr>
<tr>
<td>Heterogeneously dense</td>
<td>10.0 (8.9 – 10.3)</td>
<td>1.32 (1.21 – 1.44)</td>
</tr>
<tr>
<td>Extremely dense</td>
<td>13.0 (11.1 – 14.3)</td>
<td>1.75 (1.52 – 2.00)</td>
</tr>
</tbody>
</table>

Number of invasive and in situ breast cancers per 1000 women < 50 years old with no change in density

<table>
<thead>
<tr>
<th>Breast Tissue Composition at first screen</th>
<th>Number of invasive and in situ breast cancers per 1000 women &lt; 50 years old with no change in density</th>
<th>Number of in situ and invasive breast cancer per 1000 women &lt; 50 years old with no change in density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almost Entirely Fat</td>
<td>0.4 (0.2 – 7.7)</td>
<td>4.0 (3.1 – 5.2)</td>
</tr>
<tr>
<td>Scattered fibroglandular densities</td>
<td>3.0 (2.6 – 3.6)</td>
<td>9.0 (8.5 – 10.0)</td>
</tr>
<tr>
<td>Heterogeneously dense</td>
<td>6.0 (5.3 – 6.6)</td>
<td>12.0 (10.9 – 13.4)</td>
</tr>
<tr>
<td>Extremely dense</td>
<td>8.0 (6.6 – 10.8)</td>
<td>12.0 (9.0 – 16.3)</td>
</tr>
</tbody>
</table>
### Table 4. Critical Appraisal of Kerlikowske et al. 2007\(^5\)\(^6\)

<table>
<thead>
<tr>
<th><strong>Selection Bias</strong></th>
<th><strong>Description</strong></th>
<th><strong>Rating</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selection</strong></td>
<td>To be eligible for inclusion, participants had to have at least two screening mammograms with density assessment that were more than 9 months apart. This criterion selects for individuals that have access to regular care. Exclusion criteria included history of breast cancer or augmentation, and postmenopausal women on hormone replacement therapy. <strong>Comparability of groups:</strong> Women 70 years or older were more likely to be assigned to ‘almost entirely fat’ breast tissue composition group. <strong>Follow-up:</strong> There were no methods detailed for confirming cancer-free status, such as mail survey, review of electronic medical record, or screening mammogram with negative results.</td>
<td>Medium potential for selection bias</td>
</tr>
<tr>
<td><strong>Measurement Bias</strong></td>
<td><strong>Exposure:</strong> Investigators used ACR BI-RADS breast tissue composition assessment routinely used by radiologists in clinical practice. The ordinal scale is inherently unequally applied and unreliable due to its subjectivity. If a woman had more than two screening mammograms during the study period, they used the earliest and last breast density assessments. <strong>Outcome:</strong> Invasive and ductal carcinoma in situ breast cancers identified using pathology database, SEER database and state tumor registries. Measurement included breast cancer diagnoses within one year of the last screening examination during the study period.</td>
<td>Medium potential for measurement bias</td>
</tr>
<tr>
<td><strong>Confounding</strong></td>
<td>Investigators adjusted rates and odds ratios for age at last screening mammogram, registry site and time between mammograms.</td>
<td>Appropriate attention to confounding</td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
<td>They used logistic regression adjusted for registry, time between two screening mammograms, and age. ‘Scattered fibroglandular densities’ was used as the referent group. They were no able to adjust for body mass index (BMI) and age at first birth. They adjusted for misclassification of breast tissue composition using the Reade-Christopher and Kupper method.</td>
<td>Appropriate analysis</td>
</tr>
<tr>
<td><strong>Internal Validity</strong></td>
<td>Missing values limited investigators in adjusting for pertinent covariates, specifically BMI.</td>
<td>Fair/good</td>
</tr>
<tr>
<td><strong>External Validity</strong></td>
<td>They pooled resources from all of the existing BCSC registry sites. The majority of BCSC radiology facilities are associated with academic medical centers. There were appropriate inclusion and exclusion criteria for selection of participants from those sites.</td>
<td>Fair</td>
</tr>
</tbody>
</table>
Table 5. Data Extraction Table for Kerlikowske et al. 2010

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Prospective cohort</th>
<th>Sample size for analysis</th>
<th>587,369</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registry</td>
<td>Seven BCSC mammography registries, which included the San Francisco Mammography Registry, Group Health’s Breast Cancer Surveillance, Colorado Mammography Advocacy Project, Vermont Breast Cancer Surveillance System, New Hampshire Mammography Network, Carolina Mammography Registry, and New Mexico Mammography Registry</td>
<td>Age range</td>
<td>30 years or older</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>56.4</td>
</tr>
<tr>
<td>Dates of Recruitment</td>
<td>1996 - 2006</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Breast Tissue Composition</th>
<th>5-year risk of in situ or invasive breast cancer (95% CI)</th>
<th>5-year risk of in situ or invasive breast cancer (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-menopausal women</td>
<td>Post-menopausal women</td>
</tr>
<tr>
<td>Almost entirely fat</td>
<td>40 – 44 0.4 (0.2 – 0.5)</td>
<td>45 – 49 0.5 (0.2 – 0.7)</td>
</tr>
<tr>
<td></td>
<td>45 – 49 0.7 (0.5 – 0.9)</td>
<td>50 – 54 0.6 (0.5 – 0.8)</td>
</tr>
<tr>
<td></td>
<td>50 – 54 0.8 (0.4 – 1.1)</td>
<td>55 – 59 0.8 (0.6 – 0.9)</td>
</tr>
<tr>
<td>Scattered fibroglandular densities</td>
<td>40 – 44 0.9 (0.8 – 0.9)</td>
<td>45 – 49 0.8 (0.7 – 1.0)</td>
</tr>
<tr>
<td></td>
<td>45 – 49 1.3 (1.2 – 1.4)</td>
<td>50 – 54 1.3 (1.2 – 1.4)</td>
</tr>
<tr>
<td></td>
<td>50 – 54 1.5 (1.4 – 1.7)</td>
<td>55 – 59 1.7 (1.6 – 1.8)</td>
</tr>
<tr>
<td>Heterogeneously dense</td>
<td>40 – 44 1.5 (1.4 – 1.6)</td>
<td>45 – 49 1.6 (1.3 – 1.8)</td>
</tr>
<tr>
<td></td>
<td>45 – 49 2.1 (1.9 – 2.2)</td>
<td>50 – 54 1.8 (1.6 – 1.9)</td>
</tr>
<tr>
<td></td>
<td>50 – 54 2.3 (2.1 – 2.5)</td>
<td>55 – 59 2.4 (2.3 – 2.6)</td>
</tr>
<tr>
<td>Extremely dense</td>
<td>40 – 44 1.8 (1.7 – 2.0)</td>
<td>45 – 49 2.2 (1.6 – 2.7)</td>
</tr>
<tr>
<td></td>
<td>45 – 49 2.5 (2.3 – 2.7)</td>
<td>50 – 54 2.2 (1.8 – 2.6)</td>
</tr>
<tr>
<td></td>
<td>50 – 54 3.1 (2.7 – 3.5)</td>
<td>55 – 59 2.4 (2.0 – 2.8)</td>
</tr>
</tbody>
</table>
Table 6. Critical Appraisal of Kerlikowske et al. 2010

| Selection Bias | Selection: All women that had screening mammograms during specified time were initially selected. About 20% of the population was excluded because of missing data, specifically, body mass index (BMI), hormone use and menopausal status.  
Comparability of groups: They did not provide baseline characteristics of highlight differences in participants organized according to breast composition category. Instead, characteristics are given for the overall population.  
Follow-up: The BCSC registries had a cancer ascertainment rate of about 94% during this time. There are no details for how this rate is calculated, however. | Medium potential for selection bias |
| Measurement Bias | Exposure: Investigators used ACR BI-RADS breast tissue composition assessment routinely used by radiologists in clinical practice. The ordinal scale is inherently unequally applied and unreliable due to its subjectivity. It is unclear which mammogram was used if there were multiple recording in the registry within the study time.  
Outcome: BCSC registries link to SEER and state tumor registries. Pathology databases were also used as a resource. | Medium potential for measurement bias |
| Confounding | There was appropriate attention to the following covariates: hormone therapy use, menopausal status and age. BMI was a missing variable. BMI is a potential confounder of the relationship between exposure and outcome. | Medium potential for confounding |
| Analysis | Adjusted 5-year breast cancer risk was calculated using partly conditional Cox proportional hazard survival model, BMI was fixed at 25 kg/m2. This BMI was the average for the entire population. | Less than appropriate analysis |
| Internal Validity | Authors do pay attention to issues of follow-up by documenting the cancer ascertainment rate, which was greater than 90%. | Fair internal validity |
| External Validity | Consistent with the demographic characteristics of BCSC overall, the majority of the study population was White, concerning given that breast density thought to vary by race/ethnicity. Using an average BMI, which was also within normal range and not resembling US population, also affects its applicability to some women. | Fair external validity |
Table 7. Data Extraction Table for MacKenzie et al. 2007\textsuperscript{43}

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Sample size for analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prospective cohort</td>
<td>154,936</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Registry</th>
<th>Age range</th>
<th>Median age</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Hampshire and Vermont Breast Cancer Surveillance System</td>
<td>40 – 98</td>
<td>52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dates of study</th>
<th>Average follow-up time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994 – 2001</td>
<td>4 years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Breast Tissue Composition</th>
<th>Risk Ratio of DCIS (95% CI)</th>
<th>Risk Ratio of DCIS (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted for age</td>
<td>Adjusted for age</td>
</tr>
<tr>
<td></td>
<td>Premenopausal</td>
<td>Postmenopausal</td>
</tr>
<tr>
<td>Almost entirely fat</td>
<td>0.26 (0.03 – 1.95)</td>
<td>0.56 (0.36 – 0.90)</td>
</tr>
<tr>
<td>Scattered fibroglandular density</td>
<td>Referent</td>
<td>Referent</td>
</tr>
<tr>
<td>Heterogeneously dense</td>
<td>2.22 (1.51 – 3.28)</td>
<td>1.48 (1.18 – 1.86)</td>
</tr>
<tr>
<td>Extremely dense</td>
<td>2.68 (1.65 – 4.36)</td>
<td>1.61 (1.01 – 2.56)</td>
</tr>
</tbody>
</table>
### Table 8. Critical Appraisal of MacKenzie et al. 2007\

| Selection Bias | Selection: Investigators identified women who had mammograms documented in the NH and VT registries. About 80% of the mammography facilities in NH and 16 centers total in VT participate in the BCSC. In NH, the majority of the mammographic facilities include hospitals (58%). The minority (18%) are physicians’ private offices. In VT, 88% of the facilities are hospitals. Comparability of groups: They did not provide baseline characteristics of highlight differences in participants organized according to breast composition category. Instead, characteristics are given for the overall population. Follow-up: The outcome was presence or absence of DCIS and information was sought from registries. Confirmation of disease-free status with passive or active procedures not documented. If no cancer diagnosis documented, participant assumed to be cancer free. Given this methodology, outcomes were available for all study participants creating no lost to follow-up. | Medium potential for selection bias |
| **Measurement Bias** | Exposure: Investigators used the ACR BI-RADS breast tissue composition assessment routinely used by radiologists in clinical practice. The ordinal scale is inherently unequally applied and unreliable due to its subjectivity. Outcome: Breast cancer data was retrieved from participating pathology laboratories and supplemented with state tumor registries. They used systematized nomenclature of medicine (SNOMED) or TNM Classification of Malignant Tumors to classify cases as DCIS. | Medium potential for measurement bias |
| **Confounding** | For pre-menopausal women, investigators accounted for age, parity, BMI and family history. In post-menopausal women, they adjusted for hormonal therapy in addition to the prior. | Appropriate attention |
| **Analysis** | They used Cox proportional hazards model with age as the time-scale. Censoring occurred at the time of diagnosis of in situ or invasive breast cancer diagnosis, and at the end of the study period. Analysis using indicator variables for the missing information did not yield different result. | Appropriate analysis |
| **Internal Validity** | DCIS presents more issues for measurement bias. DCIS if often diagnosed with screening mammogram by the presence of micro-calcifications. Not all DCIS is associated with micro-calcifications, however. | Fair |
| **External Validity** | The majority of the study population had college-educated and White. The majority of the radiology facilities in this study were affiliated with hospitals. | Fair |
Table 9. Data Extraction Table for Olson et al. 2012

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Sample size for analysis</th>
<th>Source Population</th>
<th>Age range</th>
<th>Average follow-up time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nested case-control</td>
<td>2,259</td>
<td>Minnesota, Iowa, Wisconsin</td>
<td>35 years or older</td>
<td>2.4 years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dates</th>
<th>Breast Tissue Composition</th>
<th>HR of incident in situ and invasive breast cancer (95% CI)</th>
<th>HR for incident in situ and invasive breast cancer (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment – 2003 to 2006</td>
<td>Almost entirely fat</td>
<td>Referent</td>
<td>Referent</td>
</tr>
<tr>
<td>Outcome – 2010</td>
<td>Scattered fibroglandular densities</td>
<td>1.61 (1.13 – 2.29)</td>
<td>1.61 (1.11 – 2.33)</td>
</tr>
<tr>
<td></td>
<td>Heterogeneously dense</td>
<td>2.02 (1.38 – 2.95)</td>
<td>1.96 (1.30 – 2.98)</td>
</tr>
<tr>
<td></td>
<td>Extremely dense</td>
<td>2.96 (1.73 – 5.07)</td>
<td>2.56 (1.38 – 4.75)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted for age, menopausal status, body mass index, and postmenopausal hormones</td>
<td>Adjusted for all four acquisition parameters</td>
</tr>
</tbody>
</table>

C- statistic | 0.62
Table 10. Critical Appraisal of Olson et al. 2012

<table>
<thead>
<tr>
<th>Selection Bias</th>
<th>Details</th>
<th>Potential Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection: Eligibility criteria included having received screening mammography at the Mayo Clinic, being 35 years or older, residing of Minnesota, Iowa, or Wisconsin (tri-state); and having no personal history of breast cancer. The eligible women were mailed an invitation packet. The cases included all incident breast cancer cases in the prospective cohort during the follow-up period. The control group was formed by randomly selecting 10% of overall cohort.</td>
<td>Low potential for selection bias</td>
<td></td>
</tr>
<tr>
<td>Comparability of groups: They did not provide baseline characteristics of highlight differences in participants organized according to breast composition category.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow-up: Study investigators used a combination of active and passive follow-up to obtain cancer outcomes. The active follow-up to obtain cancer and vital status included a mailed questionnaire or a telephone call. Active follow-up was successful in 83.1% of participants in 2009 and 78.4% of participants in 2010. Passive follow-up included hospital and state tumor registries.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measurement Bias</th>
<th>Details</th>
<th>Potential Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure: Investigators used the ACR BI-RADS breast tissue composition assessment routinely used by radiologists in clinical practice. The ordinal scale is inherently unequally applied and unreliable due to its subjectivity.</td>
<td>Low potential for selection bias</td>
<td></td>
</tr>
<tr>
<td>Outcome: Procedures included both active and passive follow-up.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Confounding</th>
<th>Details</th>
<th>Appropriateness</th>
</tr>
</thead>
<tbody>
<tr>
<td>The cases were older and a higher proportion reported former post-menopausal hormone use and post-menopausal status. Some baseline characteristics of interest, such as socioeconomic status and race/ethnicity, are missing from reporting.</td>
<td>Appropriate attention to confounding</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Details</th>
<th>Appropriateness</th>
</tr>
</thead>
<tbody>
<tr>
<td>For residents of the tri-state area censoring occurred at diagnosis, death and study end. For non-residents, censoring occurred at diagnosis, last response to follow-up, last contact with Mayo Clinic and date last known to reside in area. The Cox proportional hazards model, using age at the time scale, was used to calculate hazard ratios. The C-statistic was also calculated to measure the degree to which density could discriminate between cases and non-cases.</td>
<td>Appropriate analysis</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Internal Validity</th>
<th>Details</th>
<th>Appropriateness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigators used appropriate analysis and procedures for reducing selection and measurement bias.</td>
<td>Good</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>External Validity</th>
<th>Details</th>
<th>Appropriateness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some of the issues that affect external validity include the trial setting, exposure, selection of participants, and exposure assessment.</td>
<td>Poor</td>
<td></td>
</tr>
</tbody>
</table>
### Table 11. Data Extraction Table for Razzaghi et al. 2012⁵⁷

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Sample size for analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-control</td>
<td>Cases = 491 Controls = 528</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Registry</th>
<th>Age range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carolina Mammography Registry (CMR), part of BCSC</td>
<td>20 – 74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dates of Recruitment</th>
<th>Time from selection/diagnosis to breast density measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996 - 2001</td>
<td>Cases Controls</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of in situ and invasive breast cancer diagnoses in the cases</th>
<th>No. of in situ and invasive breast cancer diagnoses in the controls</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fat</td>
<td>13</td>
<td>25</td>
</tr>
<tr>
<td>Scattered</td>
<td>183</td>
<td>197</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>232</td>
<td>253</td>
</tr>
<tr>
<td>Extremely</td>
<td>63</td>
<td>53</td>
</tr>
</tbody>
</table>

*Adjusted for matching factors age and race
Table 12. Critical Appraisal of Razzaghi et al. 2012\textsuperscript{57}

| Selection Bias | Selection: Eligible study participants had mammography results recorded in the CMR and were participants in the CBCS. The participants were recruited from 24 counties in North Carolina. The response rate to recruitment efforts differed between cases (76%) and controls (55%). Both cases and controls had to attend an in-person interview. Controls were identified using drivers' license and Medicare beneficiary lists.  
Comparability of groups: They did not provide baseline characteristics of highlight differences in participants organized according to breast composition category. Instead, characteristics are given for the overall population.  
Follow-up: There were no methods detailed for confirming cancer-free status, such as mail survey, review of electronic medical record, or screening mammogram with negative results. | Medium potential for selection bias |
| Measurement Bias | Exposure: Breast density measurements came from reports recorded in the registry. For cases, screening or diagnostic mammogram reports within 5 years prior to diagnosis and up to one year were accepted. For controls, screening mammogram reports up to 5 years prior and 3 years after selection data were accepted. Reading radiologist varied by site and was not blinded to patient medical history. Thus measurement was equal in that it was applied the same in both cases and controls. The measurement is neither valid nor reliable.  
Outcome: Breast cancer cases were identified using the NC Central Cancer Registry. | Medium potential for measurement bias |
| Confounding | A higher proportion of the women diagnosed with breast cancer had a family history of breast cancer, started menarche before the age of 13, and were nulliparous. The cases and controls differ in a biologically plausible manner. | Appropriate attention to confounding |
| Analysis | Cases and controls were matched based on age and race. Logistic regression used to estimate the odds ratio. | Appropriate analysis |
| Internal Validity | Without more recruitment procedure details, it is unclear how both the selected cases and controls compare to the general population and to each other in their ability to experience the outcome. Both socioeconomic and behavioral characteristics can influence cancer diagnosis. | Fair |
| External Validity | Study participants came from a limited number of counties in NC and explicit details of recruitment process are not described. | Fair |
### Table 13. Data Extraction Table for Tesic et al. 2013

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Sample size for analysis</th>
<th>52,752</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registry</td>
<td>Age range</td>
<td>50 – 69 years old</td>
</tr>
<tr>
<td>Dates of Recruitment</td>
<td>Average</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duration of follow-up</td>
<td>1 year</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Breast Tissue Composition</th>
<th>OR for in situ and invasive breast cancer (95% CI)</th>
<th>Adjusted for age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost entirely fat</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Scattered fibroglandular densities</td>
<td>1.66 (1.27 – 2.18)</td>
<td></td>
</tr>
<tr>
<td>Heterogeneously dense/Extremely dense</td>
<td>1.86 (1.29 – 2.69)</td>
<td></td>
</tr>
</tbody>
</table>
Table 14. Critical Appraisal of Tesic et al. 2013

| Selection Bias | Selection: Study participants were recruited by active invitation to participate in screening program. The authors do not provide details about the method of invitation (i.e. mail, telephone, or mass media) or about the number of invites and/or attempted extended and percentage fulfilled.  
Comparability of groups: Women in the four breast tissue composition categories differed with respect to age, body mass index (BMI), family history, history of breast procedure, and use of postmenopausal hormonal therapy among other variables. Women with ‘almost entirely fat’ breast tissue were older and had higher BMI's. A higher proportion of women with ‘extremely dense’ breast had significant family history and history of prior breast procedures, and were currently taking hormone therapy.  
Follow-up: There is no information about whether outcome information was available on all enrolled participants. | High potential for selection bias |
| Measurement Bias | Exposure: Breast density assessment performed by radiologist using “double blind” readings. This methodology increases the reliability of the measurement.  
Outcome: Data on cancer diagnosis came from pathology facilities and the registry. Only screen-detected cancers were documented. Women with interval cancers were excluded from the analysis. There was no procedure, such as review of medical records, to confirm that non-cases were in fact breast cancer free. | Medium potential for measurement bias |
| Confounding | There is appropriate attention to potential confounders with subsequent adjustment in analysis | Low potential for confounding |
| Analysis | Women with ‘scattered fibroglandular densities’ and ‘heterogeneously dense’ breast tissue composition were combined into one category for analysis and the rationale is not provided. | Less than appropriate |
| InternalValidity | The method of active invitation likely selects for participants with certain characteristics that may bias results. Issues of confounding and analysis are also present. | Fair |
| External Validity | The study was conducted in Croatia, limiting its applicability to the general US population. The assessment of the exposure also entailed a procedure, double-blind reading, which is not routinely used in clinical practice. | Poor |
Table 15. Data Extraction Table for Tice et al. 2005\textsuperscript{58}

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Sample size for analysis</th>
<th>81,777</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registry</td>
<td>Age range</td>
<td>35 years or older</td>
</tr>
<tr>
<td>San Francisco Mammography Registry</td>
<td>Average</td>
<td>55.9</td>
</tr>
<tr>
<td>Dates of Recruitment</td>
<td>Average follow-up time</td>
<td>5.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Breast Tissue Composition</th>
<th>RR of in situ and invasive breast cancer (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost entirely fat</td>
<td>0.59 (0.36 – 0.98)</td>
</tr>
<tr>
<td>Scattered fibroglandular densities</td>
<td>Referent</td>
</tr>
<tr>
<td>Heterogeneously dense</td>
<td>1.41 (1.11 – 1.78)</td>
</tr>
<tr>
<td>Extremely dense</td>
<td>1.94 (1.31 – 2.89)</td>
</tr>
</tbody>
</table>
### Table 16. Critical Appraisal of Tice et al. 2005⁴⁶

| Selection Bias | Selection: The women included in this study were part of the San Francisco Mammography Registry, which is part of the BCSC. The eighteen participating radiology facilities include hospital-based facilities (N=14), clinic based practices and a mobile clinic (N=1). Only one of the facilities is affiliated with an academic institution. Women of unknown race/ethnicity were excluded from analysis. Authors did not report what proportion of the initially selected participants was excluded from analysis. Comparability of groups: They did not provide baseline characteristics of highlight differences in participants organized according to breast composition category. Instead, characteristics are given for the overall population. **Follow-up:** There were no methods detailed for confirming cancer-free status, such as mail survey, review of electronic medical record, or screening mammogram with negative results. **Low potential for selection bias** |
| Measurement Bias | Exposure: Investigators used the ACR BI-RADS breast tissue composition assessment routinely used by radiologists in clinical practice. The ordinal scale is inherently unequally applied and unreliable due to its subjectivity. **Outcome:** Breast cancer outcomes (in situ and invasive) were obtained from SEER. Vital Status was obtained through linkage to database from the California Department of Health Services. **Medium potential for measurement bias** |
| Confounding | In the analysis, adjustments were made for race and ethnicity. Other variables have been cited as confounding the relationship between exposure and incidence that are not accounted for in this study. **Less than appropriate attention to confounding** |
| Analysis | The risk factors were categorized according to the methods used for the Gail model. Dummy variables were used for missing values. Investigators used the Cox proportional hazards model adjusted for age and ethnicity. **Appropriate analysis** |
| Internal Validity | Validity is threatened by presence of selection and inherent measurement biases. In addition, more confounders should have been considered in analysis. **Fair/Good** |
| External Validity | One of the highlights form this study is the racially and ethnically diverse study population. In addition, there is variety in type of participating radiology facilities. **Good external** |
Table 17. Data Extraction Table for Tice et al. 2008

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registry</td>
<td>Seven BCSC mammography registries, which included the San Francisco Mammography Registry, Group Health’s Breast Cancer Surveillance, Colorado Mammography Advocacy Project, Vermont Breast Cancer Surveillance System, New Hampshire Mammography Network, Carolina Mammography Registry, and New Mexico Mammography Registry.</td>
</tr>
<tr>
<td>Dates of Recruitment</td>
<td>Not reported</td>
</tr>
<tr>
<td>Age</td>
<td>Breast Tissue Composition</td>
</tr>
<tr>
<td>40 - 44</td>
<td>Almost entirely fat</td>
</tr>
<tr>
<td></td>
<td>Scattered fibroglandular</td>
</tr>
<tr>
<td></td>
<td>Heterogeneously dense</td>
</tr>
<tr>
<td></td>
<td>Extremely dense</td>
</tr>
<tr>
<td>45 – 49</td>
<td>Almost entirely fat</td>
</tr>
<tr>
<td></td>
<td>Scattered fibroglandular</td>
</tr>
<tr>
<td></td>
<td>Heterogeneously dense</td>
</tr>
<tr>
<td></td>
<td>Extremely dense</td>
</tr>
<tr>
<td>C-statistic</td>
<td></td>
</tr>
</tbody>
</table>
Table 18. Critical Appraisal of Tice et al. 2008

<table>
<thead>
<tr>
<th>Selection Bias</th>
<th>Selection: Investigators identified all women that had at least one mammogram documented in one of the seven registry sites. Women that had a personal history of breast cancer or a diagnosis within 6 months of enrollment were excluded. Other exclusion criteria included history of breast augmentation and diagnosis of DCIS. Comparability of groups: They did not provide baseline characteristics of highlight differences in participants organized according to breast composition category. Instead, characteristics are given for the overall population. Follow-up: There were no methods detailed for confirming cancer-free status, such as mail survey, review of electronic medical record, or screening mammogram with negative results.</th>
<th>Medium potential for selection bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement Bias</td>
<td>Exposure: Investigators used the ACR BI-RADS breast tissue composition assessment routinely used by radiologists in clinical practice. The ordinal scale is inherently unequally applied and unreliable due to its subjectivity. Outcome: Breast cancer outcomes (in situ and invasive) were obtained from SEER, state tumor registries and pathology databases. Vital status was obtained from state vital statistics or National Death Index.</td>
<td>Medium potential for measurement bias</td>
</tr>
<tr>
<td>Confounding</td>
<td>In the analysis, adjustments were made for race and ethnicity. Other variables have been cited as confounding the relationship between exposure and incidence that are not accounted for in this study.</td>
<td>Less than appropriate</td>
</tr>
<tr>
<td>Analysis</td>
<td>Women who were diagnosed with in situ breast cancer were censored at the time of diagnosis. Study investigators used the proportional hazards model adjusted for age and race/ethnicity. Age-specific incidence for Native Americans’ and Alaskan Native women was inconsistent with SEER, so, they were excluded from analysis. The model was developed using a random sample of 60% and the validated in the remaining population.</td>
<td>Appropriate analysis</td>
</tr>
<tr>
<td>Internal Validity</td>
<td>There are meaningful issues of both selection and measurement bias. Appropriate adjustment for confounding variables was also not performed.</td>
<td>Fair</td>
</tr>
<tr>
<td>External Validity</td>
<td>The generalizability to Native American and Alaskan Native women is threatened because this population was excluded from the analysis.</td>
<td>Fair</td>
</tr>
</tbody>
</table>
Table 19. Data Extraction Table for Vacek et al. 2004

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Sample size for analysis</th>
<th>Age range</th>
<th>Average follow-up time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prospective cohort</strong></td>
<td>61,844</td>
<td>20 – 74</td>
<td>3 years</td>
</tr>
<tr>
<td><strong>Registry</strong></td>
<td>Vermont Breast Cancer Surveillance System, part of BCSC</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dates of Recruitment</strong></td>
<td>1996 - 2000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age-adjusted RR (95% CI)</th>
<th>Age-adjusted RR (95% CI)</th>
<th>Age-adjusted RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All women</td>
<td>All women</td>
<td>All women</td>
</tr>
<tr>
<td><strong>Entirely fat</strong></td>
<td>Referent</td>
<td>Referent</td>
</tr>
<tr>
<td><strong>Scattered fibroglandular</strong></td>
<td>2.0 (1.6 – 3.0)</td>
<td>2.5 (1.0 – 6.7)</td>
</tr>
<tr>
<td><strong>Heterogeneously sense</strong></td>
<td>3.0 (2.0 – 4.0)</td>
<td>3.7 (1.4 – 10.0)</td>
</tr>
<tr>
<td><strong>Extremely dense</strong></td>
<td>4.0 (2.8 – 5.7)</td>
<td>4.6 (1.7 – 12.6)</td>
</tr>
<tr>
<td><strong>Premenopausal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Referent</strong></td>
<td>Referent</td>
<td>Referent</td>
</tr>
<tr>
<td><strong>Menopausal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Referent</strong></td>
<td>Referent</td>
<td>Referent</td>
</tr>
<tr>
<td><strong>Referent</strong></td>
<td>Referent</td>
<td>Referent</td>
</tr>
<tr>
<td><strong>Referent</strong></td>
<td>Referent</td>
<td>Referent</td>
</tr>
<tr>
<td><strong>Referent</strong></td>
<td>Referent</td>
<td>Referent</td>
</tr>
<tr>
<td><strong>Referent</strong></td>
<td>Referent</td>
<td>Referent</td>
</tr>
</tbody>
</table>
Table 20. Critical Appraisal of Vacek et al. 2004

| Selection Bias | Selection: Eligibility criteria included having at least one mammogram documented in the registry during the study period and not having a self-reported or biopsy proven personal history of breast cancer. The date of study entry was the date of first breast density measurement.  
Comparability of groups: They did not provide baseline characteristics of highlight differences in participants organized according to breast composition category. Instead, characteristics are given for the overall population.  
Follow-up: Participants that did not receive an additional mammogram or biopsy were considered lost to follow-up (20,673, 24%). Those that were lost to follow-up were younger and had a breast density distribution most similar to the pre-menopausal study participants. | Medium potential for selection bias |
|----------------|-------------------------------------------------------------------------------------------------|----------------------------------|
| Measurement Bias | Exposure: Investigators used the ACR BI-RADS breast tissue composition assessment routinely used by radiologists in clinical practice. The ordinal scale is inherently unequally applied and unreliable due to its subjectivity. BI-RADS density category documented at entry mammogram, which could have been at any of the participating mammography centers.  
Outcome: Information of breast cancer diagnoses collected from all the pathology facilities in VT, VT and NH cancer registries. The NH Cancer Registry was used because some residents of eastern Vermont receive medical care in New Hampshire. | Medium potential for measurement bias |
| Confounding | Baseline characteristics of the women in the four breast density categories differed in biologically plausible ways. A higher proportion of nulliparous and premenopausal women were categorized as “extremely dense” than “almost entirely fat”. | Low |
| Analysis | Relationship between breast density and categorical outcome assessed by Pearson’s chi-square test. Age-adjusted relative risk assessed using Cox regression model with age as the time variable. This analysis approach accounts for aging during the study period. Effect measures were reported for the entire study population, and for pre-menopausal and menopausal strata. | Low |
| Internal Validity | Both the potential for selection and measurement bias are low and analysis was appropriate with control for potential confounding variables. | Fair internal validity |
| External Validity | The study incorporated data from several registries across the US and had broad eligibility criteria. | Good |
Table 21. Data Extraction Table for Zhang et al. 2013\textsuperscript{61}

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Sample size for analysis</th>
<th>Age range</th>
<th>Age range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registry</td>
<td>Carolina Mammography Registry (CMR), which is part of the BCSC.</td>
<td>15,243</td>
<td>30 years or older</td>
</tr>
<tr>
<td>Dates of Recruitment</td>
<td>1994 - 2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average follow-up time</td>
<td>3 years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mortality probability at 5 years (95% CI)</th>
<th>Mortality probability at 10 years (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 40 – 49</td>
<td><strong>White</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Almost entirely fat/scattered fibroglandular</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heterogeneously dense/extremely dense</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05 (0.04 – 0.07)</td>
<td>0.09 (0.07 – 0.10)</td>
</tr>
<tr>
<td></td>
<td><strong>African-American</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Almost entirely fat/scattered fibroglandular</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heterogeneously dense/extremely dense</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.10 (0.08 – 0.17)</td>
<td>0.17 (0.12 – 0.22)</td>
</tr>
<tr>
<td>Age 50 – 59</td>
<td><strong>White</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Almost entirely fat/scattered fibroglandular</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heterogeneously dense/extremely dense</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05 (0.04 – 0.06)</td>
<td>0.08 (0.06 – 0.09)</td>
</tr>
<tr>
<td></td>
<td><strong>African-American</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Almost entirely fat/scattered fibroglandular</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heterogeneously dense/extremely dense</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.10 (0.07 – 0.14)</td>
<td>0.15 (0.11 – 0.20)</td>
</tr>
</tbody>
</table>
Table 22. Critical Appraisal of Zhang et al. 2013\textsuperscript{61}

| Selection Bias | Study participants were identified from the CMR database of mammography results (N = 661,866). Follow-up data were linked to breast pathology database, North Carolina Central Cancer Registry and North Carolina State Death Tapes. A substantial proportion of the study population was lost to follow-up because registry and mortality outcomes could not be linked. Comparability of groups: They did not provide baseline characteristics of participants organized according to breast composition category. Instead, characteristics are given for the overall population. Follow-up: Information was only available for about 3.5% of the women identified in the mammography database. There is no documentation of how those that were lost to follow-up compare to the initial eligible study population. | High potential for selection bias |
| Measurement Bias | Investigators used the ACR BI-RADS breast tissue composition assessment routinely used by radiologists in clinical practice. The ordinal scale is inherently unequally applied and unreliable due to its subjectivity. BI-RADS density category documented at entry mammogram, which could have been at any of the participating mammography centers. Outcome: Breast cancer diagnoses identified using pathology database and state registry. There is no mention of efforts to confirm the absence of the outcome, for example, using subsequent screening mammography results for non-cases. Mortality outcomes were identified using state record and ICD codes used to determine the cause of death. | Medium potential for measurement bias |
| Confounding | Stratification used to control for age, one of the strongest risk factors for breast cancer. Race was another strata used in the study given aim was to identify differences breast cancer risk between African-American and Caucasian women. | Appropriate |
| Analysis | Cause-specific cox proportional hazards model to analyze the effect of breast density on breast cancer mortality. They used a stratified Cox model that controlled for age and race. Cumulative incidence used to quantify mortality probabilities over time. | Appropriate analysis |
| Internal Validity | There is a high potential for selection bias in this study, but analysis was appropriate. | Fair |
| External | The study was limited to participants of certain racial groups. | Poor |