8-1-2017

Digital Breast Tomosynthesis: Outcomes and Tumor Characteristics in Women Recalled From Screening

Michelle Y. Giwerc

Yale Physician Associate Program, michelle.giwerc@gmail.com

DIGITAL BREAST TOMOSYNTHESIS: OUTCOMES AND TUMOR CHARACTERISTICS IN WOMEN RECALLED FROM SCREENING

A Thesis Presented to
The Faculty of the School of Medicine
Yale University

In Candidacy for the degree of
Master of Medical Science

August 2017

Michelle Y. Giwerc, PA-SII
Class of 2017
Yale Physician Associate Program

Liane E. Philpotts, MD, FSBI, FACR
Chief of Breast Imaging
Professor of Radiology & Biomedical Imaging
Department of Diagnostic Radiology
Yale University School of Medicine

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	Two-Dimensional	**IDC**	Invasive Ductal Carcinoma
3D	Three-Dimensional	**ILC**	Invasive Lobular Carcinoma
ACR	American College of Radiology	**IQR**	Interquartile Range
BI-RADS	Breast Imaging-Reporting and Data System	**IRB**	Institutional Review Board
CAD	Computer-Aided Detection	**PPV_1**	Positive Predictive Value from recall
CC	Craniocaudal	**MLO**	Mediolateral Oblique
CDR	Cancer Detection Rate	**mm**	Millimeters
CI	Confidence Interval	**NPV_1**	Negative Predictive Value from recall
cm	Centimeters	**OR**	Odds Ratio
DBT	Digital Breast Tomosynthesis	**OTST**	*O*slo *T*omosynthesis *S*creening *T*rial
DCIS	Ductal Carcinoma *In Situ*	**RR**	Recall Rate
FDA	Food & Drug Administration	**s-2D**	synthetic-2D (from DBT data)
FFDM	Full-Field Digital Mammography	**SD**	Standard Deviation
FP	False Positive	**STORM**	*S*creening with *T*omosynthesis *OR* standard *M*ammography
GEE	Generalized Estimating Equation	**TP**	True Positive
GLMM	Generalized Linear Mixed Model	**US(A)**	United States (of America)
HIPAA	Health Insurance Portability & Accountability Act	**USPSTF**	US Preventive Services Task Force

ABSTRACT

Breast cancer is the most frequently diagnosed and second leading cause of cancer deaths in women, accounting for 25% of cancer diagnoses and 15.4% of cancer deaths in developed countries. Thus, early detection of breast cancer through screening has become increasingly important in mortality reduction efforts. Yet, mammography has faced considerable controversy in balancing the benefits and harms associated with screening. Digital breast tomosynthesis has emerged as an important imaging technique which, compared to standard mammography alone, reduces recall rates and false positives, and improves cancer detection. Additional cancers detected with tomosynthesis have been poorly characterized in the literature to date. To assess the effectiveness of screening with adjunct tomosynthesis, we propose to utilize our large database to characterize cancers detected in true positive recalls. Our findings will help clinicians make well-informed decisions for further management of women with mammographically suspicious or inconclusive findings, and contribute to future screening guidelines.

CHAPTER I: INTRODUCTION

1.1 Background

Breast cancer is the most prevalent cancer in women worldwide, with an estimated 5-year prevalence of 6.2 million women (36.3%).[1] It is also the most frequently diagnosed cancer in women worldwide: according to the most recent global incidence data from 2012, breast cancer accounts for 25% of all cancer diagnoses, and is the most frequent cause of cancer mortality in less developed regions.[2] In developed nations such as the United States, breast cancer is second only to lung cancer as the most common cause of cancer mortality in women, accounting for 198,000 cancer deaths annually.[2] Thus, accurately identifying women at increased risk of developing breast cancer for targeted screening, in addition to effective population-based screening programs, remains a high priority in the United States and abroad.

For over a decade, conventional two-dimensional (2D) full-field digital mammography (FFDM) has been widely accepted as the most effective screening technique for the detection of breast cancer and mortality reduction in asymptomatic women.[3] Experts agree that standard screening mammography reduces breast cancer mortality by 12% to 33%.[4,5] 2D mammography is also the most common technique for measuring and classifying breast density using the Breast Imaging Reporting and Data System (BI-RADS) established by the American College of Radiology (ACR).[6] The identification of women with dense breasts has become increasingly important in breast cancer screening efforts: a recent systematic review from the United States Preventive Services Task Force (USPSTF) estimates that 27.6 million American women 40 to 74 years of age have dense breasts (43%).[7] Independent of other risk factors, women with

extremely dense breast tissue have a 3- to 5-fold increased risk of developing breast cancer, as compared to women with fatty breast tissue.[8,9]

Standard 2D mammography has several limitations, particularly in women with dense breasts. Superimposition of breast tissue can obscure areas of malignancy, and may cause otherwise normal tissue to appear mammographically suspicious.[10] This masking effect is most pronounced in women with dense breasts. Higher breast density is associated with decreased mammographic sensitivity and specificity compared to women with non-dense breasts.[11] Overall, FFDM has low sensitivity, high false positive recall rates, and limited utility for screening in women with dense breasts.[12,13] Up to 30% of breast cancers are not detected using conventional FFDM, highlighting the need to enhance the performance of screening mammography.[14]

Three-dimensional (3D) digital breast tomosynthesis (DBT), also known as 3D digital mammography, has emerged as a promising and improved technique for breast cancer screening and detection.[15,16] In 2011, the US Food and Drug Administration (FDA) approved the use of DBT in combination with FFDM for the screening and diagnosis of breast cancer.[17] In 2013, the FDA approved the use of DBT with reconstructed synthetic 2D (s-2D) images, a technique developed to address the burden of doubled radiation dose exposure in women undergoing dual acquisition DBT and FFDM, the means by which DBT has been traditionally acquired.[7] Several studies have observed non-inferior or superior performance metrics from DBT with s-2D as compared to the conventional dual acquisition DBT and FFDM.[18-20]

DBT, in conjunction with FFDM or as a standalone with reconstructed s-2D images, is associated with increased cancer detection rates and reduced false positive recall rates

from screening – regardless of age or breast density.[19-21] DBT has also been shown to increase the positive predictive value from recall.[22-24] Importantly, while DBT increases cancer detection rates, it does not appear to increase detection of *in situ* carcinomas, alleviating concerns regarding further over-diagnosis and overtreatment from screening.[25-27] Despite the growing evidence for the utility of DBT for screening, the 2016 USPSTF guidelines suggest that more evidence is needed before recommending the use of DBT as a primary screening method in practice.[4]

1.2 Statement of the Problem

While DBT has been shown to increase overall and invasive cancer detection rates, few studies have characterized the sizes, histologic types and grades, lymph node status, and receptor phenotypes of cancers detected with DBT versus FFDM alone.[10,15,22,24]

A valuable way to assess this important gap in knowledge is to determine the number of true positive cases from recall detected with DBT versus FFDM alone, and to characterize the tumors detected in this population. Evaluation of this cohort is an important way to assess the effectiveness of screening with DBT versus FFDM, since these cases represent women with indeterminate mammographic findings that require further imaging and/or biopsy. When such recalls do not lead to a cancer diagnosis, they are considered false positives, and contribute to the harms of mammography by incurring patient discomfort, risk, cost, time, and stress.[28] Although screening with DBT is known to increase the detection of cancer while reducing recall rate, current literature regarding the characterization of cancers detected with DBT versus FFDM alone is lacking.[29,30]

1.3 Goals and Objectives

The proposed study aims to compare the effectiveness of screening with DBT versus FFDM alone for the detection of breast cancer by characterizing cancers in true positive cases: biopsy-proven cancers in women recalled from screening, defined as mammographic ACR BI-RADS category 0 (inconclusive; requires recall for additional imaging).[31] Henceforth, women recalled from screening will be referred to as *BI-RADS 0 cases* for brevity.

The primary outcomes to be measured in the population of interest, BI-RADS 0 cases, are the positive predictive value from recall (PPV_1) and the cancer detection rate (CDR) per 1000 women screened, as detected with DBT versus FFDM alone. While the CDR will be measured, we are not powered to detect a statistically significant difference in cancer detection across modalities given our sample size. Secondary outcomes, which are of chief interest, include: tumor pathologic size and stage at diagnosis; histological type and grade of cancers; axillary lymph node status and receptor phenotype of invasive cancers, to be evaluated in true positive BI-RADS 0 cases. To determine our outcomes of interest, the total number of women recalled from screening (BI-RADS 0 cases) and the overall recall rate (RR, %) per modality have already been measured and reported.[32]

This study's outcomes will help to better characterize cancers detected from screening with DBT versus FFDM alone; inform future breast cancer screening efforts and guidelines; and aid clinicians in making well-informed decisions for the management of women with suspicious or inconclusive mammographic findings.

1.4 Hypotheses

Primary Hypothesis: Cancers detected in women recalled from screening with DBT versus FFDM alone will yield significantly ***higher*** PPV_1, as measured by the number of biopsy-proven true positive BI-RADS 0 cases per total BI-RADS 0 cases.[29]

Our study is not powered to detect significant changes in cancer detection, as evidenced by studies of similar size (see Chapter II for further discussion).[3] Therefore, we do not expect to detect a statistically significant increase in CDR from screening with DBT versus FFDM alone, though a non-significant increase is likely to be observed.[3,15,24]

Secondary Hypotheses: True positive BI-RADS 0 cases detected from screening with DBT versus FFDM alone will yield (statistically significant):

a) *Smaller* mean pathologic sizes of tumors.[33]
b) *Lower* grade of *in situ* cancers[31] (nuclear grade) and invasive cancers (Nottingham histologic grade) at time of diagnosis.[33]
c) *No significant difference* in the proportions of histological types of *in situ* or invasive cancers (ductal, lobular, other).[33-35]
d) *Smaller* proportions of invasive cancers with spread to axillary lymph nodes.[33,36]
e) *Lower* pathologic stage at diagnosis.
f) *No significant difference* in the proportions of receptor phenotypes for invasive cancers (ER+/PR+ or luminal, HER2+, triple negative).[10]

1.5 Definitions

- BI-RADS Mammographic Assessment Categories:[31]
  - *Category 0*: Incomplete/indeterminate – *recalled* for additional imaging
  - *Category 1*: Negative

- *Category 2*: Benign
- *Category 3*: Probably Benign
- *Category 4*: Suspicious for Malignancy
- *Category 5*: Highly Suggestive of Malignancy

- BI-RADS Description of Overall Breast Composition:[31]
  - *Category A*: Breasts are predominantly fatty
  - *Category B*: There are scattered areas of fibroglandular density in the breasts
  - *Category C*: Breasts are heterogeneously dense; this may obscure small masses
  - *Category D:* Breasts are extremely dense; this lowers mammographic sensitivity
- Nuclear Grade (for *in situ* cancers):[37]
  - *Grade 1* (Low): Well-differentiated cells; tend to grow slowly
  - *Grade 2* (Intermediate): Moderately-differentiated cells
  - *Grade 3* (High): Poorly-differentiated cells; tend to proliferate quickly
- Nottingham Histologic Grade (for invasive cancers):[38]
  - *Grade 1*: Well-differentiated malignant cells; 95% 5-year survival
  - *Grade 2*: Moderately-differentiated malignant cells
  - *Grade 3*: Poorly-differentiated malignant cells; 50% 5-year survival

Operational Definitions:

- *BI-RADS 0 cases*: Refers to women recalled after screening mammography (DBT or FFDM) for additional imaging.
- *Digital Breast Tomosynthesis (DBT)*: In this study, DBT refers to tomosynthesis adjunct to full-field digital mammography; unless otherwise specified, *DBT refers to DBT+FFDM.*

References

1. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *International journal of cancer.* 2013;132(5):1133-1145.
2. Ferlay JS, I.; Ervik, M.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer. 2013; http://globocan.iarc.fr/. Accessed December 10, 2016.
3. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology.* 2013;269(3):694-700.
4. Siu AL. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of internal medicine.* 2016;164(4):279-296.
5. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. *JAMA.* 2015;314(15):1599-1614.
6. Brennan M, Houssami N. Discussing the benefits and harms of screening mammography. *Maturitas.* 2016;92:150-153.
7. Melnikow J, Fenton JJ, Whitlock EP, et al. Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the U.S. Preventive Services Task Force. *Annals of internal medicine.* 2016;164(4):268-278.
8. Cecchini RS, Costantino JP, Cauley JA, et al. Baseline mammographic breast density and the risk of invasive breast cancer in postmenopausal women participating in the NSABP study of tamoxifen and raloxifene (STAR). *Cancer Prev Res (Phila).* 2012;5(11):1321-1329.
9. Houssami N. Digital breast tomosynthesis (3D-mammography) for screening women with dense breasts. *Expert review of medical devices.* 2016;13(6):515-517.
10. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *The Lancet. Oncology.* 2013;14(7):583-589.
11. Yang TL, Liang HL, Chou CP, Huang JS, Pan HB. The adjunctive digital breast tomosynthesis in diagnosis of breast cancer. *BioMed research international.* 2013;2013:597253.
12. Coop P, Cowling C, Lawson C. Tomosynthesis as a screening tool for breast cancer: A systematic review. *Radiography.* 2016;22(3):e190-e195.
13. Freer PE. Mammographic breast density: impact on breast cancer risk and implications for screening. *Radiographics : a review publication of the Radiological Society of North America, Inc.* 2015;35(2):302-315.
14. Lewin JM, D'Orsi CJ, Hendrick RE. Digital mammography. *Radiologic clinics of North America.* 2004;42(5):871-884, vi.
15. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology.* 2015;274(1):85-92.

16. McDonald ES, Oustimov A, Weinstein SP, Synnestvedt MB, Schnall M, Conant EF. Effectiveness of Digital Breast Tomosynthesis Compared With Digital Mammography: Outcomes Analysis From 3 Years of Breast Cancer Screening. *JAMA Oncol.* 2016;2(6):737-743.
17. US Food and Drug Administration, Summary of Safety and Effectiveness Data: Hologic DBT (P080003). *Available at* https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P080003. 2011:Accessed August 20, 2016.
18. Aujero MP, Gavenonis SC, Benjamin R, Zhang Z, Holt JS. Clinical Performance of Synthesized Two-dimensional Mammography Combined with Tomosynthesis in a Large Screening Population. *Radiology.* 2017;283(1):70-76.
19. Houssami N, Bernardi D, Pellegrini M, et al. Breast cancer detection using single-reading of breast tomosynthesis (3D-mammography) compared to double-reading of 2D-mammography: Evidence from a population-based trial. *Cancer epidemiology.* 2017;47:94-99.
20. Zuckerman SP, Conant EF, Keller BM, et al. Implementation of Synthesized Two-dimensional Mammography in a Population-based Digital Breast Tomosynthesis Screening Program. *Radiology.* 2016;281(3):730-736.
21. Houssami N, Turner RM. Rapid review: Estimates of incremental breast cancer detection from tomosynthesis (3D-mammography) screening in women with dense breasts. *Breast (Edinburgh, Scotland).* 2016;30:141-145.
22. Conant EF, Beaber EF, Sprague BL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: a cohort study within the PROSPR consortium. *Breast Cancer Res Treat.* 2016;156(1):109-116.
23. McDonald ES, McCarthy AM, Akhtar AL, Synnestvedt MB, Schnall M, Conant EF. Baseline Screening Mammography: Performance of Full-Field Digital Mammography Versus Digital Breast Tomosynthesis. *AJR. American journal of roentgenology.* 2015;205(5):1143-1148.
24. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R, Jr. Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR. American journal of roentgenology.* 2013;200(6):1401-1408.
25. Rahbar H, McDonald ES, Lee JM, Partridge SC, Lee CI. How Can Advanced Imaging Be Used to Mitigate Potential Breast Cancer Overdiagnosis? *Acad Radiol.* 2016;23(6):768-773.
26. Choi JS, Han BK, Ko EY, et al. Comparison between two-dimensional synthetic mammography reconstructed from digital breast tomosynthesis and full-field digital mammography for the detection of T1 breast cancer. *Eur Radiol.* 2016;26(8):2538-2546.
27. Gilbert FJ, Tucker L, Gillan MG, et al. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme--a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. *Health technology assessment (Winchester, England).* 2015;19(4):i-xxv, 1-136.

28. Alcusky M, Philpotts L, Bonafede M, Clarke J, Skoufalos A. The patient burden of screening mammography recall. *Journal of women's health (2002).* 2014;23 Suppl 1:S11-19.
29. McCarthy AM, Kontos D, Synnestvedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *Journal of the National Cancer Institute.* 2014;106(11).
30. Pozzi A, Corte AD, Lakis MA, Jeong H. Digital Breast Tomosynthesis in Addition to Conventional 2DMammography Reduces Recall Rates and is Cost Effective. *Asian Pacific journal of cancer prevention : APJCP.* 2016;17(7):3521-3526.
31. American Cancer Society. *ACR BI-RADS Atlas - Mammography Reporting System.* Atlanta: American Cancer Society; 2013. Available at: https://www.acr.org/~/media/ACR/Documents/PDF/QualitySafety/Resources/BIRADS/01-Mammography/02--BIRADS-Mammography-Reporting.pdf?la=en. Accessed 12/20/2016.
32. Etesami M, Giwerc MY, Harigopal M, Zaheer Kidwai W, Philpotts LE. Cancer Detection with Digital Breast Tomosynthesis (DBT) Compared with Conventional Digital Mammography in Routine Breast Cancer Screening. *Presented at: Radiological Society of North America 2017 Scientific Assembly and Annual Meeting, November 26 - December 1, 2017, Chicago, IL.* rsna2017.rsna.org/program/. Accessed July 20, 2017.
33. Wang WS, Hardesty L, Borgstede J, Takahashi J, Sams S. Breast Cancers Found with Digital Breast Tomosynthesis: A Comparison of Pathology and Histologic Grade. *Breast J.* 2016;22(6):651-656.
34. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *AJR. American journal of roentgenology.* 2014;203(3):687-693.
35. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA.* 2014;311(24):2499-2507.
36. Lang K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmo Breast Tomosynthesis Screening Trial, a population-based study. *Eur Radiol.* 2016;26(1):184-190.
37. American Cancer Society. *Cancer Facts & Figures 2015; Special Section: Breast Carcinoma in Situ.* Atlanta: American Cancer Society; 2015. Available at: http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-046202.pdf. Accessed 12/20/2016.
38. Ellsworth RE, Hooke JA, Love B, Ellsworth DL, Shriver CD. Molecular changes in primary breast tumors and the Nottingham Histologic Score. *Pathology oncology research : POR.* 2009;15(4):541-547.

CHAPTER II: REVIEW OF THE LITERATURE

2.1 Introduction

2.1.1 Literature Search

Literature was reviewed by means of PubMed between August 2016 and July 2017. All articles were published after 2007. A final search was completed on July 16, 2017. Key terms for digital breast tomosynthesis (DBT) included *tomosynth*, 3D mammography, 3-D mammography, three dimensional mammography,* and *three-dimensional mammography*. The MeSH term *Breast Neoplasm* indexed all search criteria for breast cancer. Key terms for BI-RADS 0 cases (women recalled from screening for further imaging) included *BIRADS 0, BI-RADS 0,* and *recall.* Categories were combined using *and/or* functionalities. Literature returned from the initial search (n=62) included retrospective studies, prospective studies, and systematic reviews. International articles were included in the review, but the search was narrowed to the English language only. The review was further limited to *humans* (species), *female* (sex), and *last 10 years* (publication dates), returning 59 articles. Three additional full-text articles were assessed after citation searches. Studies evaluating any of the following were excluded: women with previous diagnoses/treatment of breast cancer; women with breast cancer signs/symptoms; diagnostic/non-screening settings; one-view DBT; assessment of DBT for technological development; and review articles. Studies with fewer than 300 participants were also excluded. After full-text screening (n=62), 42 articles were excluded: eighteen reviews; ten diagnostic/non-screening studies; six technical studies; six studies assessing different DBT protocols; one study assessing women with prior history of breast cancer; and one study with fewer than 300 participants, resulting in review of twenty original reports (Figure 1).

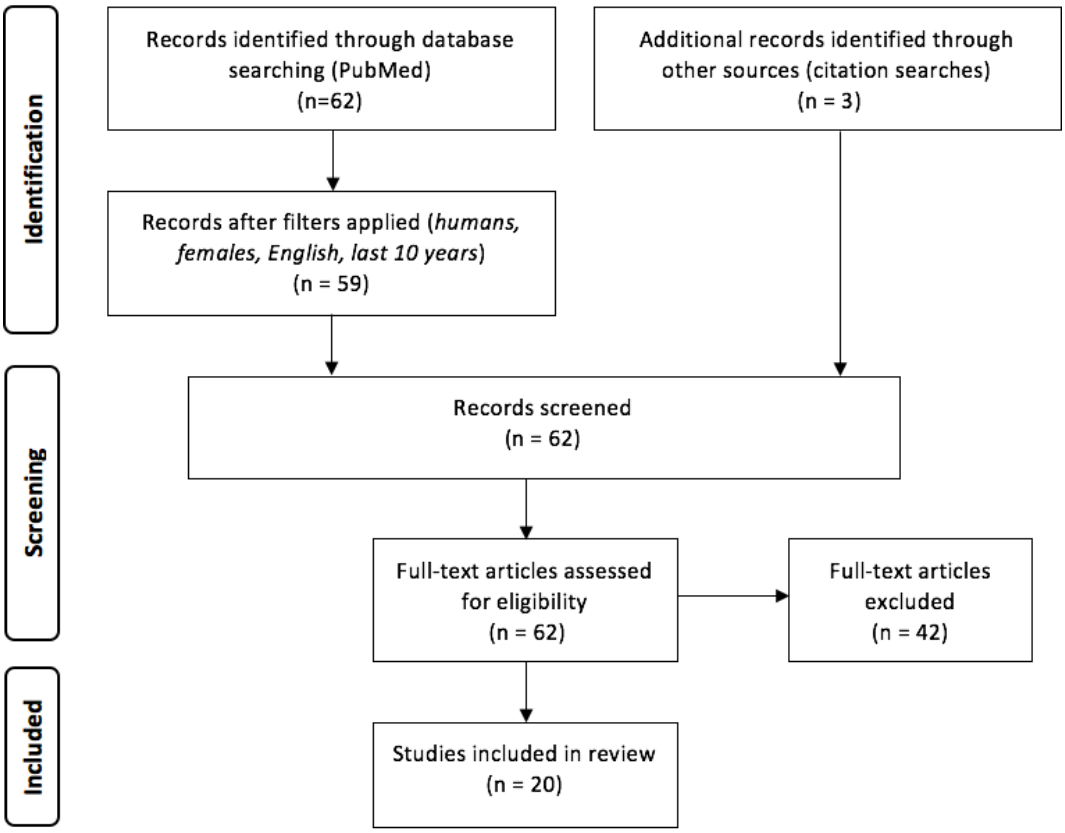

Figure 1. Literature Review Flow Diagram

2.1.2 Digital Breast Tomosynthesis

Mammography screening with two-dimensional (2D) full-field digital mammography (FFDM) has been shown to decrease mortality from breast cancer by 30% or more due to earlier detection.[1] Despite this achievement, tissue overlap observed in 2D images acquired with FFDM creates significant obstacles for interpretation.[2] Overall, FFDM has low sensitivity and specificity, leading to false positive recalls and unnecessary additional workup – particularly in women with dense breast parenchyma due to the masking effect of overlapping breast tissue.[3]

Digital breast tomosynthesis (DBT) is a form of mammography that produces quasi three-dimensional (3D) images of the breast via acquisition of a series of low dose images obtained over a limited arc, which are then reconstructed and displayed in small

(1 mm) image slices.[4] DBT images are displayed in combination with standard FFDM 2D images – or, more recently, synthetically reconstructed 2D images from the DBT data.[4] DBT provides greater imaging detail and addresses the challenges presented by overlapping breast tissue observed in FFDM, such as the concealment of true malignancy and the mimicry of cancer.[5]

As discussed in the following sections, several studies have reported significant increases in overall and invasive cancer detection rate (CDR), with no change in the *in situ* CDR, with the addition of DBT to screening. This has important implications, as the treatment of *in situ* carcinomas – specifically, ductal carcinoma *in situ* (DCIS) – is a highly controversial and heavily debated topic, with major concerns regarding over-diagnosis and overtreatment.[6] Studies assessing population screening with DBT have also reported significant reductions in recall rate (RR), in addition to an increase in positive predictive value from recall (PPV_1) with DBT versus FFDM alone, suggesting that a positive DBT screening examination is more likely to truly detect cancer.

This review will appraise prior research assessing population-based screening with DBT versus FFDM alone, with specific focus on studies that evaluate screening outcomes, including PPV_1, and detail the characteristics of tumors detected with DBT.

2.2 Review of European Studies

It is important to note some vital differences between screening studies performed in the United States and those performed in Europe. In Europe, screening mammography is performed under a double-reading protocol in which two radiologists independently review examinations; in the United States, screening interpretation is performed by a single radiologist.[7] Other distinctions involve breast cancer rates (higher in the US than in

Europe) and population demographics.[8] We will briefly review the major findings from two prospective European population screening studies comprising five reports: the STORM trial from Italy (two reports) and the OTST study from Norway (three reports).

2.2.1 STORM: Italy

The principal publication for the *S*creening with *T*omosynthesis *OR* standard *M*ammography (STORM) trial comes out of two centers in Italy. STORM prospectively compared the CDR, RR, and false positive recall rate in asymptomatic women 48 years or older (n=7,292; median age 59 years, IQR 48-71) who presented for population-based screening from 2011 to 2012. Each participant underwent two-view craniocaudal (CC) and mediolateral oblique (MLO) screening mammography with FFDM (phase 1) followed by DBT (dual-acquisition DBT+FFDM, phase 2). Images were reviewed per a double-reading protocol in which two radiologists independently assessed screening mammography exams, as is common practice in Europe. In STORM, images were reviewed in a sequential order by the two radiologists in parallel: in phase 1, 2D images from FFDM were evaluated, and each radiologist independently determined if the participant should be recalled based on those images alone. Regardless of that decision, the same two radiologists then independently reviewed the 3D images from the DBT scan later that day (phase 2), and again decided if the participant should be recalled. If either of the two radiologists decided to recall the patient at either of the two phases, the patient was recalled and the examination was considered to be a positive screen.[9]

Outcomes involved comparison of the number of cancers and false positive recalls from FFDM only (phase 1) versus the number of cancers and false positive recalls from DBT (phase 2). The authors additionally applied a conditional recall rule, whereby recall

was restricted to positive DBT screens only (phase 2), excluding FFDM-only positive screens (phase 1), in order to determine the incremental cancer detection rate attributable to DBT and to better estimate the false positive recall rate from DBT.[9]

Overall, 59 cancers were detected: 39 (66%) were seen with FFDM (phase 1) for a FFDM CDR of 5.3 per 1000 screened, while all 59 were seen on DBT+FFDM (phase 2) for a DBT CDR of 8.1 per 1000 screened ($p<0.0001$). Twenty cancers that were seen with dual-acquisition DBT were not visible on FFDM, indicating an incremental CDR of +2.8 per 1000 screened, or a 53% relative increase in CDR attributable to dual-acquisition DBT. Of the 59 cancers detected, 52 (88%) were invasive: 35 were detected on both FFDM (phase 1) and DBT (phase 2), while 17 were detected only with DBT (phase 2), yielding a FFDM invasive CDR of 4.8 per 1000 screened and a DBT invasive CDR of 7.1 per 1000 screened ($p<0.0001$), resulting in a 49% increase in the invasive CDR attributable to DBT. With the application of the conditional recall rule, the false positive recall rate was 5.5% for FFDM versus 3.5% for DBT ($p<0.0001$), resulting in a 17% reduction in the false positive recall rate attributable to DBT.[9] A secondary retrospective analysis of STORM revealed that there were no significant differences in outcome measures across the two screening centers, Trento and Verona.[10]

Six interval cancers – defined as cancers diagnosed after a negative mammography screening examination but before the next routine screen – were observed in first-year follow-up from STORM. From the 59 cancers noted in the principal STORM report and the additional six interval cancers from first-year follow-up, the authors calculated the incremental CDR attributable to the addition of DBT using retrospective models of both single-reading and double-reading methods for the 65 cancer cases. For the single-

reading strategy, 35 cancers were detected at both FFDM and DBT (phases 1 and 2); twenty detected only with DBT (phase 2); and none detected on FFDM alone ($p<0.001$). Ten cancers were not detected. For the double-reading strategy, 39 cancers were detected with both FFDM and DBT (phases 1 and 2); twenty detected only with DBT (phase 2); and none detected with FFDM alone ($p<0.001$). Six cancers were not detected. Regardless of reading strategy (single or double), the incremental CDR from the addition of DBT was +2.7 per 1000 screened ($p<0.001$).[11]

The authors additionally assessed five different strategies to minimize the false positive to true positive (FP:TP) ratio – the number of false positive recalls per true positive screen-detected breast cancers. Thus, this ratio is a measure of specificity. For both reading methods, it was determined that strategies using DBT with conditional recall yielded the highest sensitivities and lowest FP:TP ratios (highest specificities). For the single-reading method, sensitivity was 85%, with a FP:TP ratio of 3.7, and for the double-reading method, sensitivity was 91% with a FP:TP ratio of 4.3. When using a single-reading strategy without the conditional recall rule – the strategy most similar to screening protocols performed in the US – the sensitivity from DBT was 85% versus 54% for FFDM alone. While not explicitly reported in this report, the specificity, PPV_1, and negative predictive value (NPV_1) from recall have been calculated by means of two-by-two tables using data from the single-reading strategy for FFDM alone versus DBT (Table 1). While specificity and NPV_1 were nearly equal for DBT versus FFDM alone – 97% versus 96%, and 99.9% versus. 99.6%, respectively – there was a significant increase in sensitivity (85% versus 54%) and PPV_1 (21.2% versus 11.4%) with DBT versus FFDM alone, respectively.[11]

Table 1. STORM: Single-Reading for FFDM and DBT

FFDM		**Disease Status**			
		Cancer +	**Cancer -**	**Total**	
Test Results	**FFDM +**	35	272	307	**PPV_1: 11.4%**
	FFDM -	30	6,956	6,986	**NPV_1: 99.6%**
	Total	65	7,228	7,293	
		Sensitivity: 54%	**Specificity: 96%**		
DBT		**Disease Status**			
		Cancer +	**Cancer -**	**Total**	
Test Results	**DBT +**	55	204	259	**PPV_1: 21.2%**
	DBT -	10	7,024	7,034	**NPV_1: 99.8%**
	Total	65	7,228	7,293	
		Sensitivity: 85%	**Specificity: 97%**		

Sensitivity was reported by Houssami et al.[11]; Specificity, Positive Predictive Value from recall (PPV_1) and Negative Predictive Value from recall (NPV_1) were calculated for this review.

The results from the STORM reports suggest that the overall and invasive CDR increase significantly with the addition of DBT to screening, with no change in the detection of *in situ* cancers. Furthermore, there were significant reductions in both recall rate and false positive recalls from screening with DBT versus FFDM alone. DBT enhanced sensitivity and PPV_1, without compromising specificity and NPV_1. It should be noted that the 54% sensitivity detected for FFDM is considerably lower than the sensitivity of FFDM observed in the US.[1] In practice, these findings translate to a decrease in recall rate, plausibly as a result of the significant reduction in false positive recalls; an increase in true positive recalls from screening; a higher likelihood that women recalled from DBT screening will truly have cancer; and an increase in overall and invasive cancer detection, with no change in the detection of *in situ* cancers.

2.2.2 OTST: Norway

The *O*slo *T*omosynthesis *S*creening *T*rial (OTST) out of Norway prospectively compared two-view (CC and MLO) screening with DBT versus FFDM alone. Norwegian women 50-69 years of age were sent letters to participate in the trial, which ran from November 2010 through December 2011. Women scheduled for a screening examination

were asked upon arrival if they were willing to participate in the trial, pending the availability of technical staff and necessary imaging systems. Women who were not asked to participate in the trial due to lack of available resources were screened with standard two-view FFDM. Women who were unable to stand and women with breast implants were excluded from the trial. Overall, 12,621 women (average age: 59.3 years) were included in the trial.[12]

A problem posed by independent double-reading without consensus of management decisions is the substantial increase in recall rate.[12] While OTST utilized a four-arm prospective independent double-reading design, they also employed an arbitration process for double-reading prior to recalling patients, in contrast to STORM. Study arms included: Arm A, FFDM alone (2D data); Arm B, FFDM with computer-aided detection (CAD; 2D data), a software feature that augments FFDM lesion detection; Arm C, FFDM+DBT (2D+3D data); Arm D, synthetic-2D+DBT (2D+3D data). Eight radiologists participated in the study, alternating across the four arms. Because of the double-reading design, each arm was independently assessed and rated by one of four radiologists using the following standardized five-point rating scale: 1=Normal/definitely benign; 2=Probably benign; 3=Indeterminate; 4=Probably malignant; 5=Malignant. A consensus meeting was called for cases with any score of 2 or greater, involving at least two radiologists, allowing for discussion and assent of the clinical management decision to either recall or dismiss patients with at least one rating of 2 or 3; conversely, any woman with at least one score of 4 or 5 was automatically recalled and could not be dismissed at consensus. It should be noted that any screening examination rated ≥2 by at least one radiologist was considered to be a positive examination (pre-arbitration

suspicion), regardless of the consensus decision to dismiss or recall. For analyses, it was assumed that Arm A was equal to Arm B, and Arm C was equal to Arm D – in other words, the 2D paired arms (A and B) and the 2D+3D paired arms (C and D) were sufficiently similar such that independent reading of Arm A + Arm B constituted double-reading for 2D images, and independent reading of Arm C + Arm D served as double-reading for 2D+3D exams.[12] Of the 29 additional cancers detected under the 2D+3D mode (and undetected on 2D-alone mode), 24 (82.8%) were invasive, node-negative cancers. Although 2D-based imaging alone missed cancers across all four categories of breast density, the added benefits of 2D+3D (DBT) for screening most significantly improved outcomes in women with BI-RADS breast density categories B and C (scattered fibroglandular densities and heterogeneously dense breasts). There was no significant difference in the *in situ* (DCIS) CDR across modalities.[12]

A more applicable understanding of the OTST study comes from the pre-planned single-reading interim analysis of OTST, evaluating two of the four OTST arms: FFDM alone versus dual acquisition DBT. At the time of the publication, the authors reported a significantly higher CDR (27% increase, $p=0.001$); a significantly higher invasive CDR (40% increase, $p<0.001$); no significant difference in DCIS detection; and a significant 15% reduction in false positive recall rates ($p<0.001$) with DBT versus FFDM alone.[13]

The prospective European studies discussed in this review, STORM and OTST, have important strengths and limitations. Strengths include paired data and prospective study design, a powerful strategy for assessing incidence and causality. Weaknesses include lack of generalizability and limited applicability to US populations due to differences in disease prevalence, population demographics, and screening practices.

2.3 Review of United States Studies

The United States (US)-based studies are inherently more relevant to our proposed study in terms of design, demographics, and screening protocols/practices. Thus, the majority of our study methods (Chapter III) have been derived from these reports.

Study characteristics from the fourteen US reports reviewed in this chapter have been summarized in **Table 5** (Appendix A, pp. 59-60). Reports include eleven retrospective reports, two hybrid prospective/retrospective cohort studies, and one prospective study. The major outcomes for each study have been comprehensively summarized in **Table 6** (Appendix A, pp. 61-63). The ensuing discussion revolves around the study designs and specific outcomes found in the accompanying tables; thus, the reader is encouraged to utilize these resources throughout this section.

All but one of the fourteen US studies reviewed in this chapter employed independent, single-reading interpretations of images with two-view (CC and MLO) bilateral image acquisition for both FFDM and DBT screening, as is standard practice in the United States. A sole publication by Destounis et al. (2014) interpreted screening examinations under a double-reading protocol.

2.3.1 Initial Studies

In 2013, Rose et al. published a single-site retrospective study comparing 23,355 asymptomatic women 18 years and older who presented for screening both before and after a distinct FFDM-to-DBT transition period. Women who were screened with FFDM after the transition period were excluded from analyses, although most (88%) of women presenting for screening after this time elected for screening with DBT. Because the site's radiologists had varying degrees of DBT experience, the authors decided to include only

cases evaluated by radiologists who had assessed at least 500 cases pre- and post-DBT implementation. Thus, the study's sample size was derived from cases interpreted by six radiologists meeting these qualifications. *A priori*, a generalized linear model was prepared to compare outcomes across DBT and FFDM after adjusting for age at screening, baseline exams, and individual radiologists' performance via random-reader assumption. Outcomes were stratified by BI-RADS breast density categories (A, B, C, and D) and age at screening (<50 years; 50-64 years; and >64 years).[14]

Overall, Rose et al. (2013) reported a significant 37% reduction in recall rate and an 115% increase in PPV_1 with DBT versus FFDM alone, in addition to non-significant increases in the overall and invasive CDR – trends that reflected findings observed in the prospective European trials. Of note, the improved performance metrics observed with DBT were significant across all age groups and breast density subgroups, with the largest gains for women presenting for baseline (first ever) exams. These findings highlight the benefits gained by the addition of DBT to screening, such as enhanced cancer detection, in addition to the abatement of harms, as evidenced by lower recall rate and higher likelihood that recall from DBT screening results in a diagnosis of cancer, ultimately leading to fewer false positive recalls, unnecessary workups, and benign biopsies.[14]

Rose et al. (2014) sought to validate the improved performance metrics observed in their principal 2013 report. The authors subsequently published an independent reading study retrospectively interpreting FFDM screens acquired during dual acquisition DBT during the previously reported DBT time period.[14] These screening outcomes were then compared to the outcomes from the corresponding, prospectively interpreted DBT exams published in the 2013 publication; thus, all data were paired across modalities. In the

2013 report, 10,878 DBT screens from asymptomatic women were prospectively interpreted by ten radiologists, with seven radiologists now participating in the retrospective interpretation of the paired FFDM images for the 2014 study – most of whom had limited experience with DBT at the beginning of DBT transition period. Radiologists involved in the retrospective interpretation of FFDM screens were assigned by a third party to cases that they had not personally interpreted in clinic, and were blinded to the true clinical recommendations and associated outcomes. All retrospective FFDM cases rated as BI-RADS 0 (recommendation of recall) were reviewed by an independent third radiologist who had not been involved in the interpretation of the patient's screening examinations.[15]

Outcomes for recall rate were much the same as reported in the authors' 2013 publication, with a significant reduction of 33.7%. Overall and invasive CDR increased significantly with DBT versus FFDM alone – 54.3% and 63%, respectively – with no significant difference in the *in situ* CDR across modalities.[15] These findings lent credence to the authors' 2013 findings, and suggested that the previously observed non-significant increases in the overall and invasive CDR were likely limitations of the former report's design and sample size, with the latter report finding these outcomes to be significant.

Limitations of Rose et al.'s 2013 report include its retrospective nature, non-randomization of patients, and lack of power to detect significant changes in cancer detection due to the relatively small sample size. Furthermore, because women had to elect to undergo screening with DBT, self-selection bias cannot be ruled out or easily accounted for. The authors also noted that the effects of learning this new technology were difficult to quantify, though they expected that performance metrics were not likely

to have increased substantially over the short duration of the study. The 2013 study would have additionally benefited from more detailed baseline characterization across modalities to compare important and potentially confounding factors, such as race, ethnicity, and personal/family history of breast cancer. In the 2014 report, however, the use of paired data eliminated potential differences in disease prevalence and baseline characteristics across groups.

In 2013, Haas et al. published a retrospective study comparing 13,158 women presenting for screening mammography with DBT versus FFDM alone at one of four sites affiliated with Yale University. DBT screening was offered where and when it was available to consenting women at no additional cost. Women with breast implants or large breasts requiring tiled images were excluded from DBT. Per modality, recall rate was calculated by dividing the number of BI-RADS 0 women by the total number screened, and CDR was evaluated by determining the number of true positive BI-RADS 0 cases per 1000 women screened. Recall rate and CDR (overall, invasive, *in situ*) were additionally stratified by patient age at screening and breast density categories. *A priori*, the authors planned to utilize multivariate logistic regression analysis to determine any significant differences in CDRs and recall rates across modalities after controlling for patient age at screening (<40; 40-49; 50-59; 60-69; ≥70 years), BI-RADS breast density categories (A, B, C, D), and the presence of a personal history of breast cancer and/or family history of breast cancer in a first-degree relative.[16]

Comparison of baseline population characteristics revealed similarities in breast densities and ages across the two groups, but found that patients in the DBT group were more likely to have a personal history and family history of breast cancer as compared to

the FFDM group. A statistically significant reduction in recall rate of 30% was observed in women screened with DBT versus FFDM alone. Stratification by breast density revealed recall rate reduction for all subgroups, with significance observed for BI-RADS categories B, C, and D. Stratification by age demonstrated recall rate reduction for all ages, with significance seen in all groups except for women 70 years of age or older. These outcomes, reported as odds of recall, persisted after multivariate logistic regression analyses, with greatest reductions seen for women younger than 40 years old and women with BI-RADS breast density category D. Multivariate adjustment for age, breast density, and the presence of personal and/or family history (or absence of both) revealed a 38% reduction in odds of recall from DBT versus FFDM alone. The authors found age and breast density to be independent predictors for risk of recall. There was a slight, albeit non-significant, 9.6% increase in the overall CDR and 11.4% increase in the invasive CDR with DBT versus FFDM alone, with no change in the *in situ* CDR.[16]

Durand et al. expanded upon the above cohort in their 2015 publication, comparing outcomes from 17,955 women (including the 13,158 cases from Haas et al.'s 2013 report) screened with DBT versus FFDM.[17] In addition to retrospectively comparing overall recall rates and cancer detection rates across the two modalities, the authors compared recall rates for baseline examinations across the two modalities, a population previously shown to derive benefit from screening with DBT, appreciation recall rate reduction and increased cancer detection rates, as compared to FFDM.[14] In contrast to Haas et al.'s 2013 report, in which the only baseline difference across groups was the significantly higher proportion of women in the DBT group with risk factors, Durand et al. (2014) observed significant differences across all baseline measures: risk factors, age at

screening, breast density, and number of women presenting for baseline examinations ($p<0.0001$). However, the significant 36.6% recall rate reduction with DBT versus FFDM alone substantiated the 2013 findings by Haas et al. After adjusting for confounding factors, the odds of recall from FFDM were found to be 1.76 times higher than DBT. Similarly, for baseline examinations, recall rate was significantly reduced by 37% ($p<0.0001$). Non-significant increases of 3.5% and 8.8% were observed in the overall and invasive CDR, respectively, with no change in the *in situ* CDR across modalities.[17]

In their 2013 report, Haas et al. speculated that the true differences in CDR across the two modalities may have been greater than what was observed due to a confounding effect through use of DBT in the diagnostic environment at one of the four sites; consequently, women recalled from the FFDM group may have benefited from cancer detection with diagnostic DBT, falsely elevating the CDR from FFDM.[16] If accurate, this would have similarly affected the CDRs in Durand et al.'s 2015 publication. Regardless, both publications were underpowered to detect significant changes in cancer detection. Due to the discrepancies in DBT screening capability across sites in these two reports, outcomes may have been affected by selection bias due to potential differences in populations presenting to sites that both did and did not offer DBT as a screening option. The observed dissimilarity in baseline screening characteristics suggest that this is a possibility; in the 2013 report, significantly more women in the DBT group had a personal and/or family history of breast cancer, a population that is inherently at higher risk of developing breast cancer. Therefore, it is not outside the realm of reason to presume that such a population might be recalled from screening at a higher rate than the general population. Yet, despite the fact that the DBT group contained proportionally

more high-risk women than the FFDM group, a statistically significant reduction in recall rate was observed across the two modalities.[16,17] In the 2015 report, baseline characteristics across modalities further diverged, with significant dissimilarity in risk factors, age, breast density, and number of baseline examinations performed. Furthermore, the observed differences in risk factors across groups were split: the proportion of women with a strong or intermediate family history was significantly higher in the FFDM group than in the DBT group, while the proportion of women with a personal history of breast cancer was significantly higher in the DBT group than in the FFDM group.[17] Nevertheless, after adjusting for confounding factors, the odds of recall were still significantly lower for DBT versus FFDM.

2.3.2 Subgroup Benefits

Logistic regression analyses in the 2013 report by Haas et al. revealed that the women who derived the greatest benefit from the addition of DBT to screening examinations included those with dense breasts and/or those younger than 50 years of age; precisely the populations in which FFDM alone is least effective.[16] The 2015 publication by Durand et al. substantiated previous evidence published by Rose et al. (2013) suggesting that women presenting for baseline examinations derived significant benefits from screening with DBT versus FFDM alone by means of recall rate reduction.[14] The outcomes from the reports by Rose et al. (2013, 2014), Haas et al. (2013), and Durand et al. (2015) converge when considering the fact that younger women – a population with inherently higher breast density – are likely to constitute a significant proportion of women presenting for baseline screening examinations. As discussed below, several other studies

found significant benefits in specific populations from DBT screening, including young women presenting for baseline examinations.[18-21]

McCarthy et al. (2014), in their single-site retrospective study (n=26,299) comparing women screened with FFDM versus DBT, found non-significant increases of 19.6% and 22% in the overall and invasive CDR, respectively, with no change in the *in situ* CDR. When stratified by age, the CDR increased a significant 159% in women under 50 years of age screened with DBT versus FFDM alone. The overall recall rate was significantly decreased by 15.4% with DBT versus FFDM; significance persisted with univariate and multivariate adjustment for number of screening rounds, age, race, interpreting radiologist, and presence or absence of a prior mammogram, with reduced odds of recall for women screened with DBT versus FFDM. With multivariate adjustment, the odds of recall were 20% lower from DBT than from FFDM alone.[18]

Other subgroups significantly benefited from screening with DBT versus FFDM after multivariate adjustment. The odds of recall for FFDM versus DBT were highest for women presenting for baseline screening exam (adjusted OR=2.31; 95% CI: 2.04-2.61, $p<0.001$); younger women, ages 40-49 years (adjusted OR=1.85; 95% CI: 1.56-2.19, $p<0.001$) and 50-59 years (adjusted OR=1.41; 95% CI: 1.19-1.67, $p<0.001$); and black women (adjusted OR: 1.14; 95% CI: 1.04-1.26, $p=0.008$). Simply put, women with no prior mammograms (baseline examinations) had >2.5 times higher odds of recall than women with a prior mammogram; women age 40-49 and 50-59 years had 85% and 41% higher odds of recall, respectively, compared to women ≥70 years (referent age group); and black women had 14% higher odds of recall compared to white women. Further multivariable adjustment was performed in a random subset of the population (58% of

the total sample), including additional risk factors for breast cancer (Appendix F). From this adjustment model, there was a significant 23% decrease in the odds of recall from DBT versus FFDM alone (adjusted OR=0.77; 95% CI: 0.68-0.87, p<0.001). Stratification of recall rates by breast density revealed significant decreases with DBT versus FFDM in both non-dense breasts (BI-RADS categories A and B; 14.3% reduction, p=0.001) and dense breasts (BI-RADS categories C and D; 16.5% reduction, p=0.006). The positive predictive value from recall (PPV_1) significantly increased by 41% with DBT versus FFDM alone; when stratified by age, this value was a significant 17.5% higher in women under 50 years of age screened with DBT versus FFDM alone.[18]

McDonald et al.'s 2015 report is from the same population of 26,299 women as reported above. However, in attempt to better delineate the effectiveness of DBT in women with no previous screening mammography exams, the population was stratified into two cohorts across the FFDM and DBT time periods: a baseline subgroup (women with no prior mammograms) of 3,063 women and a previously-screened group, the vast majority of the population, comprising 23,236 women. Unsurprisingly, women in the baseline subgroup were, on average, nine years younger than women in the previously-screened group. The mean age was similar across modalities for both subgroups.[19]

The analyses focused primarily on the baseline subgroup, and found a significant recall rate reduction of 22% with DBT versus FFDM alone. Within the baseline subgroup, there was a significant 24.1% reduction in recall rate for women under 50 years of age with DBT versus FFDM alone. A significant (p=0.004) recall rate reduction of 24.1% was similarly observed in baseline women with non-dense breasts (BI-RADS categories A and B). Women with dense breasts (BI-RADS categories C and D) had a

non-significant 17.2% recall rate reduction (p=0.14) from DBT versus FFDM. Multivariate logistic regression analysis of recall rate (adjusting for breast density BI-RADS categories; age; race; and interpreting radiologist) resulted in significant decreases in odds of recall from DBT versus FFDM for both subgroups: 16% lower odds of recall in the previously-screened group, and 26% lower odds of recall in the baseline subgroup. The PPV_1 increased a non-significant 85% in the baseline subgroup and a significant 35.3% in the previously-screened group. Multivariate logistic regression analysis of PPV_1 (adjusting for age and breast density) was significant only in the previously-screened group, revealing a 57% increased odds of cancer detection in cases recalled from DBT versus FFDM. In the baseline subgroup, the odds of cancer detection from recall were 81% higher with DBT versus FFDM, but this was a non-significant finding.[19]

Strengths of the reports by McCarthy et al. (2014) and McDonald et al. (2015) include a swift DBT transition period of one month, and the exclusion of data acquired during this period further limited the possibility of selection bias. The same six radiologists interpreted all images pre- and post-DBT implementation. In the 2014 report by McDonald et al., the authors were able to assess several additional breast cancer risk factors that had not yet been reported in the screening population. Limitations of these studies, aside from their retrospective design, include significant disparity in distribution of baseline characteristics across women screened per imaging modality: in the DBT group, there were proportionally fewer black patients; proportionally more women categorized as BI-RADS A; and proportionally fewer patients who were previously screened. These differences were, however, accounted for in multivariate logistic regression analyses.[18]

In 2015, Sumkin et al. published a small, single-site prospective cohort study specifically comparing paired DBT and FFDM data in young asymptomatic women (ages 34-56 years) presenting for baseline screening examinations (n=1,074).[20] The only outcome of interest was recall rate, which tends to be higher in the population of interest, as noted by several studies discussed thus far.[14,16-19] Each participant was imaged with FFDM and dual acquisition DBT, and the two exams were independently reviewed in clinic by two of fourteen participating radiologists. By design, the authors included no consensus process to best model US screening practice. Therefore, a participant was recalled if either (or both) of the two independent radiologists assigned a BI-RADS 0 designation from either the FFDM-only or DBT exam. Recommended recalls were analyzed using a generalized linear mixed model for binary data in consideration of a possible correlation between exam interpretations of an individual patient and the variability of individual radiologists' performance per modality. Because of the split nature of the study (two exams under two reading modes by two radiologists), a two-sided *p* value of 0.0294 was used to test for significance. The recommendation for recall was significantly decreased by 33.6% with DBT versus FFDM alone ($p<0.001$). The false positive rate was found to be 16.4% after 176 women without breast cancer were recalled by both DBT and FFDM. Overall, in addition to a significant reduction in recall rate, the authors observed a notable decrease in benign biopsies performed with use of DBT, contributing to the evidence supporting the role of DBT in harm reduction from screening mammography.[20]

Strengths of this study include the prospective design, allowing for causal inference, and the use of paired data, eliminating differences across groups. The generalizability of

this study is limited, as it was performed out of a single institution and evaluated a small, specific group of women. Self-selection bias cannot be ruled out, especially because high risk women were recruited. Cancer prevalence is low in this age range (34-56 years),[20] and the sample size was significantly underpowered to detect significant changes in cancer detection; thus, comparison of CDR across modalities was not possible.

Sharpe et al. (2016) published a single-site prospective study of DBT examinations (n=5,703) with retrospective comparison of FFDM images (n=80,149) from a total of 85,852 asymptomatic women with no breast symptoms or diagnosis of breast cancer within five years of the study. Recall rates were calculated for the two exams overall and stratified according to patient age at screening; breast density; family history and personal history of breast cancer; history of *BRCA* gene mutation; personal history of benign breast biopsy; whether or not it was the patient's first mammographic examination; and interpreting radiologist. A mixed-effects logistic regression analysis was performed to determine the correlation between each parameter and the recall rate.[21]

Outcomes demonstrated a significant 54.3% increase in the overall CDR (including a 153% increase in the *in situ* CDR); a significant 18.8% reduction in recall rate, with significant subgroup benefit seen in women with BI-RADS categories C and D, and women ages 40-49 or 60-69 years with DBT versus FFDM alone.[21] The authors offered no plausible basis for the lack of significant recall rate reduction in the 50-59-year age group. Given the differences in baseline characteristics – namely, increased prevalence of women with risk factors for breast cancer in the DBT group – self-selection bias is a reasonable concern, and may have affected the observed differences in RR and CDR across modalities. This suspicion is heightened in light of detecting an increased overall

and *in situ* CDR, despite being significantly underpowered to do so. Because no other study has observed a statistically significant increase in the *in situ* CDR from screening with DBT versus FFDM alone, this finding likely represents an outlier.

2.3.3 Large Multisite Studies

The largest US study to date is a retrospective, multicenter, multisite trial by Friedewald et al. (2014), encompassing thirteen sites and a sample size of 454,850 exams from asymptomatic women presenting for screening mammography. The study was designed to detect changes in CDR and RR with power of 80% and >99%, respectively. All participating sites were in the process of transitioning from FFDM to DBT for screening, though this process differed across sites. Women presenting pre-transition period were screened with standard FFDM; those presenting post-transition were screened with DBT. Because sites contributed varying numbers of participants overall and per modality, outcomes were model-adjusted to account for site as a random effect.[22]

Results revealed a significant recall rate reduction of 15%; significant increases of 28.6% and 41.4% in the overall and invasive CDR, respectively, with no change in the *in situ* CDR. Of note, per 1000 women screened, DBT detected significantly more invasive ductal carcinomas (IDC; 33% increase) and invasive lobular carcinomas (ILC; 103.7% increase), the most and second most common types of invasive breast cancers, respectively.[23] There was a significant 48.8% increase in PPV_1 with DBT versus FFDM.[22]

Outcomes from this major study were consistent with findings from the smaller US studies, in addition to the prospective European trials. This is noteworthy considering the international differences in screening practices and patient demographics. Furthermore, the thirteen sites included in the meta-analyses represent geographically diverse,

academic and non-academic settings, encompassing both specially trained radiologists and non-specialists. For these reasons, outcomes from this study are more generalizable to the US population at large than many smaller, single-site retrospective studies.

This study was limited to population-level data, so adjustments were not possible for women with repeat examinations. Differences in age at screening, breast density categories, race/ethnicity, and risk factors were not described in the baseline populations. Site differences in FFDM-to-DBT conversion times were not a consideration in the adjustment model, nor were individual radiologists' performance metrics. Only two sites made a full transition, with the remaining sites operating as hybrid environments. Because most sites concurrently offered both DBT and FFDM screening throughout the study period, the authors were unable to rule out selection bias. FFDM-only screening examinations performed in the DBT period were excluded from the primary analyses. However, to address the lingering possibility of selection bias, analyses were repeated with the inclusion of the concurrent FFDM-only examinations in the DBT group to test for significant changes in CDR and RR pre- and post-DBT implementation. Outcomes did not change significantly when such cases were added, despite the fact that nearly 60% of screening exams were acquired with FFDM in the DBT study period.[22] Lastly, because women could decline DBT and opt for FFDM screening, self-selection bias is a potential confounding factor that is difficult to address (i.e., women with known risk factors might be more likely to elect for DBT screening).

That same year, Greenberg et al. (2014) published a smaller, albeit well-powered multisite retrospective study comparing screening outcomes in 59,617 women presenting for screening mammography performed with DBT versus FFDM alone. Only cases read

by one of fourteen radiologists who had interpreted more than 500 DBT images during the study period were included in analyses due to discrepancies in experience across radiologists. It is worth noting that, of the 77,833 women screened, only 30% elected for screening with DBT, while 70% opted for FFDM – possibly due to selection bias from a $50 fee for screening with DBT, though this cost was waived if a participant could not afford to pay. Despite the disparity in women who opted for screening with FFDM in lieu of DBT, there were no significant differences in baseline characteristics across groups. A generalized linear mixed model (GLMM) was employed to account for individual radiologists' performance with random reader assumption; multiple screening rounds were also evaluated, using individual patients as the unit of analysis.[24]

Outcomes revealed a significant recall rate reduction of 16% for DBT versus FFDM alone, with significance persisting after application of the GLMM. The overall and invasive CDR significantly increased 28.6% and 43.8%, respectively, for DBT versus FFDM, with no significant change in the *in situ* CDR. A significant 43.8% gain in PPV_1 was observed with DBT versus FFDM, with a yet higher estimated increase of 51.14% from the GLMM. The authors also found a significant reduction in the number of additional mammographic views obtained in women diagnosed with breast cancer after recall from DBT versus FFDM alone: 74.1% of patients in the DBT group required two or fewer views at recall, with 35.1% requiring no additional views; conversely, 51% of women in the FFDM group required two or fewer views, with no additional views needed for 6.3% ($p<0.001$).[24] In accordance with the outcomes observed in the prospective European trials and across several US studies, including a large US meta-analysis, these findings provided further support for the role of DBT in both benefit gain and harm

reduction from screening, evidencing enhanced efficiency at recall with fewer required mammographic views leading to diagnosis and, therefore, minimizing radiation exposure.

2.3.4 Sustainability of Improved Outcomes

Studies discussed in the previous subsections largely represent outcomes from first rounds of DBT screening in the United States, otherwise known as prevalence screening.[25] Generally, prevalence (first round) screening has higher rates of detection than incidence (subsequent) screening, because the cancers detected in prevalence screening are from the large pool of cancers that already exist in a given population.[25] On subsequent rounds, the cancers detected are, theoretically, incident cancers (new cancer cases). To determine whether or not the enhanced performance of DBT screening is sustainable, subsequent rounds of DBT screening (incidence rounds) must be compared to prevalence data from the same population.[26] Two 2016 publications sought to assess the sustainability of improved outcomes from subsequent DBT screening.

Conant et al. (2016) retrospectively evaluated screening outcomes from asymptomatic women 40-74 years of age screened with DBT versus FFDM. Participating women had no known history of breast cancer and no breast imaging within three months of the study. Screening was performed at three large academic research centers with several sites, all transitioning from FFDM to DBT screening, though the conversion occurred at different times and varying rates. Exams were limited to those read by radiologists who had interpreted at least 50 DBT and 50 DM exams. By these criteria, 198,881 exams from 103,401 women were eligible for analyses; the baseline group included 45,049 women with one exam, while the subsequent exam group comprised 29,041 women with two prior exams and 29,311 women with three or more prior exams. *A priori*, the authors

prepared a logistic regression model adjusting for research center, age at screening (40-49; 50-59; 60-74 years), breast density (BI-RADS categories A, B, C, D), and first exam. Recall rate was further adjusted for the interpreting radiologist in a conditional logistic regression model. The primary outcomes were subject to a generalized estimating equation (GEE) model accounting for potential correlation of screening exams within the same individual, which yielded similar ORs and 95% CIs as the multivariate model.[27]

In all, outcomes from the subsequent (incidence) screening rounds in women with at least one year of imaging follow-up followed trends in the established literature (prevalence screens), with significant recall rate reduction without increase in false negative rates; increased overall and invasive CDR; and enhanced PPV_1. Furthermore, there was a significant increase in specificity without diminished sensitivity. Lastly, after adjusting for confounding variables, the authors found similar odds of recall for women with both dense and non-dense breasts, as well as for women in both age groups, suggesting that there is no particular advantage for benefits of DBT by age or density.[27]

Perhaps the most convincing data for the longitudinal effectiveness of DBT screening comes from McDonald et al.'s 2016 publication comparing outcomes from four consecutive years of screening in a large, urban academic center: one year of FFDM (year 0; FFDM-0), followed by three years of screening post-DBT implementation (years 1-3; DBT-1, DBT-2, DBT-3). In all, the authors obtained screening outcomes from 44,468 screening exams attributable to 23,958 women with no history or clinical signs/symptoms of breast cancer. At the population level, outcomes were assessed across DBT years, and each DBT year was evaluated against FFDM-0 (the referent year); outcomes were stratified according to breast density (non-dense: BI-RADS categories A

and B versus dense: BI-RADS categories C and D) and age (<50 years versus ≥50 years). To assess the effect of prevalence and incidence screening, RR and CDR were compared at the most recent screening exam across women undergoing their first, second, or third round of screening with DBT. The screening outcomes for groups of women with only one, two, or three DBT screens were adjusted for age, race/ethnicity, breast density, and the presence or absence of a prior mammogram to estimate the odds of recall. Analysis of women with only one DBT screen was further confined to women with a prior FFDM exam available. For individual-level analysis of recall rates across the four years, GEE models with logistic regression were employed (adjusting for age, race/ethnicity, breast density, and number of prior screens using individual patients as the units of analysis).[28]

At the population level, each DBT year showed significant recall rate reduction as compared to FFDM-0, with no significant difference across DBT years despite slight increases in RR each year. The overall and invasive CDR increased non-significantly each year, and but not significantly different across DBT years or compared to the FFDM CDR. The $PPV_{1,}$ continued to rise each year compared to FFDM-0, with a non-significant 41% increase in DBT-1; a significant 47.7% increase in DBT-2; and a significant 52.3% increase in DBT-3. The PPV_1 across DBT years did not differ significantly.[28]

For the assessment of DBT prevalence and incidence screening, women with only one screen (prevalence exam), were compared to the group of women with two and three screens (incidence exams). Recall rates continued to decline with increased number of prior screens: recall rate was 13% for women with only one screen, 7.8% for two screens, and 5.9% for three screens. As compared to the one-screen group, RR reductions were significant ($p<0.001$) for women with two screens (OR=0.56; 95% CI: 0.51-0.63) and

three screens (OR=0.42; 95% CI: 0.35-0.79); CDR was significantly lower for the two-screen group (OR=0.55; 95% CI: 0.39-0.79, p<0.001), but was not significantly lower for the three-screen group (OR=0.65; 95% CI: 0.41-1.02, p=0.06). When the one-screen group was constrained to include only women with a prior FFDM exam available for comparison, similar results were obtained. The PPV_1 was significantly lower only when comparing the restricted one-screen group to the two-screen group (p=0.028).[28]

This report established continued recall rate reduction from DBT screening with subsequent screening rounds, with significant reductions each DBT year as compared to the referent FFDM year, and no significant difference across DBT years. RR reduction was amplified in women returning for second and third (incident) DBT screens. The CDR was not significantly different across DBT years, and despite increases each year, was not significantly higher as compared to FFDM, possibly due to the small sample size. The PPV_1 was significantly higher each DBT year as compared to FFDM, with no significant difference across DBT years. Overall, the benefits from DBT screening seem to extend beyond prevalence screening, and appear to be sustainable at the population level.

2.4 Tumor Characteristics

The findings from studies detailing the characteristics of screen-detected tumors are summarized in two tables: European studies, **Table 7** (Appendix B, p. 64) and US studies, **Table 8** (Appendix B, pp. 65-66). Discussion in the following subsections is heavily derived from the outcomes found in these tables; thus, the reader is encouraged to utilize these resources throughout this section. To enhance applicability to US data, the information presented from the European trial, OTST, is derived from the single-reading data. Information regarding tumor characteristics was not described separately in the

single-reading mode of STORM. One additional study that has not yet been discussed is included in the subsections below, for an aggregate review of seven reports.

Briefly, Wang et al. (2016) sought to characterize the biologic features of the additional cancers detected with DBT versus FFDM alone, expanding upon Skaane et al.'s 2013 interim OTST analysis, which revealed increased detection of primarily low-grade, invasive, node-negative cancers with DBT, in contrast to no significant differences observed by Greenberg et al. (2014). Sixty-five breast cancers detected in 63 women undergoing first-ever screening (with DBT as the first exam) were retrospectively interpreted by five radiologists, each independently evaluating FFDM versus DBT images. A cancer was considered mammographically occult if all radiologists agreed that the cancer was not visible on FFDM, for a total of ten cancers detected only with DBT.[29]

2.4.1 Tumor Size

Five studies published data on the sizes of cancers detected with DBT versus FFDM alone, four of which included size information for invasive cancers only. In the STORM trial, the mean size of invasive tumors was not significantly different for tumors detected with DBT only (FFDM occult) compared to cancers detected with both FFDM and DBT.[9] Rose et al. (2013) found that the sizes of invasive tumors did not differ significantly as detected with DBT versus FFDM.[14] Durand et al. (2015) did not provide mean or median size values, but rather the pathologic size ranges of invasive tumors, reporting that the majority of invasive cancers (75% of DBT-detected and 72% of FFDM-detected) fell in the range of less than 2 cm across both modalities, with 25% of tumor sizes ranging from 2 to 5 cm with both DBT and FFDM alone.[17] The OTST single-reading interim

publication reported tumor size means, medians, and ranges for both invasive and *in situ* cancers, with no significant differences across modalities noted for either.[13]

In the 2016 publication by Wang et al., nine of the ten cancers detected with DBT only (FFDM occult) were invasive cancers. Compared to the 32 invasive cancers visible on FFDM, the sizes of cancers detected with DBT only (FFDM occult) were smaller, trending towards but not quite achieving significance (p=0.07).[29]

2.4.2 *In Situ* Cancers

Overwhelmingly, the data suggest that there is no increase in the *in situ* CDR from screening with DBT compared to FFDM alone. Lower grade *in situ* cancers are generally slower growing and have lower probability of progression and recurrence.[30] After stratifying screen-detected DCIS by nuclear grade, Greenberg et al. (2014) found that DBT detected significantly more low- and intermediate-grade *in situ* cancers (67.6%) than did FFDM (46.7%).[24] Two other studies saw no significant difference in the distribution of DCIS grades across modalities.[13,14]

2.4.3 Invasive Cancers

Six studies characterized the histologic types of invasive cancers detected with DBT versus FFDM, and none reported significant differences in the proportions detected across the modalities. Only one study reported the proportion of receptor phenotypes (luminal, HER2-enriched, or triple negative) in invasive cancers, and found no significant differences across modalities.[9]

The Nottingham histologic grade of invasive cancers has important implications for prognosis and treatment.[31] Greenberg et al. (2014) found no significant differences in the proportions of invasive cancer grades detected across modalities.[24] The cancer grades

were similarly distributed in the prospective European trial, STORM.[9] Rose et al.'s 2013 study reported no significant differences in the grades of invasive cancers nor the nuclear grades of DCIS, but noted that more invasive cancers detected with FFDM alone were grade 3, while more invasive cancers detected with DBT were grade 2.[14] Skaane et al.'s 2013 interim OTST single-reading analysis found no significant difference in the proportions of invasive cancer grades detected across modalities; however, when evaluating the 29 additional cancers found only with adjunct DBT (FFDM occult), 48% were grade 1, 39% grade 2, and only 10% grade 3.[13] Wang et al. (2016) similarly found a significant difference in the grades of DBT-detected (FFDM occult) invasive cancers as compared to invasive cancers visible on FFDM, observing that 78% of DBT-only cancers were grade 1 compared to 47% of FFDM-detected cancers, while 22% of DBT-only cancers were grade 2 to 3 compared to 53% of FFDM-detected cancers.[29]

Perhaps the most significant prognostic factor for patients with early-stage breast cancer is axillary lymph node status.[31] Two screening studies found no difference in lymph node status across modalities.[14,17] Lymph node status was not significantly different across modalities in STORM.[9] As noted by the authors, one of the most important findings from Skaane et al.'s 2013 interim OTST analysis was that the additional cancers detected with DBT were primarily invasive and node-negative.[13] Twenty-nine of the 30 additional cancers detected by DBT (FFDM occult) were invasive (96.7%), and 80% of the invasive cancers were node negative.[13] Wang et al. (2016) found that, of the nine additional invasive cancers detected with DBT (FFDM occult), 100% were node-negative, in contrast to 87% of FFDM visible cancers, a non-significant finding.[29]

2.5 Conclusion

In consensus with the two European trials, STORM and OTST, all fourteen of the US studies included in this review appreciated significant reductions in recall rate from DBT versus FFDM alone, with relative recall rate reductions ranging from 15% to 63%. All but one of the fourteen studies reported an increase in the overall CDR with DBT versus FFDM and, although most studies were underpowered to detect significant changes in cancer detection, five publications reported significant increases in the overall CDR with the addition of DBT to screening mammography. The single report that observed a slight increase in overall CDR with FFDM alone versus DBT was far from reaching statistical significance, and was similarly the only study to observe a non-significant decrease in the invasive CDR.[32] The remaining thirteen US studies observed an increase (n=12) or no change (n=1) in the invasive CDR with DBT versus FFDM alone, though only four of these studies were powered to achieve statistical significance. A single study reported an equal invasive CDR with DBT versus FFDM alone; however, with a total sample size of only 1,048, this study was significantly underpowered to observe significant changes in cancer detection, with only one invasive cancer (1/524) detected per modality.[33] Both prospective European trials reported statistically significant increases in the invasive CDR with DBT versus FFDM alone.[9-13] Thirteen of the fourteen US studies reported no significant difference in the *in situ* CDR, with most reporting nearly equal or slightly decreased rates with DBT versus FFDM alone. The same study that reported an equal invasive CDR found a non-significant increase the *in situ* CDR with DBT versus FFDM, again due to the very small sample size resulting in detection of two cases of DCIS with DBT (2/524) and one case of DCIS with FFDM (1/524).[33] None of the reports from the

two European trials, and only one of the fourteen US studies, observed a significant increase in the *in situ* CDR, though the authors offered no explanation of this finding, which likely represents an outlier.[21] Each of the nine studies that evaluated positive predictive value from recall (PPV_1) noted increases with DBT versus FFDM alone, with seven studies achieving significance. Importantly, two recent studies found that the benefits of DBT screening are sustainable over subsequent screening rounds.[27,28]

Several studies noted preferential benefits from screening with DBT in certain subgroups – in particular, women with no prior mammograms available for comparison (baseline exams) – though it seems most likely that all women derive benefits from screening with DBT. Taken together, the evidence from the European screening trials, STORM and OTST, in addition to the US screening studies, strongly suggest that DBT, as compared to FFDM alone, significantly reduces recall rate and false positive recalls, with no significant change in the false negative recall rate; significantly increases CDR (overall and invasive) with no change the *in situ* CDR; and yields a higher PPV_1.

Finally, there is a dearth of evidence in the literature characterizing screen-detected cancer. Evidence to date suggests that DBT may have a role in detecting invasive, low-grade, node-negative cancers, but this is far from established. A detailed analysis of the characteristics of DBT-detected versus FFDM-detected cancers is warranted, and our proposed study aims to address this crucial gap in knowledge.

References

1. Lewin JM, D'Orsi CJ, Hendrick RE. Digital mammography. *Radiologic clinics of North America.* 2004;42(5):871-884, vi.
2. Park JM, Franken EA, Jr., Garg M, Fajardo LL, Niklason LT. Breast tomosynthesis: present considerations and future applications. *Radiographics : a review publication of the Radiological Society of North America, Inc.* 2007;27 Suppl 1:S231-240.
3. Skaane P. Breast cancer screening with digital breast tomosynthesis. *Breast cancer (Tokyo, Japan).* 2016.
4. Hooley RJ, Durand MA, Philpotts LE. Advances in Digital Breast Tomosynthesis. *AJR. American journal of roentgenology.* 2016:1-11.
5. Hodgson R, Heywang-Kobrunner SH, Harvey SC, et al. Systematic review of 3D mammography for breast cancer screening. *Breast (Edinburgh, Scotland).* 2016;27:52-61.
6. Rahbar H, McDonald ES, Lee JM, Partridge SC, Lee CI. How Can Advanced Imaging Be Used to Mitigate Potential Breast Cancer Overdiagnosis? *Acad Radiol.* 2016;23(6):768-773.
7. Takahashi TA, Lee CI, Johnson KM. Breast cancer screening: Does tomosynthesis augment mammography? *Cleveland Clinic journal of medicine.* 2017;84(7):522-527.
8. Cole EB, Pisano ED. Tomosynthesis for breast cancer screening. *Clinical imaging.* 2016;40(2):283-287.
9. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *The Lancet. Oncology.* 2013;14(7):583-589.
10. Caumo F, Bernardi D, Ciatto S, et al. Incremental effect from integrating 3D-mammography (tomosynthesis) with 2D-mammography: Increased breast cancer detection evident for screening centres in a population-based trial. *Breast (Edinburgh, Scotland).* 2014;23(1):76-80.
11. Houssami N, Macaskill P, Bernardi D, et al. Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading--evidence to guide future screening strategies. *Eur J Cancer.* 2014;50(10):1799-1807.
12. Skaane P, Bandos AI, Gullien R, et al. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. *Eur Radiol.* 2013;23(8):2061-2071.
13. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology.* 2013;267(1):47-56.
14. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R, Jr. Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR. American journal of roentgenology.* 2013;200(6):1401-1408.

15. Rose SL, Tidwell AL, Ice MF, Nordmann AS, Sexton R, Jr., Song R. A reader study comparing prospective tomosynthesis interpretations with retrospective readings of the corresponding FFDM examinations. *Acad Radiol.* 2014;21(9):1204-1210.
16. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology.* 2013;269(3):694-700.
17. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology.* 2015;274(1):85-92.
18. McCarthy AM, Kontos D, Synnestvedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *Journal of the National Cancer Institute.* 2014;106(11).
19. McDonald ES, McCarthy AM, Akhtar AL, Synnestvedt MB, Schnall M, Conant EF. Baseline Screening Mammography: Performance of Full-Field Digital Mammography Versus Digital Breast Tomosynthesis. *AJR. American journal of roentgenology.* 2015;205(5):1143-1148.
20. Sumkin JH, Ganott MA, Chough DM, et al. Recall Rate Reduction with Tomosynthesis During Baseline Screening Examinations: An Assessment From a Prospective Trial. *Acad Radiol.* 2015;22(12):1477-1482.
21. Sharpe RE, Jr., Venkataraman S, Phillips J, et al. Increased Cancer Detection Rate and Variations in the Recall Rate Resulting from Implementation of 3D Digital Breast Tomosynthesis into a Population-based Screening Program. *Radiology.* 2016;278(3):698-706.
22. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA.* 2014;311(24):2499-2507.
23. American Cancer Society. *Cancer Facts & Figures 2016.* Atlanta: American Cancer Society; 2016. Available at: http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf. Accessed 12/20/2016.
24. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *AJR. American journal of roentgenology.* 2014;203(3):687-693.
25. Duffy SW, Gabe R. What should the detection rates of cancers be in breast screening programmes? *British Journal of Cancer.* 2005;92(3):597-600.
26. Moon HJ, Kim EK. Characteristics of breast cancer detected by supplementary screening ultrasonography. *Ultrasonography (Seoul, Korea).* 2015;34(3):153-156.
27. Conant EF, Beaber EF, Sprague BL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: a cohort study within the PROSPR consortium. *Breast Cancer Res Treat.* 2016;156(1):109-116.
28. McDonald ES, Oustimov A, Weinstein SP, Synnestvedt MB, Schnall M, Conant EF. Effectiveness of Digital Breast Tomosynthesis Compared With Digital Mammography: Outcomes Analysis From 3 Years of Breast Cancer Screening. *JAMA Oncol.* 2016;2(6):737-743.

29. Wang WS, Hardesty L, Borgstede J, Takahashi J, Sams S. Breast Cancers Found with Digital Breast Tomosynthesis: A Comparison of Pathology and Histologic Grade. *Breast J.* 2016;22(6):651-656.
30. American Cancer Society. *Cancer Facts & Figures 2015; Special Section: Breast Carcinoma in Situ.* Atlanta: American Cancer Society; 2015. Available at: http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-046202.pdf. Accessed 12/20/2016.
31. Cianfrocca M, Goldstein LJ. Prognostic and predictive factors in early-stage breast cancer. *The oncologist.* 2004;9(6):606-616.
32. Lourenco AP, Barry-Brooks M, Baird GL, Tuttle A, Mainiero MB. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology.* 2015;274(2):337-342.
33. Destounis S, Arieno A, Morgan R. Initial experience with combination digital breast tomosynthesis plus full field digital mammography or full field digital mammography alone in the screening environment. *Journal of clinical imaging science.* 2014;4:9.

CHAPTER III: STUDY METHODS

3.1 Study Design

We will conduct a retrospective observational case-control study evaluating screening outcomes and tumor characteristics in women screened with DBT versus FFDM alone. The statistician(s) will be blinded to imaging modality during analyses.

3.2 Study Population and Sampling

We utilized convenience sampling from our existing electronic breast imaging database (PenRad; PenRad Technologies, Buffalo, MN) for all women presenting for screening mammography at a large academic center. From August 1, 2008 to August 1, 2016, a total of 44,050 women were screened with either full-field digital mammography (FFDM) alone or digital breast tomosynthesis (DBT) plus FFDM. From August 1, 2008 through July 31, 2011, all women presenting for screening underwent FFDM. In August 2011, our facility switched to DBT for screening. This was offered at no cost to all women, with the exception of those with implants and very large breasts so as not to exceed radiation dose standards. Women presenting for breast cancer screening from August 1, 2008 to July 31, 2011 (years 1-3; n=15,768) were screened with two-view craniocaudal (CC) and mediolateral (MLO) oblique FFDM alone. Women presenting for breast cancer screening from August 1, 2011 to August 1, 2016 (years 4-8; n=28,282) were imaged with combined DBT and FFDM per the current FDA-approved protocol, again consisting of standard CC and MLO of each breast.

The total number of women recalled from screening (total BI-RADS 0 cases) and recall rate (RR) have already been measured and reported, and will be discussed below. To detect the positive predictive value from recall (PPV_1) and the CDR, the population of

interest for this study includes the number of women recalled from screening (BI-RADS 0 cases) with DBT (years 4-8; n=2,194) versus FFDM alone (years 1-3; n=1,761).

3.2.1 Inclusion Criteria

For both cohorts, screening is defined as asymptomatic women (no clinical signs and/or symptoms of breast cancer). Thus, all asymptomatic women presenting for screening mammography are included in the total number screened. This also includes some women with a prior history of cancer and/or breast surgery.

3.2.2 Exclusion Criteria

Women screened with FFDM alone after the FFDM-to-DBT transition period have been excluded from analyses. As noted, women with very large breasts and/or implants were not offered screening with DBT so as not exceed radiation dose standards, and have been excluded from analyses. Diagnostic FFDM or DBT exams have also been excluded.

3.3 Subject Protection and Confidentiality

This will be an Institutional Review Board (IRB)-approved, Health Insurance Portability and Accountability Act (HIPAA)-compliant retrospective observational study evaluating data from screening examinations performed between July 2008 and August 2016. We have submitted a request for waiver of obtained written consent to the IRB (Appendix D). No identifying personal health information will be shared, as we are assessing data at a population level. All data will be kept on a private, password protected, encrypted server at YNHH.

3.4 Screening Interpretation and Data Collection

Screening mammograms were acquired with *Hologic Dimensions*^{TM} units. The imaging protocol for FFDM consisted of acquiring 2D images in the standard projections

for both breasts. The imaging protocol for DBT consisted of acquiring 2D images along with tomosynthesis images (dual acquisition DBT/FFDM) in the standard projections for both breasts. Thus, bilateral CC and MLO images were obtained for each exam. All screening mammograms were interpreted prospectively by the same group of eleven dedicated breast radiologists, the majority of whom are fellowship trained, with a median of fifteen years of breast imaging experience (range: 5-30 years). All radiologists were formally trained and certified in DBT interpretation.

The PenRad electronic breast-imaging database will be queried to quantify the total number of women screened throughout the study period from August 1, 2008 to July 31, 2011 (FFDM-only group) and August 1, 2011 to August 1, 2016 (DBT group). Population demographics for all women screened (age at screening, breast density, race/ethnicity, and presence of prior mammography) will be obtained from the PenRad database and EPIC, the electronic medical record system at Yale New-Haven Hospital (YNHH). The PenRad database will be used to extract all BI-RADS 0 cases.

For secondary analyses, the PenRad database will be specifically queried for true positive BI-RADS 0 cases over the study period. Cancer detection rate will be calculated as the number of true positive BI-RADS 0 cases per 1000 women screened. Tumor size and characteristics will be obtained from pathology reports in EPIC electronic medical records. Our data will be cross-linked with the Connecticut Tumor Registry to identify cancer cases that went to another institution, as permitted by our approved IRB protocol.[1]

3.5 Study Variables and Outcome Measures

We have retrospectively compared and reported the following outcomes from DBT versus FFDM alone for all asymptomatic women presenting for screening mammography

during 8 consecutive years: the total number of women screened (total exams); the total number of women recalled from screening (BI-RADS 0 cases); and the recall rate (RR) (%), calculated as the number of BI-RADS 0 examinations divided by the total number of screening examinations (Table 2).[1] Differences in the RR from screening with DBT versus FFDM alone were assessed using comparison of proportions and tested for significance with the Pearson chi-squared test. A *p* value of <0.05 was considered to be statistically significant. The number of true positive (biopsy-proven) BI-RADS 0 cases will determine the total cancers detected per modality, and the cancer detection rate (CDR) will be reported as the proportion of true positive BI-RADS 0 cases per 1000 women screened.

Table 2. Proposed Study: Screening Outcomes

	FFDM	**DBT**	***p*-value**
Total Screening Exams, *n*[1]	15,768	28,282	--
BI-RADS 0 Cases (Total Recalled), *n*[1]	1,761	2,194	--
Recall Rate (RR) (%)[1]	11.17	7.76	<0.001
Cancers Detected, *n*			--
CDR per 1000 screened			
Invasive, *n*			--
In situ, n			--
PPV_1 (%)			

DBT=Digital Breast Tomosynthesis; FFDM=Full-Field Digital Mammography; CDR=Cancer Detection Rate; PPV_1=Positive Predictive Value from Recall; RR=Recall Rate.

The primary focus of this study will be retrospective comparison of true positive BI-RADS 0 cases as detected with DBT versus FFDM alone for the assessment of PPV_1 and CDR, in addition to quantifying tumor size and characteristics.

3.5.1 Independent Variables

The main independent variable in this study is the modality of screening (DBT versus FFDM alone). Other independent variables, considered to be confounding factors, include breast composition (BI-RADS assessment of density); age at screening;

race/ethnicity; and number of prior screening rounds. Confounding factors will be controlled for by means of multivariate logistic regression analysis.

3.5.2 Dependent Variables

The main dependent variables (primary outcomes of interest) among the two imaging modalities include the PPV_1 and CDR. In true positive BI-RADS 0 cases, the main dependent variables include pathologic size and stage at diagnosis; histological type and grade; and axillary lymph node status and receptor phenotype of invasive cancers. Dependent variables that have already been measured (Table 2) include the total number of women screened, the total number of women recalled (BI-RADS 0 cases), and the recall rates (RR, %) for DBT versus FFDM alone.

3.5.3 Primary Outcomes

Our primary outcomes include identifying the number of true positive BI-RADS 0 cases, which will determine the PPV_1, per total number of BI-RADS 0 cases, and the cancer detection rate (CDR), per 1000 women screened.

3.5.4 Secondary Outcomes

Our secondary outcomes focus on the histologic characteristics of the true positive BI-RADS 0 cases detected with DBT versus FFDM alone, and will be derived from pathology reports in patient electronic medical records accessed through EPIC. Outcomes include tumor size in centimeters (cm); pathologic stage at diagnosis (TNM staging; see Figure 2, Appendix C); histological type (ductal, lobular, or other) and grade (low, intermediate, or high nuclear grade for *in situ* cancers; Nottingham histologic grade 1, 2, or 3 for invasive); axillary lymph node status and receptor phenotype (ER+/PR+ or luminal; HER2+; triple negative; see Figure 3, Appendix C) of invasive cancers.

3.6 Calculation of Power

Power was calculated with *PS*, version 3.1.2 (Appendix E).[2]

We will evaluate fixed sample sizes of 2,194 BI-RADS 0 cases recalled from DBT screening and 1,761 BI-RADS 0 cases recalled from FFDM screening. Prior data indicate that the positive predictive value from recall (PPV_1) among BI-RADS 0 patients recalled from FFDM is 4.4%.[3] Assuming a PPV_1 of 4.4% (0.044) for the FFDM group and an alpha of 0.05, if the PPV_1 for the DBT group is at least 6.45% (0.0645), we will be able to reject the null hypothesis that there is no difference in PPV_1 for DBT and FFDM alone with power of 0.805.

Our study is not powered to detect significant differences in cancer detection.

3.7 Analyses

3.7.1 Univariate and Bivariate Analyses

Baseline characteristics of asymptomatic women screened with DBT versus FFDM alone will be compared by means of the following statistical tests: age at screening (unpaired t-test if parametric, Mann Whitney U if non-parametric); BI-RADS breast density (category A, B, C, or D; Pearson chi-squared); race/ethnicity (Pearson chi-squared); and presence or absence of prior screening mammogram (Pearson chi-squared). All tests will be two-sided, with significance of $p<0.05$.

Primary Outcome: RR, CDR, and PPV_1 from DBT versus FFDM alone will be compared overall using a two-sided Pearson chi-squared test, with significance of $p<0.05$ (Table 2). These outcomes will additionally be stratified according to breast density, as defined by the BI-RADS breast composition categories (Table 3) and patient age (Table 4) at screening (<40 years; 40-49 years; 50-59 years; 60-69 years; ≥70 years).

Table 3. Proposed Study: Screening Outcomes by Breast Density

Breast Density Stratification												
	BI-RADS A			**BI-RADS B**			**BI-RADS C**			**BI-RADS D**		
	DM	DBT	*p*	DM	DBT	*p*	DM	DBT	*p*	DM	DBT	*p*
Total Exams, *n*			-			-			-			-
RR (%)												
Cancers, *n*			-			-			-			-
CDR per 1000												
Invasive, *n*			-			-			-			-
In situ, *n*			-			-			-			-
PPV_1 (%)												

DM=Digital Mammography, aka Full-Field DM (FFDM); DBT=Digital Breast Tomosynthesis; CDR=Cancer Detection Rate; PPV_1=Positive Predictive Value from Recall.

Table 4. Proposed Study: Screening Outcomes by Age

Age Stratification															
	<40 years			**40-49 years**			**50-59 years**			**60-69 years**			**≥70 years**		
	DM	DBT	*p*	DM	DBT	*p*	DM	DBT	*p*	DM	DBT	*p*	DM	DBT	*p*
Total Exams, *n*			-			-			-			-			-
RR (%)															
Cancers, *n*			-			-			-			-			-
CDR per 1000															
Invasive, *n*			-			-			-			-			-
In situ, *n*			-			-			-			-			-
PPV_1 (%)															

DM=Digital Mammography, aka Full-Field DM (FFDM); DBT=Digital Breast Tomosynthesis; CDR=Cancer Detection Rate; PPV_1=Positive Predictive Value from Recall.

Secondary Outcomes: Assuming parametric data, mean differences in tumor size (cm) with standard deviations will be compared using an unpaired t-test; if non-parametric, we will use the Mann Whitney U test and report tumor size (cm) as medians with interquartile ranges. The Pearson chi-squared test will be used to compare histologic types of cancer (ductal, lobular, other); Nottingham histologic grades; receptor phenotypes, percent of invasive cancers with axillary lymph node involvement; and pathologic stage at diagnosis. All tests will be two-sided with significance of $p<0.05$.

3.7.2 Multivariate Analyses

Multivariate logistic regression analysis and calculation of odds ratios (OR) with associated 95% confidence intervals (CI) will be utilized to determine the significance of the difference between the RR, CDR, and PPV_1 across FFDM and DBT groups after

controlling for possible confounding factors. *A priori*, we will adjust the logistic regression models for BI-RADS breast density category (A, B, C, or D); patient age at screening (<40; 40-49; 50-59; 60-69; ≥70); race (Caucasian, African American, Asian, Other/Unknown), ethnicity (Hispanic, non-Hispanic, Other/Unknown); and presence or absence of prior screening mammogram. We will apply a generalized linear mixed random-reader model to adjust for individual radiologist performance. For variables with more than two levels, we will apply the Bonferroni correction method to the calculation of OR and 95% CI.

In supplementary analyses (Appendix F), we will further adjust for the following additional known breast cancer risk factors in a random subset of our population: family history and/or personal history of breast cancer; *BRCA* mutation; BMI; reproductive factors; Jewish ancestry; and number of prior screens. We will obtain this information by assigning randomized, de-identified numbers to our cases and querying our databases.

3.8 Timeline and Resources

Due to the retrospective nature of this study, our data already exist and are accessible through the aforementioned databases. Personnel for this study will include one attending physician, one breast imaging fellow, one pathologist, one medical oncologist, and one physician assistant student for data retrieval, organization, and statistical analyses. Necessary resources for this study include access to the PenRad database; the Connecticut Tumor Registry; and EPIC.

References

1. Etesami M, Giwerc MY, Harigopal M, Zaheer Kidwai W, Philpotts LE. Cancer Detection with Digital Breast Tomosynthesis (DBT) Compared with Conventional Digital Mammography in Routine Breast Cancer Screening. *Presented at: Radiological Society of North America 2017 Scientific Assembly and Annual Meeting, November 26 - December 1, 2017, Chicago, IL.* rsna2017.rsna.org/program/. Accessed July 20, 2017.
2. Dupont WD PW. Power and Sample Size Calculations: A Review and Computer Program.Controlled Clinical Trials 1990; 11:116-28. Version 3.1.2.
3. McCarthy AM, Kontos D, Synnestvedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *Journal of the National Cancer Institute.* 2014;106(11).

CHAPTER IV: CONCLUSION

4.1 Advantages and Disadvantages

Our study will address a critical evidence gap in screening with DBT versus FFDM alone by characterizing the biologic characteristics of cancers detected with DBT versus FFDM alone, with complete follow-up data in women screened during the study period. We will evaluate screening outcomes and tumor characteristics in a diverse screening population from a large, urban academic center. The DBT transition period was fairly swift at our institution, minimizing possible selection bias. The participating eleven radiologists have been involved throughout the entirety of the study. Our retrospective study design is useful for studying rare outcomes, such as development of breast cancer, and several known confounding factors will be evaluated and controlled for. Furthermore, we will account for any group differences observed across modalities, with baseline characterization including age, breast density, race/ethnicity, and presence/absence of prior exams for all subjects. In a random subset of the population, we will further adjust for additional known breast cancer risk factors (Appendix F) to best account for underlying systemic population differences. Group differences will be accounted for in our multivariate regression analyses, enhancing our internal validity; minimizing potential selection bias; and allowing for evaluation of the direction and magnitude of confounding factors. Outcomes will be stratified by age, breast density, and baseline examination status in order to determine which subgroups, if any, derive preferential benefit from screening with DBT.

The major disadvantage from screening with dual acquisition DBT is the increased radiation dose, which is nearly doubled with adjunct DBT. While beyond the scope of

this study, several publications investigating the use of DBT with reconstructed synthetic 2D (s-2D) images, an imaging technique developed to address the burden of increased radiation dose exposure in women undergoing dual acquisition DBT, have observed non-inferior or superior performance metrics in screening with DBT plus s-2D as compared to the conventional DBT plus FFDM.[1-3] Our study is a natural, retrospective, observational study in a screening population that has converted from FFDM to DBT, with no randomization of patients to imaging modalities. Selection bias is a possibility and is difficult to control for, and could distort the exposure-outcome association, though DBT screening after the transition date has been offered to every woman, with the exception of women with very large breasts and/or implants so as not to exceed radiation dose standards. Since our population stems from a single institution, external validity (generalizability) is difficult to assess. Information bias is unlikely, since subjects are not self-reporting their outcomes; furthermore, if present, any information bias is likely non-differential misclassification, with similar bias across all participants, since variables were measured at the time of screening and before disease detection. Finally, while this study is not powered to detect significant differences in cancer detection, our outcomes will provide valuable insight into the characteristics of cancers detected with screening DBT versus FFDM alone.

4.2 Clinical Significance

This study will assess outcomes and tumor characteristics in women screened with DBT versus FFDM alone, providing insight into the biology of cancers detected across these modalities. Our results thus far are already in accordance with all DBT screening publications to date, with a significant reduction in recall rate.[4] To our knowledge, no

screening study to date has reported complete follow-up data regarding tumor characteristics detected with DBT versus FFDM alone, including pathologic size, grade, and stage at diagnosis; histologic type; axillary lymph node status and receptor phenotype of invasive cancers. If our hypotheses are correct and in agreement with findings discussed in our review of the literature (Chapter II) – namely, that DBT-detected cancers are found to be smaller in size,[5] primarily invasive and node-negative[5,6] – our results will have profound implications, suggesting that DBT leads to earlier detection of breast cancer than does FFDM alone, conceivably leading to less systemic treatment and improved clinical outcomes. Overall, these outcomes would validate evidence that screening with DBT reduces harms, as evidenced by decreases in recall rates with no increase in false negative cases. While higher cancer detection rates could raise potential for over-diagnosis, characterization of cancers as smaller, lower stage, and with primarily negative axillary lymph node spread suggests there would be a downstream benefit from DBT screening, theoretically reducing disease morbidity and mortality.

In future studies, it will be important to evaluate and compare the number of additional mammographic views required for cancer diagnosis for true positive cases recalled from DBT versus FFDM[7], in addition to the radiologic features (mammographic abnormalities) leading to recall across modalities.[8,9] While beyond the limits of the maximum allotted time period of two years, a large, prospective study would allow for causal inference, strengthening existing DBT screening data.

References

1. Houssami N, Bernardi D, Pellegrini M, et al. Breast cancer detection using single-reading of breast tomosynthesis (3D-mammography) compared to double-reading of 2D-mammography: Evidence from a population-based trial. *Cancer epidemiology.* 2017;47:94-99.
2. Aujero MP, Gavenonis SC, Benjamin R, Zhang Z, Holt JS. Clinical Performance of Synthesized Two-dimensional Mammography Combined with Tomosynthesis in a Large Screening Population. *Radiology.* 2017;283(1):70-76.
3. Zuckerman SP, Conant EF, Keller BM, et al. Implementation of Synthesized Two-dimensional Mammography in a Population-based Digital Breast Tomosynthesis Screening Program. *Radiology.* 2016;281(3):730-736.
4. Etesami M, Giwerc MY, Harigopal M, Zaheer Kidwai W, Philpotts LE. Cancer Detection with Digital Breast Tomosynthesis (DBT) Compared with Conventional Digital Mammography in Routine Breast Cancer Screening. *Presented at: Radiological Society of North America 2017 Scientific Assembly and Annual Meeting, November 26 - December 1, 2017, Chicago, IL.* rsna2017.rsna.org/program/. Accessed July 20, 2017.
5. Wang WS, Hardesty L, Borgstede J, Takahashi J, Sams S. Breast Cancers Found with Digital Breast Tomosynthesis: A Comparison of Pathology and Histologic Grade. *Breast J.* 2016;22(6):651-656.
6. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology.* 2013;267(1):47-56.
7. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *AJR. American journal of roentgenology.* 2014;203(3):687-693.
8. Lourenco AP, Barry-Brooks M, Baird GL, Tuttle A, Mainiero MB. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology.* 2015;274(2):337-342.
9. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology.* 2015;274(1):85-92.

APPENDIX A: United States Studies (Chapter II), Tables

Table 5. US Study Characteristics

Study: Author/Publication Year/Location/Study Period	Study *N* (Mean Age)	Radiologists/Reading Method, Views
Sharpe et al., 2016 Prospective with retrospective cohort; Single-site (Boston, MA) Two time periods: FFDM (retrospective), DBT (prospective) January 2011 - March 2014	**DBT:** 5,703 (55.68 years) **FFDM:** 80,149 (57.62 years) **Total:** 85,852	10 radiologists Average experience: 15.6 years Single-reading; Two view
Conant et al., 2016 Retrospective; Multisite (PA, VT, MA) One time period: FFDM vs. DBT in sites transitioning to DBT January 2011 - January 2015	**DBT:** 55,998 (avg. n/a) **FFDM:** 142,883 (avg. n/a) **Total:** 198,881	47 radiologists All had interpreted at least 50 DBT and FFDM screening exams Single-reading; Two-view
McDonald et al., 2016 Retrospective; Single-site (Philadelphia, PA) Four time periods: FFDM-0 (Reference), DBT-1, DBT-2, DBT-3 September 2010 - October 2014	**DBT-3:** 11,576 (56.7 years) **DBT-2:** 11,157 (56.9 years) **DBT-1:** 11,007 (56.7 years) **FFDM-0:** 10,728 (56.9 years) **Total:** 44,468	7 radiologists Experience range: 8-26 years Single-reading; Two-view
Lourenco et al., 2015 Retrospective; Single-site (Providence, RI) Two time periods: FFDM followed by DBT March 2011 - March 2013	**DBT:** 12,921 (55.3 years) **FFDM:** 12,577 (54.6 years) **Total:** 25,498	6 radiologists Experience range: 4-16 years Single-reading; Two-view
Sumkin et al., 2015 Prospective; Single-site (Pittsburgh, PA) One time period: FFDM vs. DBT (paired, independent reads) May 2010 - September 2014	**DBT:** 1,074 (42 years) **FFDM:** 1,074 (42 years) **Total:** 1,074	14 radiologists Experience range: 5-28 years (FFDM); 3-9 years (DBT) Single-reading, independent: Two-view
McDonald et al., 2015 Retrospective; Single-site (Philadelphia, PA) Two time periods: FFDM followed by DBT for two groups September 2010 - March 2013	**DBT:** 15,571 (see subgroups) **FFDM:** 10,728 (see subgroups) **Total:** 26,299[§]	6 radiologists Experience range: 3-22 years Single-reading; Two-view
	<u>Baseline Subgroup</u>: **DBT:** 1,895 (48.9 years) **FFDM:** 1,204 (49.2 years) **Total:** 3,063	**<u>Previously Screened</u>:** **DBT:** 13,712 (57.8 years) **FFDM:** 9,524 (57.9 years) **Total:** 23,236

[§]*Same population as McCarthy et al. (2014)*

Table 5 (continues)

Study: Author/Publication Year/Location/Study Period	Study *N* (Mean Age)	Radiologists/Reading Method, Views
McCarthy et al., 2014 Retrospective; Single-site (Philadelphia, PA) Two time periods: FFDM followed by DBT September 2010 - March 2013	**DBT:** 15,571 (56.7 years) **FFDM:** 10,728 (56.9 years) **Total:** 26,299[§]	6 radiologists Experience range: 3-22 years Single-reading; Two-view [§]*Same population as McDonald et al. (2015)*
Greenberg et al., 2014 Retrospective; Multisite: 6 sites total (MD, VA, DC) One time period: FFDM vs. DBT August 2011 - December 2012	**DBT:** 20,943 (59.6 years) **FFDM:** 38,674 (59.5 years) **Total:** 59,617	14 radiologists; interpreted >500 DBT images Experience range: 2-38 years Single-reading; Two-view
Friedewald et al., 2014 Retrospective; Multisite (13): 5 academic, 8 non-academic Two time periods: FFDM followed by DBT March 2010 - January 2013	**DBT:** 173,663 (56.2 years) **FFDM:** 281,187 (57 years) **Total:** 454,850	139 radiologists at 13 different sites Single-reading; Two-view *Note: all analyses adjust for site as a random effect*
Destounis et al., 2014 Retrospective; Single-site (Rochester, NY) One time period: FFDM (age-matched) vs. DBT June 2011 - December 2011	**DBT:** 524 (59 years) **FFDM:** 524 (59 years) **Total:** 1,048	6 radiologists Experience range: 1-35 years Double-reading; Two-view
Durand et al., 2015 Retrospective; Academic center: 4 sites (New Haven, CT) One time period: FFDM vs. DBT August 2011 - January 2013	**DBT:** 8,591 (avg. n/a) **FFDM:** 9,364 (avg. n/a) **Total:** 17,955*	7 radiologists Experience range: 2-25 years Single-reading; Two-view **13,158 subjects reported in Haas et al. 2013*
Haas et al., 2013 Retrospective; Academic center: 4 sites (New Haven, CT) One time period: FFDM vs. DBT October 2011 - October 2012	**DBT:** 6,100 (55.8 years) **FFDM:** 7,058 (57.5 years) **Total:** 13,158*	8 radiologists Experience range: 2-23 years Single-reading; Two-view
Rose et al., 2014 Prospective w/paired retrospective cohort; Single-site (Houston, TX) One time period: FFDM (retrospective) vs. DBT (prospective); paired May 2011 to February 2012	**DBT:** 10,878 (avg. n/a) **FFDM:** 10,878 (avg. n/a) **Total:** 10,878	10 radiologists prospectively read DBT+FFDM 7 radiologists retrospectively read corresponding FFDM exams alone Experience range: 2-32 years Single-reading; Two-view
Rose et al., 2013 Retrospective; Single-site (Houston, TX) Two time periods: FFDM followed by DBT 2010 - February 2012	**DBT:** 9,499 (54.5 years) **FFDM:** 13,856 (53.8 years) **Total:** 23,355	6 radiologists; interpreted >500 images before & after DBT transition Experience range: 2-32 years Single-reading; Two-view

Table 5

Table 6. Screening Outcomes: US Studies

Study	Recall Rate	Cancers Detected	CDR per 1000	Invasive CDR/1000	DCIS CDR/1000	PPV_1
Sharpe et al., 2016 **DBT:** 5,703 **FFDM:** 80,149 **Total:** 85,852	**DBT:** 6.1% **FFDM:** 7.51% Relative reduction: 18.8% ***p<0.0001** **Unadjusted OR:** 0.80 (0.72-0.90) ***p<0.0001** **Adjusted OR:** 0.98 (0.84-1.13) p=0.7458 *N.S.*	**DBT:** 31 Invasive: 16 (51.5%) DCIS: 15 (48.5%) **FFDM:** 280 Invasive: 197 (70.4%) DCIS: 83 (29.6%)	**DBT:** 5.4 **FFDM:** 3.5 Relative increase: 54.3% ***p<0.0018**	**DBT:** 2.81 **FFDM:** 2.46 Relative increase: 12.6% p=0.61 *N.S.*	**DBT:** 2.63 **FFDM:** 1.04 Relative increase: 153% ***p<0.0006**	n/a
Conant et al., 2016 **DBT:** 55,998 **FFDM:** 142,883 **Total:** 198,881 exams from 103,401 women	**DBT:** 8.7% **FFDM:** 10.4% Relative reduction: 16.3% ***p<0.0001** **Adjusted OR:** 0.68 (0.65-0.71) **False Negative Rate per 1000:** **DBT:** 0.60 **FFDM:** 0.46 p=0.347 *N.S.*	**DBT:** 149 Invasive: 106 (71.1%) DCIS: 43 (28.9%) **FFDM:** 499 Invasive: 378 (75.8%) DCIS: 121 (24.2%)	**DBT:** 5.9 **FFDM:** 4.4 Relative increase: 34.1% ***p=0.0026** **Adjusted OR:** 1.45 (1.12-1.88)	**DBT:** 4.2 **FFDM:** 3.3 Relative increase: 27.3% ***p=0.045** **Adjusted OR:** 1.38 (1.02-1.87)	**DBT:** 0.77 **FFDM:** 0.85 *N.S.* **Sensitivity:** **DBT:** 90.9% **FFDM:** 90.6% p=1.00 *N.S.*	**DBT:** 6.4% **FFDM:** 4.1% Relative increase: 56.1% ***p<0.0001** **Adjusted OR:** 2.02 (1.54-2.65) **Specificity:** **DBT:** 91.3% **FFDM:** 89.7% ***p<0.0001** **Adjusted OR:** 1.39 (1.30-1.48)
McDonald et al., 2016 **DBT-3:** 11,576 **DBT-2:** 11,157 **DBT-1:** 11,007 **FFDM-0:** 10,728 **Total:** 44,468 exams from 23,958 women	**DBT-3:** 9.2% OR: 0.87 (0.80-0.95) ***p=0.002** **DBT-2:** 9.0% OR: 0.85 (0.78-0.93) ***p<0.001** **DBT-1:** 8.8% OR: 0.83 (0.76-0.91) ***p<0.001** **FFDM-0 (Ref):** 10.4% *N.S. across DBT years* (p=0.55)	n/a	**DBT-3:** 6.1 OR: 1.35 (0.93-1.94); p=0.11 **DBT-2:** 5.8 OR: 1.28 (0.88-1.85); p=0.20 **DBT-1:** 5.5 OR: 1.35 (0.93-1.94); p=0.37 **FFDM-0 (Ref):** 4.6 *N.S. vs. FFDM or across DBT*	**DBT-3:** 4.1 **DBT-2:** 4.1 **DBT-1:** 3.8 **FFDM-0 (Ref):** 3.2 *N.S. vs. FFDM or across DBT*	**DBT-3:** 1.8 **DBT-2:** 1.3 **DBT-1:** 1.5 **FFDM-0 (Ref):** 1.4 *N.S. vs. FFDM or across DBT years*	**DBT-3:** 6.7% OR: 1.56 (1.07-2.26); ***p=0.02** **DBT-2:** 6.5% OR: 1.51 (1.03-2.21); ***p=0.03** **DBT-1:** 6.2% OR: 1.44 (0.98-2.12); p=0.06 *N.S.* **FFDM-0 (Ref):** 4.4% *N.S. across DBT years*
Sumkin et al., 2015 **DBT:** 1,074 **FFDM:** 1,074 **Total:** 1,074 (paired)	**DBT:** 25.5% **FFDM:** 38.4% Relative reduction: 33.6% ***p<0.001**	6 total **DBT only:** 1 **FFDM only:** 1 Both: 4	n/a	n/a	n/a	n/a
Lourenco et al., 2015 **DBT:** 12,921 **FFDM:** 12,577 **Total:** 25,498	**DBT:** 6.4% **FFDM:** 9.3% Relative reduction: 31.2% ***p<0.00001**	**DBT:** 60 Invasive: 30 (50%) DCIS: 21 (35%) **FFDM:** 68 Invasive: 41 (60.3%) DCIS: 21 (30.9%)	**DBT:** 4.6 **FFDM:** 5.4 Relative decrease: 14.8% p=0.44 *N.S.*	**DBT:** 2.32 **FFDM:** 3.26 Relative decrease: 28.8% *N.S.*	**DBT:** 1.63 **FFDM:** 1.67 *N.S.*	**DBT:** 6.2% **FFDM:** 5.2% Relative increase: 19.2% p=0.219 *N.S.* **Specificity:** **DBT:** 94% | **FFDM:** 91.1% *N.S.*

Unadjusted/Adjusted ORs: *#s in parentheses denote 95% CI intervals.* ***Acronyms/Abbreviations:*** *Breast Imaging-Reporting and Data System (BI-RADS); Cancer Detection Rate (CDR); Confidence Interval (CI); Ductal Carcinoma in situ (DCIS); Digital Breast Tomosynthesis (DBT); Full-Field Digital Mammography (FFDM); Odds Ratio (OR); Positive Predictive Value (PPV_1) from Recall; Reference (Ref)*

** = Significant; N.S. = Not significant*

Table 6 (continues)

Study	Recall Rate	Cancers Detected	CDR per 1000	Invasive CDR/1000	DCIS CDR/1000	PPV_1
McDonald et al., 2015 **Baseline subgroup:** **DBT:** 1,895 **FFDM:** 1,204 **Total:** 3,063 **Previously screened:** **DBT:** 13,712 **FFDM:** 9,524 **Total:** 23,236 **Overall:** **DBT:** 15,571 **FFDM:** 10,728 **Total:** 26,299	**Baseline subgroup:** **DBT:** 16.0% **FFDM:** 20.5% Relative reduction: 22% ***p=0.002** **Adjusted OR:** 0.74 (0.61-0.89) ***p=0.002** **Previously screened:** **DBT:** 7.8% **FFDM:** 9.1% Relative reduction: 14.3% ***p<0.001** **Adjusted OR:** 0.84 (0.76-0.92) ***p<0.001**	**Baseline subgroup:** **DBT:** 11 **FFDM:** 5 **Previously screened:** **DBT:** 74 **FFDM:** 44	**Baseline subgroup:** **DBT:** 5.9 **FFDM:** 4.2 Relative increase: 40.5% p=0.51 *N.S.* **Previously screened:** **DBT:** 5.4 **FFDM:** 4.6 Relative increase: 17.4% p=0.41 *N.S.*	n/a n/a	n/a n/a	**Baseline subgroup:** **DBT:** 3.7% **FFDM:** 2.0% p=0.25 *N.S.* **Adjusted OR:** 1.81 (0.61-5.43) p=0.29 *N.S.* **Previously screened:** **DBT:** 6.9% **FFDM:** 5.1% Relative increase: 35.3% p=0.09 *Trending, N.S.* **Adjusted OR:** 1.57 (1.05-2.33) ***p=0.03**
McCarthy et al., 2014 **DBT:** 15,571 **FFDM:** 10,728 **Total:** 26,299	**DBT:** 8.8% **FFDM:** 10.4% Relative reduction: 15.4% ***p<0.001** **Unadjusted OR:** 0.82 (0.75-0.89) ***p<0.001** **Adjusted OR:** 0.80 (0.74-0.88) ***p<0.001**	**DBT:** 85 Invasive: 60 (71%) DCIS: 23 (27%) **FFDM:** 49 Invasive: 34 (69%) DCIS: 15 (32%)	**DBT:** 5.5 **FFDM:** 4.6 Relative increase: 19.6% p=0.32 *N.S.* **Under age 50:** **DBT:** 5.7 **FFDM:** 2.2 Relative increase: 159% ***p=0.02**	**DBT:** 3.9 **FFDM:** 3.2 Relative increase: 21.9% p=0.36 *N.S.*	**DBT:** 1.5 **FFDM:** 1.4 p=0.87 *N.S.*	**DBT:** 6.2% **FFDM:** 4.4% Relative increase: 41% ***p=0.047** **Under age 50:** **DBT:** 7.4% **FFDM:** 6.3% Relative increase: 17.5% ***p=0.007**
Destounis et al., 2014 **DBT:** 524 **FFDM:** 524 **Total:** 1,048	**DBT:** 4.20% **FFDM:** 11.45% Relative reduction: 63.3% ***p<0.0001**	**DBT:** 3 Invasive: 1 (33%) DCIS: 2 (67%) **FFDM:** 2 Invasive: 1 (50%) DCIS: 1 (50%)	**DBT:** 5.7 **FFDM:** 3.8 Relative increase: 50% p=0.15 *N.S.*	**DBT:** 1.90 **FFDM:** 1.90 Relative increase: 0% *N.S.*	**DBT:** 3.82 **FFDM:** 1.90 *N.S.*	**DBT:** 13.6% **FFDM:** 3.33% Relative increase: 308% *N.S.*
Greenberg et al., 2014 **DBT:** 20,943 **FFDM:** 38,674 **Total:** 59,617	**DBT:** 13.6% **FFDM:** 16.2% Relative reduction: 16% ***p<0.0001** **Generalized Linear Mixed Model (GLMM):** 13.60% reduction ***p<0.0001**	**DBT:** 144 Invasive: 106 (73.6%) DCIS: 37 (25.7%) Other: 1 (0.7%) **FFDM:** 203 Invasive: 126 (62.1%) DCIS: 75 (36.9%) Other: 2 (1.0%)	**DBT:** 6.3 **FFDM:** 4.9 Relative increase: 28.6% ***p=0.035** **GLMM:** 27.11% increase ***p=0.0409**	**DBT:** 4.6 **FFDM:** 3.2 Relative increase: 43.8% ***p=0.006**	**DBT:** 1.6 **FFDM:** 1.7 p=0.753 *N.S.*	**DBT:** 4.6% **FFDM:** 3.0% Relative increase: 53.3% ***p=0.0003** **GLMM:** 51.14% increase ***p=0.0007**

Unadjusted/Adjusted ORs: *#s in parentheses denote 95% CI intervals.* ***Acronyms/Abbreviations:*** *Breast Imaging-Reporting and Data System (BI-RADS); Cancer Detection Rate (CDR); Confidence Interval (CI); Ductal Carcinoma in situ (DCIS); Digital Breast Tomosynthesis (DBT); Full-Field Digital Mammography (FFDM); Odds Ratio (OR); Positive Predictive Value (PPV_1) from Recall; Reference (Ref)*

** = Significant; N.S. = Not significant*

Table 6 (continues)

Study	Recall Rate	Cancers Detected	CDR per 1000	Invasive CDR/1000	DCIS CDR/1000	PPV_1
Friedewald et al., 2014 **DBT:** 173,663 **FFDM:** 281,187 **Total:** 454,850	**DBT:** 9.1% **FFDM:** 10.7% Relative reduction: 15% ***p<0.001**	**DBT:** 950 Invasive: 707 (74.4%) DCIS: 243 (25.6%) **FFDM:** 1207 Invasive: 815 (67.5%) DCIS: 392 (32.5%)	**DBT:** 5.4 **FFDM:** 4.2 Relative increase: 28.6% ***p<0.001**	**DBT:** 4.1 **FFDM:** 2.9 Relative increase: 41.4% ***p<0.001** **Invasive Ductal:** **DBT:** 3.27 **FFDM:** 2.46 Relative increase: 33% ***p<0.001** **Invasive Lobular:** **DBT:** 0.55 **FFDM:** 0.27 Relative increase: 103.7% ***p<0.001**	**DBT:** 1.4 **FFDM:** 1.4 p=0.95 *N.S.*	**DBT:** 6.4% **FFDM:** 4.3% Relative increase: 48.8% ***p<0.001**
Durand et al., 2015 **DBT:** 8,591 **FFDM:** 9,364 **Total:** 17,955[£] [£]13,158 from Haas 2013	**DBT:** 7.8% **FFDM:** 12.3% Relative reduction: 36.6% ***p<0.0001** **Adjusted OR:** 1.76 (1.58-1.96) ***p<0.0001** (1.76x higher FFDM)	**DBT:** 51 Invasive: 35 (68.6%) DCIS: 16 (31.4%) **FFDM:** 54 Invasive: 35 (64.8%) DCIS: 19 (35.2%)	**DBT:** 5.9 **FFDM:** 5.7 Relative increase: 3.5% p=0.88 *N.S.*	**DBT:** 4.07 **FFDM:** 3.74 Relative increase: 8.8% p=0.26 *N.S.*	**DBT:** 1.9 **FFDM:** 2.0 p=0.63 *N.S.*	n/a
Haas et al., 2013 **DBT:** 6,100 **FFDM:** 7,058 **Total:** 13,158[£]	**DBT:** 8.4% **FFDM:** 12% Relative reduction: 30% ***p<0.01** **Adjusted OR:** 0.62 (0.55-0.70) ***p<0.0001**	**DBT:** 35 Invasive: 24 (69%) DCIS: 11 (31%) **FFDM:** 37 Invasive: 25 (68%) DCIS: 12 (32%)	**DBT:** 5.7 **FFDM:** 5.2 Relative increase: 9.6% p=0.70 *N.S.*	**DBT:** 3.9 **FFDM:** 3.5 Relative increase: 11.4% p=0.93 *N.S.*	**DBT:** 1.8 **FFDM:** 1.7 *N.S.*	n/a
Rose et al., 2014 **DBT:** 10,878 **FFDM:** 10,878 **Total:** 10,878	**DBT:** 5.41% **FFDM:** 8.16% Relative reduction: 33.7% ***p<0.0001**	**DBT:** 59 Invasive: 48 (83.4%) DCIS: 11 (18.6%) **FFDM:** 39 Invasive: 29 (74.4%) DCIS: 10 (25.6%)	**DBT:** 5.4 **FFDM:** 3.5 Relative increase: 54.3% ***p<0.0001**	**DBT:** 4.4 **FFDM:** 2.7 Relative increase: 63% ***p<0.0001**	**DBT:** 0.92 **FFDM:** 1.02 *N.S.*	n/a
Rose et al., 2013 **DBT:** 9,499 **FFDM:** 13,856 **Total:** 23,355	**DBT:** 5.5% **FFDM:** 8.7% Relative reduction: 36.7% ***p<0.001**	**DBT:** 51 Invasive: 41 (80.4%) DCIS: 10 (19.6%) **FFDM:** 56 Invasive: 39 (69.6%) DCIS: 17 (30.4%)	**DBT:** 5.37 **FFDM:** 4.04 Relative increase: 32.9% p=0.18 *N.S.*	**DBT:** 4.32 **FFDM:** 2.81 Relative increase: 53.7% p=0.07 *Trending, N.S.*	**DBT:** 1.05 **FFDM:** 1.22 *N.S.*	**DBT:** 10.1% **FFDM:** 4.7% Relative increase: 115% ***p<0.001**

Unadjusted/Adjusted ORs: *#s in parentheses denote 95% CI intervals.* ***Acronyms/Abbreviations:*** *Breast Imaging-Reporting and Data System (BI-RADS); Cancer Detection Rate (CDR); Confidence Interval (CI); Ductal Carcinoma in situ (DCIS); Digital Breast Tomosynthesis (DBT); Full-Field Digital Mammography (FFDM); Odds Ratio (OR); Positive Predictive Value (PPV_1) from Recall; Reference (Ref)*

** = Significant; N.S. = Not significant*

Table 6

APPENDIX B: Tumor Characteristics (Chapter II), Tables

Table 7. Tumor Characteristics: European Studies

Study	Cancers Detected	CDR/1000			TNM/Tumor Size (cm)	Invasive Cancer Types	Lymph Node Status	Cancer Grade	Receptor Phenotype
STORM **DBT:** 7,294 **FFDM:** 7,294 **Total:** 7,294 (paired) *Tumor data is from double-reading.*	***Double-Read*:** **DBT total:** 59 Invasive: 52 (88%) DCIS: 7 (12%) **FFDM & DBT:** 39 Invasive: 35 (90%) DCIS: 4 (10%) ***Single-Read*:** **DBT total:** 55 **FFDM & DBT:** 35 **CDR/1000:** **DBT:** 7.5 **FFDM:** 4.8 ↑ 56.25% **Incremental CDR:** 2.7 ***p<0.001**	**Overall** ***Double-Read*:** **DBT:** 8.1 **FFDM:** 5.3 ↑ 52.8% ***p<0.0001** **Incremental CDR:** 2.7 ***p<0.0001**	**Invasive** ***Double-Read*:** **DBT:** 7.1 **FFDM:** 4.8 ↑ 47.9% ***p<0.0001**	***In situ*** ***Double-Read*:** **DBT:** 0.96 **FFDM:** 0.55 *N.S.*	***Invasive only*:** **DBT (FFDM occult):** **T1a:** 0 **T1b:** 8 (40%) **T1c:** 8 (40%) **T2:** 1 (5%) **Mean:** 1.35 **SD:** 0.67[%] **FFDM & DBT:** **T1a:** 3 (8%) **T1b:** 10 (26%) **T1c:** 20 (51%) **T2:** 2 (5%) **Mean:** 1.37 **SD:** 0.58 (*TNM classification*)	**IDC:** 37 (71%) **ILC:** 4 (7.7%) **Other:** 7 (14.5%) **Mixed:** 4 (7.7%) ***# detected with*:** **FFDM alone:** 0 **FFDM & DBT:** 29 **FFDM+DBT:** 13 **(FFDM occult)**	**DBT (FFDM occult):** **Negative:** 11 (55%) **Positive:** 3 (15%) **Micromets or isolated tumor cells:** 1 (5%) **No surgery, n/a:** 5 (25%) **FFDM & DBT:** **Negative:** 24 (62%) **Positive:** 9 (23%) **Micromets or isolated tumor cells:** 2 (5%) **No surgery, n/a:** 4 (10%)	***Invasive only*:** **DBT Total:** **Grade 1:** 6 (35%) **Grade 1-2:** 1 (6%) **Grade 2:** 7 (41%) **Grade 3:** 2 (12%) **n/a:** 1 (6%) **FFDM & DBT:** **Grade 1:** 10 (29%) **Grade 1-2:** 2 (6%) **Grade 2:** 16 (46%) **Grade 3:** 6 (17%) **n/a:** 1 (3%)	**DBT (FFDM occult):** **ER+/PR+:** 11 (92%) **HER2+:** 1 (8%) **Triple Neg:** 0 **FFDM & DBT:** **ER+/PR+:** 25 (81%) **HER2+:** 4 (13%) **Triple Neg:** 2 (6%)
OTST ***Single-Read*** **DBT:** 12,621 **FFDM:** 12,621 **Total:** 12, 621 (paired) ***DBT Total:*** *all cases detected w/DBT+FFDM & DBT alone;* ***FFDM Total:*** *all cases detected w/ DBT+ FFDM & FFDM alone.*	121 cancers 14 missed by both **DBT only:** 30 **FFDM only:** 6 **DBT+FFDM:** 71 **DBT Total:** 101 Invasive: 81 (80%) DCIS: 20 (20%) **FFDM Total:** 77 Invasive: 56 (72%) DCIS: 21 (28%) **Difference:** 24	**Overall** **DBT:** 8.0 **FFDM:** 6.1 ↑31.1% ***p=0.001** *Reader-adjusted FFDM:DBT ratio:* 1.27 (↑27%) (1.06-1.53)	**Invasive** **DBT:** 6.4 **FFDM:** 4.4 ↑45.5% ***p<0.001** *Reader-adjusted FFDM:DBT ratio:* 1.40 (↑40%) (1.13-1.71)	***In situ*** **DBT:** 1.58 **FFDM:** 1.66 *N.S.* ***DCIS Size*:** **DBT Total:** **Mean:** 2.53 **Median:** 1.85 **Range:** 0.5-8.5 **(FFDM occult)** **FFDM Total:** **Mean:** 2.10 **Median:** 1.5 **Range:** 0.5-5.0	***Invasive only*:** **DBT Total:** **Mean:** 1.32 **Median:** 1.3 **Range:** 0.1-5.0 **DBT only:** **Mean:** 1.28 **Median:** 1.3 **Range:** 0.5-5.0 **(FFDM occult)** **FFDM Total:** **Mean:** 1.32 **Median:** 1.1 **Range:** 0.1-2.7	**DBT Total:** 81 **IDC:** 49 (60.5%) **ILC:** 13 (16%) **IDC+DCIS:** 16 (20%) **Other:** 3 (3.7%) **DBT only:** 29 **IDC:** 16 (55%) **ILC:** 7 (25%) **IDC+DCIS:** 5 (17%) **Other:** 1 (3%) **(FFDM occult)** **FFDM Total:** 56 **IDC:** 35 (62.5%) **ILC:** 8 (14.3%) **IDC+DCIS:** 11 (20%) **Other:** 2 (3.6%)	**DBT Total:** 81 **Negative:** 63 (77.8%) **Positive:** 13 (16%) **n/a:** 5 (6.2%) **DBT only:** 29 **Negative:** 23 (79.3%) **Positive:** 4 (13.8%) **n/a:** 2 (6.9%) **(FFDM occult)** **FFDM Total:** 56 **Negative:** 44 (78.6%) **Positive:** 9 (16%) **n/a:** 3 (5.4%)	***Invasive*:** → **DBT Total:** 81 **Grade 1:** 32 (39.5%) **Grade 2:** 35 (43.2%) **Grade 3:** 13 (16%) **n/a:** 1 (1.2%) **DBT only:** 29 **Grade 1:** 14 (48%) **Grade 2:** 11 (39%) **Grade 3:** 3 (10%) **n/a:** 1 (3%) **(FFDM occult)** **FFDM Total:** 56 **Grade 1:** 17 (30.4%) **Grade 2:** 29 (51.8%) **Grade 3:** 9 (16%) **n/a:** 1 (1.8%)	***DCIS* Grade:** **DBT Total:** 20 **Low/Med:** 4 (20%) **High:** 16 (80%) **FFDM Total:** 21 **Low/Med:** 4 (19%) **High:** 17 (81%)

TNM Classification: See Figure 2 **| Receptor Phenotypes:** See Figure 3 **|Acronyms/Abbreviations:** *Cancer Detection Rate (CDR); Ductal Carcinoma In Situ (DCIS); Digital Breast Tomosynthesis (DBT); Centimeters (cm); Full-Field Digital Mammography (FFDM); Medium (Med); Invasive Ductal Carcinoma (IDC); Invasive Lobular Carcinoma (ILC); Not available (n/a); Standard Deviation (SD * = * = Significant; N.S. = Not significant*

Table 7

Table 8. Tumor Characteristics: US Studies

Study	Cancers Detected	CDR/1000			Tumor Size (cm)	Invasive Cancer Types	LN Status (TNM)	Invasive Cancer Grade	DCIS Grade
		Overall	**Invasive**	***In situ***					
Wang et al., 2016 **FFDM + DBT:** 12,444 FFDM → DBT **Total:** 12,444 (paired)	**DBT total:** 65 Invasive: 41 (63.1%) DCIS: 24 (36.9%) **FFDM visible:** 55 Invasive: 32 (58.2%) DCIS: 23 (41.8%) *10 of 65 cases detected were FFDM occult (seen only with DBT); 9 of 10 FFDM occult cancers were invasive (90%)*	**DBT:** 5.2 **FFDM:** 4.4 ↑18.6% *N.S.*	**DBT:** 3.3 **FFDM:** 2.6 ↑26.9% *N.S.*	**DBT:** 1.93 **FFDM:** 1.85 *23/24 DCIS cases (96%) found with FFDM alone* p=0.08 *Trending, N.S.*	***Invasive only*:** **DBT (FFDM occult):** 10 **Mean:** 0.78 **Median:** 0.6 **Range:** 0.3-1.3 **FFDM visible:** 32 **Mean:** 1.23 **Median:** 1.0 **Range:** 0.4-4.3 p=0.07 *Trending, N.S.*	**DBT (FFDM occult):** 9 **IDC:** 6 (60%) **ILC:** 1 (10%) **Tubular:** 2 (20%) **FFDM visible:** 32 **IDC:** 21 (65.6%) **ILC:** 4 (12.5%) **ILC+IDC:** 6 (18.8%) **Other:** 1 (3.1%) *N.S.*	**DBT only:** 9 **N0:** 9 (100%) **≥N1:** 0 **FFDM visible:** 32 **N0:** 28 (87.5%) **≥N1:** 4 (12.5%) p=0.6 *N.S.*	**DBT (FFDM occult):** 9 **Grade 1:** 7 (78%) **Grade 2:** 0 **Grade 3:** 2 (22%) **FFDM visible:** 32 **Grade 1:** 15 (47%) **Grade 2+3**: 17 (53%) ***p=0.02**	n/a
Greenberg et al., 2014 **DBT:** 20,943 **FFDM:** 38,674 **Total:** 59,617	**DBT:** 144 Invasive: 106 (73.6%) DCIS: 37 (25.7%) Other: 1 (0.7%) **FFDM:** 203 Invasive: 126 (62.1%) DCIS: 75 (36.9%) Other: 2 (1.0%)	**Overall** **DBT:** 6.3 **FFDM:** 4.9 ↑28.6% ***p=0.035**	**Invasive** **DBT:** 4.6 **FFDM:** 3.2 ↑43.8 ***p=0.006**	***In situ*** **DBT:** 1.6 **FFDM:** 1.7 p=0.753 *N.S.*	n/a	**DBT:** 106 **IDC:** 82 (77.4%) **ILC:** 18 (17%) **Other:** 6 (5.7%) **FFDM:** 126 **IDC:** 107 (84.9%) **ILC:** 16 (12.7%) **Other:** 3 (2.4%) p=0.255 *N.S.*	n/a	**DBT:** 106 **T1mic:** 2 (1.9%) **Grade 1:** 24 (22.6%) **Grade 2:** 61 (57.5%) **Grade 3:** 19 (17.9%) **FFDM:** 203 **T1mic:** 4 (3.2%) **Grade 1:** 20 (15.9%) **Grade 2:** 78 (61.9%) **Grade 3:** 22 (17.5%) **Unknown:** 2 (1.6%) p=0.607 *N.S.*	**DBT:** 37 **Low:** 5 (13.5%) **Interm:** 20 (54%) **High:** 12 (32.4%) **FFDM:** 75 **Low:** 3 (4%) **Interm:** 32 (43%) **High:** 40 (53.3%) ***p=0.045**
Friedewald et al., 2014 **DBT:** 173,663 **FFDM:** 281,187 **Total:** 454,850	**DBT:** 950 Invasive: 707 (74.4%) DCIS: 243 (25.6%) **FFDM:** 1207 Invasive: 815 (67.5%) DCIS: 392 (32.5%)	**Overall** **DBT:** 5.4 **FFDM:** 4.2 ↑28.6% ***p<0.001**	**Invasive** **DBT:** 4.1 **FFDM:** 2.9 ↑41.4% ***p<0.001** ***IDC*:** **DBT:** 3.27 **FFDM:** 2.46 ***p<0.001** ***ILC*:** **DBT:** 0.55 **FFDM:** 0.27 ***p<0.001**	***In situ*** **DBT:** 1.40 **FFDM:** 1.39 *N.S.*	n/a	**DBT:** 707 **IDC:** 568 (80.3%) **ILC:** 95 (13.4%) **ILC+IDC:** 29 (4.1%) **Other:** 5 (0.7%) **Unspec.:** 10 (1.4%) **FFDM:** 815 **IDC:** 693 (85%) **ILC:** 75 (9.2%) **ILC+IDC:** 39 (4.8%) **Other:** 5 (0.6%) **Unspec.:** 3 (0.4%)	n/a	n/a	n/a

TNM Classification: See Figure 2 | ***Acronyms/Abbreviations:*** *Cancer Detection Rate (CDR); Ductal Carcinoma In Situ (DCIS); Digital Breast Tomosynthesis (DBT); Centimeters (cm); Full-Field Digital Mammography (FFDM); Intermediate (Interm); Invasive Ductal Carcinoma (IDC); Invasive Lobular Carcinoma (ILC); Lymph Node (LN); Not available (n/a); Unspecified (Unspec.)*

** = Significant; N.S. = Not significant*

Table 8 (continues)

Study	Cancers Detected	CDR/1000			Tumor Size (cm)	Invasive Cancer Types	LN Status (TNM)	Invasive Cancer Grade	DCIS Grade
		Overall	**Invasive**	***In situ***					
Durand et al., 2015 **DBT:** 8,591 **FFDM:** 9,364 **Total:** 17,955	**DBT:** 51 Invasive: 35 (68.6%) DCIS: 16 (31.4%) **FFDM:** 54 Invasive: 35 (64.8%) DCIS: 19 (35.2%)	**DBT:** 5.9 **FFDM:** 5.7 p=0.88 *N.S.*	**DBT:** 4.07 **FFDM:** 3.74 p=0.26 *N.S.*	**DBT:** 1.9 **FFDM:** 2.0 p=0.63 *N.S.*	***Invasive only*:** **DBT:** data for 24 **TX:** 0 **T1:** 18 (75%) **T2:** 6 (25%) **FFDM:** data for 32 **TX:** 1 (3.1%) **T1:** 23 (71.9%) **T2:** 8 (25%) (*TNM classification*) *N.S.*	n/a	**DBT:** 35 **n/a:** 1 (2.9%) **NX:** 2 (5.7%) **N0:** 24 (68.6%) **N1:** 8 (22.9%) **FFDM:** 35 **n/a:** 3 (8.6%) **NX:** 2 (5.7%) **N0:** 25 (71.4%) **N1:** 4 (11.4%) **N2:** 1 (2.9%) p = 0.84 *N.S.*	n/a	n/a
Rose et al., 2013 **DBT:** 9,499 **FFDM:** 13,856 **Total:** 23,355	**DBT:** 51 Invasive: 41 (80.4%) DCIS: 10 (19.6%) **FFDM:** 56 Invasive: 39 (69.6%) DCIS: 17 (30.4%)	**Overall** **DBT:** 5.37 **FFDM:** 4.04 ↑33% p=0.18 *N.S.*	**Invasive** **DBT:** 4.32 **FFDM:** 2.81 ↑53.7% p=0.07 *Trending, N.S*	***In situ*** **DBT:** 1.05 **FFDM:** 1.22 *N.S.*	***Invasive only*:** **DBT:** **Mean:** 1.6 **Median:** 1.3 **FFDM:** **Mean:** 1.6 **Median:** 1.3 p=0.91 *N.S.*	**DBT:** 41 **IDC:** 34 (83%) **ILC:** 6 (14.6%) **ILC+IDC:** 1 (2.4%) **Mucinous:** 0 **FFDM:** 39 **IDC:** 31 (79.4%) **ILC:** 3 (7.8%) **ILC+IDC:** 4 (10.3%) **Mucinous:** 1 (2.5%) p = 0.69 *N.S.*	**DBT:** 41 **Nx:** 1 (2.4%) **N0:** 34 (83%) **N1:** 6 (14.6%) **N2:** 0 **FFDM:** 39 **Nx:** 2 (5%) **N0:** 33 (85%) **N1:** 2 (5%) **N2:** 2 (5%) p = 0.84 *N.S.*	**DBT:** 41 **Grade 1:** 16 (39%) **Grade 2:** 17 (41%) **Grade 3:** 9 (22%) **Unknown:** 0 **FFDM:** 39 **Grade 1:** 12 (31%) **Grade 2:** 11 (28%) **Grade 3:** 15 (38.5%) **Unknown:** 1 (2.5%) *N.S.*	**DBT:** 10 **Low:** 0 **Interm:** 3 (30%) **High:** 7 (70%) **FFDM:** 17 **Low:** 3 (17.6%) **Interm:** 6 (35%) **High:** 7 (41.2%) **Unknown:** 1 (6%) *N.S.*

TNM Classification: See Figure __ | ***Acronyms/Abbreviations:*** *Cancer Detection Rate (CDR); Ductal Carcinoma in situ (DCIS); Digital Breast Tomosynthesis (DBT); Centimeters (cm); Full-Field Digital Mammography (FFDM); Intermediate (Interm); Invasive Ductal Carcinoma (IDC); Invasive Lobular Carcinoma (ILC); Lymph Node (LN); Not available (n/a); Unspecified (Unspec.)*

** = Significant | N.S. = Not significant*

Table 8

APPENDIX C: Breast Cancer TNM Staging and Receptor Phenotypes

TNM Staging System for Breast Cancer	
T: Primary Tumor	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma *in situ*
T1	Tumor ≤ 2.0 cm in greatest dimension
T1mic	Tumor ≤ 0.1 cm in greatest dimension
T1a	Tumor > 0.1 cm but ≤ 0.5 cm in greatest dimension
T1b	Tumor > 0.5 cm but ≤ 1 cm in greatest dimension
T1c	Tumor > 1.0 cm but ≤ 2.0 cm in greatest dimension
T2	Tumor > 2.0 cm but ≤ 5 cm in greatest dimension
T3	Tumor > 5.0 cm in greatest dimension
T4	Tumor of any size with direct extension into (a) chest wall or (b) skin of breast
N: Regional Lymph Nodes	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis histologically, no additional exam for isolated tumor cells
N1	Metastasis in movable ipsilateral axillary lymph node(s)
N2	Metastases in ipsilateral axillary lymph nodes fixed or matted, or in clinically apparent ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph node metastasis
N3	Metastasis in ipsilateral infraclavicular lymph node(s), or in clinically apparent ipsilateral internal mammary nodes in the presence of clinically evident axillary lymph node metastasis; or metastasis in ipsilateral supraclavicular lymph node(s) with or without axillary or internal mammary lymph node involvement
M: Distant Metastasis	
MX	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis

Figure 2. TNM Staging System for Breast Cancer

		Estrogen Receptor (ER)	Progesterone Receptor (PR)	HER2
Luminal	**A**	ER (+) and/or PR (+)		(-)
	B	ER (+) and/or PR (+)		(+)
HER2-Enriched		(-)	(-)	(+)
Triple Negative		(-)	(-)	(-)

Figure 3. Breast Cancer Receptor Phenotypes

APPENDIX D: Institutional Review Board (IRB) Forms

Yale University Institutional Review Boards

WORKSHEET: Criteria for Approval

NUMBER	DATE	PAGE
HRP-314	10/5/2015	1 of 2

The purpose of this worksheet is to provide support for IRB members reviewing research. This worksheet must be used. It does not need to be completed or retained. (LAR = "subject's legally authorized representative")

1	**General Considerations** (Check if **"Yes"** or **"N/A"**. All must be checked)		
☒	The convened IRB (or Designated Reviewer) has, or has obtained through consultation, adequate expertise.		
☐	For initial review the principal investigator is not Restricted. **("N/A" if not initial review) N/A:** ☒		
☒	Materials are complete.		
2	**Criteria for Approval of Research:** (Check if **"Yes"** or **"N/A"**. All must be checked) (Applies to initial, continuing, modifications)		
☒	Risks to subjects are minimized by using procedures, which are consistent with sound research design and which do not unnecessarily expose subjects to risk.		
☐	Risks to subjects are minimized by using procedures already being performed on the subjects for other purposes. **("N/A" if none) N/A:** ☒		
☒	Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result.		
☒	Selection of subjects is equitable. (Consider the purpose and setting of the research, involvement of vulnerable subjects, selection criteria, and recruitment, enrollment, and payment procedures.)		
☒	The research plan makes adequate provision for monitoring the data collected to ensure the safety of subjects. **("N/A" if ≤ Minimal Risk)**		
☒	There are adequate provisions to protect the privacy of subjects.		
☒	There are adequate provisions to maintain the confidentiality of data.		
☐	Additional safeguards have been included in the study to protect the rights and welfare of subjects vulnerable to coercion or undue influence. **("N/A" if no vulnerable subjects) N/A:** ☒		
☒	The informed consent process meets one of these sections or checklists		
	☐ **Section 5: Consent Process**	☒ **Waiver or alteration of consent process (HRP-410)**	☐ **Permanently closed to enrollment**
☒	The informed consent documentation meets one of these sections, worksheets, or checklists		
	☐ **Section 6: Long Form**	☒ **Waiver of documentation (HRP-411)**	☐ **Permanently closed to enrollment**
	☐ **Short Form (HRP-317)**	☒ **Waiver or alteration of consent process (HRP-410)**	
☐	Additional applicable criteria[1] are met **("N/A" if none) N/A:** ☒		
3	**Additional Considerations** (Check all that apply.)		
☒	Does the research involve no more than Minimal Risk to subjects?		
☐	Should review take place more often than annually?[2] If so, specify period.		
☐	Is verification needed from sources other than the investigator that no material changes have occurred since prior review?[3] **("N/A" if initial) N/A:** ☒		
☐	Does information need to be provided to subjects because it may affect their willingness to continue participation? **("N/A" if initial) N/A:** ☒		
4	**Primary Reviewer Criteria for Initial review** (Check if **"Yes"** or **"N/A"**. All must be checked; May be determined by a primary reviewer)		
☒	The research has the resources necessary to protect subjects. (Time to conduct and complete the research; adequate facilities, subject pool, and medical/psychosocial resources; qualified investigators and research staff; appropriate qualifications for international research.)		
☐	There are no inconsistencies between the DHHS grant and protocol. **("N/A" if there is no DHHS grant.) N/A:** ☒		
☒	The plan for communication among sites is adequate to protect subjects. **("N/A" if not a multicenter trial where PI is the lead or not initial) N/A:** ☒		
	Complete remaining items when applicable		
5	**Consent Process** (Check if **"Yes"**. All must be checked)		
☐	The investigator will obtain the legally effective informed consent of the subject or LAR.		
☐	The circumstances of consent provide the prospective subject or LAR sufficient opportunity to consider whether or not to participate.		
☐	The circumstances of consent minimize the possibility of coercion or undue influence.		
☐	Information to be given to the subject or LAR will be in language understandable to the subject or LAR.		
☐	There is no exculpatory language through which the subject or LAR is made to waive or appear to waive the subject's legal rights, or releases or appears to release the investigator, the sponsor, the institution or its agents from liability from negligence.		
☐	Consent will disclose the elements in **Section 7: Elements of Consent Disclosure**		

[1] Advertisements (HRP-315); Payments (HRP-316); Additional Federal Agency Criteria (HRP-318); Pregnant Women (HRP-412); Non-Viable Neonates (HRP-413); Neonates of Uncertain Viability (HRP-414); Prisoners (HRP-415); Children (HRP-416); Cognitively Impaired Adults (HRP-417); Non-Significant Risk Device (HRP-418)

[2] Consider nature and level of risks; degree of uncertainty regarding the risks; subject vulnerability; investigator experience; IRB's experience with investigator or sponsor; projected rate of enrollment; and whether study involves novel procedures.

[3] Implement when the veracity of the information provided is questioned.

3	**Waiver of the Consent Process for FDA-Regulated Research Involving Anonymous Tissue Specimens[3]** (Check if **"Yes"**. All must be checked)
☐	The research does not involve Human Subjects as Defined by DHHS.
☐	The study involves an in vitro diagnostic device investigation.
☐	The testing is noninvasive.
☐	The testing does not require an invasive sampling procedure that presents significant risk.
☐	The testing does not by design or intention introduce energy into a subject.
☐	The device is not used as a diagnostic procedure without confirmation of the diagnosis by another, medically established diagnostic product or procedure.
☐	For a product in the laboratory research phase of development, and not represented as an effective in vitro diagnostic product, all labeling bears the statement, prominently placed: "For Research Use Only. Not for use in diagnostic procedures."
☐	For a product being shipped or delivered for product testing prior to full commercial marketing (for example, for use on specimens derived from humans to compare the usefulness of the product with other products or procedures which are in current use or recognized as useful), all labeling bears the statement, prominently placed: "For Investigational Use Only. The performance characteristics of this product have not been established."
☐	The study uses one of more of the following: **(Check all boxes that are true. One must be checked)** ☐ Specimens collected for routine clinical care or analysis that would have been discarded. ☐ Specimens obtained from specimen repositories. ☐ Leftover specimens that were previously collected for other research purposes.
☐	The identity of the subject is not known to the investigator or any other individuals associated with the investigation, including the sponsor meaning neither the investigator nor any other individuals associated with the investigation, including the sponsor can readily ascertain the identity of the subject.
☐	One of the following is true: **(Check all boxes that are true. One must be checked)** ☐ Specimens are not coded where "Coded" means that 1) a number, letter, symbol, or combination thereof (i.e., the code) has replaced identifying information (such as name or social security number) that would enable the investigator or any other individuals associated with the investigation, including the sponsor to readily ascertain the identity of the individual to whom the specimen pertains; and 2) a key to decipher the code exists, enabling linkage of the identifying information to the specimen. ☐ Neither the investigator(s) nor any other individuals associated with the investigation or the sponsor can link the specimen to the subject from whom the specimen was collected, either directly or indirectly through coding systems.
☐	One of the following is true: **(Check all boxes that are true. One must be checked)** ☐ The specimens are not accompanied by clinical information. ☐ Clinical information that accompanies the specimens does not make the specimen source identifiable to the investigator or any other individual associated with the investigation, including the sponsor.
☐	The individuals caring for the patients are different from those conducting the investigation and do not share information about the patient with those conducting the investigation.
☐	The individuals caring for the patients do not share information about the patient with those conducting the investigation.
☐	The specimens are provided to the investigator(s) without identifiers.
☐	The supplier of the specimens has established policies and procedures to prevent the release of personal information.
4	**Waiver of Informed Consent for Planned Emergency Research[4]**
☐	The research meets the criteria in "**CHECKLIST: Waiver of Consent for Emergency Research (HRP-419)**."

[3] Guidance on Informed Consent for In Vitro Diagnostic Device Studies Using Leftover Human Specimens that are Not Individually Identifiable – April 25, 2006

[4] 21 CFR §50.24 and 45 CFR §46 Waiver of Informed Consent Requirements in Certain Emergency Research – November 1, 1996

Yale University
Institutional Review Boards

CHECKLIST: Waiver or Alteration of Consent Process

NUMBER	DATE	PAGE
HRP-410	10/5/2015	1 of 2

The purpose of this checklist is to provide support for IRB members or the Designated Reviewer following the WORKSHEET: Criteria for Approval (HRP-314) when research involves waiver or alteration of the consent process. This checklist must be used for all reviews (initial, continuing, modification, review by the convened IRB, and review using the expedited procedure.)

- For initial review using the expedited procedure and modifications and continuing reviews where the determinations relevant to this checklist made on the previous review have changed, the Designated Reviewer completes this checklist to document determinations required by the regulations along with protocol specific findings justifying those determinations. The Designated Reviewer attaches this checklist to "Submit Non-Committee Review" activity. The IRB Office retains this checklist in the protocol file.
- For initial review using the convened IRB and for modifications and continuing reviews where the determinations relevant to this checklist made on the previous review have changed, one of the following two options may be used:
  1. The convened IRB completes the corresponding section of the meeting minutes to document determinations required by the regulations along with protocol specific findings justifying those determinations, in which case this checklist does not need to be completed or retained.
  2. The convened IRB completes this checklist to document determinations required by the regulations along with protocol specific findings justifying those determinations and the IRB Office uploads this checklist in the "Submit Committee Review" activity and retains this checklist in the protocol file.

The research must meet one of the following four sets of criteria

	1 Waiver or Alteration of Consent Process[1] (Check if "**Yes**". All must be checked)
☒	The research is **NOT** FDA-regulated.
☒	The research does **NOT** involve non-viable neonates.
☒	The research does **NOT** involve newborn dried blood spots.
☒	The research involves no more than Minimal Risk to the subjects. *Provide protocol specific findings justifying this determination:* **Retrospective review of de-identified data**
☒	The waiver or alteration will **NOT** adversely affect the rights and welfare of the subjects. *Provide protocol specific findings justifying this determination:* **De-identified data**
☒	The research could **NOT** practicably be carried out without the waiver or alteration *Provide protocol specific findings justifying this determination:* **Retrospective review of >40,000 records**
☒	Whenever appropriate, the subjects will be provided with additional pertinent information after participation. *Provide protocol specific findings justifying this determination:* **Retrospective study; unapplicable**
	2 Waiver or Alteration of Consent Process[2] (Check if "**Yes**". All must be checked)
☐	The research is **NOT** FDA-regulated.
☐	The research does **NOT** involve non-viable neonates.
☐	The research or demonstration project is to be conducted by or subject to the approval of state or local government officials. *Provide protocol specific findings justifying this determination:*
☐	The research or demonstration project is designed to study, evaluate, or otherwise examine one or more of the following: **(Check all boxes that are true. One must be checked)** ☐ Public benefit or service programs. ☐ Procedures for obtaining benefits or services under those programs. ☐ Possible changes in or alternatives to those programs or procedures. ☐ Possible changes in methods or levels of payment for benefits or services under those programs. *Provide protocol specific findings justifying this determination:*
☐	The research could **NOT** practicably be carried out without the waiver or alteration. *Provide protocol specific findings justifying this determination:*

[1] 45 CFR §46.116(d)
[2] 45 CFR §46.116(c)

3	**Waiver of the Consent Process for FDA-Regulated Research Involving Anonymous Tissue Specimens[3]** (Check if **"Yes"**. All must be checked)
☐	The research does not involve Human Subjects as Defined by DHHS.
☐	The study involves an in vitro diagnostic device investigation.
☐	The testing is noninvasive.
☐	The testing does not require an invasive sampling procedure that presents significant risk.
☐	The testing does not by design or intention introduce energy into a subject.
☐	The device is not used as a diagnostic procedure without confirmation of the diagnosis by another, medically established diagnostic product or procedure.
☐	For a product in the laboratory research phase of development, and not represented as an effective in vitro diagnostic product, all labeling bears the statement, prominently placed: "For Research Use Only. Not for use in diagnostic procedures."
☐	For a product being shipped or delivered for product testing prior to full commercial marketing (for example, for use on specimens derived from humans to compare the usefulness of the product with other products or procedures which are in current use or recognized as useful), all labeling bears the statement, prominently placed: "For Investigational Use Only. The performance characteristics of this product have not been established."
☐	The study uses one of more of the following: **(Check all boxes that are true. One must be checked)** ☐ Specimens collected for routine clinical care or analysis that would have been discarded. ☐ Specimens obtained from specimen repositories. ☐ Leftover specimens that were previously collected for other research purposes.
☐	The identity of the subject is not known to the investigator or any other individuals associated with the investigation, including the sponsor meaning neither the investigator nor any other individuals associated with the investigation, including the sponsor can readily ascertain the identity of the subject.
☐	One of the following is true: **(Check all boxes that are true. One must be checked)** ☐ Specimens are not coded where "Coded" means that 1) a number, letter, symbol, or combination thereof (i.e., the code) has replaced identifying information (such as name or social security number) that would enable the investigator or any other individuals associated with the investigation, including the sponsor to readily ascertain the identity of the individual to whom the specimen pertains; and 2) a key to decipher the code exists, enabling linkage of the identifying information to the specimen. ☐ Neither the investigator(s) nor any other individuals associated with the investigation or the sponsor can link the specimen to the subject from whom the specimen was collected, either directly or indirectly through coding systems.
☐	One of the following is true: **(Check all boxes that are true. One must be checked)** ☐ The specimens are not accompanied by clinical information. ☐ Clinical information that accompanies the specimens does not make the specimen source identifiable to the investigator or any other individual associated with the investigation, including the sponsor.
☐	The individuals caring for the patients are different from those conducting the investigation and do not share information about the patient with those conducting the investigation.
☐	The individuals caring for the patients do not share information about the patient with those conducting the investigation.
☐	The specimens are provided to the investigator(s) without identifiers.
☐	The supplier of the specimens has established policies and procedures to prevent the release of personal information.
4	**Waiver of Informed Consent for Planned Emergency Research[4]**
☐	The research meets the criteria in **"CHECKLIST: Waiver of Consent for Emergency Research (HRP-419)."**

[3] Guidance on Informed Consent for In Vitro Diagnostic Device Studies Using Leftover Human Specimens that are Not Individually Identifiable – April 25, 2006

[4] 21 CFR §50.24 and 45 CFR §46 Waiver of Informed Consent Requirements in Certain Emergency Research – November 1, 1996

Yale University
Institutional Review Boards

CHECKLIST: Waiver of Written Documentation of Consent

NUMBER	DATE	PAGE
HRP-411	10/5/2015	1 of 1

The purpose of this checklist is to provide support for IRB members or the Designated Reviewer following the WORKSHEET: Criteria for Approval (HRP-314) when research involves the waiver of written documentation of consent. This checklist must be used for all reviews (initial, continuing, modification, review by the convened IRB, and review using the expedited procedure.)

- For initial review using the expedited procedure and modifications and continuing reviews where the determinations relevant to this checklist made on the previous review have changed, the Designated Reviewer completes this checklist to document determinations required by the regulations along with protocol specific findings justifying those determinations. The Designated Reviewer attaches this checklist to "Submit Non-Committee Review" activity. The IRB Office retains this checklist in the protocol file.
- For initial review using the convened IRB and for modifications and continuing reviews where the determinations relevant to this checklist made on the previous review have changed, one of the following two options may be used:
  1. The convened IRB completes the corresponding section of the meeting minutes to document determinations required by the regulations along with protocol specific findings justifying those determinations, in which case this checklist does not need to be completed or retained.
  2. The convened IRB completes this checklist to document determinations required by the regulations along with protocol specific findings justifying those determinations and the IRB Office uploads this checklist in the "Submit Committee Review" activity and retains this checklist in the protocol file.

The research must meet one of the following two sets of criteria

1 Waiver of Written Documentation of Consent[1] (Check if "Yes". All must be checked)

☒	The written script of the information to be provided orally (if consent is obtained in person) and all written information to be provided or electronically displayed include all required and appropriate additional elements of consent disclosure in **Section 7: ELEMENTS OF CONSENT DISCLOSURE** in the WORKSHEET: Criteria for Approval (HRP-314).
☒	The research presents no more than Minimal Risk of harm to subjects.
☒	The research involves no procedures for which written consent is normally required outside of the research context.
☒	The research does **NOT** involve newborn dried blood spots.

Select one of the following: **(One must be checked)**

- ☐ Written information describing the research **is to be provided** to the subject or the subject's legally authorized representative.
- ☒ Written information describing the research **does not need to be provided** to the subject or the subject's legally authorized representative.

2 Waiver of Written Documentation of Consent[2] (Check if "Yes". All must be checked)

☐	The research is not FDA-regulated.
☐	The written script of the information to be provided orally and all written information to be provided include all required and appropriate additional elements of consent disclosure in **Section 7: ELEMENTS OF CONSENT DISCLOSURE** in the WORKSHEET: Criteria for Approval (HRP-314).
☐	The only record linking the subject and the research would be the consent document.
☐	The principal risk of a signed consent document would be the potential harm resulting from a breach of confidentiality.
☐	Each subject will be asked whether the subject wants documentation linking the subject with the research, and the subject's wishes will govern.

Select one of the following: **(One must be checked)**

- ☐ Written information describing the research **is to be provided** to the subject or the subject's legally authorized representative.
- ☐ Written information describing the research **does not need to be provided** to the subject or the subject's legally authorized representative.

[1] 21 CFR §56.109(c)(1) and 45 CFR §46.117(c)(2)
[2] 45 CFR §46.117(c)(1)

APPENDIX E: Calculation of Power

Dupont WD, Plummer WD: 'Power and Sample Size Calculations: A Review and Computer Program', Controlled Clinical Trials 1990; 11:116-28. PS, Version 3.1.2.

Input variables on the *Dichotomous* dialogue box:

- α (alpha): The Type I error probability for a two-sided test. This is the probability that we will falsely reject the null hypothesis.
- *n*: For case-control studies, *n* is the number of case patients. For prospective studies, *n* is the number of patients receiving the experimental treatment.
- *Power*: The probability of correctly rejecting the null hypothesis that the relative risk (odds ratio) equals 1 given *n* case patients, *m* control patients per experimental patient, and a Type I error probability α.
- p_0: For case-control studies, p_0 is the probability of exposure in controls. In prospective studies, p_0 is the probability of the outcome for a control patient.
- p_1: For case-control studies, p_1 is the probability of exposure in cases. In prospective studies, p_1 is the probability of the outcome for an experimental subject.
- *m*: For independent prospective studies, *m* is the ratio of control to experimental subjects. For matched prospective studies, *m* is the number of control subjects matched to each experimental subject. For independent case-control studies, *m* is the ratio of control to case patients. For matched case-control studies, *m* is the number of control patients matched to each case.

Description output: We are planning a study with 2,194 (*n*) experimental subjects and 1,761 (*n***m*) control subjects. Prior data indicate that the probability of exposure among controls is 0.044 (p_0). If the true probability of exposure among cases is 0.0645 (p_1), we will be able to reject the null hypothesis that the exposure rates for case and controls are equal with probability (power) .805 (*Power for uncorrected chi-squared test*). The Type I error probability associated with this test of this null hypothesis is 0.05 (α). We will use an uncorrected chi-squared statistic to evaluate this null hypothesis.

Table 9. Power Calculation

Calculation of Power: Design and Inputs	**PPV_1**
Type of study	Dichotomous
Requested output	Power
How is the alternative hypothesis expressed?	Two proportions
Matched or Independent?	Independent
Case-Control?	Case-Control
Uncorrected chi-squared or Fisher's exact test?	Uncorrected chi-squared test
α (alpha)	0.05
n	2,194
p_0	0.044
p_1	0.0645
m	0.80264357
Power for uncorrected chi-squared test	**0.805**

APPENDIX F: Supplementary Analyses

Table 10. Extended Risk Data in Study Subset

	FFDM	DBT	*p*-value*
Age Categories, n (%)			
<40 years			
40-49 years			
50-59 years			
60-69 years			
≥70 years			
Race, N (%)			
Caucasian			
African American			
Asian			
Other/Unknown			
Ethnicity, N (%)			
Hispanic			
Non-Hispanic			
Other/Unknown			
Breast Density, N (%)			
BI-RADS A			
BI-RADS B			
BI-RADS C			
BI-RADS D			
Prior Mammogram(s), N (%)			
0			
1			
2			
≥3			
BMI, Mean (SD)			
Age at Menarche, N (%)			
7-11 years			
12-13 years			
≥14 years			
Unknown/Missing			
Age at First Birth, N (%)			
No births			
<20 years			
20-24 years			
25-29 years			
≥30 years			
Unknown/Missing			
First Degree Fam History, N (%)			
Yes			
No			
Unknown/Missing			
Jewish Ancestry, N (%)			
Yes			
No			
Unknown/Missing			

**p-value will be calculated from chi-squared test, excluding unknown/missing data. All tests will be two sided. p-values of <0.05 will be considered statistically significant.*

BIBLIOGRAPHY

1. American Cancer Society. *Cancer Facts & Figures 2016.* Atlanta: American Cancer Society; 2016. Available at: http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf. Accessed 12/20/2016.
2. American Cancer Society. *Cancer Facts & Figures 2015; Special Section: Breast Carcinoma in Situ.* Atlanta: American Cancer Society; 2015. Available at: http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-046202.pdf. Accessed 12/20/2016.
3. American Cancer Society. *ACR BI-RADS Atlas - Mammography Reporting System.* Atlanta: American Cancer Society; 2013. Available at: https://www.acr.org/~/media/ACR/Documents/PDF/QualitySafety/Resources/BIRADS/01-Mammography/02--BIRADS-Mammography-Reporting.pdf?la=en. Accessed 12/20/2016.
4. US Food and Drug Administration, Summary of Safety and Effectiveness Data: Hologic DBT (P080003). *Available at* https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P080003. 2011:Accessed August 20, 2016.
5. Alcusky M, Philpotts L, Bonafede M, Clarke J, Skoufalos A. The patient burden of screening mammography recall. *Journal of women's health (2002).* 2014;23 Suppl 1:S11-19.
6. Aujero MP, Gavenonis SC, Benjamin R, Zhang Z, Holt JS. Clinical Performance of Synthesized Two-dimensional Mammography Combined with Tomosynthesis in a Large Screening Population. *Radiology.* 2017;283(1):70-76.
7. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *International journal of cancer.* 2013;132(5):1133-1145.
8. Brennan M, Houssami N. Discussing the benefits and harms of screening mammography. *Maturitas.* 2016;92:150-153.
9. Caumo F, Bernardi D, Ciatto S, et al. Incremental effect from integrating 3D-mammography (tomosynthesis) with 2D-mammography: Increased breast cancer detection evident for screening centres in a population-based trial. *Breast (Edinburgh, Scotland).* 2014;23(1):76-80.
10. Cecchini RS, Costantino JP, Cauley JA, et al. Baseline mammographic breast density and the risk of invasive breast cancer in postmenopausal women participating in the NSABP study of tamoxifen and raloxifene (STAR). *Cancer Prev Res (Phila).* 2012;5(11):1321-1329.
11. Choi JS, Han BK, Ko EY, et al. Comparison between two-dimensional synthetic mammography reconstructed from digital breast tomosynthesis and full-field digital mammography for the detection of T1 breast cancer. *Eur Radiol.* 2016;26(8):2538-2546.
12. Cianfrocca M, Goldstein LJ. Prognostic and predictive factors in early-stage breast cancer. *The oncologist.* 2004;9(6):606-616.

13. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *The Lancet. Oncology.* 2013;14(7):583-589.
14. Cole EB, Pisano ED. Tomosynthesis for breast cancer screening. *Clinical imaging.* 2016;40(2):283-287.
15. Conant EF, Beaber EF, Sprague BL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: a cohort study within the PROSPR consortium. *Breast Cancer Res Treat.* 2016;156(1):109-116.
16. Coop P, Cowling C, Lawson C. Tomosynthesis as a screening tool for breast cancer: A systematic review. *Radiography.* 2016;22(3):e190-e195.
17. Destounis S, Arieno A, Morgan R. Initial experience with combination digital breast tomosynthesis plus full field digital mammography or full field digital mammography alone in the screening environment. *Journal of clinical imaging science.* 2014;4:9.
18. Duffy SW, Gabe R. What should the detection rates of cancers be in breast screening programmes? *British Journal of Cancer.* 2005;92(3):597-600.
19. Dupont WD PW. Power and Sample Size Calculations: A Review and Computer Program.Controlled Clinical Trials 1990; 11:116-28. Version 3.1.2.
20. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology.* 2015;274(1):85-92.
21. Ellsworth RE, Hooke JA, Love B, Ellsworth DL, Shriver CD. Molecular changes in primary breast tumors and the Nottingham Histologic Score. *Pathology oncology research : POR.* 2009;15(4):541-547.
22. Etesami M, Giwerc MY, Harigopal M, Zaheer Kidwai W, Philpotts LE. Cancer Detection with Digital Breast Tomosynthesis (DBT) Compared with Conventional Digital Mammography in Routine Breast Cancer Screening. *Presented at: Radiological Society of North America 2017 Scientific Assembly and Annual Meeting, November 26 - December 1, 2017, Chicago, IL.* rsna2017.rsna.org/program/. Accessed July 20, 2017.
23. Ferlay JS, I.; Ervik, M.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer. 2013; http://globocan.iarc.fr/. Accessed December 10, 2016.
24. Freer PE. Mammographic breast density: impact on breast cancer risk and implications for screening. *Radiographics : a review publication of the Radiological Society of North America, Inc.* 2015;35(2):302-315.
25. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA.* 2014;311(24):2499-2507.
26. Gilbert FJ, Tucker L, Gillan MG, et al. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme--a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with

digital mammography alone. *Health technology assessment (Winchester, England).* 2015;19(4):i-xxv, 1-136.
27. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *AJR. American journal of roentgenology.* 2014;203(3):687-693.
28. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology.* 2013;269(3):694-700.
29. Hodgson R, Heywang-Kobrunner SH, Harvey SC, et al. Systematic review of 3D mammography for breast cancer screening. *Breast (Edinburgh, Scotland).* 2016;27:52-61.
30. Hooley RJ, Durand MA, Philpotts LE. Advances in Digital Breast Tomosynthesis. *AJR. American journal of roentgenology.* 2016:1-11.
31. Houssami N. Digital breast tomosynthesis (3D-mammography) for screening women with dense breasts. *Expert review of medical devices.* 2016;13(6):515-517.
32. Houssami N, Bernardi D, Pellegrini M, et al. Breast cancer detection using single-reading of breast tomosynthesis (3D-mammography) compared to double-reading of 2D-mammography: Evidence from a population-based trial. *Cancer epidemiology.* 2017;47:94-99.
33. Houssami N, Macaskill P, Bernardi D, et al. Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading--evidence to guide future screening strategies. *Eur J Cancer.* 2014;50(10):1799-1807.
34. Houssami N, Turner RM. Rapid review: Estimates of incremental breast cancer detection from tomosynthesis (3D-mammography) screening in women with dense breasts. *Breast (Edinburgh, Scotland).* 2016;30:141-145.
35. Lang K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmo Breast Tomosynthesis Screening Trial, a population-based study. *Eur Radiol.* 2016;26(1):184-190.
36. Lewin JM, D'Orsi CJ, Hendrick RE. Digital mammography. *Radiologic clinics of North America.* 2004;42(5):871-884, vi.
37. Lourenco AP, Barry-Brooks M, Baird GL, Tuttle A, Mainiero MB. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology.* 2015;274(2):337-342.
38. McCarthy AM, Kontos D, Synnestvedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *Journal of the National Cancer Institute.* 2014;106(11).
39. McDonald ES, McCarthy AM, Akhtar AL, Synnestvedt MB, Schnall M, Conant EF. Baseline Screening Mammography: Performance of Full-Field Digital Mammography Versus Digital Breast Tomosynthesis. *AJR. American journal of roentgenology.* 2015;205(5):1143-1148.
40. McDonald ES, Oustimov A, Weinstein SP, Synnestvedt MB, Schnall M, Conant EF. Effectiveness of Digital Breast Tomosynthesis Compared With Digital

Mammography: Outcomes Analysis From 3 Years of Breast Cancer Screening. *JAMA Oncol.* 2016;2(6):737-743.
41. Melnikow J, Fenton JJ, Whitlock EP, et al. Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the U.S. Preventive Services Task Force. *Annals of internal medicine.* 2016;164(4):268-278.
42. Moon HJ, Kim EK. Characteristics of breast cancer detected by supplementary screening ultrasonography. *Ultrasonography (Seoul, Korea).* 2015;34(3):153-156.
43. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. *JAMA.* 2015;314(15):1599-1614.
44. Park JM, Franken EA, Jr., Garg M, Fajardo LL, Niklason LT. Breast tomosynthesis: present considerations and future applications. *Radiographics : a review publication of the Radiological Society of North America, Inc.* 2007;27 Suppl 1:S231-240.
45. Pozzi A, Corte AD, Lakis MA, Jeong H. Digital Breast Tomosynthesis in Addition to Conventional 2DMammography Reduces Recall Rates and is Cost Effective. *Asian Pacific journal of cancer prevention : APJCP.* 2016;17(7):3521-3526.
46. Rahbar H, McDonald ES, Lee JM, Partridge SC, Lee CI. How Can Advanced Imaging Be Used to Mitigate Potential Breast Cancer Overdiagnosis? *Acad Radiol.* 2016;23(6):768-773.
47. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R, Jr. Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR. American journal of roentgenology.* 2013;200(6):1401-1408.
48. Rose SL, Tidwell AL, Ice MF, Nordmann AS, Sexton R, Jr., Song R. A reader study comparing prospective tomosynthesis interpretations with retrospective readings of the corresponding FFDM examinations. *Acad Radiol.* 2014;21(9):1204-1210.
49. Sharpe RE, Jr., Venkataraman S, Phillips J, et al. Increased Cancer Detection Rate and Variations in the Recall Rate Resulting from Implementation of 3D Digital Breast Tomosynthesis into a Population-based Screening Program. *Radiology.* 2016;278(3):698-706.
50. Singletary SE, Allred C, Ashley P, et al. Revision of the American Joint Committee on Cancer staging system for breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2002;20(17):3628-3636.
51. Sioshansi S, Huber K, Wazer D. The Implications of Breast Cancer Molecular Phenotype for Radiation Oncology. *Frontiers in Oncology.* 2011;1(12).
52. Siu AL. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of internal medicine.* 2016;164(4):279-296.
53. Skaane P. Breast cancer screening with digital breast tomosynthesis. *Breast cancer (Tokyo, Japan).* 2016.
54. Skaane P, Bandos AI, Gullien R, et al. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a

population-based screening programme using independent double reading with arbitration. *Eur Radiol.* 2013;23(8):2061-2071.
55. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology.* 2013;267(1):47-56.
56. Sumkin JH, Ganott MA, Chough DM, et al. Recall Rate Reduction with Tomosynthesis During Baseline Screening Examinations: An Assessment From a Prospective Trial. *Acad Radiol.* 2015;22(12):1477-1482.
57. Takahashi TA, Lee CI, Johnson KM. Breast cancer screening: Does tomosynthesis augment mammography? *Cleveland Clinic journal of medicine.* 2017;84(7):522-527.
58. Wang WS, Hardesty L, Borgstede J, Takahashi J, Sams S. Breast Cancers Found with Digital Breast Tomosynthesis: A Comparison of Pathology and Histologic Grade. *Breast J.* 2016;22(6):651-656.
59. Yang TL, Liang HL, Chou CP, Huang JS, Pan HB. The adjunctive digital breast tomosynthesis in diagnosis of breast cancer. *BioMed research international.* 2013;2013:597253.
60. Zuckerman SP, Conant EF, Keller BM, et al. Implementation of Synthesized Two-dimensional Mammography in a Population-based Digital Breast Tomosynthesis Screening Program. *Radiology.* 2016;281(3):730-736.