Is full field digital mammography equivalent in performance to screen film mammography? UK perspective

Gaurav J Bansal 1, Iain Lyburn 2, Thomas Jones 3

1. The Breast centre, Llandough University Hospital, Llandough, UK. 2. Cheltenham Imaging centre, Cheltenham, UK. 3. Department of Radiology, Taunton and Somerset NHS trust, Somerset, UK.

Correspondence: G J Bansal. Address: The Breast centre, Llandough University Hospital, Llandough, CF642XX, UK. Email: gjbansal@gmail.com

Received: May 15, 2014 Accepted: June 11, 2014 Online Published: June 17, 2014

DOI: 10.5430/ijdi.v1n2p106 URL: http://dx.doi.org/10.5430/ijdi.v1n2p106

Abstract

Aim: The purpose of this project was to compare the performance of full-field digital imaging (FFDM) with screen film mammography (SFM).

Methods: Data from 24,876 and 32,142 women in SFM and FFDM group respectively was collected retrospectively and comparison was made of the recall rate, breast cancer detection rate and positive predictive value (PPV) for the recalls for the two screening modalities, screen film mammography (SFM) and full field digital mammography (FFDM). Difference in radiologic characteristics of the tumours and tumour type (Ductal carcinoma in situ (DCIS) versus invasive) was also evaluated. Statistical software SPSS (PASW18) was used for data analysis with Pearson chi-square and Mann-Whitney-U test used to compare the two groups. A p value of less that 0.05 was considered to indicate a statistically significant difference.

Results: In the year 2008-2009(SFM group), there was a recall rate of 4.5% and a cancer detection rate of 7.5/1000 in the recalled group. For the year 2009-2010 (FFDM group), there was a recall rate of 4.9% and cancer detection rate of 8.8/1000 in the recalled group. There was no statistically significant difference between the recall rate and cancer detection rate between the two groups (p = 0.157). Within the digital group, there was 88(88%) invasive cancer and 12(12%) non-invasive cancer, whereas within the analogue group, there were 74 (92.5%) invasive cancer and 6(7.5%) non-invasive cancer. Most of the non-invasive cancers in the FFDM group were high grade with only one mixed low-intermediate grade. In the analogue group, 50% were high grade, rest being low or intermediate grade. The histological distribution of cancer between two groups was not found to be statistically significant (p = 0.108).

Conclusion: Our study demonstrated FFDM was similar to SFM in terms of cancer detection rate, recall rate and positive predictive value for recalls.

Keywords
Screen-film mammography, Full-field digital mammography, Ductal carcinoma in situ, Breast neoplasms, Cancer screening
1 Introduction

There is continuing controversy about the performance of breast screening. The Independent UK Panel on Breast Cancer Screening published a review [1] in October 2012 which concluded that breast screening extends lives. Allowing for a great deal of uncertainty, there is a 20% reduction in mortality in women invited to screening (Independent UK Panel on Breast Cancer Screening). This is mainly due to its ability to detect subtle abnormalities and microcalcifications that may represent early cancer [1, 2]. There has been replacement of screen-film mammography (SFM) with full field digital mammography (FFDM) in many centres because of consistent image quality of digital images, ability to post-process digital images and improved storage and transmission capabilities [3]. A number of studies have been conducted to compare screen-film with digital mammography and results have shown that digital imaging is as good as SFM in population based screening practices [4-8], although most of these studies rely on surrogate end points such as cancer detection rate, recall rate and tumour characteristics at diagnosis. Few studies have been done within an operational national screening programme [9-12]. Accurate evaluation of risks and benefits of digital imaging in the screening environment is necessary at a time when the value of screening itself is questioned [13, 14].

Digital screening was installed in our National Health Service (NHS) Breast cancer Screening Centre in a single switch over replacing the analogue system overnight, following which there were two unmixed cohorts of patients to compare. We have compared performance one year prior to digital changeover and a year after the changeover. The purpose of this project was to directly compare these two cohorts of patients, in terms of the recall rate, cancer detection rate and positive predictive value for the recalls. It was also the purpose of this study to compare imaging characteristics and histological distribution of the cancers detected in the two groups and to identify any difference whatsoever. Lastly, the lymph-node positivity rate of the two groups was compared.

2 Material and methods

This study was conducted at a NHS Breast Screening centre within the context of an ongoing population based screening programme (National health service Breast screening programme-NHSBSP) for asymptomatic women aged 50 to 70. This programme is based on the European guidelines for quality assurance in mammography screening [15] and their results met the required standards. The centre is covered by general ethical approval of the NHSBSP. The project was granted approval by the Audit Committee. Informed consent was not considered necessary for this retrospective study with anonymised data. Participation was on the basis of written invitation by mail according to information provided by population registry. All invited women have three yearly screening mammograms. Screening examinations performed with either FFDM or SFM included two standard views, mediolateral oblique and cranio-caudal at all screening visits. Image acquisition was done with Molybdenum target and filter (600T; GEHealth care, Buc, France), Hologic CCD-OT Lo-Rad (Hologic, Bedford, Mass) and a Min- R 2000/Min- R 2190 Fast screen film combination(Kodak, Rochester,NY) for SFM. Digital mammograms were acquired with Hologic Selenia-MG (Hologic, Bedford, Mass) systems. Technique factors and breast doses for SFM and FFDM were monitored and found to be compliant with national guidelines and were below the dose that constitutes an acceptable risk as suggested by mammography quality standard act limit for a two view screening mammographic study [15]. The average glandular dose was 2.17mGy for SFM and 1.25mGy for FFDM. The small differences in dose were due to small differences in machine performances over time and variation in the size of breasts. Mammograms were processed with commercially available processing algorithms, which were regularly updated.

All mammograms were read independently by two experienced readers and final decision about recall resolved by arbitration of a third reader. These procedures were not specifically implemented for the purpose of the study and followed NHSBSP guidelines. All readers had experience of reading digital mammograms for at least 2 years prior to reading the digital screening images. Diagnostic assessments were performed by experienced breast imagers with more than 2 years experience of working in a digital radiology environment in a clinical setting before the study started, although none of them had experience of FFDM in screening.
Conventional mammograms were read in a darkened room by using mammograms alternators with a luminance of at least 2500cd/m² with prior mammograms always mounted with current screening mammograms. FFDM cases were interpreted in a dedicated darkened room, with a dedicated mammography workstation equipped with two 5-megapixel displays, which were regularly calibrated, as per departmental protocol. Image manipulation tools including contrast manipulation and image inversion could be used and imagers were able to use the tools for adjustment of brightness or contrast for FFDM images. Computer aided detection (CAD) was not used for FFDM.

Data was retrospectively collected regarding the total number of women screened during the study period and the number of women recalled after screening by doing a National Breast Screening service (NBSS) database search. If recall led to a cancer diagnosis, results of histological examination were recorded. The recall rate was calculated by dividing the number of patients recalled by the number of screening examinations within both groups separately. Cancer detection rates were calculated by dividing the number of cancers detected in recalled women by the number of screening examinations. Positive predictive value for the recalls was calculated by dividing the number of cancers detected by the number of patients recalled. We compared the recall rate, breast cancer detection rate and positive predictive value (PPV) for the recalls for the two screening modalities. Difference in radiologic characteristics of the tumours and tumour type was also evaluated. Finally, lymph-node positivity rate was compared between the two groups. In cancers with multiple imaging characteristics, the predominant imaging feature was used for further evaluation. Statistical software SPSS (PASW18) was used for data analysis with Pearson chi-square test used to compare the two groups-SFM and FFDM for differences in categorical variables (Tumour morphology, pathology, lymph node positivity and tumour grade) and Mann-Whitney U test for continuous variables (size of cancer, recall rate, cancer detection rate and PPV of recall), due to non-parametric distribution of data. A p value of less that 0.05 was considered to indicate a statistically significant difference.

3 Results
In the year 2008-2009 (SFM group), 24,876 women were invited for screening, out of which, 20,557 (78%) attended screening. Out of the women who attended screening, 944 were recalled for assessment and 155 were diagnosed with cancer. This gives a recall rate of 4.5% and a cancer detection rate of 7.5/1000. PPV for the recalls was 16.4%.

For the year 2009-2010 (FFDM group), 32,143 women were invited for screening, out of which 25,088(82%) were screened. Out of the women who attended screening, 1230 were recalled for assessment and 221 were diagnosed with cancer. This gives a recall rate of 4.9% and cancer detection rate of 8.8/1000. PPV for the recalls was 17.9%. There was no statistically significant difference between the recall rate and cancer detection rate between the two groups ($p = 0.157$) (see Figure 1).

![Figure 1. Comparison of recall rate/100 and cancer detection rate/1000 in analogue and digital groups respectively](image-url)
Similarly, there was no statistically significant difference between the PPV for the recalls between the two groups 16.4% versus 17.9% \((p = 0.225)\).

First consecutive 100 patients with cancer within the FFDM group were compared to first consecutive 80 cases with cancer within the SFM group as a representative sample of all cancers detected. These two groups were comparable with respect to the age of the patients with the mean age in the FFDM group being 61.2 years (range 50-72 years) and mean age in the SFM group being 59.7, range 49-73 \((p = 0.169)\). All patients within both groups included initial screening examination as well as subsequent screening examinations. There was equal distribution of patients between the incident and prevalent rounds within the SFM group. In the FFDM group, 82% patients were at their subsequent round of screening.

The radiological characteristics of lesions on the basis of which patients were recalled and a diagnosis of cancer made is shown in Table 1.

**Table 1.** Radiological characteristics of cancers within the two groups, SFM (n = 80) and FFDM (n = 100)

<table>
<thead>
<tr>
<th></th>
<th>SFM No. (%)</th>
<th>FFDM No. (%)</th>
<th>(p) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>47 (58.7%)</td>
<td>49 (49%)</td>
<td>0.066</td>
</tr>
<tr>
<td>Microcalcification</td>
<td>15 (18.7%)</td>
<td>30 (30%)</td>
<td>0.420</td>
</tr>
<tr>
<td>Asymmetric density</td>
<td>5 (6.2%)</td>
<td>11 (11%)</td>
<td>-*</td>
</tr>
<tr>
<td>Distortion</td>
<td>13 (16.2%)</td>
<td>10 (10%)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Note. SFM= Screen film mammography; FFDM= Full-field Digital mammography

* Numbers in this group were too small to give an accurate \(p\) value.

Within the SFM group, 47 patients (58.7%) presented with masses versus 49 (49%) in the FFDM group. This was not found to be statistically significant \((p = 0.066)\). Similarly, there was no statistically significant difference in the number of cancers presenting with micro-calcification between the two groups, 15 (18.7%) in the SFM group versus 30 (30%) in the digital group, \((p = 0.420)\). Numbers presenting with asymmetric densities, 5 (6.2%) in the SFM group versus 11 (11%) in the FFDM group, were too small to give an accurate result of significance. The only significant difference between the two groups was with respect to the number of cases presenting with distortion, 13(16.2%) in the SFM group versus 10 (10%) in the FFDM group \((p = 0.002)\).

There was no difference in the size of cancers between the two groups. Within the digital group, pathological size of invasive cancer was 17.9mm versus 20.5mm in the analogue group \((p = 0.414)\). Similarly there was no statistical difference in the sizes of masses, microcalcification and distortion identified between the two groups. See Table 2.

**Table 2.** Comparison of mean size of cancers in mms between the two groups, SFM (n = 80) and FFDM (n = 100)

<table>
<thead>
<tr>
<th></th>
<th>SFM Mean +/-SD</th>
<th>FFDM +/-SD</th>
<th>(p) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass size (mm)</td>
<td>16.8 +/-2.3</td>
<td>15.06 +/-2.0</td>
<td>0.461</td>
</tr>
<tr>
<td>Microcalcification size (mm)</td>
<td>31.21 +/-4.2</td>
<td>32.12 +/-4.3</td>
<td>0.918</td>
</tr>
<tr>
<td>Distortion size (mm)</td>
<td>20.77 +/-3.5</td>
<td>27.2 +/-3.2</td>
<td>0.31</td>
</tr>
<tr>
<td>Pathological invasive cancer size (mm)</td>
<td>20.5 +/-3.8</td>
<td>17.9 +/-2.2</td>
<td>0.414</td>
</tr>
</tbody>
</table>

Note. SFM= Screen film mammography; FFDM= Full-field Digital mammography
Surprisingly, there were more patients with more than 2 lymph nodes positive in the analogue group when compared to digital (10% versus 1%) and therefore carrying a worse prognosis. This was found to be statistically significant ($p = 0.002$).

We found no statistically significant difference in the tumour characteristics between the two groups. Within the digital group, there was 88% invasive cancer and 12% non-invasive cancer, whereas within the analogue group, there was 92.5% invasive cancer and 7.5% non-invasive cancer. Most of the non-invasive cancers in the FFDM group were high grade with only one mixed low-intermediate grade. In the analogue group, 50% were high grade, rest being low or intermediate grade. Ductal carcinoma formed 80% of cancer in the digital group versus 85% in the analogue group, followed by lobular at 5% in the digital group versus 3.75% in the analogue group. The histological distribution of cancer between two groups was not found to statistically significant ($p = 0.108$). Receptor positivity of invasive cancer was similar between two groups with approximately 70% being ER positive. PR positivity ranged from 29% to 43% and HER positivity ranged from 8.1 to 17.9%. Similarly, there was equal distribution of grades of invasive cancer amongst the two groups with most cancers being grade 2 (58.6% in the digital group versus 47.5% in the analogue group), followed by grade 3 (23.2% in the digital group versus 27.5% in the analogue).

4 Discussion

In this study, we avoided potential biases due to change in screening centre or population, which has hampered some of the previous studies. By making concurrent comparison in the same population within the centre, we avoided variation in the group of readers or changes in their criteria of recall over time, which remained very stable during the study period. As same set of readers were involved in both SFM and FFDM screenings, any biases due to reading skill differences was minimised.

There were initial concerns about the lower spatial resolution of digital mammography leading to limited ability to detect microcalcifications, but these have been rapidly discounted due to higher detective quantum efficiency (DQE) and dynamic range, leading to improved contrast resolution. Digital mammographic imaging screening trial (DMIST) [16] found digital mammography had significant increase in performance in younger women with dense breasts, while film based mammography performed better in women aged 65 years or older and with fatty breasts. Similar results were found by Pisano et al. (10) who found higher detection levels by FFDM in patients less than 50 years with dense breasts. In our study, there was no difference in the age of patients between the two groups with mean age of about 60 years in both groups. Cancer detection rate was 8.8/1000 in the FFDM group versus 7.5/1000 in the SFM group. Although more cancers were detected within the digital group, we found no statistical difference between detection rates of cancer between digital and analogue group. Oslo 2 study [4], showed significantly higher detection rate with FFDM in women between 45 and 69 years of age (5.9 versus 3.8 per 1000 ($p = 0.002$). Studies [5, 6, 8, 17] done in the context of routine screening mammograms showed similar results. Del Turco et al. [5] and Vigeland et al. [6] compared SFM and FFDM in a local population based programme and found statistically higher detection rate for FFDM, which was marked for cancers presenting with microcalcification. The Norwegian study by Vigeland et al. [6] was done over a nine year period and read by different radiologists in 18 different counties. By contrast, in our study, experienced readers remained constant over the period of the study and although more microcalcification was detected in FFDM group in our study, this was not found to be statistically significant. Vinnecombe et al. [17] found similar cancer detection rates for SFM and FFDM using hard copy image reading in a UK breast screening setting. The authors also performed a metaanalysis of published data and showed that FFDM performed as well as conventional SFM.

Karssemeijer et al. [18] found detection rate to be higher, though not statistically significant with FFDM. Significantly more DCIS were found in FFDM with increased detection of DCIS related to better detection of microcalcification, although they found interpretation of micro-calculations more difficult as PPV associated with micro-calcification strongly decreased. In our study we found no statistical difference in the micro-calcification detection rate between the two groups,
although more micro-calcification (30%) was detected in FFDM group versus 18.7% in the SFM group. Furthermore, in our study, although more non-invasive cancers were detected in the FFDM group (12.5%) versus 7.5% in the SFM group, this was not found to be statistically significant. Most of the non-invasive cancers in the FFDM group were high grade with only one mixed low-intermediate grade. In the analogue group, 50% were high grade, rest being low or intermediate grade. With low grade DCIS being one of the presumed reasons for overdiagnosis, our study demonstrated that FFDM does not cause overdiagnosis of low grade DCIS. Various factors could account for these differences. Use of CAD could lead to more sensitive detection of micro-calcifications by Karssemeijer et al. [18]. The fact that previous studies [5, 6] without CAD also reported increased detection of micro-calcifications, suggest that this result should not be attributed to CAD alone. Another factor may be the distribution of patients between the first screening round and subsequent rounds. In our study, although we selected a random group of people from the two cohorts, there were more people having their first screen in the analogue group and more people in their subsequent screening round in the digital group. This could explain the non-significant difference in the detection rate of microcalcification between the two groups as the effect of FFDM on microcalcification detection rate after the first screening round remain small [18]. This may be because low grade DCIS represented as micoclacification could have been already detected at the first screening round. We did not consider worthwhile doing subgroup analysis based on screening round, as numbers were too small within the subgroups to give accurate statistical power.

The proportion of DCIS was also higher in digital mammography group than SFM in a study by Domingo et al. [12], although this difference was not significant. This is in agreement with many previous studies [5, 8] as well as our own study.

Detection of DCIS is crucial as this type of tumour might be immediate precursor of invasive cancer, although 25%-30% of DCIS may not progress to invasive disease [19]. Since we cannot determine for certain, which tumours will progress and which will not, screening mammography may currently be benefiting some women with DCIS and potentially harming some women, whose low grade DCIS would never be associated with subsequent invasive cancer and who are always treated with surgery and adjuvant therapies for lack of good prognostic indicators [20, 21]. In our study, most of the DCIS in the FFDM group was high grade.

In our study, the difference in the recall rate between the two cohorts was not found to be statistically significant. There was 4.9% recall rate in the FFDM group versus 4.5% in the SFM group. This recall rate is well below the NHSBSP target of less than 7% for prevalent screening and 5% for incident screening examinations. Recall rate was higher for FFDM group in three [4, 5, 8] of the four studies showing higher detection for FFDM [4, 5, 8, 16]. Skaane et al. [4] had very low recall rate for SFM group, which might overestimate FFDM-SFM difference in detection rate. Recall rate was similar for two groups in a UK study by Vinnicombe et al. [17], which might explain the lack of FFDM-SFM detection rate. Karssemeijer et al. [18] had significantly higher recall rates with FFDM (4.4% versus 2.3%), which however led to lower PPV of recall rate. This was especially significant for micro-calcifications at the time of first screening. We did not find this to be true in our study, where PPV for recalls was in fact higher for FFDM group at 17.9% versus 16.4% for the SFM group. In our study, we found significantly more cases presenting with distortion, 13(16.2%) in the SFM group versus 10 (10%) in the FFDM group. We have not found many studies so far which shows difference in detection rate of distortion between SFM and FFDM, apart from a study by Hambly et al. [8] which showed higher detection rate of FFDM secondarily to improved detection of distortion and microcalcification. This discrepancy could be related to small numbers presenting with distortion. With increasing use of digital tomosynthesis in screening, cancers presenting as distortion will evolve, as tomosynthesis has been shown to be better at detecting distortion.

In our study, more cancers presented with positive lymph nodes in the analogue group, the reasons for which are difficult to elucidate and may be many fold, but these results are in favour of digital mammographic screening. Although invasive cancer size was in general bigger in the analogue group (20.5mm versus 17.9 mm), this was not statistically significant.
Skaane [22] has reviewed studies comparing digital with SFM. He compared 10 studies and found overall higher cancer detection rates with FFDM approaching borderline significance. For some sub-groups, including women with dense breasts and those presenting with microcalcification, this is close to statistically significant. However, his meta-analysis found no difference in the positive predictive value of recalls between FFDM and SFM.

One potential limitation of our study is lack of assessment of false negatives and interval cancers. However, we do not think this would have influenced the results as the main premise of the study is cancer detection rate, and although we cannot evaluate for false negatives, this is true for both cohorts of patients. Another important limitation, as elucidated before was the unequal distribution between the first screening round and subsequent rounds between the two groups of women, with more women in their subsequent rounds in the FFDM group. This could lead to potential bias in our microcalcification detection rate between two groups. We have slightly overcome this issue by making sure the age distribution is homogenous across both groups with mean age of women within the two groups being comparable. However, this potential bias must be borne in mind when interpreting results. Another limitation of our study is the lack of use of CAD, as we believe this will become standard practice in screening programmes especially in detection of clustered micro-calcifications. However measurement of CAD performance was not a subject of this study. Also, following the switch to the digital technology, there may have been a learning curve, which was not excluded from the analysis. However, since our readers have been working in a digital environment before, we consider this not to be a limitation. In our study, we have got rid of confounding variable which might affect results in a cohort study, making sure women who underwent FFDM were not different to women who had SFM, as all women who attended screening after 5th October, 2009 underwent digital imaging. However, it is not possible to completely exclude unknown confounding. In terms of evaluating the specific imaging features reviewing a proportion rather than the entire group of cancers in each group introduces a potential variable, however selection of the cases was completely random and there is unlikely to be any significant selection bias.

Another potential limitation of this study is that, we did not record breast density and did not include patients less than 50 years in our study (the group that has been shown to benefit most from FFDM by Pisano et al. [16]. Our study result therefore includes a broader category of women, and shows that a broader category of women might benefit from the use of digital mammography in breast screening.

In summary, our study demonstrated FFDM was as good as SFM in terms of cancer detection rate, recall rate and positive predictive value of recall rate. There were more cancers presenting as distortion in the SFM group, which might evolve with increased use of digital tomosynthesis in the screening setting and should be subject of further studies.

References


