Performance characteristics of digital vs. film screen mammography in community practice

Firas Dabbous, Ph.D.,
James R. & Helen D. Russell Institute for Research and Innovation at Advocate Lutheran General Hospital, Park Ridge, IL

Therese A. Dolecek, Ph.D.,
School of Public Health, University of Illinois at Chicago

Sarah M. Friedewald, M.D.,
Lynn Sage Comprehensive Breast Center, Northwestern Medicine, Prentice Women's Hospital

Katherine Y. Tossas-Milligan,
University of Illinois at Chicago

Tere Macarol, RT(R)(M)(QM)CN-BI,
Advocate health care

Wm. Thomas Summerfelt, Ph.D, and
Advocate Health Care

Garth H Rauscher, Ph.D
University of Illinois at Chicago, School of Public Health, Division of Epidemiology and Biostatistics

Abstract

We compared the performance characteristics of 297,629 full field digital (FFDM) and 416,791 screen film mammograms (SFM). Sensitivity increased with age, decreased with breast density, and was lower for more aggressive and lobular tumors. While sensitivity did not differ significantly by modality, specificity was generally 1–2 percentage points higher for FFDM than for SFM across age and breast density categories. The lower recall rate for FFDM vs. SFM in our study may partially explain performance differences by modality. In this large healthcare organization, modest gains in performance were achieved with the introduction of FFDM as a replacement for SFM.

Corresponding Author: Garth H. Rauscher, PhD, University of Illinois at Chicago, School of Public Health, Division of Epidemiology and Biostatistics, 1603 West Taylor Street (SPHPI, M/C 923), Room 952, Chicago, IL 60612, Phone: 312-413-4317, FAX: 312-996-0064, garthr@uic.edu.
DR. FIRAS M DABBOUS (Orcid ID : 0000-0002-0823-6099)

INTRODUCTION

Data from the Digital Mammography Imaging Screening Trial (DMIST) supported the use of Full field digital mammography (FFDM) [1, 2]. This trial found that FFDM was more accurate in women with heterogeneously dense or extremely dense breasts, for women under age 50, and for pre- or peri-menopausal women of any age. However, facilities that participated in DMIST tended to be academic facilities and mammograms were likely to be read by breast imaging specialists. Therefore, it was not entirely clear whether the findings of DMIST would translate into community practice. A recent comparative effectiveness study of FFDM versus Screen-Film mammography (SFM) in the Breast Cancer Surveillance Consortium (BCSC) found that overall, cancer detection rates and tumor characteristics were similar between the two modalities, but differences were apparent in both sensitivity and specificity within patient subgroups defined by age, menopausal status, breast density and tumor aggressiveness subtypes. When compared to SFM, sensitivity for FFDM was roughly 7 points higher for women in their 40s and 60s but 4.5 points lower among women in their 50s. Sensitivity for FFDM was roughly 15 points higher among women with extremely dense breasts and roughly 12 points higher for patients diagnosed with more aggressive, estrogen receptor (ER) negative breast cancer compared to SFM. Specificity for FFDM was nearly 2 points lower than for SFM among women in their 40s [3]. We sought to re-examine the comparative effectiveness of these two screening modalities in a single large health care organization that lacks an academic affiliation or fellowship program in breast imaging, and to compare results to those from the BCSC.

METHODS

Data came from a single large healthcare delivery organization in the Greater Metropolitan Chicago Area. All sites are connected with a radiology database (PenRad Technologies, Inc., Buffalo, MN) [4]. The study was reviewed and approved by the institutional review boards at all participating institutions. A screening mammogram was defined as a bilateral mammogram with a description of screening, in women without a prior history of breast cancer, mastectomy, or breast implants, and without any breast imaging in the 9 months prior to the screen. We linked 761,908 (297,629 FFDM and 416,791 SFM) screening examinations from women aged 40–79, conducted between 2001–2010 to the population-based Illinois State Cancer Registry and identified 4,829 breast cancers diagnosed between 2001 and 2011 and within 12 months of a screen. The linkage was performed using probabilistic methods with Automatch, version 4 (Matchware Technologies, Inc., Silver Spring, MD). Each mammogram was interpreted using the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) such that: 0 = need additional imaging evaluation, 1 = negative finding, 2 = benign finding, 3 = probably benign finding, 4 = suspicious abnormality, and 5 = finding highly suggestive of malignancy.

Statistical analyses

We conducted logistic regression models overall and by patient subgroups while controlling for age, exam year, race/ethnicity, timing of last screen and indicator variables for screening facility. For specificity, we also used generalized estimating equations to account for clustering of exams by patient. Model-based standardization (predictive margins) was then used to estimate the overall and stratum-specific performance characteristics [5,6]. Ninety-five percent confidence intervals were estimated using the delta method as implemented in the margins command within Stata. All analyses were conducted using Stata statistical software, version 12 (Stata Corp, College Station, TX). All p-values are two-sided.

RESULTS

Patient characteristics by imaging modality were roughly similar except that nL White women were more likely than nL black women to be screened with FFDM, as were women with a more recent prior screen (data not shown).

Overall performance characteristics

In the adjusted analyses, FFDM was associated with a slightly lower cancer detection rate (4.3 vs. 4.8, p=0.07), fewer false negative results (0.54 vs. 0.63, P=0.39), slightly higher specificity (89% vs. 87%, p<0.0001), and a lower recall rate (11% vs. 14%, p<0.0001) compared to SFM. No differences in sensitivity or positive predictive value were apparent (Table 1).

Subgroup analyses

Generally for both SFM and FFDM, sensitivity increased with age and decreased with breast density. Sensitivity was lower for tumors displaying more aggressive characteristics and lower for lobular versus other histologies (Table 2). Within subgroups defined by age, menopausal status, breast density and tumor characteristics, there were no statistical differences in sensitivity of FFDM vs. SFM. Sensitivity of FFDM was qualitatively higher than SFM for postmenopausal women and across all age groups. FFDM sensitivity was also qualitatively higher among women with lower breast density, but lower among women with extremely dense breasts, compared to SFM. FFDM sensitivity was 3 percentage points higher than for SFM with respect to more aggressive ER negative breast cancers, and 7 percentage points higher for higher grade disease, though none of these associations were statistically evident (Table 2).

Overall, specificity increased with age and menopausal status but varied non-monotonically with increasing breast density. By modality, specificity was generally one or two percentage points higher for FFDM than for SFM across age, menopausal status, and breast density categories (Table 2).

DISCUSSION

The BCSC found that sensitivity for FFDM was roughly 7 points higher than for SFM among women in their 40s and 60s but 4.5 points lower among women in their 50s. In

Author Manuscript

contrast, we did not find substantial differences in sensitivity by imaging modality across age groups, though our sensitivity estimates were more constrained by the relatively smaller number of breast cancers compared to the BCSC study. While BCSC reported a 15 point higher sensitivity for FFDM than for SFM among women with extremely dense breasts; we found that sensitivity was 3 points lower for FFDM than for SFM in these women. While BCSC reported higher sensitivity FFDM than for SFM among more aggressive tumors, we found more modest, not significant or marginally significant differences.

While BCSC reported slightly lower (1 point) specificity for FFDM vs. SFM among women in their 40s but similar for other ages, we found slightly higher specificity (about 2 points) for FFDM vs. SFM for all age groups. While BCSC found slightly lower specificity for FFDM vs. SFM regardless of breast density; we found slightly better specificity (by roughly 2 points) for FFDM vs. SFM in all women except those with extremely dense breasts.

The lower recall rate for FFDM vs. SFM in our study may partially explain differences in our results compared to those from the BCSC. Higher recall rates are generally associated with higher sensitivity and PPV but lower specificity. In our study, recall rates were lower for FFDM than for SFM (11% vs. 14%, respectively), whereas in the BCSC study, recall rates were higher for FFDM. In both studies specificity was higher for FFDM than for SFM; however, PPV were similar by modality in our study, whereas in the BCSC study PPV was higher for FFDM than for SFM.

Prior research has shown that performance characteristics vary by geographic region and across different cohorts from multiple facilities [7]. Randomized controlled trials (RCTs) have demonstrated a survival benefit for women screened on film and it is important to show that a universal switch to FFDM will be at minimum equal to the performance of SFM. A newer screening technique called Digital Breast Tomosynthesis (DBT) has been shown to significantly improve the PPV but was not clinically applicable until the digital detector was available. In this large healthcare organization, modest gains in performance were achieved with the introduction of FFDM as a replacement for SFM. Therefore, FFDM is an appropriate substitution for SFM and the digital technique will enable newer technologies such as DBT to further improve screening performance.

Acknowledgments

References

1. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening.[Erratum appears in N Engl J Med. 2006 Oct 26;355(17):1840]. N Engl J Med. 2005; 353:1773–1783. [PubMed: 16169887]

2. Pisano ED, Hendrick RE, Yaffe MJ, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology. 2008; 246:376–383. [PubMed: 18227537]

3. Kerlikowske K, Hubbard RA, Miglioretti DL, et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study.[Summary for patients in *Ann Intern Med.* 2011 Oct 18;155(8):I30]. Ann Intern Med. 2011; 155:493–502. [PubMed: 22007043]

4. PenRad. 2015. Available from: http://www.penrad.com/about_profile.html

5. Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. Am J Epidemiol. 2004; 160:301–305. [PubMed: 15286014]

6. Ahern J, Hubbard A, Galea S. Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. Am J Epidemiol. 2009; 169:1140–1147. [PubMed: 19270051]

7. Taplin S, Abraham L, Barlow WE, et al. Mammography facility characteristics associated with interpretive accuracy of screening mammography. J Natl Cancer Inst. 2008; 100:876–887. [PubMed: 18544742]

Table 1

Performance Measures of Screening Mammography Among Women Undergoing Digital and Film-Screen Examinations [*]

Performance Measure[†]	FFDM	SFM	P-Value
Cases of cancer per 1000 examinations[‡]	4.8 (4.1, 5.5)	5.4 (4.7, 6.1)	0.04
Cancer detection per 1000 examinations[§]	4.3 (3.5, 5.0)	4.8 (4.1, 5.4)	0.07
False-negative results per 1000 examinations[//]	0.54 (0.31, 0.76)	0.63 (0.42, 0.84)	0.39
Sensitivity[¶]	0.89 (0.84, 0.93)	0.88 (0.84, 0.92)	0.69
Specificity	0.89 (0.89, 0.89)	0.87 (0.87, 0.87)	<0.0001
Recall rate, %	0.11 (0.11, 0.12)	0.14 (0.13, 0.14)	<0.0001
Positive predictive value	0.04 (0.03, 0.04)	0.04 (0.03, 0.04)	0.28

[*] Values in parentheses are 95% CIs.

[†] Adjusted for site, age, ethnicity, year, and time between screening examinations.

[‡] Invasive cancer or DCIS within 12 months of a screening examination

[§] Invasive cancer or DCIS within 12 months of a positive screening examination.

[//] Invasive cancer or DCIS within 12 months of a negative screening examination

[¶] Sensitivity to detect invasive cancer or DCIS within 12 mos. of a screening examination.

Table 2

Sensitivity and Specificity of screening mammography by imaging modality and patient characteristics

	FFDM[1] (N=1,475)		SFM[1] (N=2,196)		FFDM[2] (N=296,154)		SFM[2] (N=414,595)	
	N	Sensitivity[3]	N	Sensitivity[3]	N	Specificity[4]	N	Specificity[4]
Age								
40–49 y	301	0.85 (0.74, 0.96)	480	0.83 (0.75, 0.92)	93,797	0.85 (0.85, 0.86)	135,251	0.84 (0.83, 0.84)**
50–59 y	406	0.87 (0.78, 0.96)	635	0.88 (0.81, 0.95)	95,828	0.89 (0.88, 0.89)	132,830	0.87 (0.87, 0.88)**
60–69 y	460	0.93 (0.87, 0.99)	596	0.90 (0.84, 0.96)	66,689	0.91 (0.90, 0.91)	86,939	0.89 (0.88, 0.89)**
70–79 y	308	0.92 (0.85, 0.98)	485	0.88 (0.80, 0.96)	39,840	0.92 (0.91, 0.93)	59,575	0.90 (0.89, 0.91)**
Postmenopausal					296,154		414,595	
No	214	0.84 (0.74, 0.93)	320	0.84 (0.70, 0.97)	70,763	0.85 (0.84, 0.85)	81,268	0.83 (0.82, 0.84)**
Yes	1261	0.91 (0.86, 0.95)	1876	0.88 (0.84, 0.92)	225,391	0.90 (0.90, 0.90)	333,327	0.88 (0.88, 0.88)**
BI-RADS breast density								
Almost entirely fat	151	0.95 (0.89, 1.00)	192	0.94 (0.89, 0.98)	24,567	0.92 (0.90, 0.93)	30,529	0.90 (0.89, 0.91)**
Scattered fibroglandular	522	0.94 (0.89, 0.99)	867	0.90 (0.84, 0.97)*	112,837	0.90 (0.90, 0.91)	173,747	0.88 (0.88, 0.89)**
Heterogeneously dense	677	0.86 (0.79, 0.94)	988	0.86 (0.80, 0.92)	130,958	0.87 (0.86, 0.87)	182,238	0.85 (0.84, 0.85)**
Extremely dense	124	0.75 (0.55, 0.96)	146	0.78 (0.64, 0.93)	27,750	0.89 (0.88, 0.90)	27,930	0.89 (0.88, 0.90)
Tumor Characteristics								
Histology								
Ductal	950	0.89 (0.84, 0.95)	1490	0.88 (0.83, 0.92)	--	--	--	--
Lobular	138	0.86 (0.72, 1.00)	189	0.84 (0.72, 0.96)	--	--	--	--
Mixed Ductal/Lobular	61	0.89 (0.80, 0.98)	94	0.85 (0.76, 0.94)	--	--	--	--
Other	37	0.70 (0.45, 0.94)	100	0.84 (0.76, 0.91)	--	--	--	--
ER Status								
Negative	215	0.82 (0.65, 0.99)	296	0.79 (0.65, 0.93)	--	--	--	--
Positive	1215	0.89 (0.84, 0.95)	1262	0.90 (0.85, 0.94)	--	--	--	--
Grade								
Low	336	0.91 (0.82, 1.00)	454	0.91 (0.83, 0.98)	--	--	--	--
Moderate	624	0.91 (0.85, 0.98)	917	0.90 (0.84, 0.95)	--	--	--	--

	FFDM[1] (N=1,475)		SFM[1] (N=2,196)		FFDM[2] (N=296,154)		SFM[2] (N=414,595)	
	N	Sensitivity[3]	N	Sensitivity[3]	N	Specificity[4]	N	Specificity[4]
High	438	0.89 (0.81, 0.96)	660	0.82 (0.74, 0.91)[*]	--	--	--	--

*
P-values <0.15.

**
P-values <0.0001

[1]Number of breast cancers.

[2]Number of screening mammograms.

[3]All logistic regression models adjusted for age, individual screening facility, exam year and timing of last screen.

[4]All logistic regression models conducted with generalized estimating equations (exchangeable correlation matrix) to account for clustering by screening facility, and adjusted for age, exam year and timing of last screen