Burden and Morbidity of Breast Screening and Diagnostic Work up: Results from a Community Based Approach

BY

Firas Dabbous
B.S., Beirut Arab University, Beirut, Lebanon 2004
M.S., University of Southern Maine, Portland, Maine 2009

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
in the Graduate College of the
University of Illinois at Chicago, 2016
Chicago, Illinois

Defense Committee:

Garth Rauscher, Epidemiology and Biostatistics, Chair and Advisor
Therese Dolecek, Epidemiology and Biostatistics
Michael Berbaum, Institute for Health Research and Policy
Anthony LoSasso, Health Policy & Administration
Kent Hoskins, UIC School of Medicine
Sarah M. Friedewald, Northwestern Medicine

This thesis is dedicated to my parents, Mohammad and Rajaa.

ACKNOWLEDGMENTS

I would like to extend my sincere appreciation and gratitude to my advisor and committee chair Dr. Garth Rauscher for his continuous support, guidance, encouragement and mentoring throughout my PhD program. Dr. Rauscher, thank you for being a great mentor and for spending countless hours working with me on this project. Working with you has made my learning journey an exciting and a joyful one.

I am very thankful to professors, Therese Dolecek, Michael Berbaum, Anthony LoSasso, Kent Hoskins and Dr. Sally Friedewald, for serving on my PhD committee and for their guidance. Special thanks to Dr. Dolecek for all her help and encouragement, Dr. Berbaum for his support with the statistical methods and Dr. Friedewald for her valuable contributions to this work.

This work would not have been completed without the support from the Advocate HealthCare team, Dr. Tom Summerfelt and Tere Macarol who are an integral part of this project.

I would not have achieved this degree without the support of many wonderful people who I will always be indebted to, my uncle Dr. Omar Dabbous and Aunt Olivia, my cousins Hussein and Monique, Wael, Marwan, Mohammad, and my friends, Marwan Rihaoui, Rabih Makki, Ayman Dayeh, Ali Homsi, William Schoellkopf, Ben Booth, Jakcob Kresovich, Katherine Tossas-Milligan, Joelle Hallak and Katie Brewer.

I want to express my deepest appreciation to my parents Mohammad and Rajaa, my brothers Salim, Aboudi and my sister Enas.

Shana, thank you for you continuous support and for being such a wonderful person in my life. I am looking forward to many more experiences with you in the future.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

LIST OF ABBREVIATIONS

CI	Confidence Interval
ISCR	Illinois State Cancer Registry
FFDM	Full Field Digital Mammogram
SFM	Screen Film Mammogram
FP	False Positive
FPR	False Positive Rate
TN	True Negative
FN	False Negative
TP	True Positive
FNR	False Negative Rate
cFNR	Complete False Negative Rate
NPV	Negative Predicted Value
DCIS	Ductal Carcinoma *in-situ*
ADH	Atypical Ductal Hyperplasia
RR	Relative Risk
BCSC	Breast Cancer Surveillance Consortium
nH	Non-Hispanic
FDA	Food and Drug Administration
DMIST	Digital Mammography Imaging Screening Trial
USPSTF	United States Preventive Services Task Force
RCT	Randomized Control Trial
FNA	Fine Needle Aspiration
USA	United States of America
UK	United Kingdom
CNB	Core Needle Biopsy
VA	Vacuum Assisted
BIRADS	Breast Imaging Reporting and Data System
MRI	Magnetic Resonance Imaging
IGF	Insulin-like Growth Factor
NCCN	National Comprehensive Cancer Network
NAACCR	North American Association of Central Cancer Registries
GEE	Generalized Estimating Equations
OR	Odds Ratio
HR	Hazards Ratio
ER-	Estrogen Receptor negative
AHRQ	Agency for Healthcare Research and Quality

SUMMARY

Breast cancer is a major public health problem which inflicts substantial burden and morbidity and is a major cause of death among women. In the last two decades, breast cancer mortality rates have been on the decline which is attributed to improvements in treatment, early detection and diagnostic technologies. Mammography screening is designed to detect tumors at an early stage and plays a major role in reducing mortality and improving breast related health outcomes. Similarly, breast biopsies have been designed to have high accuracy to rule in or rule out breast cancer diagnosis. Many prior studies have examined the effectiveness of mammography screening and biopsy procedures as well as their limitations. The main objective of this thesis is to examine potential limitations with respect to screening mammography and diagnostic breast biopsy. With respect to screening mammography, Aim 1 of this thesis examines the false positive rate (FPR) and burden of diagnostic workup associated with Full Field Digital mammography (FFDM). Full field digital mammography is associated with increased screening accuracy for younger women and those with dense breasts when compared to conventional screen film mammography (SFM), but the increased accuracy may come at a cost of decreased specificity and/or increased diagnostic burden. In addition, Aim 2 examines more generally the potential impact of a false positive (FP) screening result on causing women to delay returning for subsequent screening. Unlike screening where unnecessary diagnostic burden is created by a FP result, in the case of a biopsy, it is the potential for false negative (FN) results that creates an unnecessary diagnostic burden by prompting a second biopsy. Aim 3 examines the probability of a missed cancer detection on image guided core needle biopsies (CNB) using a large and diverse cohort of women from multiple facilities within a large healthcare organization.

For the first two Aims we linked the screening mammogram level data with the Illinois cancer registry data to define FP and TN mammography exams. A screening mammogram was considered FP if the exam had a BIRADS score of (0, 4, 5) without an evidence of breast cancer in the subsequent 12

months after the exam date. True negative mammogram was defined as any screening mammogram with a BIRADS score (1, 2, 3) and without an evidence of breast cancer in the subsequent 12 months after the exam date. In the first aim we compared the FPR, FP burden (work-up rate including additional mammography, ultrasound or Magnetic resonance imaging (MRI) imaging) and FP morbidity (risk of receipt of a biopsy) between FFDM and SFM. The results suggested that FFDM had a similar FPR, FP burden but slightly higher risk of receipt of biopsy compared to SFM. In Aim 2, we examined whether the experience of a FP mammogram delays the return to the subsequent round of screening. We compared the probability of returning to subsequent screening between women who received a FP mammogram compared to a TN mammogram. Our results suggest that there is a compelling evidence that the women who received a TN mammogram had significantly higher chance of returning to their next screening mammogram compared to women with a FP mammogram HR=1.35 (95% CI: 1.34-1.36).

In the third aim, we estimated the risk of a FN finding for stereotactic CNB, vacuum-assisted ultrasound CNB and non-vacuum assisted ultrasound CNB using a large cohort of women from multiple facilities within a single healthcare organization. Additionally, we examined which patient characteristics were associated with increased risk of a FN biopsy. Similar to the first two Aims, we linked the biopsy data with the cancer registry to define true positive (TP) and FN biopsies. A TP biopsy was defined as any biopsy with a malignant finding. A FN biopsy was defined as any biopsy with a benign finding associated with breast cancer diagnosis in the subsequent 12 months following the biopsy date. A complete FN biopsy was defined as any biopsy with a benign or high-risk benign finding associated with a cancer diagnosis in the subsequent 12 months following the biopsy date. In this study, FN rate for image-guided CNBs were 5.4% for stereotactic-guided, 4.8% for vacuum-assisted ultrasound-guided and 3.8% for non-vacuum-assisted ultrasound-guided biopsies. The complete FN rate was 11.3% for stereotactic-guided, 7% for vacuum-assisted ultrasound-guided and 5.1% for non-vacuum-assisted ultrasound-guided

biopsies. We also found that the FN rate and the complete FN rate were higher in minority women, younger women and women with denser breasts.

Although the evidence from this thesis highlights the shortcomings of mammography screening and CNBs, it should not discourage women from receiving mammography screening and following up with suspicious findings. It is beyond doubt that mammography screening has reduced mortality over the past decades. It is important for women to follow the guidelines that were set forth by country-specific experts.

1. BACKGROUND AND AIMS

1.1 Epidemiology of Breast Cancer

Breast cancer is a multifactorial disease with a strong interaction of genes, life style factors and environment (1). While genetic factors include the mutations of BRCA1 and BRCA2, non-genetic risk factors include modifiable and non-modifiable risk factors. Modifiable risk factors include exposure to hormonal therapies such as oral contraceptives, hormone replacement therapy, obesity, smoking, dietary habits, and alcohol intake. Non-modifiable risk factors include age, race/ ethnicity, family history of breast cancer, circulating hormones such as estrogen, Insulin-Like Growth Factor (IGF), and mammographic breast density and bone mineral density. Environmental factors include exposure to radiation prior to the age of 30 and chemicals in food, water, and plastics (2).

Breast cancer is a major public health problem and is a cause of substantial morbidity and deaths with 231,840 expected new cases and 40,290 expected deaths in 2015 in the United States (3). It is the most commonly diagnosed cancer among women and accounts for 14% of all new cancer cases in the U.S. In terms of cancer mortality it attributes to 41000 deaths among women annually, ranking second after lung cancer. Death rates have been on the decline since 1989. For the period between 2007 and 2011, death rates declined annually by 3.2% among women younger than 50 years of age in whites and 2.4% in blacks. Among women who are 50 years or older, mortality rates declined by 1.8% in whites and 1.1% in blacks per year (4). This decline in mortality rates can be attributed to treatment improvements and early detection.

1.2 Mammography Screening

The main premise of screening mammography is to detect breast cancer at an early stage to maximize survival. Therefore it is imperative for screening mammography to have high sensitivity to detect abnormalities when they exist and high specificity to rule out any malignancies when

malignancies do not exist. In light of the benefits and increased awareness of the importance of screening, prevalence of self-reported receipt of mammography within the last two years among women older than 40 years of age has improved from 29% in 1987 to 70% in 2000 but dropped to 65.7% in 2013 (5). An earlier racial disparity in mammography screening between non-Hispanic (nH) Blacks and nH-Whites appeared to narrow but has not disappeared. Hispanics consistently continued to have the lowest mammography rates among race/ethnic groups of women. Generally, women who are in the lower socioeconomic spectrum, i.e. those without health insurance, less than high school education and recent immigrants to the United States are the least likely to have received a mammogram in the past two years compared to their more affluent counterparts (6).

In healthcare, medical technologies are constantly changing. These changes can be in different forms such as new knowledge, treatment or healthcare services and have contributed to extending longevity of life (7). Healthcare providers and patients consider aspects such as benefits and effectiveness, risks and safety and costs before adopting any new technology. In an ideal setting, medical technology is diffused equitably such that risks and benefits are uniformly distributed across all patient groups. Screening mammography is a form of technology that is continuously changing. Full Field Digital Mammography (FFDM) was developed to address some of the limitations of Screen-Film Mammography (SFM) and was approved by U.S. Food and Drug Administration (FDA) in 2000. Digital mammography was introduced to improve the quality of display and to enable contrast adjustment so that the mammography images can be modified according to breast density (8, 9). Another attribute of FFDM mammography is the flexibility in sharing the images and storing them in an electronic format. However, with regards to the overall performance characteristics, FFDM and SFM are similar. In the recent years, FFDM has supplanted SFM mammograms such that the majority (>95%) of the accredited U.S mammography facilities in 2014 were using FFDM units (10).

1.2.1 Full Field Digital vs. Screen Film Mammogram

Despite the quick adoption of FFDM few studies in the USA have compared the effectiveness of film screen and digital mammography. Two studies that have made such comparisons were paired design clinical trials in which patients were screened using SFM and FFDM (9, 11) and another study was community–based (12). The Digital Mammography Imaging Screening Trial (DMIST) included a large sample of women between the ages of 47 and 62 from both Canada and US. Overall, the study found similar diagnostic accuracy of FFDM and SFM; however in sub group analysis, FFDM had significantly higher sensitivity than SFM for younger women (<50 years of age), women with heterogeneously dense or extremely dense breasts and premenopausal or peri-menopausal women. For older women with non-dense breasts, SFM had higher sensitivity than FFDM. Of the exams performed, 14% were recalled, a rate that is higher than what has been reported in literature because women underwent two screening exams, according to the authors (9, 13). Another clinical trial found a non-significant lower detection rate of cancer using FFDM vs SFM (11). Studying a large cohort of women from the Breast Cancer Surveillance Consortium (BCSC) Kerlikowske et al reported an overall similar performance measure between SFM and FFDM. However, recall rates were statistically higher in FFDM compared to SFM (12). European studies conducted in Oslo, London, Italy and the Netherlands reported higher detection rates of cancer as well as higher recall rates when using FFDM compared to SFM (14-17) and studies from Spain reported higher positive predictive value in FFDM vs SFM, though consistent detection rates of ductal carcinoma *in-situ* (DCIS) for both modalities (18, 19).

1.2.2 Mammography and Breast Cancer Mortality

Since the early 1970's, reports have shown the importance of mammography in early detection and reducing morbidity and mortality (20). Since then several randomized control trials (RCT) (21-26) have examined the effect of screening mammography on breast cancer related mortality. The results from these studies were pooled in several meta-analyses and all have shown a significant reduction in

breast cancer mortality. The latest systematic review and meta-analysis was published as part of the United States Preventive Services Task force (USPSTF) update to the 2009 guideline process (27). The meta-analysis included over 600,000 women with median follow-up time from 11.2 to 21.9 years. The authors reported pooled Relative Risks (RR) of breast cancer mortality by decades of age. The pooled RR for mortality was 0.92 (95% CI: 0.75-1.02) for women in their 40's, 0.86 (95% CI: 0.68-0.97) for women in their 50's, 0.67 (95% CI: 0.54-0.83) for women between 60's, and 0.80 (95% CI: 0.51-1.28) for women aged 70-74. Other reviews have estimated similar risk reduction in mortality with RR ranging from 0.79 to 0.85 (Table I).

Table I.

LIST OF REVIEWS AND META-ANALYSES OF THE BREAST CANCER SCREENING TRIALS AND BREAST CANCER DEATHS AFTER 13 YEARS OF FOLLOW-UP

Study	N studies	RR[1] (95% CI)
UK independent Panel Review (28)	11 RCTs	0·80 (0·73–0·89)
Cochrane Review (29)		
All women	9 RCTs	0·81 (0·74–0·87)
Women 50 years or older	7 RCTs	0·77 (0·69–0·86)
Adequately designed RCTs	4 RCTs	0·90 (0·79–1·02)
Sub-optimally randomized RCTs	5 RCTs	0.75 (0·67–0·83)
USPSTF (27)	9 RCTs	
Women in their 40's		0.88 (0.73- 1.003)
Women in their 50's	7 RCTs	0·86 (0.68-0.97)
Women in their 60's	5 RCTs	0·67 (0·54– 0·83)
Women between 70 and 74	3 RCTs	0.80 (0.51, 1.28)
Canadian Task Force (30)		
Women in their 50's and 60's	7 RCTs	0·79 (0·68–0·90)
Duffy et al(31)	2 RCTs	0·79 (0·73–0·86)

[1] Pooled relative risk of mortality

1.2.3 False Positive Mammography

False positive (FP) mammography is a major concern of screening that it not only increases economic burden on women, but it can also impact women's psychological well-being and adversely influence adherence to subsequent mammography among those who experience a FP breast mammogram. A FP mammogram is defined as one that is interpreted as abnormal with evidence of an actionable lesion that requires further diagnostic work-up (usually a diagnostic imaging exam) in women without breast cancer. False positive rates decline with increasing age and lower breast density and may be lower for mammograms performed by technologists and read by radiologists with greater expertise and experience (32-34), but may also be higher in the absence of available prior films for comparison. Women with a history of prior invasive breast work-up and a family history of breast cancer may also experience higher FP rates (35-37).

The risk of experiencing a FP mammogram varies across international regions due most likely to differences in the respective practices and guidelines across regions. In Europe, the cumulative risk of experiencing a FP mammogram was among 50-60 years old women undergoing 10 biennial screens were 19.7% and the cumulative risk of receipt of a biopsy was 2.9% (38). In contrast, the cumulative risk of receipt of a false positive mammogram in the U.S. was 47.3% among women in their 50's who underwent 10 screening mammograms (39). Another study from the U.S. has estimated the cumulative risk of a FP among women who started their mammography screening between the ages of 40 and 59. Hubbard et al. (40) included 169,456 women from seven US registries that participated in the BCSC and estimated the FPR to be 16.3% among first screeners and 9.6% at a subsequent screen. Over a 10 years period, the cumulative risk of receiving a FP mammogram was 52.4% with annual and 37.8% with biennial screening among women who started their screening at the age of 40. Similar findings were reported for women who started their screening at 50. The cumulative risk of receipt of a FP related

biopsy was 7% with annual and 4.8% with biennial for starting age 40 and 9.4% and 6.4% for starting age 50.

1.2.4 Psychological Burden of False Positive Mammogram

The psychological impact among women who experienced a FP test is well documented in the literature. One literature review (41) included 54 studies from 13 countries and examined the psychological impact of FP mammography screening. The majority of these studies in the report were cohort studies and half of the studies used validated psychological measurement scales. While some studies reported a persistent anxiety among recalled women even after receiving negative findings, other studies showed a short-term anxiety. Another meta-analysis published in 2007 that included 23 studies examined the long term effects of experiencing a FP mammogram with respect to mammography screening rates and psychological burden (42). Increased symptoms of distress were associated with FP mammograms in four studies, while no statistical difference was reported in three studies and two studies reported mixed findings. With regards to anxiety, FP was associated with an increased mood of anxiety that was specific to breast cancer. However, FP mammography was not linked to Depression. Studies conducted after the publication of the meta-analysis have also been mixed. A more recent study included women 50-69 years of age in Denmark reported that women with FP had statistically higher negative psychological outcomes at 1 month which persisted up to 36 months after the final diagnosis compared to those with true negative (TN) findings (43). Other studies have reported that the risk of adverse psychological outcomes increased was directly associated with the extent of work up such that women who had biopsies had the highest risk of adverse psychological outcomes, followed by fine needle aspiration (FNA) and early recalls (44, 45).

1.2.5 Economic Burden of False Positive Experience

Few studies have estimated the cost associated with the work-up following a false positive mammogram. Except for a single study (46), all other studies were performed before the availability of

FFDM or at early phases of adoption and did not stratify by screening modality despite the fact that recall rates vary according to the modality used (39, 47-50). Also the cost of an FFDM mammogram is significantly higher than that of the SFM because of expenses associated with purchasing and maintaining the machines and the additional time needed to read and manipulate the digital image (51, 52).

Table II shows a summary of the available studies on the economic burden of the false positive experience. False positive screening mammograms create an additional financial cost to patients and insurers. In a Medicare population, 8.4% of the screened women underwent additional diagnostic breast imaging studies with additional cost per screened patient (per capita) of $361 (in 2004 USD). Breast biopsy costs accounted for an additional $90 per capita (48). Other studies from the U.S. and Europe have also estimated a substantial increase in cost per patient post false positive findings (39, 47-49). Hendersen et al (46) estimated the costs of additional work up among a large cohort of women who were Medicare beneficiaries during a period of time when FFDM was first introduced. Higher recall rates among FFDM exams were reported and costs for any follow-up procedures were observed to be consistently higher but not statistically significant for SFM compared to FFDM for the calendar years 2002-2005 (Table II).

Table II.
STUDIES SHOWING THE COST OF FALSE POSITIVE BREAST SCREENING PROCEDURES

Author	Study Period	Age	Size	Results	Summary
Lee, 2008 (48)	2003-2008	65+	731, 666 women	On average, women accrue $360.89 in work up costs after receiving a positive screen. Breast biopsies account for 25% of the average cost.	Diagnostic workup among women with positive mammograms costs Medicare $679 million annually and false positive costs $250 million
Lidbrink, 1996 (50)	1983-1985	40-64	60,261women in the 1st round, 58,633 in the 2nd round	False positive screening costs (£250, 000) in the first round and (£84 000) in the second round	Substantial cost associated with false positive work-up
Chubak, 2010 (47)	1998-2002	40-80	21,125 women	$499 ($487-$512) more in false positive compared to true negative	Significant cost which substantially contribute to the US healthcare spending
Henderson, 2012 (46)	2001-2005	66+	354,731 Exams	Higher cost of subsequent FFDM vs SFM mammograms in 2001-2004. In 2001, for women receiving follow up biopsies the costs were $566.5 for SFM and $292.6 for FFDM, respectively	Overall, non-significant differences in follow-up cost after screening mammogram between SFM and FFDM. Higher cost associated with false positive compared to true negative
Poplack, 2005 (49)	1996-2000	All ages	99,064 women	Per capita costs per women who underwent screening only, diagnostic examination and biopsies were $99, $286 and $993, respectively.	At individual level, high costs were incurred by women who underwent diagnostic examination and biopsy work up.
Elmore, 1998 (39)	1983-1995	40-69	9762 mammograms and 10,905 clinical breast exam	For every $100 spent on breast screening, an additional $33 is spent on work up to evaluate false positives	Highlights the need to develop ways to lower false positive and therefore reduce psychological and economic burden

1.2.6 Impact of False Positive on subsequent screening

False positive results have been associated with increased anxiety and worry and increased anxiety has been found to last for up to three years after the FP was resolved as normal (41, 43). These results suggest the possibility that the FP experience could lead women to alter their screening behavior, either by delaying their next screening mammogram or forgoing them altogether. Studies that have examined the potential impact of a false positive screening mammogram on subsequent screening mammography guideline adherence have yielded inconsistent findings. Several studies have found that re-screening rates were actually higher among women who experienced a FP as opposed to a TN mammogram (53-56). Other studies found no difference in re-screening behavior (57-61), and yet others reported lower re-screening rates among FP than among TN (62-64). A 2007 meta-analysis found that in Europe and Canada, women who experience a FP screening mammogram were less likely to return for their next screen compared to women experiencing a TN screen. Conversely, among women in the U.S., FP results were associated with greater subsequent screening mammography adherence (42). The inconsistency in results between Europe and USA may be attributed to variations in screening practices such as screening intervals are shorter in the US than in Europe, higher emphasis on accuracy in Europe by using double reading which results in 3 to 5% lower recall rates compared to the U.S. and differences between national programs in Europe and both public and private screening providers in the U.S. In addition, increased acknowledgement of the potential burdens that screening mammography places on women due to false positive results, and the recent debate surrounding the effectiveness of screening mammography, may alter how a FP result impact the timing of a woman's next screen.

1.2.7 Over-diagnosis

Over-diagnosis of breast cancer is defined as a screen detected cancer which can be invasive or non-invasive that would not have otherwise come to the attention of the woman during her lifetime. Over-diagnosis is a side effect of screening for early forms of disease and it can occur under different

scenarios, 1) if mammography screening detects cancers that will never become symptomatic or progress during a woman's life or 2) if cancer was detected early but the woman dies from other causes before the development of symptoms (65). Currently, there is no biological test to identify which of the screen detected cancers are life threating and which are not. Therefore, all diagnosed cancers are treated accordingly which leads to additional burden and morbidity to the woman if the screen detected cancer was truly over diagnosed. A panel of experts from the United Kingdom (U.K.) acknowledged the existence of over diagnosis with the caveat that there is a great deal of uncertainty surrounding the extent of over diagnosis (due to difficulty identifying when over-diagnosis actually occurs). (28).

An estimate of over-diagnosis would require the ability to differentiate those screen detected lesions that would progress on a malignant disease course and ultimately to cause-specific death from those that would not. A literature review identified 13 studies from multiple countries across Europe and reported an over-diagnosis rate ranging between 0% and 54% (65). Two reviews that estimated over-diagnosis using data from clinical trials the U.K. independent panel review (28) and the Cochrane review (29). The U.K. independent panel review (28) included two RCTs (26, 66) that did not screen the control group at the end of the screening program, and reported that 10.7% (95% CI: 9.3-12.2) of all cancers diagnosed over the entire follow-up period were over-diagnosed, and 19.0% (95% CI: 15.2-22.7) of all cancers detected during the screening period among women invited to screening were over-diagnosed. The Cochrane review (29) included 7 trials (23, 66-69) and estimated a higher over-diagnosis rate of 29% (95% CI: 23-35) than that reported in the U.K. review. Myers et al. (70) published a systematic review on the benefits and harms of mammography screening and in addition to the aforementioned meta-analyses they included 17 cohort studies that estimated over-diagnosis, with estimates similar to what was reported in the Puliti el al (65) review paper. Both reviews did not report a pooled estimate from these studies due to study heterogeneity with respect to study design,

inclusion/exclusion of DCIS, adjustment of lead time and accounting for breast cancer incidence secular trends.

1.3 Breast biopsy

Breast core needle biopsy is a widely accepted technique that is used to rule in or rule out breast malignancies. It has been recommended as an alternative to open surgical biopsies which are associated with increased humanistic and economic burden (71, 72). Image guided biopsy is often used after the detection of an abnormality at mammography screening or physical breast examination. In general, women with abnormalities identified at screening receive follow-up imaging that may include diagnostic mammography, ultrasound and Magnetic Resonance Imaging (MRI) before the receipt of a breast biopsy if indicated. In the U.S., roughly 50% of women who receive annual mammography screening over 10 years will undergo at least one additional imaging and of these 7%-17% will have at least one biopsy (39, 40) with the majority of biopsies (77%) resulting in the diagnosis of a non-malignant lesion that does not require further work up or treatment (73).

There are three biopsy techniques currently available to sample cells from suspicious breast lesions: fine needle aspiration (FNA), core needle biopsy (CNB) and open surgical biopsy (incisional or excisional breast biopsy). Fine needle aspiration was the main diagnostic procedure for non-palpable lesions in the 1980's; it was however criticized for its modest sensitivity 92.7% (95% CI: 92.1-93.3) and specificity, 94.8% (95% CI: 94.3-95.2) (74) and high proportion of procedures with insufficient samples (35%) (8). Therefore, the diagnostic direction was shifted to CNB which has higher sensitivity and specificity than the FNA.

CNB is a procedure that requires the insertion of a hollow probe to remove small samples of breast tissue. The lesion can be located via different types of imaging guidance such as stereotactic, ultrasound and MRI. Although associated with high sensitivity 97% (95% CI: 95-99), CNB is also associated with high histological underestimation rate, with as many as 40% of CNB that diagnose a

lesion with Atypical Ductal Hyperplasia (ADH) later upgraded to malignant upon a subsequent biopsy. In addition, 15% of CNB that diagnose a lesion with in-situ breast cancer are later upgraded to invasive breast cancer upon subsequent biopsy (75). Core needle biopsy can also result in collection of an insufficient amount of tissue from 5-10% of non-palpable lesions (76). For these reasons, add-on devices were developed to address some of the limitations of CNB.

The FDA approved image-guided vacuum assisted (VA) biopsy in 1995 to increase CNB sensitivity and to reduce the burden associated with multiple insertions and the probability of epithelial displacement (77, 78). Briefly, VA biopsy works by inserting an image (MRI, Ultrasound or X-Ray) guided hollow probe through a small cut in the breast with a knife inside the probe to cut the specimen. Multiple samples can be obtained through one insertion and compared to core biopsy, larger amounts of tissue can be obtained (79). By reducing the likelihood of histological underestimation, VA biopsy reduces the need for a second, usually surgical (incisional or excisional) biopsy.

Open (excisional or incisional) biopsy is considered the gold standard for evaluating suspicious lesions because of its very high accuracy in diagnosing breast lesions. The procedure collects a large amount of samples and in some cases completely removes the lesion; it is however only applicable to the evaluation of palpable masses. This surgical procedure is associated with increased economic burden and morbidity on patients (80, 81) which can be reduced with the use of CNB. As a result, the National Comprehensive Cancer Network (NCCN) recommended CNB as the preferred alternative to open surgical biopsy (82). Similarly, in 2009, another interdisciplinary group of physicians specializing in the diagnosis and treatment of breast disease stated that CNB is an “optimal initial tissue-acquisition method and the procedure of choice for image-detected breast abnormalities” (72). Despite these recommendations and the benefits of CNB, excisional biopsy is still used in 24-36% of patients with suspicious lesions (83-86), a proportion that is much higher than the proportion of patients who have a preference for excisional biopsy or that have an unfavorable lesion position for CNB (72).

Consequently, image-guided CNB biopsies have been used as an alternative to surgical biopsies for the histologic assessment of non-palpable breast lesions in recent years. The Agency for Healthcare Research and Quality published a comparative effectiveness report of core-needle and open surgical biopsy for the diagnosis of breast lesions in 2009 (87) which was updated in 2014 (88). Similar findings were reported from both reports. Most of the included studies compared the results of CNB to results obtained from open surgical biopsies and/ or patient follow up for at least six months to estimate the accuracy of the CNB by estimating sensitivity, specificity and underestimation rates of DCIS and high risk benign lesions. These performance measures were estimated for each biopsy by imaging modality and whether VA was used or not. The latest AHRQ report included 27 cohorts with 16,287 biopsies that used US CNB, 12 cohorts with 1,543 VA-US CNB, 37 cohorts with 9,535 stereotactic CNB, 43 cohorts with 14,667 VA- stereotactic CNB, two cohorts with MRI CNB and one cohort with VA-MRI CNB. Table III summarizes the performance characteristics for CNB by imaging guidance and VA. The reports concluded that image guided CNB was an accurate method for the diagnosis of breast cancer and could be used as an alternative to open surgical biopsy. However, the strength of evidence was considered weak because of the poor reporting and low internal validity of the studies on which the report was based. In addition several questions remained un-answered including lack of information on patient and tumor factors that might affect the accuracy of CNB.

Table III.

POOLED PERFORMANCE CHARACTERISTICS ESTIMATES OF IMAGE-GUIDED CNB AS REPORTED IN THE UPDATED AHRQ REPORT (88)

Performance Characteristics	Stereotactic CNB[1]	VA-Stereotactic CNB[1]	Ultrasound CNB[1]	VA- Ultrasound CNB[1]	MRI CNB[1]	VA-MRI CNB[1]
Sensitivity	0.97 (0.95-0.98)	0.99 (0.98-0.99)	0.99 (0.98-0.99)	0.97 (0.92-0.99)	0.90 (0.57-0.99)	1.00 (0.98-1.00)
Specificity	0.97 (0.96-0.98)	0.92 (0.89-0.94)	0.97 (0.95-0.98)	0.98 (0.96-0.99)	0.99 (0.91-1.00)	0.91 (0.54-0.99)
DCIS Underestimation	0.26 (0.19-0.36)	0.11 (0.08-0.14)	0.38 (0.26- 0.51)	0.09 (0.02-0.26)	--	--
High Risk Underestimation	0.47 (0.37-0.58)	0.18 (0.13-0.24)	0.25 (0.16-0.36)	0.11 (0.02-0.33)	--	--

[1] Core Needle Biopsy;

Estimates are presented as percentages with 95% confidence intervals in parenthesis;

1.4 Research Contribution

Mammography has been established as a successful screening and diagnostic tool to detect and diagnose breast cancer. Similarly, image-guided biopsy techniques are considered the gold standard in diagnosing breast cancers. These techniques along with improved treatment options have played an integral part in reducing breast cancer deaths. Although the benefits have been established, mammography screening and image-guided biopsies have been criticized due to major limitations that have sparked debate in literature whether benefits outweigh harms. With this project we aim to expand the research on burdens and morbidity of screening and diagnosing breast cancer. The first aim is to examine whether Full Field Digital Mammography results in greater burden and morbidity to the patient when compared to Screen-Film Mammography. This study will expand on previous studies that have quantified the burden and morbidity of mammography screening by stratifying the results by Full Field Digital Mammography vs Screen-Film Mammography and by patient characteristics such as by age, breast density, and other factors. The second aim examines whether receipt of a false positive screening mammogram is associated with delays in obtaining the subsequent screening mammogram. Several studies have estimated the rate to return to screening after receipt of false positive but with the exception of one study, all studies were done in early 2000's and reported inconsistent findings. Similar to study 1, this study will also report stratified results by patient and clinical characteristics. The third aim will examine whether the diagnostic accuracy is higher in VA- image-guided core needle biopsies compared to non-VA image-guided core needle biopsies and whether the underestimation of DCIS and invasive cancer is lower in vacuum-assisted biopsies vs. non-VA biopsies. This study will add to the literature on estimating the performance characteristics of image guided biopsies using community based and diverse dataset from multiple facilities. Additionally, this study will estimate the complete false negative rate, an outcome that includes biopsies with benign and high risk benign findings associated with a cancer diagnosis in the subsequent 12 months following the biopsy date. Complete false negative rate is a measure of the missed cancer detection which is not estimated in the majority of the studies

included in the AHRQ report. All of the three projects will use data from a single health care organization which was linked to the Illinois State Cancer Registry to rule in/out cancer diagnosis.

1.5 Specific Aims

Aim 1 is to examine whether FFDM results in greater burden and morbidity to the patient when compared to SFM. While FFDM is more accurate than SFM in younger women and women with denser breasts, SFM may or may not be more effective among older women. Digital mammography has been associated with slightly higher recall rates in previous studies. We will explore whether FFDM is associated with increased diagnostic burden and morbidity when compared to SFM, for specific patient groups defined by age, breast density, and other factors. Diagnostic burden will be defined in terms of the total number of procedures obtained, the number of needle and excisional biopsies obtained, and the time from index mammogram to diagnostic resolution as a negative finding.

Aim 2 is to examine whether receipt of a false positive screening mammogram is associated with delays in obtaining the next scheduled screening mammogram. Women may be discouraged from receiving a subsequent mammogram after experiencing a false positive mammogram. Recalls and work ups may undermine their confidence in the benefits of mammography. Alternatively, false positive mammograms may lead to worry and subsequent over-utilization of medical care and mammography. We will examine whether, among patients who are screened and found not to have breast cancer, a false positive result is associated with delayed subsequent screening mammography, and whether greater diagnostic burden is associated with greater delay. Among patients subsequently diagnosed with breast cancer, we will also examine whether receipt of a false positive mammogram is associated with higher rate of interval and late stage breast cancer and if so, whether these associations are mediated by delayed subsequent rescreening.

Aim 3 is to examine whether the diagnostic accuracy is superior with vacuum-assisted image guided core needle biopsies compared to non-vacuum-assisted image guided core needle biopsies and whether the underestimation of DCIS and invasive cancer differs by vacuum assistance status. Vacuum-assistance was approved to improve the diagnostic accuracy for core needle biopsies. It also has the potential to replace surgical biopsies in the diagnosis of breast cancer. Our hypothesis is that

vacuum-assisted core needle biopsies have better accuracy in detecting breast cancer and have lower underestimation rates compared to non-vacuum-assisted core needle biopsies.

2. Study 1:
Burden of false positive in mammography screening: Results from a large community based study

2.1 Introduction

Screening mammography is based on the premise that early detection of breast tumors can maximize survival. To optimally achieve this goal, mammography screening must be highly sensitive to detect tumors when they exist and highly specific to avoid false diagnoses and subsequent physical, emotional and economic burdens associated with a false positive (FP) event. Although screening mammography has been shown to reduce mortality (28, 89, 90), the occurrence of FP is considered a major limitation of mammography screening (39, 91). A FP mammography outcome is defined as a mammogram that is interpreted as abnormal with evidence of an actionable lesion that requires further diagnostic work-up such as diagnostic imaging (including diagnostic mammogram, ultrasound or MRI) or biopsy in women without breast cancer. False positive rates decline with increasing age and lower breast density and may be lower for mammograms performed by technologists and read by radiologists with greater expertise and experience (32-34), but may be higher in the absence of available prior films for comparison. Women with a history of prior breast work-up and a family history of breast cancer may also experience higher FP rates (35-37).

Screen-Film Mammography (SFM) was the mainstay technology for breast cancer screening until Full Field Digital Mammography (FFDM) was introduced in early 2000's. Digital mammography was developed to address some of the limitations of SFM by enhancing the quality of imaging to improve early detection. In addition, FFDM are easier to share and store in an electronic format (9). Despite the quick adoption of FFDM, few studies in the USA have compared the effectiveness of SFM and FFDM (9, 11, 12). The DMIST trial included a large sample of women between the ages of 47 and 62 from both Canada and US. Overall, the study found similar diagnostic accuracy of FFDM and SFM mammography; however in sub group analysis, FFDM had significantly higher sensitivity than SFM for younger women (<50 years of age), women with heterogeneously dense or extremely dense

breasts and premenopausal or peri-menopausal women. For older women with non-dense breasts, SFM had higher sensitivity than FFDM (9, 13). Another clinical trial found a non-significant lower detection rate of cancer using FFDM vs. SFM and a lower recall rate in FFDM than SFM (92). Using a large cohort of women from the Breast Cancer Surveillance Consortium (BCSC), Kerlikowske et al. reported overall similar performance measures between SFM and FFDM. However, recall rates were statistically higher in FFDM compared to SFM (12). Studies conducted in Oslo, London, Italy and the Netherlands reported higher detection rates of cancer and higher recall rates using FFDM (14-17); however, studies from Spain reported higher positive predictive value in FFDM vs SFM suggesting that detection rate of cancer is higher in FFDM, though consistent detection rates of DCIS (18, 19).

Similarly, few studies have examined the extent of work-up after receipt of positive screening exam (93-96). Apart from Hubbard et al. (95) other studies preceded the wide adoption of FFDM. Actions associated with FP follow up can result in humanistic and economic burden on the patients which involves additional costs, diagnostic imaging and in some instances biopsies in addition to stress and anxiety. Therefore, it is important to understand the potential influence of the introduction of FFDM on the burden and morbidity post screening in a large community setting.

The aim of this study is to examine whether false positive rates differ between the uses of FFDM versus SFM within the screening mammography facilities of a large community-based healthcare organization. In addition, the cumulative burden and morbidity associated with false positive screening mammography is compared.

2.2 Methods

The study was conducted using two large population-based health data sources. Mammography screening data on women were obtained from a large Health Care Organization with multiple facilities in the Greater Metropolitan Chicago Area (97). Facilities within this healthcare organization used PenRad to collect radiology information and patient characteristics (98). PenRad was first introduced in 2001 and had been implemented at all facilities by 2005. Breast cancer incidence data were obtained from the Illinois State Cancer Registry (ISCR) which collects information on all incident

cancer cases in the state of Illinois (99). Annually since 1999 ISCR has been awarded gold certification as an incidence registry, the highest quality registry status by the North American Association of Central Cancer Registries (NAACCR).

The radiology data set included information pertaining to patient-level data on demographic characteristics and risk factors, and exam-level data on procedure types and results, in addition to unique identifiers for facility and interpreting radiologist for screening and diagnostic procedures that were performed between January 1st, 2001 and December 31st, 2014. Family history was self-reported and was defined as none, weak, moderate and strong. Age was determined by taking the difference between date of index mammogram and date of birth. Race/ethnicity was self-reported as Non-Hispanic (nH) White, nH-Black, Hispanic, other and unknown. Personal history of prior biopsy was defined as present if a prior biopsy existed in the radiology dataset or if it was self-reported. Time since last mammogram was defined as 9-18 months, 19-30 months, >30 months and no prior mammogram based on the radiology dataset. Breast density was defined following the American College of Radiology classification as entirely fatty, fibroglandular density, heterogeneously dense and extremely dense.

Each mammogram was interpreted by the reading radiologist and was given a score using the American College of Radiology Breast Imaging Reporting and Data System (BIRADS). BIRADS assessment for screening and diagnostic mammography ranges from 0 to 5 such that 0= need additional imaging evaluation, 1= negative finding, 2= benign finding, 3= probably benign finding, 4= suspicious abnormality, and 5= finding highly suggestive of malignancy.

A linkage of the screening mammograms performed between 2001 and 2010 to ISCR breast cancer cases resulted in a 98-99% match rate for diagnosis years 2001-2011. To allow 12 months of follow up for cancer diagnosis, we restricted our analytic dataset to include bilateral screening mammograms that were performed January 1, 2001 and December 31, 2010. The unit of analysis was the screening mammogram. Based on a comparison of the screening mammogram interpretation (normal vs. abnormal) and cancer status within 12 months of the screen, screening

mammograms were defined as true positive (TP), true negative (TN), false positive (FP) and false negative (FN) screens.

2.2.1 Measures and Definitions

Screening mammograms were defined using the BCSC criteria (100). In brief, the BCSC defines screening mammograms as any bilateral mammogram that was indicated for screening. Screening exams which were preceded with any radiologic exam in the 9 past months were excluded or exams from women with prior history of breast cancer, mastectomy, breast implants, and exams with BIRADS 6 were excluded. In the case of multiple exams on the same day, only the first exam in the sequence will be included. A false positive screen was defined as a screening mammogram with an abnormal interpretation (BIRADS 0,4,5) conducted on a women without a breast cancer diagnosis in the subsequent 365 days. The index exposure was defined as the receipt of a FP mammogram, and the referent condition was defined as the receipt of a TN screening exam (BIRADS 1, 2, 3). Screening exams that were associated with a breast cancer diagnosis in the subsequent 365 days (TP or FN exams) were excluded from these analyses.

False positive diagnostic burden was defined as the total number of diagnostic imaging (diagnosis mammogram, ultrasound and MRI) and biopsy exams in the 12 months following a false positive mammogram or before the next screening mammogram, whichever came first. False positive morbidity was defined as the receipt of one or more core or surgical breast biopsies within 365 days of the false positive mammogram. True negative exams were assumed to have no additional work-up post index mammogram in the 12 months follow-up period.

2.2.2 Defining the Sub-cohort

The study design is depicted in Figure 1. The date of FFDM adoption at each facility was defined as the date of the first FFDM examination at each site. Facilities which performed only one screening modality during the study period were excluded. Screen-Film exams that were performed more than 24 months prior to the date of adoption of FFDM and FFDM mammograms that were

performed more than 30 months after the date of adoption of FFDM were excluded from the analysis in order to limit any potential effects of secular trends on our results. Consistent with the literature, FFDMs conducted within the first 6 months of adoption of FFDM screening were excluded, to limit the potential effect of a learning curve in use of FFDM mammography (Figure 1). All SFMs that were performed after the date of adoption of FFDM mammography were excluded to limit the effect of late adopters. To examine whether the selection of the time windows for both modalities accounted for the secular trends and learning curve effects, we performed a set of sensitivity analyses by including different scenarios. The first set of analyses were designed to examine whether the learning curve effect was adequately accounted for by excluding all FFDMs exams that occurred between 0 and 12 months and shifted the time frame for FFDM to 1) 12 to 36 months in one analysis and 2) 18 to 42 in a second analysis while including all SFMs that were performed between -24 and 0 months. The second set of sensitivity analyses were designed to examine whether the secular trend effects were accounted for appropriately by including one year of data for each modality in the analyses. In these analyses, all SFMs that were performed between -12 and 0 months were included and varied the time frame for FFDMs such that we included all FFDMs that were performed between 6 and 18 months, 12 – 24 and 18-30 months.

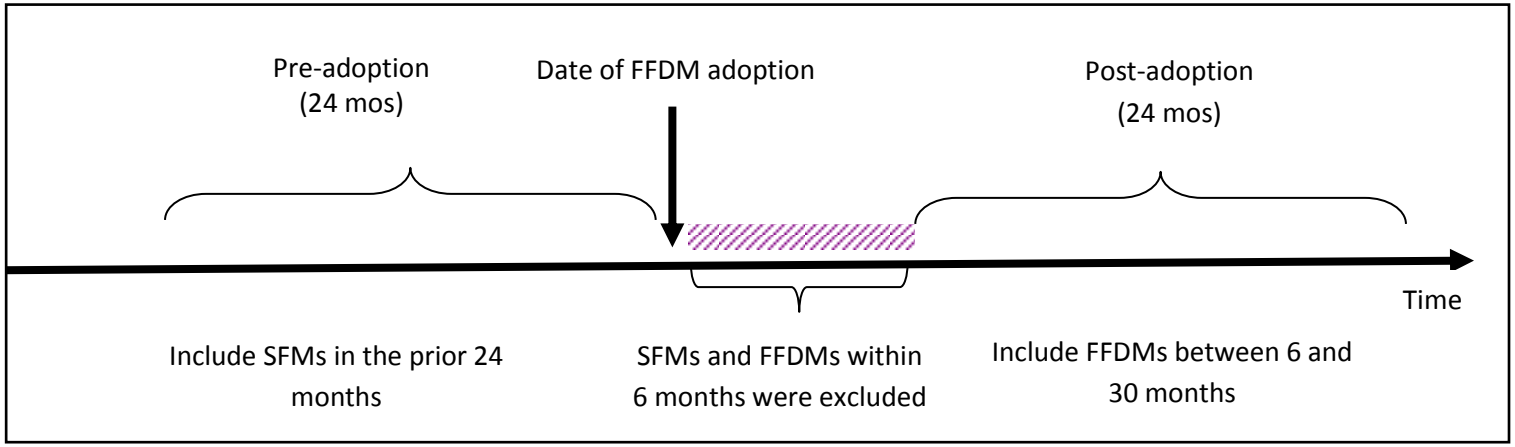


Figure 1: Study design overview FFDM vs SFM. The arrow represents the date at which FFDM was first introduced at a given facility. The shaded area represents the six months period after adoption of FFDM during which all exams were excluded from the analysis. Post-adoption period included only FFDM exams that were performed between 6 and 30 months from adoption. Pre-adoption period represents a two years period which included only SFM.

2.2.3 Statistical Analysis

The distribution of each dependent variable by imaging modality (SFM vs. FFDM) were compared across levels of patients and clinical characteristics using chi-square, fisher exact tests and trend tests as appropriate. To examine whether FPR and FP morbidity varied by imaging modality, Generalized Estimating Equations (GEE) models with a binomial distribution and logit link were fitted. In addition, GEE models with negative binomial distribution and logit link function were used to estimate FP burden by imaging modality. All models were adjusted for screening modality, decade of age, race/ethnicity, family history of breast cancer, mammographic breast density, prior biopsy, time since adoption of FFDM mammography, time since last screening mammogram and facility of exam. To estimate the stratum-specific risks by modality, product terms were included a-priori in all models between screening modality and age, breast density, prior biopsy, time since last mammogram, race/ethnicity and family history. Other product terms among patient characteristics were included in the model at an alpha level of 0.05. All models used an exchangeable variance covariance structure to account for correlations within patients. For all multivariable models, main effects and interaction terms between modality and patient characteristics were forced to remain in the models at all times. The models were allowed to go through a backward selection process to exclude interaction terms between patient characteristics not including modality if these were not statistically significant at an alpha of 0.05. To estimate the average population risks and rates by type of screening modality, marginal standardization in STATA was used (margins command).

2.3 Results

The evaluation was based on 226,210 screening exams (117,281 SFMs and 108,929 FFDMs) from 123,308 women. Women undergoing SFM were similar to FFDM in age, race/ethnicity, breast density, family history of breast cancer, menopausal status, availability of comparison film and receipt of prior biopsy as shown in (Table IV). However, women in the SFM group were more likely to be first screeners (27.1%) than women who received FFDM (20.0%).

2.3.1 False Positive Rate Comparisons

Overall, the FPR was 0.9% percentage point higher for SFM than FFDM in the unadjusted analysis (12% vs 12.9% for FFDM vs SFM, respectively). In the fully adjusted model the risk difference in FPR attenuated to 0.5% (12.8% vs. 12.3% for SFM vs. FFDM, respectively, p=0.02).

In general and regardless of screening modality, the FPR was highest among nH Black women, decreased with increasing age, increased with greater breast density, and was lowest among women with less time between screens as well as those receiving mammograms that were interpreted with the aid of a comparison film (Table V).

The FPR was generally higher for SFM compared to FFDM within strata of patient and clinical factors but the differences were generally in the range of one percentage point or less. Most notably, compared to SFM, FFDM was associated with a lower FPR for women with less dense breasts, for women undergoing their first screening mammogram, and for screening mammograms that were interpreted without the aid of a comparison film.

2.3.2 False Positive Burden (Work-up)

The overall average work-up per 1000 screening mammogram was 240±75 and 220±70 for SFM and FFDM (P<0.001), respectively. In the fully adjusted model, the predicted overall rate of work-up was 233 additional exams per 1000 SFM exams compared to 225 additional exams per 1000 FFDM exams (P=0.001).

For the vast majority of selected characteristics and respective strata, the rate of work-up for the SFM modality was observed to exceed the observed rate for the FFDM modality. The magnitude of the difference was particularly striking for nH Black and Hispanic women and time since last screen of greater than 19 months and first screens. Conversely, the FPBR was substantially higher for FFDM than SFM among women with 9-18 months since last screen (Table VI).

Among only false positive exams shown in (Table VII), the average number of additional diagnostic exams per 1000 false positive mammograms was 1880 for SFM and 1830 for FFDM, respectively. After accounting for patient and facility characteristics the rates of additional diagnostic

work-up per false positive mammogram were similar between the two modalities (1860 per 1000 false positive screens for both SFM and FFDM). As shown, work-up rates did not differ between the two modalities within strata of patient characteristics.

2.3.3 Risk of Biopsy

The total number of biopsies performed for the duration of the study was 3,163, of which 1,964 were in SFM and 1,199 were in the FFDM. Women with denser breasts, younger women, women with family history of breast cancer, first screeners, women with prior history of biopsy and without availability of prior screen film for comparison were more likely to have received diagnostic biopsy. As shown in (Table VIII), results from the adjusted model suggest that the overall risk of receipt of biopsy was 1.24% (95% CI: 0.74-1.73) with FFDM compared to 1.04% (95% CI: 0.7-1.6) with SFM, (p-value=0.05). Stratum-specific risk comparison estimates of biopsy were consistently higher in FFDM compared to SFM. However, the only significant difference was observed among nH-whites, the age group 50-59 years, no family history, first screeners, no history of prior biopsy, and those with available comparison film.

Table IX shows the overall risk of biopsy after a FP mammogram was 2% higher in FFDM 10.1% (95% CI: 6.4-13.8) compared to 8.1% (95% CI: 5.1-11.0) in SFM (p=0.002). Similarly, patient characteristic findings consistently indicated the risk of biopsy to be higher for FFDM compared to SFM.

2.4 Discussion

This study showed that SFM was associated with slightly higher risk of FP and rate of work-up but lower risk of biopsy compared to FFDM in a community based sample of 123,308 women who received their mammography screening at a single large healthcare organization. The results also revealed that regardless of type of modality, the risk of FP result, work-up and biopsy are highest among first screeners and among younger age groups and women with radiographically denser breasts. In addition, there was an observed disparity in false positive and work-up rates, such that

nH-whites had lower false positive rates and lower rate of work-up compared to the Black and Hispanic minorities. In addition, African Americans were at higher risk of receipt of biopsy compared to nH-Whites.

Most studies that have examined the performance characteristics of mammography screening found that the majority of recalled exams are FP exams which do not result in cancer diagnosis in the subsequent months. To rule out any possible malignancy, women without breast cancer who are nonetheless recalled due to an abnormal screening mammogram undergo an array of diagnostic exams which in some instances lead to invasive or surgical biopsies. Generally, 10% of exams are recalled for work-up and 20% of these lead to biopsy work-up. Approximately one in five biopsies will result in a cancer diagnosis. These additional exams come at a high price: women who receive a FP result may incur additional financial costs, lost time and psychological and physical morbidity (41, 43, 47).

False positive mammograms have been associated with increased worry and anxiety which has been found to last for up to three years after the false positive was resolved as normal (41, 43). In this cohort the FPR was found to be 12.3% for FFDM and 12.8% for SFM. These estimates are slightly higher to those reported (9.6% for FFDM and 9% for SFM) in an earlier study utilizing data from the BCSC (12). These differences can be attributed to differences in the study designs between our study and the BCSC study. The BCSC study included data from multiple registries while our study included date from a single healthcare organization. Overall, FPR vary across cohorts and between countries. Compared to the USA, studies from Europe have reported lower FP rates. In the UK, the breast screening program reported false positive rates to range from 7.9% at first screens and 3.2% for frequent screens (101). In Italy false positive rates were 4.5-4.7% (102) and studies from Canada have shown that recall rates can be 9.5% for fist screeners and as low as 4.6% for subsequent screeners (103). These inconsistencies in FPRs between the USA, Europe and Canada may be attributed to variations in screening practices such as screening intervals are shorter in the US than

in Europe, and higher emphasis on accuracy in Europe by using double reading and differences between national programs in Europe and both public and private screening providers in the US (42).

It is very important from the patient and societal perspective to reduce the burden and morbidity on women undergoing mammography screening as well as recognize the risks of receiving a FP mammogram among women complying with breast cancer screening recommendations. Roughly one in eight women without breast cancer received a FP screening mammogram in this study, and this did not vary between the screening modalities. Women who received a FP mammogram received an average of nearly two additional diagnostic procedures and the rate of work-up to resolve the positive mammogram was identical for both modalities. These results are consistent with a study which pooled data from five mammography registries from the BCSC and did not find a substantial difference between SFM and FFDM in the rates of work-up after receipt of a positive result (95).

FP morbidity as measured by the probability of receipt of any type of biopsy after a FP mammogram was not different in FFDM compared to SFM (1.2% vs 1.12%, P=0.15). However, among women with a FP screen, FFDM was associated with higher probability of receiving a biopsy (8.6% vs 10.3%, P=0.01). These results are similar to what has been reported from the Nova Scotia Breast Cancer Screening program which included a total of 608,088 screening mammograms. Similar to our study, the authors of this study reported the probability of receipt of biopsy by age group and the risk of biopsy across all age groups was consistently higher in FFDM compared to SFM (104).

This study has some limitations. First, we used data from a single, large healthcare organization, and while screened patients appeared highly likely to obtain diagnostic follow-up within the organization, some loss to follow-up is inevitable as patients might choose to receive care outside of the organization causing an underestimation of the count of diagnostic procedures obtained. However, any small underestimation is not likely to be differential by screening modality. Finally, this study did not account for radiologist characteristics such as experience in mammography

interpretations which can impact the performance characteristics of mammography in a way that more experienced radiologists tend to have better outcomes than those with less experience.

We went to great efforts to restrict our sample in order to limit the potential impact of secular changes in application of screening, learning curves in the adoption of FFDM, and late adopter effects. We included SFM exams that were performed within two years from the date of adoption of FFDM at each facility, in included only FFDM exams performed 6 months after the adoption of FFDM at each facility and up to 30 months after adoption. By restricting our timeframe to 4.5 years we intended to limit any secular effects on our estimates. By excluding FFDM performed in the first 6 months we intended to remove learning curve effects, and by excluding FFDM exams within the first 6 months of FFDM adoption we intended to limit late adopter effects. To further account for any potential effect of time on the outcomes of interest, we included months since adoption in the model. To validate our approach, a set of sensitivity analysis were conducted and the results were consistent with the main findings from our base case scenario such that the difference in FPR for SFM ranged from 12.2% to 13.0% and the FPR for FFDM ranged from 11.9% to 12.8% (Table X).

We found a modest difference in the rate of work-up and false positive rates between SFM and FFDM. The risk of biopsy was slightly higher in FFDM vs SFM. This is suggesting that the transition to Full Field Digital Mammography did not impact recall rates and therefore did not introduce additional burden to the patients.

Table IV.

DISTRIBUTION OF MAMMOGRAMS BY SELECTED CHARACTERISTICS AND BY SCREENING MODALITY (N=226, 210)

	SFM[1]		FFDM[2]	
	N[3]	%	N[3]	%
Overall	117,281	51.8	108,929	48.2
Ethnicity				
nH White	54,638	46.6	51,753	47.5
nH Black	39,541	33.7	36,352	33.4
Hispanic	2,824	2.4	3,151	2.9
Other	6,232	5.3	8,301	7.6
Unknown	14,046	12	9,372	8.6
Age				
<40	3,897	3.3	3,091	2.8
40-49	33,519	28.6	29,457	27
50-59	33,675	28.7	32,102	29.5
60-69	23,921	20.4	23,598	21.7
70-79	17,037	14.5	15,465	14.2
80+	5,232	4.5	5,216	4.8
Breast Density[4]				
Fatty	8,671	7.4	9,068	8.3
Scattered	47,410	40.5	43,368	39.8
Heterogeneous	49,829	42.5	46,364	42.6
Dense	11,251	9.6	10,112	9.3
Family history				
None	79,150	67.5	73,578	67.6
Weak	18,231	15.5	17,180	15.8
Moderate	14,142	12.1	12,855	11.8
Strong	5,758	4.9	5,316	4.9
Time since last screen				
First Screen	31,770	27.1	21,767	20
9-18 months	58,181	49.6	59,600	54.7
18-30 months	18,838	16.1	17,136	15.7
>30	8,492	7.2	10,426	9.6
Prior biopsy				
Yes	19,515	16.6	18,067	16.6
None	97,766	83.4	90,862	83.4
Comparison Film				
Yes	95,462	81.4	87,446	80.3
None	21,819	18.6	21,483	19.7
Parity				
Nulliparous	14,027	12	14,390	13.2
Parous	99,309	84.7	91,294	83.8
Unknown	3,945	3.4	3,245	3
Menopause				
Pre-menopausal	35,903	30.6	32,161	29.5
Post-menopausal	81,378	69.4	76,768	70.5

[1] Screen Film; [2] Full Field Digital Mammography; [3] Number of true negative and false positive screening mammography; [4] 137 exams had missing mammographic density; nH-non Hispanic

Table V.

RISK OF EXPERIENCING A FALSE POSITIVE MAMMOGRAM AMONG WOMEN UNDERGOING SFM VS. FFDM, OVERALL AND BY SELECTED CHARACTERISTICS

	SFM[1]		FFDM[2]		
	N	Risk(CI)[3]	N	Risk (CI)[3]	P-value[4]
Modality	117,161	12.8 (11.6, 13.9)	108,912	12.3 (11.3, 13.3)	0.02
Race/Ethnicity					
nH White	54,596	12.7 (11.6, 13.8)	51,742	12.9 (11.9, 14.0)	0.20
nH Black	39,479	14.5 (13.0, 15.9)	36,351	13.5 (12.2, 14.8)	<0.01
Hispanic	2,819	13.3 (11.5, 15.1)	3,149	10.8 (9.3, 12.3)	0.01
Other	6,226	12.7 (11.2, 14.1)	8,300	12.3 (11.0, 13.6)	0.60
Unknown	14,041	7.5 (6.6, 8.3)	9,370	7.1 (6.2, 7.9)	0.30
Age group					
<40	3,881	11.8 (10.2, 13.4)	3,088	12.1 (10.47, 13.8)	0.09
40-49	33,475	14.0 (12.7, 15.3)	29,455	13.2 (12.07, 14.4)	0.12
50-59	33,656	13.0 (11.8, 14.2)	32,097	12.4 (11.35, 13.5)	0.05
60-69	23,900	11.4 (10.3, 12.6)	23,595	10.9 (9.8, 11.9)	0.01
70-79	17,023	10.1 (9.0, 11.2)	15,461	9.9 (8.89, 10.9)	0.09
80+	5,226	9.2 (8.0, 10.5)	5,216	9.0 (7.87, 10.2)	0.42
Breast Density					
Fatty	8,671	8.8 (7.7, 9.8)	73,563	7.2 (6.3, 8.0)	<0.01
Scattered	47,410	11.3 (10.2, 12.4)	17,180	10.8 (9.8, 11.8)	0.03
Heterogeneous	49,829	13.6 (12.4, 14.9)	12,853	13.3 (12.1, 14.4)	0.07
Dense	11,251	11.6 (10.4, 12.9)	5,316	11.8 (10.6, 13.0)	0.43
Family History					
None	79,059	12.6 (11.4, 13.7)	73,563	12.0 (10.97, 13.0)	<0.01
Weak	18,215	12.8 (11.6, 14.1)	17,180	12.2 (11.08, 13.3)	0.08
Moderate	14,132	12.9 (11.6, 14.2)	12,853	13.3 (12.07, 14.5)	0.66
Strong	5,755	14.9 (13.3, 16.5)	5,316	14.5 (12.97, 16.0)	0.50
Time since last screen					
9-18	58,180	9.7 (8.7, 10.6)	59,600	8.8 (7.95, 9.6)	<0.01
19-30	18,838	11.3 (10.1, 12.5)	17,136	10.2 (9.19, 11.2)	0.01
30+	8,489	14.7 (13.17, 16.2)	10,426	13.4 (12.07, 14.7)	0.02
First Screen	31,654	15.1 (13.68, 16.6)	21,750	16.3 (14.87, 17.8)	0.41
Prior biopsy					
No	97,659	12.5 (11.3, 13.6)	90,848	12.0 (11.0, 13.1)	0.04
yes	19,502	14.3 (13.0, 15.7)	18,064	13.7 (12.47, 14.9)	0.05
Comparison Film					
Yes	95,446	11.1 (10.0, 12.1)	87,440	10.8 (9.85, 11.8)	0.15
No	21,715	14.8 (13.4, 16.2)	21,472	13.5 (12.2, 14.7)	0.02

[1] Screen Film Mammography;

[2] Full Field Digital Mammography;

[3] Generalized Estimating Equations with binomial distribution and log link adjusted for modality, race, decade of age, breast density, family history, time since last screen, personal history of biopsy, comparison film, site, time since adoption and product terms modality*ethnicity, modality*age, modality*density, modality*family history, modality*time since last screen, modality*prior biopsy, modality*comparison film, modality*site, time since adoption*site, age*ethnicity, age *density, age*comparison film, ethnicity*density, density*time since last screen, ethnicity*comparison film and time since last screen*comparison film;

[4] P-values were generated after excluding all other interaction terms with modality; nH- non Hispanic; CI-confidence interval

Table VI.

RATE OF WORK-UP PER 1000 SCREENS AMONG WOMEN UNDERGOING SFM VS. FFDM, OVERALL AND BY SELECTED CHARACTERISTICS

	SFM[1]		FFDM[2]		
	N	Rate (CI)[3]	N	Rate (CI)[3]	P-value[4]
Modality	117,161	233 (221, 244)	108,912	225 (206,245)	0.001
Race/Ethnicity					
nH White	54,596	238 (226, 250)	51,742	239 (219, 260)	0.94
nH Black	39,479	277 (260, 295)	36,351	257 (232, 282)	<0.001
Hispanic	2,819	239 (214, 264)	3,149	185 (161, 208)	0.01
Other	6,226	253 (234, 272)	8,300	238 (213, 262)	0.28
Unknown	14,041	118 (110, 127)	9,370	111 (98, 123)	<0.001
Age group					
<40	3,881	223 (200, 246)	3,088	227 (198, 256)	0.23
40-49	33,475	261 (247, 275)	29,455	247 (225, 269)	0.03
50-59	33,656	236 (224, 249)	32,097	226 (206, 247)	0.01
60-69	23,900	204 (192, 216)	23,595	193 (175, 211)	<0.001
70-79	17,023	169 (158, 180)	15,461	165 (148, 181)	0.01
80+	5,226	136 (123, 150)	5,216	149 (130, 167)	0.33
Breast Density					
Fatty	8,671	147 (135, 159)	73,563	131 (116, 145)	0.01
Scattered	47,410	205 (194, 216)	17,180	198 (180, 216)	0.003
Heterogeneous	49,829	250 (237, 262)	12,853	243 (221, 264)	0.01
Dense	11,251	212 (197, 226)	5,316	206 (185, 226)	0.27
Family History					
None	79,059	227 (216, 238)	73,563	217 (198, 236)	<0.001
Weak	18,215	242 (228, 256)	17,180	224 (203, 245)	<0.001
Moderate	14,132	236 (221, 250)	12,853	253 (229, 278)	0.18
Strong	5,755	280 (259, 301)	5,316	277 (247, 306)	0.28
Time since last screen					
18-Sep	58,180	190 (179, 200)	59,600	328 (298, 358)	<0.001
19-30	18,838	220 (207, 234)	17,136	174 (159, 190)	<0.001
>30	8,489	274 (254, 293)	10,426	191 (173, 209)	<0.001
First Screen	31,654	307 (292, 322)	21,750	248 (224, 273)	0.51
Prior biopsy					
Yes	97,659	273 (258, 289)	90,848	264 (239, 288)	0.02
No	19,502	226 (215, 237)	18,064	219 (200, 238)	0.002
Comparison Film					
Yes	95,446	212 (202, 223)	87,440	203 (186, 221)	0.16
No	21,715	323 (305, 342)	21,472	311 (282, 339)	<0.001

[1]Screen Film Mammography;

[2]Full Field Digital Mammography;

[3]Generalized Estimating Equations with negative binomial distribution and log link adjusted for modality, race, decade of age, breast density, family history, time since last screen, personal history of biopsy, comparison film, site, time since adoption of FFDM and product terms modality*ethnicity, modality* age, modality*breast density, modality*family history, modality*time since last screen, modality*prior biopsy, modality*comparison film, modality*site, time since adoption*site, age*ethnicity, age *density, age*comparison film, ethnicity*breast density, ethnicity*time since last screen, ethnicity*comparison film, breast density*time since last screen, breast density*comparison film and time since last screen*comparison film,

[4]P-values were generated after excluding all other interaction terms with modality.

nH-Non Hispanic; CI- confidence interval

Table VII.

RATE OF WORK-UP PER 1000 SCREENS AMONG WOMEN WITH A FALSE POSITIVE SFM VS. FFDM, OVERALL AND BY SELECTED CHARACTERISTICS

	SFM[1]		FFDM[2]		
	N[3]	Risk (CI)[4]	N[3]	Risk (CI)[4]	P-value[5]
Modality	15,008	1,860 (1,650, 2,070)	13,032	1,860 (1,650, 2,060)	0.87
Race/Ethnicity					
nH-White	6,221	1,860 (1,660, 2,060)	5,685	1,870 (1,670, 2,080)	0.76
nH-Black	6,236	1,910 (1,670, 2,160)	5,262	1,900 (1,650, 2,140)	0.97
Hispanic	392	1,840 (1,520, 2,160)	381	1,740 (1,430, 2,050)	0.8
Other	820	2,000 (1,720, 2,290)	996	1,940 (1,680, 2,210)	0.91
Unknown	1,339	1,520 (1,310, 1,730)	708	1,560 (1,320, 1,800)	0.96
Age group					
<40	759	1,930 (1,650, 2,210)	658	1,960 (1,660, 2,260)	0.98
40-49	5,214	1,900 (1,680, 2,120)	4,425	1,910 (1,690, 2,130)	0.88
50-59	4,178	1,870 (1,650, 2,090)	3,751	1,860 (1,650, 2,080)	0.86
60-69	2,696	1,860 (1,630, 2,090)	2,382	1,830 (1,600, 2,050)	0.75
70-79	1,693	1,750 (1,520, 1,970)	1,387	1,710 (1,490, 1,940)	0.84
80+	468	1,580 (1,320, 1,840)	429	1,700 (1,420, 1,980)	0.29
Breast Density					
Fatty	1,009	1,780 (1,530, 2,030)	792	1,850 (1,590, 2,120)	0.4
Scattered	5,288	1,860 (1,640, 2,080)	4,546	1,870 (1,650, 2,090)	0.91
Heterogeneous	7,234	1,870 (1,650, 2,080)	6,350	1,860 (1,650, 2,080)	0.75
Dense	1,477	1,850 (1,610, 2,090)	1,344	1,780 (1,550, 2,010)	0.37
Family History					
None	10,059	1,840 (1,630, 2,050)	8,669	1,830 (1,630, 2,040)	0.89
Weak	2,403	1,910 (1,680, 2,150)	2,073	1,860 (1,630, 2,090)	0.41
Moderate	1,735	1,870 (1,630, 2,100)	1,574	1,930 (1,690, 2,170)	0.56
Strong	811	1,920 (1,650, 2,180)	716	1,940 (1,660, 2,210)	0.95
Time since last screen					
9-18	5,473	1,860 (1,640, 2,070)	5,138	1,850 (1,640, 2,070)	0.6
19-30	2,093	1,880 (1,640, 2,110)	1,738	1,790 (1,560, 2,020)	0.81
>30	1,234	1,840 (1,600, 2,080)	1,388	1,840 (1,600, 2,070)	0.26
First Screen	6,208	1,860 (1,640, 2,080)	4,768	1,880 (1,660, 2,110)	0.79
Prior biopsy					
None	12,530	1,840 (1,630, 2,050)	10,907	1,840 (1,630, 2,050)	0.88
yes	2,478	1,940 (1,700, 2,170)	2,125	1,940 (1,700, 2,170)	0.88
Comparison Film					
Yes	10,115	1,860 (1,650, 2,080)	8,425	1,880 (1,660, 2,090)	0.78
No	4,893	1,870 (1,640, 2,090)	4,607	1,830 (1,610, 2,050)	0.95

[1] Screen Film Mammography; [2] Full Field Digital Mammography

[3] True negative exams were excluded.

[4] Generalized Estimating Equations with negative binomial distribution and log link adjusted for modality, race, decade of age, breast density, family history, time since last screen, personal history of biopsy, site, comparison film, time since adoption and product terms modality*ethnicity, modality* age, modality*breast density, modality*family history, modality*time since last mammogram, modality*prior biopsy, modality*comparison film, modality*site, months since FFDM adoption*site and ethnicity*comparison film.

[4] P-values were generated after excluding all other interaction terms with modality

nH-non Hispanic; CI-confidence Interval

Table VIII.

RISK OF RECEIPT OF BIOPSY AMONG WOMEN UNDERGOING SFM VS. FFDM, OVERALL AND BY SELECTED CHARACTERISTICS

	SFM[1]		FFDM[2]		
	N	Risk (CI)[3]	N	Risk (CI)[3]	P-value[4]
Modality	117,161	1.04 (0.61, 1.48)	108,912	1.24 (0.74, 1.73)	0.05
Race/Ethnicity					
nH-White	54,596	1.06 (0.64, 1.47)	51,742	1.33 (0.81, 1.86)	0.01
nH-Black	39,479	1.16 (0.65, 1.68)	36,351	1.4 (0.80, 2.00)	0.22
Hispanic	2,819	1.17 (0.51, 1.83)	3,149	0.86 (0.35, 1.37)	0.34
Other	6,226	1.46 (0.79, 2.14)	8,300	1.51 (0.84, 2.19)	0.68
Unknown	14,041	0.19 (0.08, 0.30)	9,370	0.19 (0.07, 0.31)	0.92
Age group					
<40	3,881	0.95 (0.46, 1.45)	3,088	1.03 (0.50, 1.57)	0.63
40-49	33,475	1.1 (0.63, 1.56)	29,455	1.26 (0.75, 1.78)	0.20
50-59	33,656	1.09 (0.63, 1.56)	32,097	1.37 (0.81, 1.93)	0.04
60-69	23,900	0.98 (0.55, 1.41)	23,595	1.13 (0.66, 1.60)	0.26
70-79	17,023	0.76 (0.41, 1.10)	15,461	0.92 (0.51, 1.32)	0.22
80+	5,226	0.68 (0.31, 1.06)	5,216	0.79 (0.38, 1.21)	0.50
Breast Density					
Fatty	8,671	0.76 (0.39, 1.12)	73,563	0.95 (0.50, 1.39)	0.23
Scattered	47,410	0.99 (0.57, 1.40)	17,180	1.16 (0.69, 1.64)	0.10
Heterogeneous	49,829	1.07 (0.62, 1.53)	12,853	1.26 (0.75, 1.77)	0.10
Dense	11,251	0.9 (0.47, 1.32)	5,316	1.1 (0.61, 1.60)	0.25
Family History					
None	79,059	1.00 (0.58, 1.42)	73,563	1.22 (0.73, 1.70)	0.03
Weak	18,215	1.12 (0.64, 1.60)	17,180	1.09 (0.63, 1.55)	0.69
Moderate	14,132	1.13 (0.63, 1.62)	12,853	1.39 (0.80, 1.97)	0.14
Strong	5,755	1.23 (0.65, 1.82)	5,316	1.73 (0.96, 2.51)	0.07
Time since last screen					
9-18	58,180	0.73 (0.41, 1.05)	59,600	0.76 (0.45, 1.07)	0.37
19-30	18,838	0.93 (0.51, 1.34)	17,136	1.05 (0.61, 1.49)	0.25
30+	8,489	1.35 (0.73, 1.97)	10,426	1.48 (0.85, 2.10)	0.42
First Screen	31,654	1.56 (0.89, 2.24)	21,750	2.17 (1.31, 3.04)	0.02
Prior biopsy					
None	97,659	0.98 (0.57, 1.39)	90,848	1.17 (0.70, 1.63)	0.05
Yes	19,502	1.41 (0.81, 2.02)	18,064	1.64 (0.97, 2.32)	0.22
Comparison Film					
Yes	95,446	0.91 (0.53, 1.29)	87,440	1.15 (0.69, 1.61)	0.03
None	21715	1.41 (0.82, 2.01)	21472	1.45 (0.85, 2.04)	0.25

[1] Screen Film Mammography; [2] Full Field Digital Mammography

[3] Risk of receipt of biopsy (probability) was generated from the Generalized Estimating Equations model with binomial distribution and log link adjusted for modality, race, decade of age, breast density, family history, time since last screen, personal history of biopsy, comparison film, site, time since adoption of FFDM and product terms modality*ethnicity, modality*age, modality*breast density, modality*family history, modality*time since last screen, modality*prior biopsy, modality*comparison film, modality*site, time since adoption*site, ethnicity*breast density, ethnicity*time since last screen and breast density*time since last screen.

[4] P-values were generated after excluding all other interaction terms with modality

nH-non Hispanic; CI-Confidence Interval

Table IX.

RISK OF RECEIPT OF BIOPSY AMONG WOMEN WITH FALSE POSITIVE SFM VS FFDM, OVERALL AND BY SELECTED CHARACTERISTICS

	SFM[1]		FFDM[2]		
	N[3]	Risk (CI)[4]	N[3]	Risk (CI)[4]	p-value[5]
Modality	15,008	8.1 (5.1, 11.0)	13,045	10.1 (6.4, 13.8)	0.002
Race/Ethnicity					
nH White	6221	8.0 (5.2, 10.9)	5,694	10.4 (6.7, 14.1)	0.01
nH Black	6236	8.2 (4.9, 11.4)	5,263	10.5 (6.4, 14.6)	0.01
Hispanic	392	9.5 (4.7, 14.3)	381	8.4 (3.7, 13.2)	0.74
Other	820	11.6 (7.0, 16.3)	997	12.5 (7.5, 17.5)	0.51
Unknown	1339	3.2 (1.4, 4.9)	710	3.1 (1.1, 5.1)	0.98
Breast Density					
Fatty	759	7.6 (4.3, 10.9)	792	11.4 (6.8, 16.1)	0.03
Scattered	5214	8.5 (5.3, 11.7)	4,546	10.4 (6.5, 14.3)	0.02
Heterogeneous	4178	7.9 (4.9, 10.8)	6,350	9.8 (6.2, 13.4)	0.01
Dense	2696	7.8 (4.6, 11.0)	1,344	9.6 (5.7, 13.5)	0.23
Age group	1693				
<40	468	7.2 (3.9, 10.5)	661	8.6 (4.5, 12.7)	0.60
40-49		8.0 (5.0, 11.0)	4,427	10.0 (6.3, 13.8)	0.03
50-59	1009	8.4 (5.2, 11.5)	3,754	11.0 (7.0, 15.1)	0.004
60-69	5288	8.6 (5.3, 11.9)	2,384	10.3 (6.4, 14.3)	0.05
70-79	7234	7.4 (4.4, 10.4)	1,390	8.9 (5.3, 12.5)	0.09
80+	1477	7.2 (3.7, 10.7)	429	8.6 (4.4, 12.8)	0.31
Family History					
None	10059	7.8 (4.9, 10.7)	8,681	10.2 (6.4, 13.9)	0.01
Weak	2403	8.8 (5.4, 12.1)	2,073	9.0 (5.4, 12.5)	0.89
Moderate	1735	8.5 (5.2, 11.8)	1,575	10.5 (6.5, 14.6)	0.07
Strong	811	8.7 (5.0, 12.3)	716	12.1 (7.2, 17.0)	0.05
Time since last screen					
9-18	6,291	6.8 (4.1, 9.4)	4,781	8.5 (5.2, 11.7)	0.01
19-30	2,093	7.7 (4.6, 10.7)	1,738	10.1 (6.2, 14.0)	0.02
30+	1,236	8.9 (5.3, 12.5)	1,388	10.9 (6.6, 15.2)	0.12
First Screen	5,473	9.3 (5.9, 12.8)	5,138	11.5 (7.3, 15.8)	0.03
Prior biopsy					
None	12,609	7.7 (4.9, 10.6)	10,919	9.8 (6.2, 13.4)	0.01
yes	2,484	9.3 (5.9, 12.8)	2,126	11.9 (7.5, 16.3)	0.06
Comparison Film					
Yes	10,126	8.0 (5.1, 11.0)	8,429	10.4 (6.6, 14.2)	0.05
No	4,967	8.2 (5.1, 11.3)	4,616	9.7 (6.0, 13.4)	0.002

[1]Screen Film Mammography; [2] Full Field Digital Mammography

[3] True negative exams were excluded from this analysis.

[4] Risk of receipt of biopsy (probability) was generated from the Generalized Estimating Equations with binomial distribution, adjusted for modality, race, decade of age, breast density, family history, time since last screen, personal history of biopsy, comparison film, site, time since FFDM adoption and product terms modality*decade of age, modality*ethnicity, modality*density, modality*time since last mammogram, modality*family history, modality* comparison film, modality*site, months since FFDM adoption*site, ethnicity*comparison film.

[5] P-values were generated after excluding all other interaction terms with modality

nH-Non Hispanic; CI-Confidence interval

Table X.
UNADJUSTED AND ADJUSTED RISK OF FALSE POSITIVE RESULT BY SCREENING MODALITY

	Unadjusted[1]		Adjusted[2]	
Timeframe in months	SFM[3]	FFDM[4]	SFM[3]	FFDM[4]
-24 to 0 and 6 to 30	**12.9 (12.7, 13.1)**	**12.0 (11.8, 12.2)*****	**12.8 (11.6, 13.9)**	**12.3 (11.3, 13.3)***
-24 to 0 and 12 to 36	12.9 (12.7,13.1)	11.5 (11.3, 11.7)***	12.5 (11.4, 13.7)	11.9 (10.6, 13.2)***
-24 to 0 and 18 to 42	12.9 (12.7,13.1)	11.3 (11.1, 11.5)***	12.4 (10.9, 13.6)	12.0 (10.3, 13.7)
-12 to 0 and 6 to 18	13.0 (12.8,13.3)	12.2 (12.0, 12.5)***	12.7 (11.3, 14.2)	12.8 (11.4, 14.3)
-12 to 0 and 12 to 24	13.0 (12.8,13.3)	12.0 (11.7, 12.2)***	13.0 (11.3, 14.8)	12.1 (10.7, 13.5)**
-12 to 0 and 18 to 30	13.0 (12.8,13.3)	11.7 (11.4, 12.0)***	12.2 (9.2, 15.2)	12.8 (9.4, 16.4)

*P<0.05;**P <0.01;***P <0.001; P-values in the adjusted analysis were generated using the model below without the interaction terms with modality

[1] Unadjusted logistic regression

[2] GEE with binomial distribution and log link adjusted for modality, race, decade of age, breast density, family history, time since last screen, personal history of biopsy, comparison film, site, time since adoption of FFDM and product terms modality*ethnicity, modality*decade of age, modality*breast density, modality*family history, modality*time since last screen, modality*prior biopsy, modality*comparison film, modality*site, time since adoption*site, ethnicity*breast density, ethnicity*time since last screen and breast density*time since last screen

[3] Screen Film Mammography;

[4] Full Field Digital Mammography

Estimates are presented as percentages with confidence intervals in parenthesis

3. Study 2:
Impact of a false positive mammogram experience on subsequent screening behavior and stage at breast cancer diagnosis

3.1 Introduction

Screening mammography is an established routine public health procedure for the early detection of breast cancer and has been shown to reduce mortality from the disease (105). Along with the benefits of early detection, mammography screening is also associated with FP results that lead to unnecessary additional imaging and biopsy procedures as well as associated financial costs, lost time and psychological and physical morbidity (41, 43, 47). False positive rates have been estimated to be as high as 10% of screening mammograms(12) and roughly 50% of women who screen annually for 10 years can expect at least one false positive mammogram finding of which 7-17% will require biopsy (106, 107). The FP screening mammogram issue is part of an ongoing debate regarding the extent to which the risk of mammography screening might outweigh the benefits in certain women (29, 108). Therefore, the most recent guidelines set forth by the USPSTF advised against routine screening in women aged 40-49 years of age and women 75 years or older (89).

Furthermore, a FP mammogram could lead women to alter their future screening behavior, either by delaying the next scheduled mammogram or foregoing the exam altogether. Studies that have examined the potential impact of a FP mammogram on subsequent adherence to screening mammography recommendations have yielded inconsistent findings. Several studies found that re-screening rates were actually higher among women who experienced a FP as opposed to a TN (53-56) while other studies found no difference in re-screening rates based on screening mammography outcome (57-61). Still other reports document lower re-screening rates for women experiencing FP compared with those with TN mammograms (41, 64, 109). A 2007 meta-analysis which pooled data from the above studies found that in Europe and Canada, women who experienced a FP screening mammogram were less likely to return for their next screen compared to women with TN screen finding. Conversely, women in the U.S. were associated with greater subsequent screening

mammography adherence after experiencing a FP (42). The need to provide additional insights into patterns of adherence to mammography screening recommendations by women after the false positive experience prompted this respective analysis. The primary study objective was to examine the impact of a FP screening mammogram on the receipt of subsequent screening mammography among a racially diverse population in a network of mammography centers within a large health care organization. The secondary objective was to determine whether the experience of a FP result at index mammogram increases the risk of subsequent late stage disease for those women subsequently diagnosed with breast cancer.

3.2 Methods

The study was conducted using two large population-based health data sources. Mammography screening data on women were obtained from a large Health Care Organization with multiple facilities in the Greater Metropolitan Chicago Area. Facilities within this healthcare organization used PenRad to collect radiology information and patient characteristics (98). PenRad was first introduced in 2001 and had been implemented at all facilities by 2005. Breast cancer incidence data were obtained from the Illinois State Cancer Registry (ISCR) (99) which collects information on all incident cancer cases in the state of Illinois. Annually since 1999 ISCR has been awarded gold certification as an incidence registry, the highest quality registry status by the North American Association of Central Cancer Registries (NAACCR).

The radiology data set included information pertaining to patient-level data on demographic characteristics and risk factors, and exam-level data on procedure types and results, in addition to unique identifiers for facility and interpreting radiologist for screening and diagnostic procedures that were performed between January 1st, 2001 and December 31st, 2014. Family history was self-reported and was defined as none, weak, moderate and strong. Age was determined by taking the difference between date of index mammogram and date of birth. Race/ethnicity was self-reported as Non-Hispanic (nH) White, nH-Black, Hispanic, other and unknown. Personal history of prior biopsy was defined as present if a prior biopsy existed in the radiology dataset or if it was self-reported.

Time since last mammogram was defined as 9-18 months, 19-30 months, >30 months and no prior mammogram based on the radiology dataset. Breast density was defined following the American College of Radiology classification as entirely fatty, fibroglandular density, heterogeneously dense and extremely dense.

Each mammogram was interpreted by the reading radiologist and was given a score using the American College of Radiology Breast Imaging Reporting and Data System (BIRADS). BIRADS assessment for screening and diagnostic mammography ranges from 0 to 5 such that 0 = need additional imaging evaluation, 1 = negative finding, 2 = benign finding, 3 = probably benign finding, 4 = suspicious abnormality, and 5 = finding highly suggestive of malignancy.

Women with a prior history of breast cancer or who developed breast cancer anytime during the study period were excluded from these analyses as were women with a history of breast reduction, breast implants and breast reconstruction or mastectomy. Screening mammograms which were preceded by any radiologic exam in the prior 9 months were also excluded. In the case of multiple exams on the same day, only the first exam in the sequence was used in the analysis. Screening mammograms were only included if at least three years of follow-up were available in order to allow at least 3 years of observation.

A linkage of 761, 908 screening mammograms performed between 2001 and 2010 to ISCR breast cancer cases resulted in a 98-99% match rate for diagnosis years 2001-2011. To allow 12 months of follow up for cancer diagnosis, we restricted our analytic dataset to include bilateral screening mammograms that were performed January 1, 2001 and December 31, 2010. The unit of analysis was the screening mammogram. Based on a comparison of the screening mammogram interpretation (normal vs. abnormal) and cancer status within 12 months of the screen, screening mammograms were defined as true positive (TP), true negative (TN), false positive (FP) and false negative (FN) screens. For these analyses, we compared the experiences of women with false positive (FP) and true negative mammogram.

A TN mammogram was defined as any mammogram with BIRADS (1,2,3) and that cancer was not detected in the subsequent 12 months from date of screening mammogram. Whereas, a false positive mammograms was defined as any mammogram with BIRADs (0,4,5) and that cancer was not detected in the subsequent 12 months from date of screening mammogram. The burden of FP was defined as the number of additional imaging after a FP mammogram and morbidity was defined as the receipt of biopsy after a FP mammogram. Women with a TN mammogram were assumed to have no additional work up during the follow up period.

Because the recommended interval for routine screening is at least 12 months, we defined the index date (T=0) as 365 days after the index screening date. Therefore, any index screening mammograms that were followed with a subsequent screening mammogram prior to 12 months were excluded (N=68,289, 9%). Follow up period was defined as the number of months between the index date and the date of the subsequent screening mammogram among those who returned to screening within the network. Women who did not return to screening at our network were considered right censored and their follow up time was estimated as the difference between index date and December 31, 2014. This date was used because our data included all screening mammograms that were performed on or before December 31, 2014. The dependent variable for the primary analysis was the number of months (T) after index date for both TN and FP mammograms.

In an additional analysis we adjusted the follow up time to account for the time required to resolve a positive mammogram by setting the index date to be the date of the last diagnostic procedure.

For the primary objective, we excluded exams from women who were diagnosed with breast cancer at any point during our study period. The unit of analysis was the mammography screening exam. For the secondary analysis in which we examined the impact of false positive on stage at diagnosis, we included women who were diagnosed with breast cancer and for whom data on at least two screening mammograms prior to diagnosis were available. Similar to the analysis of the

primary objective, the index screening mammogram result can be true negative or false positive. The subsequent mammogram for the secondary analysis, on the other hand, was either true positive (if the breast cancer was screen-detected) or false negative (if the breast cancer was an interval cancer) (Figure 2). Stage at diagnosis was defined according to the American Joint Committee on Cancer (AJCC). In the event of missing stage data, we used TNM clinical stage and TNM pathology stages to impute the data. The dependent variable was late stage at diagnosis which was defined as stages 2, 3, 4 vs. early stage at diagnosis (stage 0 or 1).

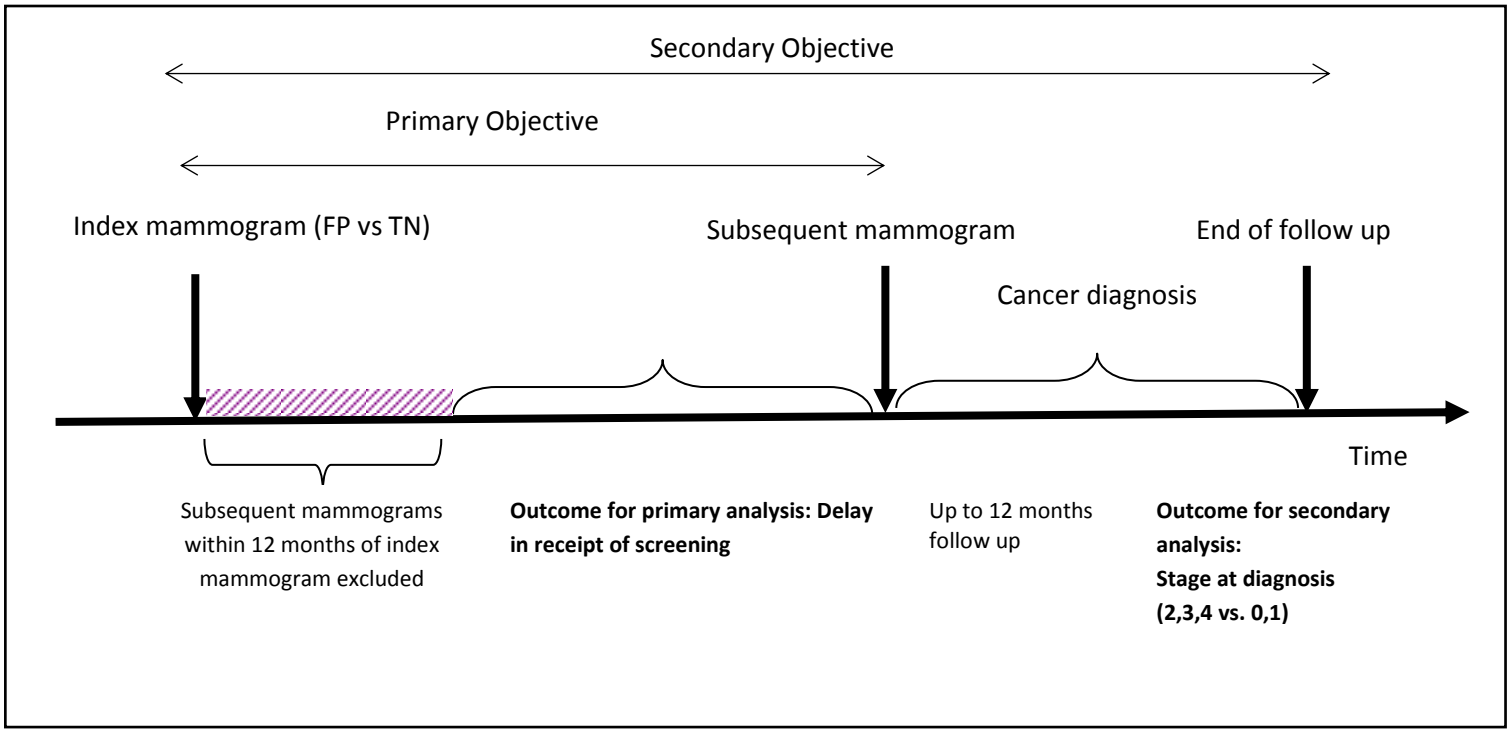


Figure 2: Primary and secondary analyses design overview FFDM vs SFM. The first arrow represents the index mammogram (true negative and false positive mammograms only). The shaded area represents the first 12 months after index mammogram. Mammograms that were followed by a screening mammogram within 12 months were excluded. The second arrow represents the subsequent mammogram after index mammogram. For secondary analyses the subsequent mammogram can be a TP or FN mammogram (must be associated with a breast cancer diagnosis).

3.2.1 Statistical Analysis

Patient characteristics by mammogram result (TN vs FP) and by stage at diagnosis were compared based on a chi-square test of association for nominal categorical variables and based on a test for trend for ordinal variables by including them as independent continuous variables in logistic regression model. The Kaplan-Meier (product limit) estimator was used to estimate the overall unadjusted delay in return to screening by mammogram result (TN vs FP), and log-rank tests were used to compare the differences between the two curves. Cox proportional hazard models were used to estimate the hazard ratio for delay in the receipt of subsequent screening mammogram within the next 5 years from the index mammogram date. Women who did not return to screening at this network were right censored as well as women who returned to screening after 5 years from index mammogram date. In addition to mammogram result (TN vs FP), the model included variables for age, race/ethnicity, family history of breast cancer, mammographic breast density, parity, prior history of biopsy, time since last screening mammogram, calendar year, availability of comparison film and facility. Stratum-specific hazard ratios were generated using the same model as above with the addition of each individual product term between the index mammogram result and the variable of interest.

In addition to multivariable models described above, a propensity score matching technique was used to match on the probability of a FP result. Logistic regression modeling was used to predict the probability of being FP vs TN adjusting for decade of age, race/ethnicity, family history of breast cancer, mammographic breast density, parity, prior history of biopsy, time since last screening mammogram, calendar year, availability of comparison film at interpretation, facility and any possible interaction terms that were significant at an alpha 0.05 level. Off support probabilities were excluded and greedy matching algorithm without replacement was used to match 1-1 TN and FP mammograms (110). The matched dataset was then analyzed using Kaplan-Meier estimator to estimate the probability of returning to screening by index mammogram result. Proportional hazards

modeling was used to estimate the risk of not returning to recommended screening by index mammogram result.

3.2.1.1 Secondary Analyses

To estimate the probability of late stage at diagnosis, we used logistic regression adjusting for the same variables as above. Two-way product terms were included if they were statistically significant at p less than 0.05 level. Marginal standardization was used in order to estimate the average population risk of late stage at diagnosis by prior mammogram result, and 95 percent confidence intervals were obtained using the delta method.

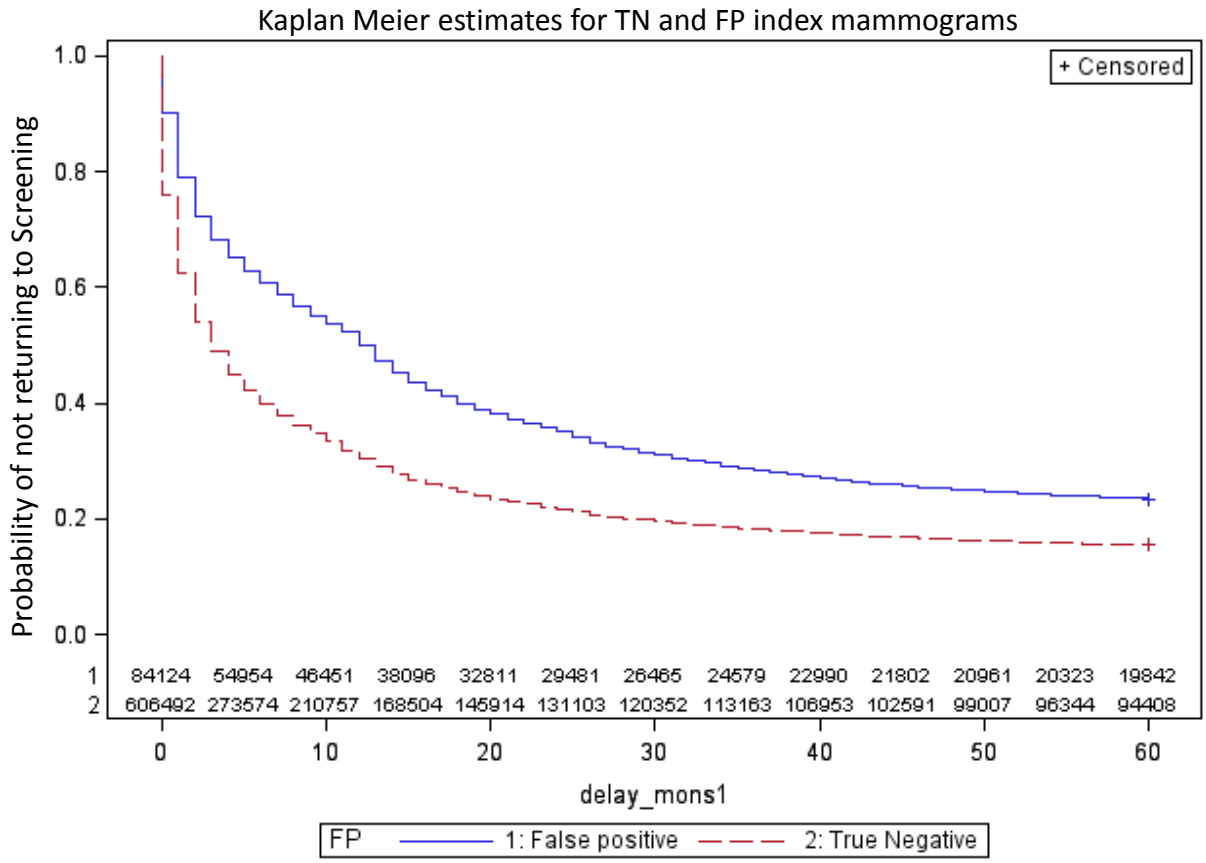
3.3 Results

A total of 690,610 screening mammograms (FP=84,118, TN=606,492) from 247,361 women were included in this study. The overall false positive rate was 12.1%. Women experiencing a FP result were more likely to not have a subsequent screen in the database than women experiencing a TN result (21.1% vs 14.3%, P-value <0.001). Women who did not return for screening at these facilities may have forgone screening altogether (a substantively important result of this study) or may have sought subsequent screening elsewhere (may have been lost to follow-up). Table XI summarizes the characteristics of women contributing index mammograms to these analyses. Women with false positive mammograms were younger, premenopausal and were more likely to be experiencing their first mammogram screening. Also, they were more likely to be nH-Black, have denser breasts and were less likely to have a comparison film available at interpretation (Table XI).

Regardless of index screen result, younger and premenopausal women as well as women who were obtaining their first screening mammogram or whose prior mammogram occurred more than 30 months before the index screen were more likely to delay subsequent screening. The median delay in return to screening was higher for FP than for TN mammograms (13 months vs 3 months, P-value <0.001) (Figure 3). Delays in returning for subsequent screening were consistently

longer after a FP mammogram than after a TN mammogram across strata of patient characteristics (Table XII).

Figure 3: Kaplan-Meier estimates for time to next screen in months by index mammogram result



In the adjusted proportional hazards model, women with TN result were 34% more likely to return to screening compared to women with a FP result HR=1.35 (95% CI: 1.34-1.36). In addition, a FP result was consistently associated with delays in subsequent screening within strata of patient characteristics and screening history (Table XIII).

3.2.2 Adjusted Index Date

The results after resetting the index date to account for the time required to resolve a FP mammogram were similar to the results that were observed when using the actual mammogram date as the index date. Briefly, the median delay in return to screening was higher for FP than for TN mammograms (12 months vs 2 months, P-value <0.001). Delays in returning for subsequent screening were consistently longer after a FP mammogram than after a TN mammogram across strata of patient characteristics. In the adjusted proportional hazards model, women with TN result were 36% more likely to return to screening compared to women with a FP result HR=1.36 (95% CI: 1.35-1.3) (Data not shown).

3.2.3 Burden and Morbidity

The extent of the burden and morbidity of FP were also associated with delays in the receipt of subsequent screening mammography. Compared to women who did not receive additional work up, women who received additional imaging were 24% less likely to return to screening HR=1.31 (95% CI:1.29-1.32) and women who received imaging and biopsy were 34% less likely to return to screening HR=1.51 (95% CI:1.49- 1.56) (P-value for trend <0.001). Among false positives only, women who experienced additional imaging and biopsy were 16% less likely to return to screening compared to women who received imaging only HR=1.19 (95% CI: 1.15-1.22).

3.2.4 Propensity Score matching results

We reanalyzed our primary results using propensity score matching. We matched 83,467 (99.2 % of all FPs in the dataset) false positive index mammograms to a similar number of true negative mammograms. The proportion of women who did not return to screening was slightly

higher among women who experienced a false positive mammogram compared to women with a TN exam (21.6% vs 18.2%, P-value <0.001). The two cohorts were balanced in terms of women's characteristics (Table XIV). Similar to the analysis which included all exams, delay in return to subsequent mammograms was longer among women with FP compared to women with TN mammograms. The median delay was 7 months among FP compared to 3 months among TN (P-value <0.001). After adjusting for patient characteristics, the chance of returning to screening was 34% higher in women with TN exams compared to women with FP mammograms HR=1.34 (95% CI: 1.33-1.35).

3.2.5 Late Stage at Diagnosis

We identified 2,170 breast cancer cases meeting study criteria that were diagnosed within 12 months from the subsequent screening mammogram, of which 548 (25.3%) were late stage at diagnosis. The overall false negative rate was 12.6% and late stage at diagnosis was prevalent in 52.2% of interval cancers and 21.4% of true positives (P=0.01). Women who had a prior FP mammogram, with denser breasts, weak or no family history, and parous women had increased risk of late stage at diagnosis (Table XV). Delaying the receipt of subsequent mammogram did not seem to have an effect on the risk of late stage at diagnosis (P-value =0.9). After adjusting for age, race/ethnicity, family history, parity, breast density, facility and exam year, the prevalence of late stage at diagnosis was 3% higher in women with prior FP 27.9% (95% CI: 22.2-33.5) compared to women with TN mammogram 25.2% (22.6, 27.8%) but this difference was not statistically significant (P-value =0.3) (Table XVI).

In a separate analysis, we identified 6,244 exams from women who had a breast cancer diagnosis within 3 years from the index mammogram. This sample includes exams from women who came back and from women who were lost to follow up but developed breast cancer within 3 years from index mammogram. Late stage at diagnosis was significantly associated with increased delays such that the prevalence of late stage was increasing with increasing delays (27%, 28%, 30%, 32 and 34%) for 18 months, 24 months, 30 months , >30 months and never returned to screening,

respectively. False Positive experience did not seem to have an impact on the diagnosis of late stage at diagnosis.

3.4 Discussion

We sought to examine how the experience of a false positive mammogram might impact adherence to subsequent mammography screening in a large cohort of women from a single healthcare organization. The FP rate in our study was 12.2%, slightly higher than the rate that was reported to be 9.6% for digital and 9% for screen film mammography in a study utilizing data from the Breast Cancer Surveillance Consortium (BCSC) (12) but within the range (6%-15%) of reported recall rates in the US, for which most are false positive results (35, 106, 111).

In our study, women who had a FP mammogram were less likely to return for screening within the following 5 years compared to those with TN mammogram results. This finding is consistent with another US-based study that used secondary data from telephone interviews and medical claims records for calendar years 2005-2008 on 2406 women which were followed for 36 months. This study found that 21% of women with FP mammogram compared to 16% of women with TN mammogram delayed their receipt of the subsequent screening (112). Conversely, studies conducted more than a decade ago using data from the 1990s found that women who experienced FP mammogram had better adherence to subsequent screening compared to women with a true negative mammogram exam outcome(54-56). When we stratified our results by calendar year, we observed that delays in returning to screening were increasing by calendar year (Table XIII).

Several other studies from Europe and Canada found no difference in re-screening (53, 58-61), and yet others have reported lower re-screening rates among false positives than among true negatives (41, 57, 64, 109, 113, 114). These inconsistent results suggest both secular and geographic variation in the impact of FP mammography on adherence to screening recommendations among the USA, Europe and Canada (42). The conflicting results for international comparisons may be attributed to variations in screening practices such as screening intervals are shorter in the US than in Europe, greater emphasis on accuracy in Europe by double readings which have been reported to

result in 3 to 5% lower recall rates compared to the US, and differences in national mammography programs for Europe and US public and private screening providers. The inconsistency between the majority of the USA studies and our study might be explained by secular changes in how women perceive and adapt to a false positive mammogram, which may be related to changes in guidelines (USPSTF guidelines 2002 and 2009) and increased awareness of the balance of benefits and harms of mammography screening over the last decade.

Our study findings suggest that the delay in returning to recommended mammography screening practices qualitatively increased the risk of subsequent diagnoses with late stage breast cancer. A similar observation was reported from a study in the United Kingdom which found an increased likelihood of late stage at diagnosis among women with FP compared to those with TN mammogram results OR=1.37 (95% CI: 0.67-2.28) (57). However, potential impact on stage at diagnosis appeared to be small.

Strengths of this study include the longitudinal design, the availability of screening and diagnostic records of prior exams that were conducted within our network and the large number of exams from a diverse community-based cohort. Other studies have used women as the unit of analysis to estimate the probability of returning to the subsequent screening mammography (54, 56).

This study has several limitations as well. First, we could not account for insurance status in our analysis as these data were not available in our data collection system. Women who are uninsured or underinsured may be more likely to be truly lost to follow-up if they lack a medical home. Alternatively, underinsured women may be more likely to delay or forgo altogether subsequent screening as a result of a false positive screen, perhaps due to the concern regarding high out of pocket costs.

Some women who experience a false positive result might decide to get their next screening mammogram 12 months after the completion of their diagnostic work-up, rather than 12 months after their last screen. When we adjusted the follow-up time for women with a FP screen to begin at

the date of the last diagnostic procedure, our results were similar to the results generated when using the screening mammogram date as the index date. Thus, the potential for a perceived shift in the appropriate date for the next screen among women with a FP index mammogram could not account for the association of a FP result with delayed subsequent screening.

In these analyses, we included the 14% of exams that were not followed by a subsequent screening mammogram within our network as right censored. It is possible that these women may have never returned to screening or could have received their mammography screening somewhere else outside our network. It is also possible that some women who appeared to delay their subsequent screen may have obtained another screen elsewhere in the interim, outside this healthcare organization and thus not captured by our radiology database. Given the high percentage (86%) of index screens were associated with a subsequent screen, loss to follow-up would appear to be modest, but this is could not be determined empirically.

In conclusion, our study found that women who experienced a false positive mammogram were more likely to delay their subsequent screening compared to women with a TN mammogram. The finding is important in that women who experience a FP mammogram result should be provided with more information about the continued benefits of mammography screening and encouraged to maintain adherence to screening mammography recommendations.

Table XI.

PATIENT CHARACTERISTICS OF 690,610 SCREENING MAMMOGRAMS BY MAMMOGRAM RESULT FOR THE PERIOD 2001-2010

	TN		FP	
	N[1]	%	N[1]	%
Loss to Follow up				
Yes	86,932	14.3	18,139	21.6
No	519,560	85.7	65,979	78.4
Age				
<40	20,149	3.3	4,925	5.9
40-49	173,614	28.6	30,390	36.1
50-59	179,280	29.6	24,079	28.6
60-69	123,252	20.3	14,082	16.7
70-79	83,114	13.7	8,240	9.8
80+	27,083	4.5	2,402	2.9
Ethnicity				
nH White	333,513	55.0	44,205	52.6
nH Black	143,672	23.7	24,325	28.9
Hispanic	14,215	2.3	2,285	2.7
Other	43,370	7.2	6,065	7.2
Unknown	71,722	11.8	7,238	8.6
Breast Density*				
Fatty	50,004	8.2	5,140	6.1
Scattered	249,507	41.1	29,743	35.4
Heterogeneous	255,321	42.1	41,666	49.5
Dense	51,606	8.5	7,412	8.8
Family history				
None	410,285	67.6	56,115	66.7
Weak	94,079	15.5	13,427	16.0
Moderate	72,503	12.0	9,943	11.8
Strong	29,625	4.9	4,633	5.5
Parity				
Nulliparous	74,908	12.4	10,434	12.4
Parous	483,609	79.7	64,368	76.5
Missing	47,975	7.9	9,316	11.1
Menopause				
Pre-menopausal	130,168	21.5	25,440	30.2
post-menopausal	476,324	78.5	58,678	69.8
Prior Biopsy				
Yes	107,515	17.7	14,632	17.4
No	498,977	82.3	69,486	82.6
Time since last screen				
9-18	298,398	49.2	29,854	35.5
19-30	78,042	12.9	9,334	11.1
>30	35,189	5.8	5,529	6.6
First Screen	194,863	32.1	39,401	46.8
Comparison Film				
Yes	513,559	84.7	57,544	68.4
No	92,933	15.3	26,574	31.6

TN – True negative; FP – False positive.

[1] Includes exams from women who never returned to screening within our network.

[2] Chi-square test

*211 exams were missing breast density

nH-non Hispanic

Table XII.

MEAN AND MEDIAN DELAY IN MONTHS FOR FOLLOW-UP SCREENING MAMMOGRAPHY BY PATIENT CHARACTERISTICS AND MAMMOGRAM FINDINGS

	TN				FP			
	N[1]	Mean	Median	P-value[2]	N[1]	Mean	Median	P-value[2]
Age				<0.001				<0.001
<40	14,741	21	14		3,474	23	17	
40-49	148,233	9	3		23,808	14	8	
50-59	157,502	7	2		19,347	12	6	
60-69	109,702	6	2		11,531	10	5	
70-79	70,773	5	1		6,382	9	4	
80+	18,612	5	1		1,438	7	3	
Ethnicity				<0.001				<0.001
nH White	293,725	7	2		35,854	13	6	
nH Black	126,873	8	3		19,573	12	6	
Hispanic	7,872	8	3		986	13	9	
Other	40,685	9	2		5,276	16	10	
Unknown	50,408	8	3		4,291	13	6	
Breast Density*				<0.001				<0.001
Fatty	41,291	7	2		3,598	13	7	
Scattered	213,749	7	2		23,309	12	6	
Heterogeneous	220,106	8	2		33,151	13	7	
Dense	44,373	8	2		5,812	15	9	
Family history				<0.001				<0.001
None	345,270	8	2		42,806	13	7	
Weak	82,894	8	2		11,029	13	6	
Moderate	65,183	6	2		8,411	12	5	
Strong	26,216	7	2		3,734	13	6	
Parity				<0.001				0.13
Nulliparous	66,771	7	2		8,576	13	7	
Parous	422,104	8	2		52,532	13	7	
Missing	30,688	8	2		4,872	12	6	
Menopause				<0.001				<0.001
Pre-menopausal	98,285	11	4		17,255	15	9	
post-menopausal	421,278	7	2		48,725	12	6	
Prior Biopsy				<0.001				<0.001
No	423,812	8	2		53,882	13	7	
Yes	95,751	6	1		12,098	11	5	
Time since last screen				<0.001				<0.001
18-Sep	273,707	4	1		26,406	8	3	
19-30	67,094	9	4		7,754	12	7	
>30	27,686	12	6		4,156	15	10	
First Screen	151,076	12	5		27,664	17	11	
Comparison Film				<0.001				<0.001
Yes	17,607	18	12		48,373	11	5	
No	66,505	15	7		453,058	7	2	

TN – True negative; FP – False positive

[1]105,087 exams were lost to follow up;

[2] Log Rank Test;

NH-non Hispanic

Table XIII.

OVERALL AND STRATIFIED HAZARDS RATIOS OF RETURNING TO SCREENING AMONG TRUE NEGATIVE COMPARED TO FALSE POSITIVE MAMMOGRAMS

	HR (TN vs. FP)[1]	P-value[2]
Overall	1.35 (1.34, 1.36)	<0.001
Stratum-Specific		
Calendar Year		
≤ 2005	1.29 (1.28, 1.31)	<0.001
> 2005	1.37 (1.36, 1.39)	<0.001
Race/ Ethnicity		
nH-Whites	1.38 (1.36, 1.40)	<0.001
nH-Blacks	1.28 (1.26, 1.30)	<0.001
Hispanics	1.35 (1.26, 1.44)	<0.001
Other	1.39 (1.35, 1.44)	<0.001
Unknown	1.29 (1.25, 1.34)	<0.001
Age group		
<40	1.31 (1.29, 1.32)	<0.001
40-50	1.39 (1.37, 1.42)	<0.001
50-60	1.4 (1.37, 1.42)	<0.001
60-70	1.41 (1.37, 1.44)	<0.001
70-80	1.31 (1.24, 1.38)	<0.001
80+	1.06 (1.02, 1.10)	0.01
Time Since Screen		
First Screen	1.28 (1.26, 1.29)	<0.001
18-Sep	1.23 (1.20, 1.26)	<0.001
19-30	1.46 (1.45, 1.48)	<0.001
>30	1.22 (1.18, 1.26)	<0.001
Family History		
None	1.34 (1.32, 1.35)	<0.001
weak	1.38 (1.35, 1.42)	<0.001
Moderate	1.44 (1.39, 1.49)	<0.001
Strong	1.33 (1.30, 1.36)	<0.001
Prior Biopsy		
Yes	1.46 (1.43, 1.49)	<0.001
No	1.32 (1.31, 1.33)	<0.001
Comparison film		
Yes	1.23 (1.21, 1.25)	<0.001
No	1.38 (1.37, 1.40)	<0.001

TN – True negative; FP – False positive.

[1] Proportional hazards regression model adjusted for mammogram result (FP vs TN), decade of age, race/ethnicity, calendar year, breast density, family history, time since last screen, history of prior biopsy, parity, availability of comparison film and site.

2 P-values were generated after removing all product terms with the index mammogram except for the interaction of interest

Table XIV.
PATIENT CHARACTERISTICS AFTER MATCHING ON THE PROBABILITY OF FALSE POSITIVE

	TN		FP	
	N	%	N	%
Loss to Follow up				
Yes	15,203	18.2	18,070	21.6
No	68,264	81.8	65,397	78.4
Age				
<40	5,141	6.16	4,808	5.76
40-49	29,189	34.97	30,147	36.12
50-59	23,180	27.77	23,932	28.67
60-69	14,083	16.87	13,986	16.76
70-79	8,860	10.61	8,202	9.83
80+	3,014	3.61	2,392	2.87
Ethnicity				
nH White	41,713	49.98	43,919	52.62
nH Black	24,018	28.78	24,078	28.85
Hispanic	2,812	3.37	2,257	2.70
Other	6,842	8.20	5,991	7.18
Unknown	8,082	9.68	7,222	8.65
Breast Density*				
Fatty	5,990	7.18	5,067	6.07
Scattered	28,864	34.58	29,641	35.51
Heterogeneous	40,005	47.93	41,430	49.64
Dense	8,608	10.31	7,329	8.78
Family history				
None	51,899	62.18	55,789	66.84
Weak	14,675	17.58	13,310	15.95
Moderate	11,348	13.60	9,833	11.78
Strong	5,545	6.64	4,535	5.43
Parity				
Nulliparous	12,891	15.44	10,285	12.32
Parous	60,692	72.71	63,975	76.65
Unknown	9,884	11.84	9,207	11.03
Prior Biopsy				
Yes	16,711	20.02	14,460	17.32
No	66,756	79.98	69,007	82.68
Time since last screen				
9-18	27,243	32.64	29,832	35.74
19-24	10,459	12.53	9,318	11.16
25-36	6,405	7.67	5,471	6.55
First Screen	39,360	47.16	38,846	46.54
Comparison Film				
Yes	56,818	68.07	57,422	68.80
No	26,649	31.93	26,045	31.20

TN – True negative; FP – False positive; nH-non Hispanic

Table XV.

PREDICTORS OF LATE STAGE AT DIAGNOSIS WITHIN 12 MONTHS FROM A SUBSEQUENT MAMMOGRAM

	Early Stage (0,1) n=1,618		Late Stage (2,3,4) n=555		
	N	%	n	%	P-value
Prior Mammogram					0.2
TN	1423	75.2	477	24.8	
FP	195	71.6	77	28.4	
Age					0.04
<50	259	73.5	95	26.5	
50-59	420	74.2	150	25.8	
60-69	454	75.2	149	24.8	
70-79	340	72.3	132	27.7	
80+	145	84.0	28	16.0	
Ethnicity					0.9
nH White	977	75.3	326	24.7	
nH Black	487	74.0	175	26.0	
Hispanic	14	77.8	4	22.2	
Other	80	74.3	27	25.7	
Unknown	60	72.1	22	27.9	
Breast Density					0.01
Fatty	81	80.4	20	19.6	
Scattered	679	78.0	196	22.0	
Heterogeneous	728	71.8	290	28.2	
Dense	130	72.9	48	27.1	
Family history					0.05
None	927	74.0	328	26.0	
Weak	252	71.1	105	28.9	
Moderate	297	77.9	86	22.1	
Strong	142	80.8	35	19.2	
Parity					0.0022
Nulliparous	199	84.0	42	16.0	
Parous	1281	73.6	462	26.4	
Missing	138	73.5	50	26.5	
Prior Biopsy					0.7
None	1149	74.5	401	25.5	
Yes	469	75.4	153	24.6	
Delay in months					0.97
0-6	1099	74.9	375	25.1	
6-12	203	73.9	70	26.1	
12-18	127	73.9	47	26.1	
>18	189	75.6	62	24.4	
BIRADS					<0.001
(1,2,3)	125	47.8	142	52.2	
(0,4,5)	1493	78.6	412	21.4	

TN True negative, FP-False Positive; nH-non –Hispanic

Table XVI.

ADJUSTED RISK OF LATE STAGE AT DIAGNOSIS WITHIN 12 MONTHS FROM A SUBSEQUENT MAMMOGRAM

	N	Risk (95% CI)[1]	P-value[2]
Prior Mammogram			
TN	1,899	25.2 (22.6 ,27.8)	0.30
FP	271	27.9 (22.2 ,33.5)	
Age			0.09
<50	355	24.9 (20.1 ,29.8)	
50-59	566	25.6 (21.7 ,29.5)	
60-69	604	25.7 (21.8 ,29.6)	
70-79	470	28.9 (24.3 ,33.5)	
80+	175	17.0 (11.1 ,22.9)	
Ethnicity			0.97
nH White	1,307	25.3 (22.3 ,28.3)	
nH Black	654	26.1 (21.9 ,30.4)	
Hispanic	18	20.5 (2.4 ,38.6)	
Other	105	24.2 (16.1 ,32.4)	
Unknown	86	27.1 (17.2 ,36.9)	
Breast Density			0.01
Fatty	97	19.0 (11.2 ,26.8)	
Scattered	876	22.2 (19.0 ,25.4)	
Heterogeneous	1,016	28.7 (25.4 ,32.0)	
Dense	181	27.7 (20.7 ,34.8)	
Family history			0.06
None	1,257	26.2 (23.3 ,29.2)	
Weak	357	28.9 (23.9 ,33.9)	
Moderate	384	23.0 (18.5 ,27.6)	
Strong	172	19.1 (13.2 ,25.1)	
Prior Biopsy			0.60
Yes	466	24.6 (21.0 ,28.0)	
None	1,156	25.5 (23.0 ,29.0)	
Parity			<0.001
Nulliparous	243	16.8 (11.9 ,21.8)	
Parous	1,742	26.3 (23.7 ,29.0)	
unknown	185	29.5 (22.1 ,36.9)	

TN True negative, FP-False Positive; nH-non-Hispanic

[1] Marginal Standardization after adjusting for prior mammogram, delay in days between screens, age, ethnicity, breast density, family history, parity, exam year and site.

[2] type III analysis

4. Study 3:
Accuracy of image-guided biopsy: comparative effectiveness by imaging modality and vacuum assistance

4.1 Introduction

Image guided core needle biopsy is a widely accepted technique that is used to rule in or rule out breast malignancies. It has been recommended as an alternative to open surgical biopsies which are associated with increased humanistic and economic burden on the patient. In general, women with abnormalities receive imaging including diagnostic mammography, ultrasound and MRI before the receipt of a breast biopsy. In the USA, about 50% of women who receive annual mammography screening over 10 years will undergo at least one additional imaging and of these 7%-17% will have biopsies (39, 40) with the majority of biopsies (77%) (73) resulting in the diagnosis of non-malignant lesions that do not require further work up or treatment.

There are three biopsy techniques currently available to sample cells from suspicious breast lesions: fine needle aspiration (FNA), core needle biopsy (CNB) and open surgical biopsy (incisional or excisional breast biopsy). Fine needle aspiration was often criticized for its modest sensitivity 92.7% (95% CI: 92.1-93.3) and specificity, 94.8% (95% CI: 94.3-95.2) (74) and high proportion with insufficient samples (35%) (8). Core needle biopsy is a procedure that requires the insertion of a hollow probe to remove small samples of breast tissue and locates the lesions via different types of imaging guidance such as stereotactic, ultrasound and MRI. It is associated with high sensitivity 97% (95% CI: 95 -99), but also associated with high atypical ductal hyperplasia (ADH) underestimation of 40% (95% CI: 26-56) and high ductal carcinoma *in-situ* (DCIS) underestimation of 15% (95% CI: 8-26) (75). Similar to FNA but to a lesser extent, CNB results in collection of an insufficient amount of tissue from about 5-10% of non-palpable lesions (76). For these reasons, the FDA approved image-guided vacuum-assisted (VA) biopsy in 1995 to increase CNB sensitivity and to reduce the burden associated with multiple insertions and the probability of epithelial displacement (77, 78).

Open (excisional or incisional) biopsy is considered the gold standard for evaluating suspicious lesions because of its very high accuracy. The procedure collects a large amount of

samples and in some cases completely removes the lesion; it is however only applicable to for palpable masses. This surgical procedure is associated with increased economic burden and morbidity (80, 81) which are reduced with the use of CNB. Consequently, both the National Comprehensive Cancer Network (NCCN) and an interdisciplinary group of physicians specializing in the diagnosis and treatment of breast disease recommended CNB as the preferred alternative to open surgery (71, 82). Despite these recommendations and the benefits of CNB, excisional biopsy is still used in 24-36% of patients with suspicious lesions (83-86), a proportion that is much higher than the expected 10% (72) of patients who have a preference for excisional biopsy or that have an unfavorable lesion position for CNB.

The Agency for Healthcare Research and Quality published a comparative effectiveness report of core-needle and open surgical biopsy for the diagnosis of breast lesions in 2014 (88) The report concluded that image guided CNB was an accurate method for the diagnosis of breast cancer and could be used as an alternative to open surgical biopsy. However, the strength of evidence was considered weak in that the referenced studies had poor reporting and low internal validity (88). In addition several questions remained un-answered including a lack of information on patient and tumor factors that might affect the accuracy of CNB.

In this study, we sought 1) to compare the accuracy of vacuum-assisted stereotactic core needle biopsy versus vacuum-assisted ultrasound guided core needle biopsy, 2) to compare the accuracy of vacuum-assisted versus non-vacuum-assisted ultrasound-guided biopsy and 3) to examine how patient and tumor characteristics might affect the chance that an invasive breast cancer would be missed or underestimated at biopsy.

4.2 Methods

This study was conducted using two large data sources containing information on biopsy procedures and findings and incident breast cancer data. The biopsy data were obtained from a large Health Care Organization with multiple facilities in the Greater Metropolitan Chicago Area (97). Facilities within this healthcare organization used PenRad (98), an administrative software program

designed to collect radiology information, pathology and patient characteristics (98). PenRad was first introduced in 2001 and had been implemented at all facilities within this organization by 2005. Breast cancer incidence data were obtained from the Illinois State Cancer Registry (ISCR) (99) which collects information on all incident cancer cases in the state of Illinois. Annually since 1999 ISCR has been awarded gold certification as an incidence registry, the highest quality registry status by the North American Association of Central Cancer Registries (NAACCR).

The biopsy data set included information pertaining to patient-level data on demographic characteristics and risk factors, and exam-level data on procedure types and results for procedures performed between January 1st, 2001 and December 31st, 2014. Race/ethnicity was self-reported as Non-Hispanic (nH) White, nH-Black, Hispanic, other and unknown. Age was determined by taking the difference between date of biopsy and date of birth. Breast density from the most recent mammogram prior to the actual biopsy was used and was defined following the American College of Radiology classification as entirely fatty, fibroglandular density, heterogeneously dense and extremely dense. Family history of breast cancer was defined as none (no first or second degree relatives affected), weak (only second degree relatives affected), moderate (one first degree relatives over age 50 affected), and strong (multiple first degree relatives affected or one under age 50). Personal history of prior biopsy was defined as present if it was self-reported or documented in the radiology database. Mode of lesion detection was defined as screen detected (asymptomatic) if the biopsy was preceded in the prior 12 months with a screening mammogram with Breast Imaging Reporting and Data System (BIRADS) score of (0 = need additional imaging evaluation, 4 = suspicious abnormality, and 5 = finding highly suggestive of malignancy) and symptomatic if otherwise.

Women with a prior history of breast cancer were excluded from these analyses as were women with a history of breast reduction, breast implants and breast reconstruction or mastectomy. Biopsy findings were defined as benign, high risk benign and malignant. Benign findings included diagnosis such as fibroadenoma, papilloma, ductal hyperplasia, fibrosis and fibrocystic changes and calcifications. High risk benign findings included lobular carcinoma *in-situ*, atypical hyperplasia,

phyllodes tumor and other unusual histologic entities. Malignant findings included invasive tumors, and DCIS.

A linkage of 22,297 image guided biopsies performed between 2001 and 2010 to ISCR breast cancer incident cases resulted in a 99% match rate for diagnosis years 2001-2011. To allow a 12-months period of follow up for cancer diagnosis, we restricted our analytic dataset to include image guided biopsies that were performed between January 1, 2001 and December 31, 2010. Breast cancer diagnoses that occurred in the 90 days prior to the biopsy through 365 days after the biopsy were included. We included the 90 days period prior to the biopsy date because the date at diagnosis as recorded in the cancer registry can be the date of initial detection through clinical detection of a palpable mass, or through an abnormal screening or diagnostic imaging result, all of which could occur some months prior to the breast biopsy. A definition of 90 days prior to biopsy appeared to capture virtually all diagnoses in these data.

We rolled up multiple biopsies with the same guidance and laterality that were performed on the same day (a very rare occurrence, n of biopsies =37) into one biopsy with the most severe finding as the final result of the biopsy. Similarly, we rolled up multiple lesions per biopsy per woman into one record with the most severe finding being considered as the final result of the biopsy (n=3,464). Therefore the unit of analysis was the biopsy. Based on the comparison of the biopsy findings (benign, high risk benign and malignant) and cancer status within -3 to 12 months of the biopsy, biopsies were defined as true positive (TP), true negative (TN), false positive (FP) and false negative (FN) biopsies. To validate the classification of biopsies, we abstracted records for 20 patients who we defined as false negatives. The results were 100% in agreement with the abstraction.

4.2.1 Definition of Outcomes

A biopsy was considered FN if malignancy was found during the follow up period after the biopsy had shown a benign finding. Similarly, a complete FN biopsy was defined if malignancy was found during the follow up period after the biopsy had shown a benign or high risk benign finding.

The complete false negative rate (cFNR) outcome was estimated to describe the results of the whole diagnosis procedure (115). High risk benign underestimation was defined as the proportion of biopsies with high risk benign finding where an ipsilateral malignancy (DCIS or invasive breast cancer) was diagnosed during the follow up period. Ductal carcinoma *in-situ* Underestimation was defined as the proportion of biopsies with DCIS findings but where ipsilateral invasive breast cancer was diagnosed during the follow up period. Negative predicted value (NPV) was defined as the proportion of benign biopsies without breast cancer diagnosis during the follow up period. Cancer detection rate was defined as the proportion of biopsies associated with a cancer diagnosis during the follow up period.

Within this healthcare organization, all stereotactic CNB were performed with the vacuum assistance while ultrasound CNBs were performed with or without the vacuum assistance. All surgical biopsies were excluded from the study.

4.2.2 Statistical Analysis

Patient characteristics and clinical factors by image guidance and vacuum assistance status were compared using chi-square, Fisher's exact tests and tests for trend as appropriate. The overall FNR, cFNR, NPV, cancer detection rate and underestimation rates were estimated for each modality. To examine whether the performance characteristics varied by image guidance or vacuum assistance, logistic regression with generalized estimating equations (GEE) using binomial distribution and logit link function was fitted to account for facility clusters. The model adjusted for imaging modality, decade of age, race/ethnicity, mammographic breast density, mode of detection, and calendar year of the procedure. To estimate the average population risks and rates by type of imaging modality, marginal standardization in STATA was used (margins command).

4.3 Results

The sample included a total of 22,297 ultrasound and stereotactic imaged guided biopsies. We excluded 609 ultrasound CNBs which had a missing vacuum assistance status and 616 biopsies

that did not have a concordant laterality with ISCR data. Of the remaining 21,072 biopsies, 10,024 were stereotactic CNB and 11,048 were ultrasound CNB (3239 were VA, 7809 were non-VA) performed on 19,524 women.

Compared to stereotactic CNB, women who received VA-ultrasound CNB were younger, premenopausal, tend to have denser breasts, less likely to have a history of prior biopsy, more parous and were more likely to be minorities. Further, women receiving VA-ultrasound CNB were more likely to have symptomatic lesions and malignant findings than those with stereotactic CNB (32% vs 19%). Additionally, tumor characteristics such as grade and progesterone receptor were similar in both biopsy procedures; however, VA-ultrasound CNBs were more likely to identify tumors with estrogen receptor negative (ER-) status (Table XVII).

Women receiving VA-ultrasound CNB and non-VA-ultrasound CNB were similar in family history and history of prior biopsy. Vacuum-assisted-ultrasound CNB women, however, were more likely to be non-Hispanic (nH)-Black, older, post-menopausal, parous, with lower breast density and their lesions were more likely to be screen detected compared to women who received non-VA-ultrasound CNB. Tumor characteristics were similar between VA-ultrasound CNB and non-VA-ultrasound CNB except for ER- which was higher in VA-ultrasound CNB (Table XVII).

Table XVIII presents the unadjusted analysis for the predictors of FNR and cFNR. Similar patterns by patient characteristics were observed for both outcomes. The risk of FNR or cFNR was highest in nH-Whites compared to nH-Backs and Hispanics, younger age groups, heterogeneous and dense breasts, women with history of a prior biopsy and pre-menopausal women. Further, the risk of FNR and cFNR was higher in early stage tumors, low grade tumors and tumors with a DCIS behavior.

After adjusting for patient characteristics, compared to VA-ultrasound CNB FNR was 0.6% point higher for stereotactic CNB (5.4% vs 4.8%, p=0.5) and was 1% lower for non-VA-ultrasound CNB (3.8% vs 4.8%, p=0.4). Regardless of image guidance modality, FNR was highest for women in

their 50's and 60's, women with denser breasts, nH-Blacks compared to nH-Whites and Hispanics and symptomatic women (Table XIX).

Similarly, compared to VA-ultrasound CNB, cFNR was higher for stereotactic CNB (11.3% vs 7.0%, p<0.01) and was 1.9% lower for non-VA-ultrasound CNB (5.1% vs 7.0%, p=0.22). With regards to the patient characteristics, cFNR was highest in women with dense breasts, younger women and in nH-Blacks compared to nH-Whites and Hispanics (Table XIX).

Table XX summarizes the adjusted performance characteristics for each of the modalities. With regards to the NPV, stereotactic CNB was associated with the highest NPV followed by VA-ultrasound CNB and non-VA-ultrasound CNB (98.4%, 97.4% and 98.1%, respectively) but had lower DCIS underestimation (20%, 48% and 65.9% for stereotactic, VA-ultrasound and non-VA ultrasound CNB, respectively) and lower cancer detection rate per 1000 biopsies (169.5, 293.6 and 273.7, for stereotactic, VA-ultrasound and non-VA-ultrasound CNB, respectively) . The high risk underestimation was similar across all three biopsy techniques (Table XX).

4.4 Discussion

This study is one of the few large studies that have examined the accuracy of image guided CNB using data from a community setting which collected information from multiple sites. We found that among women with breast cancer a false negative (benign) finding was highest in stereotactic CNB (5.4%), followed by VA-ultrasound CNB (4.8%) and non-VA-ultrasound CNB (3.8%). When including high-risk benign lesions in the definition of cFNR, we found that a breast cancer was more often missed by stereotactic CNB (11.3%) and to a lesser degree with VA-ultrasound CNB (7.0%) and non-VA-ultrasound CNB (5.1%), respectively.

The sensitivity (1-FNR) of CNB by imaging modality and vacuum assistance has been estimated in several studies, which were subsequently pooled in three different systematic literature reviews (75, 87, 88, 116) all of which reported similar results. The AHRQ updated review (88) reported higher sensitivities (lower FNR) than the results from this study, with a sensitivity of 99% (95% CI: 98-99) for stereotactic CNB, 97% (95% CI: 92-99) for VA-ultrasound CNB and 99% (95% CI:

98-99) for non-VA-ultrasound CNB. While our study used population-based cancer registry data as the gold standard for ruling in and ruling out a breast cancer diagnosis, most of the prior studies relied on the surgical biopsy results as the gold standard thus assuming that open surgery has a 100% accuracy which is not always true. The FNR for open surgery has been estimated to be at 1-2% (117) which would tend to artificially reduce the estimated FNR (and artificially raise the sensitivity) in these pooled studies. In addition, these studies relied on their pathology data such as surgical biopsy as their gold standard in ruling the cancer diagnosis during a short follow up period which may have underestimated their FNRs due to their limited ability in capturing all cancer cases that were diagnosed outside their network. Another limitation of the studies that were included in the AHRQ report is that the majority of these studies came from a single healthcare facility (clinic or hospital) which might impact the external validity of the data and is in contrary to our study which included a diverse sample of women from multiple facilities conducted over a 10 year span (2001-2010)

With respect to our definition of cFNR, our results are in line with the published literature, the majority of which are from Europe, which have defined sensitivity alternatively as absolute sensitivity and complete sensitivity. Absolute sensitivity refers to the proportion of breast cancers that are identified as malignant on breast biopsy and complete sensitivity refers to the proportion of breast cancers that are identified as uncertain malignant potential, suspicious of malignancy, and malignant. Therefore, the compliment of absolute sensitivity (1-absolute sensitivity) is similar to our definition of cFNR. Our estimates for cFNR with ultrasound CNB and stereotactic CNB are in line with the published literature (80% - 96%) (115, 118-122).

The high risk underestimations for both modalities and by VA status are consistent with the published literature. Our estimates of high risk underestimation for stereotactic CNB, VA-ultrasound CNB and non-VA-ultrasound CNB (23.7%, 23.8% and 22.0%, respectively) are well within the 95% confidence intervals from the pooled estimates reported in the AHRQ report of 18% (95% CI: 13-24),

11% (95% CI: 2-33) and 25% (95% CI: 16-36) for stereotactic, VA-ultrasound and non-VA-ultrasound CNB, respectively.

The DCIS underestimation of 20% (95% CI: 7-33) for stereotactic CNB is similar to the pooled estimate in the AHRQ review 26% (95% CI: 19 -36); however that for VA-ultrasound CNB in our study is higher. This is possibly due to the small sample of DCIS cancers that were diagnosed with ultrasound CNB in our study. Negative predicted value for both modalities and by VA- status from our study was also consistent with the estimates published in the AHRQ report.

In our study, we examined whether patient factors were associated with FNR which the majority of the published studies did not address. Our results found that both FNR and missed detection increase with increased breast density. This is consistent with findings from a study of 180 core biopsies which found that missed detection were more likely to occur in women with denser breasts (123). We also found that FNR and cFNR are higher for pre-menopausal women and decrease with age while, contrary to expectation FNR and cFNR were lower for women with screen-detected findings versus symptomatic presentation. We were unable to examine whether the number of cores had an impact on the accuracy of CNB. It was reported that taking multiple cores improves the accuracy of the diagnosis (123-125).

Our study included a large sample of biopsies that were conducted in a large community practice with multiple facilities spanning more than 10 years. Although we accounted for biopsy facility in our analysis we did not account for differences in the radiologist who performed the biopsy, such as level of expertise which has been shown to have impact upon the accuracy of the procedure (126, 127). The relatively small number of the false negative biopsies limited our ability to examine the accuracy of CNB within levels of patient and tumor characteristics.

We relied on population-based cancer registry data to serve as the gold standard for cancer diagnosis which minimizes attrition and drop outs that may have been an issue in earlier studies that relied on open surgery as the gold standard for cancer. Reliance on population-based cancer registry data comes with limitations as well, due to the inability in some instances to directly map the breast

cancer diagnosis in the registry with the specific lesions or biopsies involved. Stemming from this issue, there are three scenarios which could cause some misclassification in our estimates of FNR and cFNR. First, we defined a biopsy result as a false negative or complete false negative rate when any breast cancer was diagnosed within 12 months of the biopsy and in the same breast as the non-malignant biopsy result. Therefore a true interval breast cancer (one that was truly undetectable in the breast at biopsy) would be called a false negative in our study, which would artificially increase our FNR and cFNR estimates, and could explain the increased FNR among more aggressive tumors. However, in a sensitivity analysis in which we reduced the follow-up period to 6 months while including cancers that were diagnosed 90 days prior to the biopsy date, both FNR and cFNR were slightly reduced. The FNR for stereotactic CNB attenuated from 5.4% to 5.1% and from 4.8% to 4.5% for VA-ultrasound CNB and from 3.8% to 3.7% for non-VA-ultrasound CNB. The attenuated from 11.3% to 10.9%, 7.0 to 6.6% and 5.1 % to 4.9% for stereotactic CNB, VA-ultrasound CNB, and non-VA-ultrasound CNB, respectively). These analyses suggest that the interval cancers had minimal effect on our estimates for FNR and cFNR

Second, because we did not have lesion level data from the cancer registry, we rolled up biopsy results for 3,464 biopsies with multiple lesions (16.4% out of a total of 21,072 biopsies) and took the most malignant finding to be the overall biopsy finding. Of these 2,619 biopsies with multiple lesions 48(1.8%) had discordant findings. The overall FNR for the biopsies with multiple lesions was (5.8%) and the cFNR was (9.8%), both estimates are comparable to the overall estimates of FNR (5.1%) and cFNR (9%) when including all biopsies in the analysis. Further, excluding biopsies with multiple lesions had a minimal effect on the FNR for stereotactic CNB and non-VA-ultrasound CNB but decreased the FNR by 1 point for VA-ultrasound CNB (3.8% vs 4.8%) (Table XXI). Our approach in handling these discordant findings would protect us from artificially overestimating the FNR and cFNR. Third, for patients with multiple biopsy procedures with discordant results within a few weeks or months of each other and in the same breast, we were unable to tell which biopsy result should be linked to the subsequent breast cancer diagnosis in that breast. For example, if

there were two biopsies a couple months apart (one benign and one malignant) and both within 12 months of a breast cancer diagnosis, the malignant result would be called a true positive and the benign a false negative when it might have been a true negative. In our sample there were 1525 biopsies that were subsequently followed by another biopsy during the follow up period. Of all cancers that were diagnosed during the follow-up period, 508 (9.7%) had multiple biopsies and 74(4.8%) and 108 (7.1%) of these biopsies were classified as false negatives and complete false negatives. In an additional set of analyses we examined the effect of multiple lesions and repeated biopsies within 12 months on the FNR. Removing biopsies with multiple biopsies within a 12 months period reduced the FNR for stereotactic CNB by 1.6 points (3.8% vs 5.4%), 0.5% for VA-ultrasound CNB (4.3% vs 4.8%) and 0.3% for non-VA-ultrasound CNB (3.5% vs 3.8%). Finally, removing biopsies with multiple lesions and multiple biopsies reduced the FNR by 1.3 points 1.5 points for stereotactic CNB and ultrasound CNB but did not impact the FNR for non-VA-ultrasound CNB (Table XXI). These analyses suggest that including repeated biopsies and rolling up multiple lesions may have inflated the FNR estimates for the CNB however excluding patients with multiple biopsies and biopsies with multiple lesions may reduce the internal validity of our data.

In conclusion, we estimated the performance of image guided biopsies using data from a community setting. Both stereotactic and ultrasound image guided biopsies demonstrate high accuracy in detection of breast cancers. Results from this study supplement the literature on the topic and support the recommendations for using image guided biopsies as an alternative to the open surgery to reduce burden and morbidity on women.

Table XVII.

DISTRIBUTION OF CORE NEEDLE BIOPSIES BY IMAGE GUIDANCE, VACUUM ASSISTED STATUS AND SELECTED CHARACTERISTICS (N=21,072)

	Stereotactic CNB[a] (N=10, 024)		VA-Ultrasound CNB[a] (N=3239)		Non-VA-Ultrasound CNB[a] (N=7809)	
	N	%	N	%	N	%
Findings						
Benign	7,638	76	2,063	64	5,544	71
High Risk Benign	541	5	158	5	155	2
DCIS	664	7	61	2	70	1
Invasive	1,181	12	957	30	2,040	26
Ethnicity						
nH-White	6,434	64	1,177	36	5,872	75
nH-Black	2,524	25	1,722	53	487	6
Hispanic	358	4	171	5	569	7
Other	457	5	143	4	632	8
Unknown	251	3	26	1	249	3
Age						
<50	3,279	33	1,434	44	3,962	51
50-59	3,145	31	707	22	1,763	23
60-69	2,050	21	557	17	1,073	14
70+	1,550	16	541	17	1,011	13
Breast Density[1]						
Fatty	743	7	251	9	493	7
Scattered	3,509	36	976	33	1,912	28
Heterogeneous	4,588	47	1,350	46	3,413	49
Dense	948	10	365	12	1,135	16
Family history						
None	6,263	62	2,036	63	4,899	63
Weak	1,683	17	569	18	1,418	18
Moderate	1,390	14	427	13	967	12
Strong	688	7	207	6	525	7
Prior biopsy						
Yes	2,664	27	754	23	1,845	24
None	7,360	73	2,485	77	5,964	76
Parity						
Nulliparous	1,342	13	378	12	1,099	14
Parous	7,320	73	2,552	79	5,234	67
Unknown	1,362	14	309	10	1,476	19
Menopause						
Pre-menopausal	4,148	41	1,646	51	4,682	60
Post-menopausal	5,876	59	1,593	49	3,127	40
Mode of Lesion detection						
Symptomatic	2,960	30	1,664	51	4,543	58
Screen detected	7,064	70	1,575	49	3,266	42
Stage						
Early	286	16	472	49	952	48
Late	1,473	84	486	51	1,046	52
Behavior						
DCIS	836	47	45	5	63	3
Invasive	950	53	931	95	1,983	97
Grade						
Low	349	21	192	20	433	22
Moderate	723	44	417	44	868	44
High	587	35	336	36	681	34
Estrogen Receptor						
Negative	224	15	228	24	332	19
Positive	1,223	85	726	76	1,388	81
Progesterone Receptor						
Negative	354	25	254	27	428	25
Positive	1,084	75	701	73	1,282	75

[a] Core needle biopsy; VA vacuum assisted; NH-non Hispanic; [1] 1,389 biopsies were missing mammographic density.

Table XVIII.

PREDICTORS OF FALSE NEGATIVE RATE AND COMPLETE FALSE NEGATIVE RATE BY SELECTED CHARACTERISTICS

	N[1]	FNR[2] (CI)	P-value[3]	N[1]	cFNR[4] (CI)	P-value[3]
Overall	5,285	5.2 (4.7 ,5.9)		5,285	9.0 (8.3 ,9.8)	
Modality						
Stereotactic CNB[a]	2,030	5.7 (4.8 ,6.8)	0.13	2,030	12.1 (10.7 ,13.6)	0.01
Non-VA-Ultrasound CNB[a]	2,201	5.1 (4.4 ,6.2)	0.35	2,201	7.0 (6.0 ,8.2)	0.70
VA-ultrasound CNB[a]	1,054	4.5 (3.4 ,5.9)		1,054	7.4 (5.9 ,9.1)	
Ethnicity			0.01			0.20
nH White	3,577	5.5 (4.8 ,6.3)		3,577	9.1 (8.2 ,10.1)	
nH Black	1,291	3.9 (2.9 ,5.1)		1,291	7.4 (6.1 ,8.9)	
Hispanic	216	4.2 (2.2 ,7.8)		216	8.8 (5.7 ,13.4)	
Other	243	7.8 (5.0 ,11.9)		243	12.8 (9.1 ,17.6)	
Unknown	114	9.7 (5.4 ,16.6)		114	14.0 (8.8 ,21.7)	
Age			0.004			<0.001
<50	1,296	6.3 (5.1 ,7.7)		1,296	10.3 (8.8 ,12.1)	
50-59	1,394	5.8 (4.7 ,7.2)		1,394	10.7 (9.2 ,12.4)	
60-69	1,279	5.2 (4.1 ,6.6)		1,279	8.5 (7.1 ,10.2)	
70+	1,472	3.9 (3.1 ,5.1)		1,472	6.5 (5.3 ,7.8)	
Breast Density[2]			0.03			<0.001
Fatty	591	5.1 (3.6 ,7.2)		591	8.0 (6.0 ,10.4)	
Scattered	1,799	4.3 (3.4 ,5.3)		1,799	7.2 (6.1 ,8.5)	
Heterogeneous	2,150	5.3 (4.5 ,6.4)		2,150	9.8 (8.6 ,11.1)	
Dense	448	7.8 (5.7 ,10.7)		448	14.3 (11.3 ,17.8)	
Family history[3]			0.60			0.50
None	3,285	5.2 (4.5 ,6.1)		3,285	8.7 (7.8 ,9.7)	
Weak	892	6.1 (4.7 ,7.8)		892	10.3 (8.5 ,12.5)	
Moderate	862	5.0 (3.7 ,6.7)		862	8.4 (6.7 ,10.4)	
Strong	402	4.5 (2.8 ,7.0)		402	9.2 (6.7 ,12.4)	
Prior biopsy			<0.001			<0.001
Yes	1,342	6.6 (5.7, 8.4)		1,342	11 (9.4, 12.8)	
None	4,099	3.2 (2.5, 3.6)		4,099	6.4 (5.6, 7.1)	
Parity			0.24			0.50
Nulliparous	676	6.4 (4.8 ,8.5)		676	9.9 (7.9 ,12.4)	
Parous	4,072	5.2 (4.6 ,5.9)		4,072	8.9 (8.1 ,9.8)	
Unknown	693	4.6 (3.3 ,6.5)		693	8.2 (6.4 ,10.5)	
Menopause			0.03			0.12
Pre-menopause	2,119	6.1 (5.1 ,7.2)		2,119	9.8 (8.6 ,11.1)	
Post-menopause	3,322	4.8 (4.1 ,5.5)		3,322	8.4 (7.5 ,9.4)	
Mode of Lesion detection			0.06			0.61
Symptomatic	2,772	5.8 (5.0 ,6.8)		2,772	8.8 (7.8 ,10.0)	
Screen detected	2,669	4.7 (3.9 ,5.6)		2,669	9.1 (8.0 ,10.2)	
ER			0.22			0.003
Negative	853	3.9 (2.8 ,5.4)		853	5.3 (4.0 ,7.0)	
Positive	3,710	4.9 (4.2 ,5.6)		3,711	8.6 (7.7 ,9.5)	
Pr			0.50			<0.001
Negative	1,132	4.3 (3.3 ,5.7)		1,132	5.7 (4.5 ,7.3)	
Positive	3,410	4.8 (4.1 ,5.6)		3,411	8.7 (7.8 ,9.7)	
Stage at Diagnosis			<0.001			<0.001
Early Stage	3,372	6.2 (5.4 ,7.0)		3,372	11.6 (10.6 ,12.8)	
Late Stage	1,966	3.8 (3.1 ,4.8)		1,966	4.6 (3.7 ,5.6)	
Grade			0.16			<0.001
Low	1,138	5.8 (4.6 ,7.3)		1,139	11.9 (10.1 ,13.9)	
Moderate	2,238	4.7 (3.9 ,5.7)		2,237	7.8 (6.7 ,9.0)	
High	1,752	4.6 (3.7 ,5.6)		1,752	5.5 (4.6 ,6.7)	
Behavior			<0.001			<0.001
DCIS	1,211	9.4 (7.9 ,11.2)		1,211	21.5 (19.2 ,23.9)	
Invasive	4,230	4.1 (3.5 ,4.7)		4,230	5.4 (4.7 ,6.1)	

[1] Total number of cancer cases within 12 months from the biopsy date;

[2] False negative rate includes cancers with benign biopsy findings

[3] P-value for linear trend for ordinal variables age, time since last mammogram and density, and chi-square test for nominal categorical variables

[4] Complete False Negative Rate includes cancers with a benign or high risk benign biopsy findings

[a] Core needle biopsy; NH-non Hispanic; CI-Confidence interval

Table XIX.

ADJUSTED PREDICTORS OF FALSE NEGATIVE RATE AND COMPLETE FALSE NEGATIVE RATE BY SELECTED CHARACTERISTICS[1]

	N[2]	FNR[3] (CI)	P-value[5]	N[2]	cFNR[4] (CI)	P-value
Modality[a]						
Stereotactic CNB[b]	1,920	5.4 (3.5 ,7.3)	0.5	1,920	11.3 (7.6, 15.1)	<0.01
Non-VA-Ultrasound CNB	1,894	3.8 (2.1 ,5.4)	0.4	1,894	7.0 (3.7, 10.3)	0.22
VA-ultrasound CNB	978	4.8 (2.6 ,7.0)		978	5.1 (2.5, 7.8)	
Ethnicity			0.3			0.5
nH-White	3,189	4.1 (2.4 ,5.8)		3,189	7.2 (4.3 , 10.1)	
nH-Black	1,134	5.5 (3.1 ,8.0)		1,134	9.3 (5.6 , 13.1)	
Hispanic	183	3.8 (0.6 ,7.0)		183	7.7 (2.9 , 12.6)	
Other	214	5.0 (1.6 ,8.4)		214	7.0 (2.7 , 11.3)	
Unknown	72	6.8 (1.0 ,12.6)		72	14.7 (5.9 , 23.5)	
Age[e]			0.7			0.3
<50	1,089	4.5 (2.4 ,6.5)		1,089	7.7 (4.4 , 11.1)	
50-59	1,239	4.9 (2.9 ,6.9)		1,239	9.2 (5.7 , 12.7)	
60-69	1,145	4.8 (2.8 ,6.9)		1,145	7.7 (4.4 , 10.9)	
70+	1,319	4.1 (2.2 ,5.9)		1,319	6.6 (3.5 , 9.6)	
Breast Density			0.3			0.1
Fatty	564	5.0 (2.4 ,7.6)		564	7.4 (3.7, 11.2)	
Scattered	1,721	3.9 (2.2 ,5.7)		1,721	6.6 (3.7, 9.5)	
Heterogeneous	2,072	4.5 (2.7 ,6.2)		2,072	8.4 (5.1, 11.6)	
Dense	435	7.0 (3.7 ,10.2)		435	11.6 (6.9, 16.2)	
Menopause status[f]			0.14			0.6
Pre-Menopausal	1,606	4.5 (2.7 ,6.3)		1,606	8.2 (5.1 ,11.3)	
Post-Menopausal	3,186	4.6 (2.8 ,6.4)		3,186	7.7 (4.7 ,10.8)	
Mode of lesion detection			0.2			0.6
Symptomatic	2,243	5.1 (3.2 ,7.0)		2,243	8.1 (5.0 ,11.2)	
Asymptomatic	2,549	4.2 (2.5 ,5.8)		2,549	7.7 (4.8 ,10.7)	

nH-non Hispanic; CI-Confidence Interval

[1] GEE model adjusted for imaging guidance, race/ethnicity, age, breast density, menopause, mode of lesion detection and year of biopsy

[2] Number of cancers 3 months before and 12 months after biopsy

[3] False Negative Rate; CI confidence interval

[4] Missed cancer detection includes cancers with a benign or high risk benign biopsy findings

[5] P-value for linear trend for ordinal variables age and breast density and type III analysis for nominal categorical variables (all other variables)

[a] Each modality was compared to VA-USCNB

[b] Core needle biopsy

[e] GEE model adjusted for imaging guidance, race/ethnicity, age, breast density mode of lesion detection and year of biopsy

[f] GEE model adjusted for imaging guidance, race/ethnicity, breast density, menopause mode of lesion detection and year of biopsy

Table XX.

ADJUSTED PERFORMANCE MEASURES BY IMAGING MODALITY AND VACUUM ASSISTED STATUS

	Stereotactic CNB[a]		VA-Ultrasound CNB[a]		Non-VA-Ultrasound CNB[a]	
	N[7]	% (CI)[8]	**N[7]**	% (CI)[8]	**N[7]**	% (CI)[8]
False Negative Rate[1]	1,920	5.4 (3.5, 7.3)	978	4.8 (2.6, 7.0)	1,894	3.8 (2.1, 5.4)
Complete False Negative Rate[2]	1,920	11.3 (7.6, 15.1)***	978	7.0 (3.7, 10.3)	1,894	5.1 (2.5, 7.8)
Negative Predicted Value[3]	7,502	98.6 (98.9, 98.2)**	1,846	97.4 (98.3, 96.5)	4,967	98.1 (98.6, 97.6)*
Cancer Detection per 1000 biopsies[4]	9,756	169.5 (154.0, 185.0)***	2,937	293.6 (265.8, 321.4)	6,917	273.7 (254.3, 293.0)+
ADH Underestimation5 [5]	519	23.7 (10.7, 36.7)	147	23.8 (8.7, 38.9)	138	22.0 (9.0, 35.0)
DCIS Underestimation[6]	635	20.0 (7.0, 33.0)***	61	48.0 (24.0, 71.5)	59	65.9 (45.9, 85.9)***

+P<0.1; *P<0.05; **P<0.01; ***P<0.001

[1] Percentage of cancers with a benign biopsy finding

[2] Percentage of cancers with a benign or high risk benign biopsy finding

[3] Percentage of benign biopsies without cancer diagnosis in the subsequent 12 months

[4] Rate per 1000 of all biopsies associated with a cancer diagnosis in the subsequent 12 months

[5] Among biopsies with an ADH or ALH finding, the percentage associated with a diagnosis of DCIS or invasive cancer in the subsequent 12 months

[6] Among biopsies with DCIS findings, the percentage associated with a diagnosis of invasive cancer in the subsequent 12 months

[7] Total number of cancer cases within 12 months from the biopsy date;

[8] GEE model adjusted for imaging guidance, race/ethnicity, age, breast density, menopause, year of biopsy

[b] Core needle biopsy; CI-confidence Interval

Table XXI.

SENSITIVITY ANALYSIS TO EXAMINE THE EFFECT OF MULTIPLE LESIONS AND MULTIPLE BIOPSIES ON THE FNR

	Stereotactic CNB[a]		VA-Ultrasound CNB[a]		Non-VA-Ultrasound CNB[a]	
	N[1]	FNR[2]	N[1]	FNR[2]	N[1]	FNR[2]
All lesions /All biopsies	1,920	5.4 (3.5 ,7.3)	978	4.8 (2.6, 7.0)	1,894	3.8 (2.1, 5.4)
Single lesion / All biopsies	1,704	5.7 (4.1, 7.3)[+]	799	3.8 (2.1, 5.5)	1,546	3.9 (2.5, 5.3)
Single lesion /Single biopsy	1490	4.1 (3.1, 5.1)	799	3.3 (1.9, 4.7)	1546	3.8 (2.8, 4.7)
All lesions/Single biopsy	1681	3.8 (2.5, 5.2)	905	4.3 (2.5, 6.1)	1746	3.5 (2.2, 4.7)

+P<0.1

[1] Number of cancers during the follow-up period

[2] Percentage of cancers with a benign or high risk benign biopsy finding estimated using GEE model adjusted for imaging guidance, race/ethnicity, age, breast density, menopause, year of biopsy and mode of lesion detection

[a] Core needle biopsy; CI –Confidence Interval

5. Conclusions

5.1 Public Health Significance

Although mammography screening has been established as a routine public health procedure for early detection and a major contributor to the reduction of breast cancer mortality rates, it is also associated with substantial false positive results. In the USA, the false positive rates have been estimated to be as high as 10% and 50% of women who screen annually for 10 years will experience at least one false positive. In most cases resolving a positive mammogram includes additional imaging in the form of a diagnostic mammogram, ultrasound or MRI. In about 20% of the cases a biopsy will be performed. In addition to the unnecessary work up, receiving a FP result impacts the psychological wellbeing of the women by increasing acute and chronic anxiety, distress, and worry of having breast cancer. In some cases, anxiety and worry can last up to 3 years after the experience of a false positive mammogram. In addition to the psychological burden there is an increased economic burden on both the women and the public healthcare systems to resolve a positive mammogram. Another consequence of experiencing a false positive mammogram is delaying or relinquishing subsequent mammography screening among women who experience a false positive mammogram.

Similar to mammography screening, image guided biopsies have been designed to accurately diagnose breast tumors. Although these diagnostic procedures have been established as highly accurate, a missed cancer detection or a false negative biopsy can lead to additional burden and morbidity which includes additional biopsies. In the literature, the false negative rate varies according to the imaging technique such that it is about 1% with stereotactic image guided, 3% with vacuum-assisted ultrasound and 1% with non-vacuum-assisted ultrasound image guided biopsy according to the AHRQ review. In comparison, in the present study the false negative rate was 5.4% with stereotactic, 4.8% with vacuum-assisted ultrasound and 3.8% with non-vacuum-assisted ultrasound core needle biopsy. A complete false negative rate, a measure that is not usually

reported in the USA, ranges between 4%-15% based on studies from Europe. In the present study, we found results similar to those reported in Europe.

5.2 Future Directions

This thesis was designed to examine the burden and morbidity associated with the mammography screening and image guided breast biopsies. Although the evidence suggests that digital mammography and screen film mammography share similar performance characteristics, digital mammography supplanted screen film nationwide. As was discussed in Aim 1 of this thesis, digital mammography marginally reduced the false positive rate and burden but not the morbidity of false positive compared to screen-film mammography. These findings along with the evidence in literature have sparked the debate whether the added cost and learning curve associated with the adoption of digital mammography justify the need to adopt it. The findings presented in this dissertation are generally reassuring that there is little concern regarding the potential for an increased false positive burden due to establishing FFDM into practice. Tomosynthesis is a newer technology that generates a three-dimensional image of the breast and is attracting attention due to its improved accuracy in detecting abnormal lesions and reducing burden and morbidity in comparison to two-dimensional digital mammography. One future study would be to examine the effectiveness, burden and morbidity of adopting Tomosynthesis using a large national dataset. Results from such a study will provide evidence with regards to the added benefit of Tomosynthesis which will elucidate the decision to adopt this technology.

Another potential study that could be viewed as an extension to this thesis is to replicate Aim 2 using a national dataset. In Aim 2, the impact of receiving a false positive mammogram on women's screening behavior was examined. Our findings from this study suggested that the risk of delaying or skipping the subsequent screening mammography is higher in women with false positive compared to those with true negative mammography. This result was inconsistent with the majority of literature in the U.S, except for one study which used medical claims data and found similar results to our study. Future studies are needed to examine whether similar results can be replicated

using other data sources or different study designs. A prospective study design or registry would be a potential design which collects complete routine mammography data from all participants across different facilities and specifically when participants switch between healthcare organizations.

A third potential study would be to estimate the accuracy and morbidity associated with MRI image guided biopsy. In Aim 3 we examined the accuracy of stereotactic and ultrasound core needle biopsy but not MRI guided biopsy due to the small number of MRI biopsies that were performed. Currently, the evidence in literature is scant and the AHRQ report highlights the need to conduct studies that describe the benefits and harms of MRI guided biopsy. For example, one could use national data from the Breast Cancer surveillance Consortium to examine the accuracy of core needle biopsies including MRI guided, ultrasound guided and stereotactic guided biopsies.

5.3 Summary of Findings

The research work in this thesis assessed the burden and morbidity associated with mammography screening and breast cancer diagnostic resembled by image guided biopsy. The main findings from the three studies are as follows: (i) A modest reduction in the false positive rate and rate of work up was associated with digital mammogram compared to screen film, (ii) digital mammography was associated with an increased risk of receipt of a breast biopsy compared to screen-film, (iii) women who experience a false positive mammogram have a higher risk of delaying or skipping their subsequent screening as compared to women with a true negative mammogram, (iv) false negative rate and the complete false negative rate, which are a major driver of additional morbidity, were found to be higher for stereotactic core needle biopsy compared to vacuum-assisted ultrasound image-guided biopsy and (v) and were higher in vacuum-assisted compared to non-vacuum assisted ultrasound core needle biopsy.

Although the evidence from this thesis highlights the shortcomings of mammography screening and core needle biopsies, it should not discourage women from receiving their mammography screening on schedule or follow-up to resolve any suspicious findings. It is beyond doubt that early detection and accurate diagnosis have reduced mortality over the past decades. It is

important for women to follow the guidelines that were set forth by country-specific experts. Additionally, image guided biopsies are considered highly accurate and are recommended to be used as an alternative to excisional biopsies or open surgery.

REFRENCES

1. Martin AM, Weber BL. Genetic and hormonal risk factors in breast cancer. J Natl Cancer Inst. 2000 Jul 19;92(14):1126-35.

2. Breast Cancer Risk Factors [Internet].; 2016 [updated 01/15/2016; cited 02/10/2016]. Available from: http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-risk-factors.

3. **SEER Stat Fact Sheets: Female Breast Cancer** [Internet]. [cited 02/10/2016]. Available from: http://seer.cancer.gov/statfacts/html/breast.html.

4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin. 2015 Jan-Feb;65(1):5-29.

5. Breen N, Gentleman JF, Schiller JS. Update on mammography trends: comparisons of rates in 2000, 2005, and 2008. Cancer. 2011 May 15;117(10):2209-18.

6. National Center for Health Statistics (US). **Health, United States, 2012 With Special Feature on Emergency Care**
. Hyattsville, MD: Health, United States; 2013. Report No.: 2013-1232.

7. Cutler DM, Deaton AS, Lleras-Muney A, National Bureau of Economic Research. The Determinants of Mortality. 2006;w11963.

8. Pisano ED, Fajardo LL, Caudry DJ, Sneige N, Frable WJ, Berg WA, et al. Fine-needle aspiration biopsy of nonpalpable breast lesions in a multicenter clinical trial: results from the radiologic diagnostic oncology group V. Radiology. 2001 Jun;219(3):785-92.

9. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005 Oct 27;353(17):1773-83.

10. **Screen-film mammography: Headed the way of xeromammography?** [Internet]. U.S. Food and Drug Administration; 2015 [updated 09/30/2015; cited 02/10/2016]. Available from: http://www.fda.gov/radiation-emittingproducts/mammographyqualitystandardsactandprogram/facilityscorecard/ucm465058.htm.

11. Lewin JM, Hendrick RE, D'Orsi CJ, Isaacs PK, Moss LJ, Karellas A, et al. Comparison of full-field digital mammography with screen-film mammography for cancer detection: results of 4,945 paired examinations. Radiology. 2001 Mar;218(3):873-80.

12. Kerlikowske K, Hubbard RA, Miglioretti DL, Geller BM, Yankaskas BC, Lehman CD, et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. Ann Intern Med. 2011 Oct 18;155(8):493-502.

13. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology. 2008 Feb;246(2):376-83.

14. Skaane P, Skjennald A, Young K, Egge E, Jebsen I, Sager EM, et al. Follow-up and final results of the Oslo I Study comparing screen-film mammography and full-field digital mammography with soft-copy reading. Acta Radiol. 2005 Nov;46(7):679-89.

15. Del Turco MR, Mantellini P, Ciatto S, Bonardi R, Martinelli F, Lazzari B, et al. Full-field digital versus screen-film mammography: comparative accuracy in concurrent screening cohorts. AJR Am J Roentgenol. 2007 Oct;189(4):860-6.

16. Perry NM, Patani N, Milner SE, Pinker K, Mokbel K, Allgood PC, et al. The impact of digital mammography on screening a young cohort of women for breast cancer in an urban specialist breast unit. Eur Radiol. 2011 Apr;21(4):676-82.

17. Karssemeijer N, Bluekens AM, Beijerinck D, Deurenberg JJ, Beekman M, Visser R, et al. Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. Radiology. 2009 Nov;253(2):353-8.

18. Domingo L, Romero A, Belvis F, Sanchez M, Ferrer J, Salas D, et al. Differences in radiological patterns, tumour characteristics and diagnostic precision between digital mammography and screen-film mammography in four breast cancer screening programmes in Spain. Eur Radiol. 2011 Sep;21(9):2020-8.

19. Sala M, Salas D, Belvis F, Sanchez M, Ferrer J, Ibanez J, et al. Reduction in false-positive results after introduction of digital mammography: analysis from four population-based breast cancer screening programs in Spain. Radiology. 2011 Feb;258(2):388-95.

20. Strax P, Venet L, Shapiro S. Value of mammography in reduction of mortality from breast cancer in mass screening. Am J Roentgenol Radium Ther Nucl Med. 1973 Mar;117(3):686-9.

21. Moss SM, Wale C, Smith R, Evans A, Cuckle H, Duffy SW. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. Lancet Oncol. 2015 Sep;16(9):1123-32.

22. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet. 2002 Mar 16;359(9310):909-19.

23. Bjurstam N, Bjorneld L, Warwick J, Sala E, Duffy SW, Nystrom L, et al. The Gothenburg Breast Screening Trial. Cancer. 2003 May 15;97(10):2387-96.

24. Tabar L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A, et al. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. Cancer. 1995 May 15;75(10):2507-17.

25. Habbema JD, van Oortmarssen GJ, van Putten DJ, Lubbe JT, van der Maas PJ. Age-specific reduction in breast cancer mortality by screening: an analysis of the results of the Health Insurance Plan of Greater New York study. J Natl Cancer Inst. 1986 Aug;77(2):317-20.

26. Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. BMJ. 2014 Feb 11;348:g366.

27. Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of Breast Cancer Screening: Systematic Review and Meta-analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med. 2016 Jan 12.

28. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. Br J Cancer. 2013 Jun 11;108(11):2205-40.

29. Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2013 Jun 4;6:CD001877.

30. Canadian Task Force on Preventive Health Care, Tonelli M, Connor Gorber S, Joffres M, Dickinson J, Singh H, et al. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. CMAJ. 2011 Nov 22;183(17):1991-2001.

31. Duffy SW., Ming-Fang Yen A., Hsiu-Hsi Chen T. Long-term benefi ts of breast screening. 2012; 1: 31–38. Breast Cancer Management. 2012;1:31-8.

32. Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. Radiology. 2009 Dec;253(3):641-51.

33. Miglioretti DL, Gard CC, Carney PA, Onega TL, Buist DS, Sickles EA, et al. When radiologists perform best: the learning curve in screening mammogram interpretation. Radiology. 2009 Dec;253(3):632-40.

34. Kan L, Olivotto IA, Warren Burhenne LJ, Sickles EA, Coldman AJ. Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. Radiology. 2000 May;215(2):563-7.

35. Smith-Bindman R, Ballard-Barbash R, Miglioretti DL, Patnick J, Kerlikowske K. Comparing the performance of mammography screening in the USA and the UK. J Med Screen. 2005;12(1):50-4.

36. Roman M, Hubbard RA, Sebuodegard S, Miglioretti DL, Castells X, Hofvind S. The cumulative risk of false-positive results in the Norwegian Breast Cancer Screening Program: updated results. Cancer. 2013 Nov 15;119(22):3952-8.

37. Esserman L, Cowley H, Eberle C, Kirkpatrick A, Chang S, Berbaum K, et al. Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst. 2002 Mar 6;94(5):369-75.

38. Hofvind S, Ponti A, Patnick J, Ascunce N, Njor S, Broeders M, et al. False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. J Med Screen. 2012;19 Suppl 1:57-66.

39. Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med. 1998 Apr 16;338(16):1089-96.

40. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Ann Intern Med. 2011 Oct 18;155(8):481-92.

41. Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. Psychooncology. 2005 Nov;14(11):917-38.

42. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med. 2007 Apr 3;146(7):502-10.

43. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. Ann Fam Med. 2013 Mar-Apr;11(2):106-15.

44. Brett J, Austoker J. Women who are recalled for further investigation for breast screening: psychological consequences 3 years after recall and factors affecting re-attendance. J Public Health Med. 2001 Dec;23(4):292-300.

45. Ong G, Austoker J, Brett J. Breast screening: adverse psychological consequences one month after placing women on early recall because of a diagnostic uncertainty. A multicentre study. J Med Screen. 1997;4(3):158-68.

46. Henderson LM, Hubbard RA, Onega TL, Zhu W, Buist DS, Fishman P, et al. Assessing health care use and cost consequences of a new screening modality: the case of digital mammography. Med Care. 2012 Dec;50(12):1045-52.

47. Chubak J, Boudreau DM, Fishman PA, Elmore JG. Cost of breast-related care in the year following false positive screening mammograms. Med Care. 2010 Sep;48(9):815-20.

48. Lee DW, Stang PE, Goldberg GA, Haberman M. Resource use and cost of diagnostic workup of women with suspected breast cancer. Breast J. 2009 Jan-Feb;15(1):85-92.

49. Poplack SP, Carney PA, Weiss JE, Titus-Ernstoff L, Goodrich ME, Tosteson AN. Screening mammography: costs and use of screening-related services. Radiology. 2005 Jan;234(1):79-85.

50. Lidbrink E, Elfving J, Frisell J, Jonsson E. Neglected aspects of false positive findings of mammography in breast cancer screening: analysis of false positive cases from the Stockholm trial. BMJ. 1996 Feb 3;312(7026):273-6.

51. Haygood TM, Wang J, Atkinson EN, Lane D, Stephens TW, Patel P, et al. Timed efficiency of interpretation of digital and film-screen screening mammograms. AJR Am J Roentgenol. 2009 Jan;192(1):216-20.

52. Haygood TM, Wang J, Lane D, Galvan E, Atkinson EN, Stephens T, et al. Why does it take longer to read digital than film-screen screening mammograms? A partial explanation. J Digit Imaging. 2010 Apr;23(2):170-80.

53. Lampic C, Thurfjell E, Sjoden PO. The influence of a false-positive mammogram on a woman's subsequent behaviour for detecting breast cancer. Eur J Cancer. 2003 Aug;39(12):1730-7.

54. Burman ML, Taplin SH, Herta DF, Elmore JG. Effect of false-positive mammograms on interval breast cancer screening in a health maintenance organization. Ann Intern Med. 1999 Jul 6;131(1):1-6.

55. Lipkus IM, Halabi S, Strigo TS, Rimer BK. The impact of abnormal mammograms on psychosocial outcomes and subsequent screening. Psychooncology. 2000 Sep-Oct;9(5):402-10.

56. Pinckney RG, Geller BM, Burman M, Littenberg B. Effect of false-positive mammograms on return for subsequent screening mammography. Am J Med. 2003 Feb 1;114(2):120-5.

57. McCann J, Stockton D, Godward S. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. Breast Cancer Res. 2002;4(5):R11.

58. Lerman C, Trock B, Rimer BK, Boyce A, Jepson C, Engstrom PF. Psychological and behavioral implications of abnormal mammograms. Ann Intern Med. 1991 Apr 15;114(8):657-61.

59. Pisano ED, Earp J, Schell M, Vokaty K, Denham A. Screening behavior of women after a false-positive mammogram. Radiology. 1998 Jul;208(1):245-9.

60. O'Sullivan I, Sutton S, Dixon S, Perry N. False positive results do not have a negative effect on reattendance for subsequent breast screening. J Med Screen. 2001;8(3):145-8.

61. Andersen SB, Vejborg I, von Euler-Chelpin M. Participation behaviour following a false positive test in the Copenhagen mammography screening programme. Acta Oncol. 2008;47(4):550-5.

62. Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. Psychooncology. 2005 Nov;14(11):917-38.

63. Hofvind SS, Wang H, Thoresen S. The Norwegian Breast Cancer Screening Program: re-attendance related to the women's experiences, intentions and previous screening result. Cancer Causes Control. 2003 May;14(4):391-8.

64. Chiarelli AM, Moravan V, Halapy E, Majpruz V, Mai V, Tatla RK. False-positive result and reattendance in the Ontario Breast Screening Program. J Med Screen. 2003;10(3):129-33.

65. Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen. 2012;19 Suppl 1:42-56.

66. Yen AM, Duffy SW, Chen TH, Chen LS, Chiu SY, Fann JC, et al. Long-term incidence of breast cancer by trial arm in one county of the Swedish Two-County Trial of mammographic screening. Cancer. 2012 Dec 1;118(23):5728-32.

67. Johns LE, Moss SM, Age Trial Management Group. False-positive results in the randomized controlled trial of mammographic screening from age 40 ("Age" trial). Cancer Epidemiol Biomarkers Prev. 2010 Nov;19(11):2758-64.

68. Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmo Mammographic Screening Program. J Natl Cancer Inst Monogr. 1997;(22)(22):63-7.

69. Frisell J, Lidbrink E. The Stockholm Mammographic Screening Trial: Risks and benefits in age group 40-49 years. J Natl Cancer Inst Monogr. 1997;(22)(22):49-51.

70. Myers ER, Moorman P, Gierisch JM, Havrilesky LJ, Grimm LJ, Ghate S, et al. Benefits and Harms of Breast Cancer Screening: A Systematic Review. JAMA. 2015 Oct 20;314(15):1615-34.

71. Silverstein MJ, Lagios MD, Recht A, Allred DC, Harms SE, Holland R, et al. Image-detected breast cancer: state of the art diagnosis and treatment. J Am Coll Surg. 2005 Oct;201(4):586-97.

72. Silverstein MJ, Recht A, Lagios MD, Bleiweiss IJ, Blumencranz PW, Gizienski T, et al. Special report: Consensus conference III. Image-detected breast cancer: state-of-the-art diagnosis and treatment. J Am Coll Surg. 2009 Oct;209(4):504-20.

73. Lacquement MA, Mitchell D, Hollingsworth AB. Positive predictive value of the Breast Imaging Reporting and Data System. J Am Coll Surg. 1999 Jul;189(1):34-40.

74. Yu YH, Wei W, Liu JL. Diagnostic value of fine-needle aspiration biopsy for breast mass: a systematic review and meta-analysis. BMC Cancer. 2012 Jan 25;12:41,2407-12-41.

75. Verkooijen HM, Peeters PH, Buskens E, Koot VC, Borel Rinkes IH, Mali WP, et al. Diagnostic accuracy of large-core needle biopsy for nonpalpable breast disease: a meta-analysis. Br J Cancer. 2000 Mar;82(5):1017-21.

76. Bernik SF, Troob S, Ying BL, Simpson SA, Axelrod DM, Siegel B, et al. Papillary lesions of the breast diagnosed by core needle biopsy: 71 cases with surgical follow-up. Am J Surg. 2009 Apr;197(4):473-8.

77. Grady I, Gorsuch H, Wilburn-Bailey S. Ultrasound-guided, vacuum-assisted, percutaneous excision of breast lesions: an accurate technique in the diagnosis of atypical ductal hyperplasia. J Am Coll Surg. 2005 Jul;201(1):14-7.

78. Parker SH, Klaus AJ, McWey PJ, Schilling KJ, Cupples TE, Duchesne N, et al. Sonographically guided directional vacuum-assisted breast biopsy using a handheld device. AJR Am J Roentgenol. 2001 Aug;177(2):405-8.

79. Park HL, Kim LS. The current role of vacuum assisted breast biopsy system in breast disease. J Breast Cancer. 2011 Mar;14(1):1-7.

80. Liberman L, Feng TL, Dershaw DD, Morris EA, Abramson AF. US-guided core breast biopsy: use and cost-effectiveness. Radiology. 1998 Sep;208(3):717-23.

81. Parker SH, Jobe WE, Dennis MA, Stavros AT, Johnson KK, Yakes WF, et al. US-guided automated large-core breast biopsy. Radiology. 1993 May;187(2):507-11.

82. Bevers TB, Anderson BO, Bonaccio E, Buys S, Daly MB, Dempsey PJ, et al. NCCN clinical practice guidelines in oncology: breast cancer screening and diagnosis. J Natl Compr Canc Netw. 2009 Nov;7(10):1060-96.

83. Adepoju L, Qu W, Kazan V, Nazzal M, Williams M, Sferra J. The evaluation of national time trends, quality of care, and factors affecting the use of minimally invasive breast biopsy and open biopsy for diagnosis of breast lesions. Am J Surg. 2014 Sep;208(3):382-90.

84. Zimmermann CJ, Sheffield KM, Duncan CB, Han Y, Cooksley CD, Townsend CM,Jr, et al. Time trends and geographic variation in use of minimally invasive breast biopsy. J Am Coll Surg. 2013 Apr;216(4):814,24; discussion 824-7.

85. Clarke-Pearson EM, Jacobson AF, Boolbol SK, Leitman IM, Friedmann P, Lavarias V, et al. Quality assurance initiative at one institution for minimally invasive breast biopsy as the initial diagnostic technique. J Am Coll Surg. 2009 Jan;208(1):75-8.

86. Gutwein LG, Ang DN, Liu H, Marshall JK, Hochwald SN, Copeland EM, et al. Utilization of minimally invasive breast biopsy for the evaluation of suspicious breast lesions. Am J Surg. 2011 Aug;202(2):127-32.

87. Bruening W, Schoelles K, Treadwell J, Launders J, Fontanarosa J, Tipton K. **Comparative Effectiveness of Core-Needle and Open Surgical Biopsy for the Diagnosis of Breast Lesions [Internet]**. AHRQ Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2009 Dec. Report No.: 10-EHC007-EF.

88. Dahabreh IJ, Wieland LS, Adam GP, Halladay C, Lau J, Trikalinos TA. **Core Needle and Open Surgical Biopsy for Diagnosis of Breast Lesions: An Update to the 2009 Report** . AHRQ Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Sep. Report No.: 14-EHC040-EF.

89. US Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009 Nov 17;151(10):716,26, W-236.

90. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. JAMA. 1995 Jan 11;273(2):149-54.

91. Christiansen CL, Wang F, Barton MB, Kreuter W, Elmore JG, Gelfand AE, et al. Predicting the cumulative risk of false-positive mammograms. J Natl Cancer Inst. 2000 Oct 18;92(20):1657-66.

92. Lewin JM, D'Orsi CJ, Hendrick RE, Moss LJ, Isaacs PK, Karellas A, et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. AJR Am J Roentgenol. 2002 Sep;179(3):671-7.

93. Carney PA, Abraham LA, Miglioretti DL, Yabroff KR, Sickles EA, Buist DS, et al. Factors associated with imaging and procedural events used to detect breast cancer after screening mammography. AJR Am J Roentgenol. 2007 Feb;188(2):385-92.

94. Carney PA, Kasales CJ, Tosteson AN, Weiss JE, Goodrich ME, Poplack SP, et al. Likelihood of additional work-up among women undergoing routine screening mammography: the impact of age, breast density, and hormone therapy use. Prev Med. 2004 Jul;39(1):48-55.

95. Hubbard RA, Zhu W, Horblyuk R, Karliner L, Sprague BL, Henderson L, et al. Diagnostic imaging and biopsy pathways following abnormal screen-film and digital screening mammography. Breast Cancer Res Treat. 2013 Apr;138(3):879-87.

96. H. Gilbert Welch, Elliott S. Fisher. **Diagnostic Testing Following Screening Mammography in the Elderly**. Journal of national cancer institure. 1998;90(18):1389.

97. Advocate Health Care [Internet]. Downers Grove, Illinois, USA: Advocate Health Care [cited 02/16/2016]. Available from: http://www.advocatehealth.com/overview-of-advocate.

98. PenRad [Internet].: Technologies of Plymouth, MN; 2015 []. Available from: http://www.penrad.com/.

99. Illinois State Cancer Registry, public dataset, 1986-2012 [Internet].: Illinois Department of Public Health; 2014 [updated November, 2014;]. Available from: http://www.idph.state.il.us/cancer/statistics.htm.

100. The Breast cancer Surveillance Consortium. BCSC Glossary of Terms [Internet].; 09/16/2009 []. Available from: http://breastscreening.cancer.gov/.

101. Bennett RL, Blanks RG, Patnick J, Moss SM. Results from the UK NHS Breast Screening Programme 2000-05. J Med Screen. 2007;14(4):200-4.

102. Bucchi L, Falcini F, Baraldi GP, Bondi A, Bonsanto R, Bravetti P, et al. Integrating self-referral for mammography into organised screening: results from an Italian experience. J Med Screen. 2003;10(3):134-8.

103. Paquette D, Snider J, Bouchard F, Olivotto I, Bryant H, Decker K, et al. Performance of screening mammography in organized programs in Canada in 1996. The Database Management Subcommittee to the National Committee for the Canadian Breast Cancer Screening Initiative. CMAJ. 2000 Oct 31;163(9):1133-8.

104. Payne JI, Martin T, Caines JS, Duggan R. The burden of false-positive results in analog and digital screening mammography: experience of the Nova Scotia Breast Screening Program. Can Assoc Radiol J. 2014 Nov;65(4):315-20.

105. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005 Oct 27;353(17):1784-92.

106. Mushlin AI, Kouides RW, Shapiro DE. Estimating the accuracy of screening mammography: a meta-analysis. Am J Prev Med. 1998 Feb;14(2):143-53.

107. Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM, et al. Performance benchmarks for screening mammography. Radiology. 2006 Oct;241(1):55-66.

108. Woloshin S, Schwartz LM. The benefits and harms of mammography screening: understanding the trade-offs. JAMA. 2010 Jan 13;303(2):164-5.

109. Hofvind S, Ursin G, Tretli S, Sebuodegard S, Moller B. Breast cancer mortality in participants of the Norwegian Breast Cancer Screening Program. Cancer. 2013 Sep 1;119(17):3106-12.

110. Reducing Bias in a Propensity Score Matched-Pair Sample Using Greedy Matching Techniques [Internet].: SUGI []. Available from: http://www2.sas.com/proceedings/sugi26/p214-26.pdf.

111. Yankaskas BC, Schell MJ, Miglioretti DL. Recall and detection rates in screening mammography. Cancer. 2004 Dec 1;101(11):2710,1; author reply 2711-2.

112. DeFrank JT, Rimer BK, Bowling JM, Earp JA, Breslau ES, Brewer NT. Influence of false-positive mammography results on subsequent screening: do physician recommendations buffer negative effects? J Med Screen. 2012 Mar;19(1):35-41.

113. Alamo-Junquera D, Murta-Nascimento C, Macia F, Bare M, Galceran J, Ascunce N, et al. Effect of false-positive results on reattendance at breast cancer screening programmes in Spain. Eur J Public Health. 2012 Jun;22(3):404-8.

114. Roman R, Sala M, De La Vega M, Natal C, Galceran J, Gonzalez-Roman I, et al. Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. Breast Cancer Res Treat. 2011 Nov;130(2):543-52.

115. Brancato B, Crocetti E, Bianchi S, Catarzi S, Risso GG, Bulgaresi P, et al. Accuracy of needle biopsy of breast lesions visible on ultrasound: audit of fine needle versus core needle biopsy in 3233 consecutive samplings with ascertained outcomes. Breast. 2012 Aug;21(4):449-54.

116. Fahrbach K, Sledge I, Cella C, Linz H, Ross SD. A comparison of the accuracy of two minimally invasive breast biopsy methods: a systematic literature review and meta-analysis. Arch Gynecol Obstet. 2006 May;274(2):63-73.

117. Antley CM, Mooney EE, Layfield LJ. A comparison of accuracy rates between open biopsy, cutting-needle biopsy, and fine-needle aspiration biopsy of the breast: A 3-year experience.

Breast J 1998;4(1):3-8. . 1998;4(1):3-8.

118. Kurita T, Tsuchiya S, Watarai Y, Yamamoto Y, Harada O, Yanagihara K, et al. Roles of fine-needle aspiration and core needle biopsy in the diagnosis of breast cancer. Breast Cancer. 2012 Jan;19(1):23-9.

119. Clarke D, Sudhakaran N, Gateley CA. Replace fine needle aspiration cytology with automated core biopsy in the triple assessment of breast cancer. Ann R Coll Surg Engl. 2001 Mar;83(2):110-2.

120. Shannon J, Douglas-Jones AG, Dallimore NS. Conversion to core biopsy in preoperative diagnosis of breast lesions: is it justified by results? J Clin Pathol. 2001 Oct;54(10):762-5.

121. Pilgrim S, Ravichandran D. Fine needle aspiration cytology as an adjunct to core biopsy in the assessment of symptomatic breast carcinoma. Breast. 2005 Oct;14(5):411-4.

122. Westenend PJ, Sever AR, Beekman-De Volder HJ, Liem SJ. A comparison of aspiration cytology and core needle biopsy in the evaluation of breast lesions. Cancer. 2001 Apr 25;93(2):146-50.

123. Abdsaleh S, Azavedo E, Lindgren PG. Semiautomatic core biopsy. A modified biopsy technique in breast diseases. Acta Radiol. 2003 Jan;44(1):47-51.

124. Koskela AK, Sudah M, Berg MH, Karja VJ, Mustonen PK, Kataja V, et al. Add-on device for stereotactic core-needle breast biopsy: how many biopsy specimens are needed for a reliable diagnosis? Radiology. 2005 Sep;236(3):801-9.

125. Lomoschitz FM, Helbich TH, Rudas M, Pfarl G, Linnau KF, Stadler A, et al. Stereotactic 11-gauge vacuum-assisted breast biopsy: influence of number of specimens on diagnostic accuracy. Radiology. 2004 Sep;232(3):897-903.

126. Pfarl G, Helbich TH, Riedl CC, Wagner T, Gnant M, Rudas M, et al. Stereotactic 11-gauge vacuum-assisted breast biopsy: a validation study. AJR Am J Roentgenol. 2002 Dec;179(6):1503-7.

127. Ciatto S, Houssami N, Ambrogetti D, Bianchi S, Bonardi R, Brancato B, et al. Accuracy and underestimation of malignancy of breast core needle biopsy: the Florence experience of over 4000 consecutive biopsies. Breast Cancer Res Treat. 2007 Mar;101(3):291-7.

VITA

NAME:	Firas M. Dabbous
EDUCATION:	B.S., Mathematics, Beirut Arab university, Beirut, Lebanon 2004
	M.S., Statistics, University of Southern Maine, Portland, Maine, 2009
	PhD, Epidemiology, University of Illinois at Chicago, Chicago, IL, 2016
RESEARCH EXPERIENCE:	Graduate Research Assistant, Performance Characteristics of Breast Imaging Modalities in Women Using Large Dataset from a Major Health Organization in Chicago Illinois Study, University of Illinois at Chicago, Aug. 2011-Present
	Consultant Health Economics and Outcomes Research, Epidemiology and Biostatistics, February 2013-Present
	Graduate Research Assistant, Adiposity And Outcomes of Clinically Localized Prostate Cancer Study, 2010-August 2011
	Graduate Research Assistant, Methodology Research Core, 2010-December 2011
	Graduate Research Assistant, Maine Center for Business and Economic Research, August 2007-May 2009
	Research Fellow, Global Health Economics and Outcomes Research Department, UIC/Takeda Pharmaceuticals International, Inc. August 2009-August 2010
TEACHING EXPERIENCE:	Teaching Assistant, Department of Epidemiology and Biostatistics, University of Illinois at Chicago, BIOSTAT 505: Logistic Regression and Survival Analysis, August. 2010-December 2010
	Co-instructor, Department of Epidemiology and Biostatistics, University of Illinois at Chicago, Pub 120: Introduction to Public Health, January 2013-May 2013
	Teaching Assistant, Department of Business, University of Southern Maine, BUS 370: Management Science, August 2008-May 2009
	Teaching Assistant, Department of Business, University of Sothern Maine, MBA 670 Management Science, January 2009-May 2009

ABSTRACTS, POSTERS

Dabbous F, J Khan-Gates, T Macarol, TA Dolecek, S Friedewald, AM Murphy, Wm. Thomas Summerfelt, GH Rauscher. Comparison of false negative probabilities for digital and film mammography within a large healthcare organization [abstract]. American Society of Clinical Oncology. May-June 21014.

Dabbous F, J Khan-Gates, T Macarol, TA Dolecek, S Friedewald, AM Murphy, Wm. Thomas Summerfelt, GH Rauscher. Is digital mammography associated with a higher risk of false positive results? Findings from a community setting [abstract]. American Society of Clinical Oncology. May-June 21014.

Dabbous F, Dorey J, Thokagevistk K, Toumi M. Impact of health insurance on health care resource utilization among diabetic patients in the USA [abstract]. International Society for Pharmacoeconomics and Outcomes Research. May-June 21014.

Aballea S, Dabbous F, Thokagevistk K, Methodological Choices for Propensity Score Matching and Their Consequences [work shop]. International Society for Pharmacoeconomics and Outcomes Research.

Dabbous F, Rauscher G, Macarol T, Khan J, Dolecek TA, Friedewald S, Francois T, Summerfelt Wm. Examining racial/ethnic disparities in mammography screening performance in a single, large healthcare organization [abstract]. American Association for Cancer Research. December 2013.

Dookeran K, Dabbous F, Booth B. Race-based mortality as determined by intrinsic breast tumor subtype and p53 status [abstract]. American Society of Clinical Oncology. May-June 2013.

Dabbous F, Al-alem U, Springfield S, M. Mahmoud A, Batai K, Shah E, Ahaghotu C, Kittles R. Dietary zinc and prostate cancer: a case-control study in African Americans men [abstract]. American Association for Cancer Research; April 2013

Freeman VL, Flanigan RC, Hollowell CM, Branch J, Deane L, Quek M, Vidal PP, Friedman N, Dabbous F, Cintron D, Millbrand L, Garabedian A, Cleofe J, Gomez SL. Body weight status predicts pre-treatment risk of PSA failure in men with clinically localized prostate cancer [abstract]. American Association for Cancer Research. Chicago, IL March-April 2012

Mody R, Beaulieu N, Wu E, Dabbous F, McHale J, Chey WD. Impact of Nighttime Symptoms (NTS) on Treatment Patterns and Healthcare Costs among Gastroesophageal Reflux Disease (GERD) Patients [abstract]. Digestive Disease week. New Orleans, LA, May 2010

Sun S, Spiegel B, Mody R, Dabbous F, Dabbous O; Chey W. Economic and Humanistic Burden of Chronic Constipation [abstract]. American College of Gastroenterology. San Antonio, TX, October 2010

Chey W, Schoenfeld P, Sun S, Dabbous F, Baum C, Mody R. Management of

Chronic Constipation: Results from the EXCCEED Registry. American College of Gastroenterology [abstract]. San Antonio, TX, October 2010

Sun S, Lembo A, Mody R, Dabbous F, Baum C, Chey WD. Characteristics of US Patients with Chronic Constipation Encountered in Primary Care: Data from the Examination of Chronic Constipation Treatment and Outcomes in Everyday Medical Practice (EXCCEED) Registry [abstract]. American College of Gastroenterology. San Antonio, TX, October 2010