HHS Public Access

Author manuscript

Eur Radiol. Author manuscript; available in PMC 2017 March 01.

Published in final edited form as:

Eur Radiol. 2016 March ; 26(3): 874–883. doi:10.1007/s00330-015-3885-y.

Breast cancer detection rates using four different types of mammography detector

Alistair Mackenzie, PhD[a,b,*], **Lucy M. Warren, PhD**[a,b], **Matthew G. Wallis, MD, MB, ChB, FRCR**[c], **Julie Cooke, MB, BS, MRCP, FRCR**[d], **Rosalind M. Given-Wilson, MA, MB, BS, MRCP, FRCR**[e], **David R. Dance, PhD**[a,b], **Dev P. Chakraborty, PhD**[f], **Mark D. Halling-Brown, PhD**[g], **Padraig T. Looney, PhD**[a], and **Kenneth C. Young, PhD**[a,b]

[a]National Coordinating Centre for the Physics in Mammography, Royal Surrey County Hospital, Guildford, GU2 7XX, UK

[b]Department of Physics, University of Surrey, Guildford GU2 7XH, UK

[c]Cambridge Breast Unit, Cambridge University Hospitals NHS Foundation Trust, Cambridge & NIHR Cambridge Biomedical Research Centre, Cambridge, UK

[d]Jarvis Breast Screening and Diagnostic Centre, Guildford, UK

[e]Department of Radiology, St George's Healthcare NHS Trust, London, UK

[f]Department of Radiology, University of Pittsburgh, Pittsburgh, USA

[g]Scientific Computing, Department of Medical Physics, Royal Surrey County Hospital, Guildford, UK

Abstract

Objectives—To compare the performance of different types of detectors in breast cancer detection.

Methods—A mammography image set containing subtle malignant non-calcification lesions, biopsy-proven benign lesions, simulated malignant calcification clusters and normals was acquired using amorphous-selenium (a-Se) detectors. The images were adapted to simulate four types of detectors at the same radiation dose: digital radiography (DR) detectors with a-Se and caesium iodide (CsI) convertors, and computed radiography (CR) detectors with a powder phosphor (PIP) and a needle phosphor (NIP). Seven observers marked suspicious and benign lesions. Analysis was undertaken using jackknife alternative free-response receiver operating characteristics weighted figure of merit (*FoM*). The cancer detection fraction (CDF) was estimated for a representative image set from screening.

Results—No significant differences in the *FoMs* between the DR detectors were measured. For calcification clusters and non-calcification lesions, both CR detectors' *FoM*s were significantly lower than for DR detectors. The calcification cluster's *FoM* for CR NIP was significantly better

[*]Correspondence author: Alistair Mackenzie, NCCPM, Medical Physics, Level B, St Luke's Wing, Royal Surrey County Hospital, Guildford, GU2 7XX, United Kingdom, alistairmackenzie@nhs.net, tel: +44(0)1483-571122 ext, 2459 fax: +44(0)1483 406742.

than for CR PIP. The estimated CDFs with CR PIP and CR NIP detectors were up to 15% and 22% lower respectively than for DR detectors.

Conclusion—Cancer detection is affected by detector type and the use of CR in mammography should be reconsidered.

Keywords

Breast cancer screening; Digital mammography; Computed radiography; image quality; observer study

INTRODUCTION

Prospective (1, 2) and retrospective reviews (3–10) of screening studies have been undertaken to compare the clinical performance of digital imaging systems and screen film mammography (SFM). The switch from SFM is now permanent, but only a few retrospective screening review studies have compared cancer detection performance with different digital detectors (6–10). There are a number of different digital detectors for mammography on the market, and these have a wide range of image quality as measured by physical metrics (11). Two retrospective studies (9, 10) demonstrated that the cancer detection rate using powder image plate (PIP) computed radiography (CR) was lower than with digital radiography (DR) systems. In particular there were large differences in the detection of in-situ cancers. A third retrospective study (8) claimed comparable detection rates for CR PIP with DR but at a cost of 60% higher mean glandular dose (MGD) for CR. To date no clinical studies have reported on the effect on cancer detection of using the potentially better needle image plate (NIP) CR technology.

Judging the effect of using different types of imaging detector on cancer detection from retrospective reviews of data from breast cancer screening programmes is difficult because of the large number of confounding factors. These include differences in women, radiologists, X-ray equipment, dose and radiographic practice. Prospective studies might be better but would still have many confounding factors and take years to complete because of the low prevalence of breast cancer in screening.

A faster clinical evaluation method for digital mammography detectors with fewer confounding factors is required. To address this, the present work comprises an observer study based on a method (12) to adapt one set of mammograms to appear with the image quality of a range of detectors: DR detectors with amorphous-selenium (a-Se) and caesium iodide (CsI) convertors, and CR detectors with PIP and NIP phosphors. This study builds on a previous study of this type (13) with improved simulation, use of both breasts, wider range of cancer types and improved reporting software with the purpose of comparing the performance of different types of detectors in the detection of breast cancers.

MATERIALS AND METHODS

The study protocol was approved by the regional research ethics committee. Individual consent was not required as the study used anonymous mammographic images and data.

Image selection

Mammograms were collected from one screening centre (14), where women between the ages of 50 and 70 years old are invited for mammography screening triennially. During screening, both breasts are imaged using mediolateral oblique (MLO) and craniocaudal (CC) views. The centre has five Hologic Selenia and two Hologic Dimensions X-ray units (Hologic inc., Bedford, USA). From March 2011, all available mammograms with a malignancy and a random selection of biopsy-proven benign lesions and normals were collected to create a database. An expert radiologist, who did not participate in the study, marked a region of interest (ROI) around the cancers and benign lesions in the database and classified their appearance (mass, distortion, focal asymmetry, calcification clusters) and conspicuity (very subtle, subtle, or obvious) in each view. The images used in this study were acquired between April 2011 and December 2012. The cases and images were chosen as follows:

Normal images—The database contained 580 cases which had been classified as not requiring further investigation. 80 cases were randomly selected from the database and classified as normals.

Insertion of calcification clusters into images—89 malignant calcification clusters were inserted into a further 80 pairs of normal images randomly selected from the database. There was a maximum of three clusters inserted per case. The original clusters were extracted from images of mastectomy slices acquired at high magnification in a specimen cabinet. The process of extraction and insertion of calcification clusters has been validated such that the calcifications matched real malignant clusters in image quality and appearance (15, 16). The locations for the insertions were chosen to ensure that the calcifications were seen in a wide range of glandularities. The expert radiologist visually reviewed the final images for realism of the location and appearance of the calcification clusters.

Images containing malignant lesions (non-calcification)—80 pairs of images were selected from the database containing 459 cases with marked non-calcification lesions. Each selected case contained at least one biopsy-proven malignant lesion that was described as subtle or very subtle in the view chosen. Ideally, simulated lesions would have been used to remove any selection bias. However, the full range of these types of lesions cannot be accurately simulated.

Images containing biopsy-proven benign lesions—Twenty nine pairs of images with biopsy-proven benign lesions were added to the image set.

In total 269 cases were used in this study. For each selected case, one view (either CC or MLO) of both breasts was randomly selected for the study. For 44 of the non-calcification cases, only one view met the subtlety criterion and was selected. Only one view from each case was used in this study as realistic insertion of calcifications into two views was not possible.

Due to time constraints, the study was started before the women whose images had been used in the study had returned for their three yearly breast screen examination. Therefore,

the cases had not been confirmed as cancer free. Subsequently, 134 out of the 160 women classified as not having cancer and 21 out of the 29 women with a biopsy-proven benign lesion returned for routine screening. Two cases used as normals and one case used for insertion of calcification clusters were found to have a cancer at the subsequent screening. These three cases were removed from the analysis.

Image conversion

A validated methodology (12, 17) was used to adjust the 'for processing' images to appear as if they had been acquired using four detector types: a-Se photoconductor, CsI phosphor, CR NIP and CR PIP detectors (table 1).

The simulation methodology adjusts the sharpness of the image using measured modulation transfer functions. Extra noise is added to the images to match the required image quality. The extra noise is calculated from noise power spectra measurements, taking into account the pixel values in the unprocessed image since this is related to the energy absorbed in the detector at that point. Other factors also included are the tube voltage, anode/filter combination and compressed breast thickness. The pixel pitch in the simulated images is the same as the original images, corresponding to a detector pixel pitch of 70μm. The final image quality is representative of the type of detector being simulated rather than being an exact match for a specific detector. The characterization of the image quality of these study arms is published elsewhere (18). To be able to adjust the images to simulate the CsI Arm, it was necessary during simulation to apply a dose reduction of 20% to all images. The MGD for the original images for breast thickness between 50 and 60mm was 1.36mGy (19), which was reduced to 1.08mGy for all arms of the study, which is still within the range of clinical doses (20). The images in the study were post-processed using 'MUSICA^2' (Agfa Healthcare NV, Mortsel, Belgium) image processing. This software is designed to work for a variety of detectors and dose levels and was suitable for each study arm without any software adjustments. Figures 1 and 2 show examples of the appearance of non-calcification lesions and calcification clusters in the four study arms.

Observer study

Seven observers were recruited from two hospitals. Each observer was an accredited reader working in the NHS breast screening programme with between 3 and 21 years' experience and reading over 3,500 screening mammography cases annually.

Images were reviewed on mammography workstations equipped with two 5 megapixel monitors: Barco MDMG (Barco NV, Kontrijk, Belgium) at one hospital and Eizo Radiforce GS520 monitors (Eizo, Ishikawa, Japan) at the other hospital. Each case was shown to the readers using MedXviewer software (21) in the following order. [1] Both images were displayed de-magnified to fit on one monitor; [2] The images were then scaled to fit one per monitor; [3] The images were shown in quadrant zoom at magnification 1.2. Panning the image and use of a magnification tool were the only controls available to the observers, as in clinical practice other tools are not commonly used.

To reduce bias due to learning effects each observer viewed the images in a different order with a minimum interval of 14 days before seeing the same case again at a different image

quality. The observers were not informed of the underlying purpose of the study or of the prevalence of cancers or number of lesions per case. The observers were asked to mark any suspicious or benign lesions. The locations of the marks and the responses were automatically recorded. The following question was posed for each lesion marked:

Indicate whether you would recall this patient on the basis of this lesion, along with your confidence in this decision, i.e. your confidence that a recalled lesion will be malignant

1. No, very confident
2. No, moderately confident
3. No, slightly confident
4. Yes, slightly confident
5. Yes, moderately confident
6. Yes, very confident

Data analysis of observer study

A mark was considered a lesion localisation if it was within the ROI of the malignant lesion defined by the expert radiologist. Marks outside of the ROIs or within the ROI of a benign lesion were classed as non-lesion localisations. The data were analysed using JAFROC-4.2 software (www.devchakraborty.com) (22). The analysis was performed for random observers and random cases, i.e. the observers and cases were assumed to be random samples from their respective populations. As part of the jackknife alternative free-response receiver operating curves (JAFROC) analysis, a global F-test using the null hypothesis that all the arms are equal was undertaken. Observer performance was characterised by the equally weighted JAFROC figure of merit (*FoM*). Significance testing was performed using Dorfman-Berbaum-Metz analysis of variance (23).

The study question splits between recall (rating ≥ 4) and not recall (rating ≤ 3) and provides a natural operating point for investigating the correctly or incorrectly recalled cancers. The false recall fraction (FRF) was calculated as the fraction of cases (without cancers) with at least one non-lesion localisation mark with a rating of 4 or greater. The cancer detection fraction (CDF) was calculated as the fraction of cancers with a lesion localisation mark with a rating of 4 or greater.

The cancer images in this study are a selected subset from the database of screen detected cancers. It is of interest to generalise the study results to the whole image database by estimating the CDF_{db}, if the study had been undertaken using all cases in the image database. The CDF_{db} was estimated using the measured CDFs, the proportion of calcification clusters and non-calcification lesions in the database and an estimate of the CDF for the type of cancers not represented in the study. By March 2014, the image database contained 1381 annotated screen detected malignant lesions (456 calcification clusters (33%) and 925 non-calcification lesions (67%)). Warren *et al.* (16) showed that the inserted calcification clusters were representative of approximately 60% of calcification clusters seen in the image database. The non-calcification cancers used in the study were

judged to be equivalent to 26% (241/925) of non-calcifications cancers in the database on the basis of conspicuity grading (i.e. subtle or very subtle in both views). The CDF for the lesions in the database not represented in this study (40% of calcification clusters and 74% of non-calcifications) were estimated using the alternate assumptions that they are all detected irrespective of the image quality or at the same rate as in the study. A range in the CDF_{db} was then estimated using the above information.

RESULTS

Observer study

The results of the global F-test gave *F*-values equal to 18.1 (degrees of freedom (DF) of 3 and 46) and 7.4 (DF of 3 and 43) for the calcification clusters and non-calcification lesions respectively. Therefore, there are significant differences between the study arms for both calcification clusters ($P<0.0001$) and non-calcification lesions ($P=0.0004$). Figures 3a and 3b show the alternative free-response receiver operating characteristics (AFROC) curves for the marking and scoring of calcification clusters and non-calcification lesions and figures 3c and 3d show their corresponding JAFROC *FoM*. The percentage differences between the JAFROC *FoM*s are shown in table 2. For the calcification clusters, there were significant differences between each of the arms except between the a-Se and CsI arms. For the non-calcification lesions the only significant differences were between the DR and the CR detectors.

The FRF for all marks and the CDF for calcification clusters and non-calcification lesions for each arm are shown in figures 4 and 5 respectively. The DR arm was created by averaging the results of the a-Se and CsI arms, as there was no significant difference between these arms. The percentage differences for the two CR arms compared to the DR arms for FRF and CDF are shown in tables 3 and 4 respectively. The FRF for the CR PIP detector was significantly lower than that for either DR detector, while there was no significant difference in the FRF for the CR NIP detector compared to the DR detectors.

Table 4 shows that the fraction of cancers detected was significantly lower for CR technologies compared to DR technologies, except between a-Se and CR NIP arms for non-calcification lesions. The CDFs of calcification clusters were more affected by detector type than the non-calcifications lesions.

Table 4 also shows the estimated effect on CDF_{db} for a representative set of screening images for CR PIP and CR NIP compared to DR. The maximum estimated differences in cancer detection in screening were 15% and 22% lower for CR NIP and CR PIP respectively compared to DR at the same dose. The CDF_{db} for CR NIP was up to 11% higher than CR PIP (Table 5), which was almost entirely due to differences in detection of calcification clusters.

DISCUSSION

This study examined types of detectors that are used internationally in breast screening programs. We have found significant differences between the detectors for the detection of

both calcification clusters and non-calcification lesions without a change in dose. Overall the DR detectors perform significantly better than CR NIP which in turn is significantly better than CR PIP.

The estimated CDF for CR PIP was estimated to be between 10% and 22% lower than for DR for a representative set of screening images. It is impractical to run an observer study that exactly matches screening, but there is evidence that the performance of experienced mammography readers in observer studies correlates with their performance in screening (24). Two retrospective screening studies (9, 10) calculated the cancer detection rate (CDR) for two detector types: DR (equivalent to combination of a-Se and CsI arms) and CR PIP. They showed that the cancer detection rate for using CR PIP was 31% (9) and 23% (10) lower than for using DR detectors. The upper estimate of the measurements for CR PIP arm is lower than these results, but it does provide evidence that the major cause of reduced cancer detection for CR PIP was due to the detector type used. Some large differences in cancer detection were seen in retrospective screening studies (8–10) although they were not always significant. Our study was more sensitive and significant differences were measured by using a set of subtle cancers and by avoiding differences between the patient samples that occurred in other studies. The effect of using CR NIP detectors in breast screening has not been published previously. Our study provides evidence the CDR of CR NIP is expected to be better than that for CR PIP but less than for DR without a change in MGD for each detector. Our study examines cancer detection and does not investigate whether there is a clinical benefit to the higher cancer detection rate associated with DR systems. However, Weigel *et al* (25) have shown that where there were variations in overall cancer detection rates in screening, this was due to differences in all types of cancers and not just in low risk groups. This suggests that high cancer detection in screening is beneficial.

Warren *et al* (13) have previously shown a relationship between calcification detection and image quality as measured using the CDMAM contrast detail test object as described in European Guidelines (26). This is expected as the detection of the details in the phantom and calcifications are affected by noise associated with the detector. The non-calcification lesions are much larger than individual calcifications and detection of these lesions is likely to be affected more by anatomical structure than the noise associated with the detector and so a strong relationship with standard quality assurance test objects is not expected (27,28).

Although the study was designed to examine cancer detection, it was of interest to study the FRF. The FRF was high due to the decreased specificity of readers in observer studies (29) and as 27% (29/107) of the cases without cancers contained a biopsy-proven benign lesion. The FRF for CR PIP was significantly lower than for other detectors. It is likely that the relatively poor image quality for CR PIP reduced the visibility of suspicious areas rather than aiding the observer to distinguish between malignant and benign lesions.

In this study there was no difference in the doses between the study arms. A previous study (13) showed that the detection of calcification clusters is sensitive not only to the detector type but also to dose. Therefore, it is expected that increasing the MGD will increase calcification detection. Table 6 shows the difference in dose between the DR and CR PIP imaging systems in the literature (8–10). The results in the table indicate that dose for CR

PIP should be at least 60% higher than for DR. Even at higher dose the detection of in-situ cancers was 27% lower compared to DR (8). The data from Warren et al (13) suggest that the dose should be more than doubled when using CR PIP compared to using an a-Se detector for calcification detection. It is expected that a smaller dose increase would be required for CR NIP for equivalent cancer detection with DR detectors. The evidence shows that CR must be used at higher doses than DR, but it was not the aim of this study to estimate an appropriate dose.

Our study was different from clinical practice as only one view was used, no previous images were available and the image set was enriched with cancers. There were many cancers marked that did not appear sufficiently suspicious for the observers to classify them for recall. The use of a second view could have increased the number of cancers correctly recalled and reduced the number of false positives. Image processing must be used to ensure that the image quality is adequate for reporting. 'MUSICA^2' processing was used for all study arms as it is flexible and as an attempt to reduce differences between the arms in this study not due to the detector. Other manufacturers have their own specific processing, which may affect cancer detection. There is a potential for selection bias in this study as all of the original non-calcification lesions were found during screening using Hologic imaging systems. However, the bias may be reduced as the appearance of the a-Se arm images were changed from that originally seen during screening, due to a change in dose and image processing. Also many of these lesions would have been primarily identified using the other view where the lesion was classified with a conspicuity of 'obvious'. The calcifications clusters do not have a selection bias. A study weakness is that the radiographic factors are the same for the simulated images as the original images, which may not be optimal for other detectors. The image quality levels used were for well set-up equipment and so this study does not account for deterioration and artefacts that may occur. For example, more technically inadequate mammograms were found using CR PIP compared to DR or SFM (10) and distracting dust artefacts were noted in CR PIP (11). Ultimately, it must be realised that the study arms are representative of detectors types rather than specific detectors. Also, the x-ray system, radiographic factors and image processing used with the detector will also influence cancer detection.

The study compared the difference in performance between detector types using one set of clinical images and a clinically relevant question, while confounding factors such as differences in women, readers, X-ray systems, post-processing and anti-scatter grid were removed. We have shown significant differences in the detection of lesions between detector types when the same dose was used. The detector type and its potential effect on cancer detection must be considered during the procurement process for systems. The use and set up of CR imaging systems in mammography should be reconsidered.

Acknowledgments

The scientific guarantor of this publication is Prof Kenneth Young. The authors of this manuscript declare relationships with the following companies: Agfa Healthcare, Carestream Health Inc, MIS healthcare. This study has received funding by Cancer research-UK (CRUK) – Optimam2 grant, number: C30682/A17321. One author (DPC) was supported in part by a grant from the Department of Health and Human Services, National Institutes of Health (R01-EB005243). One of the authors has significant statistical expertise. Institutional Review Board approval was obtained. Written informed consent was waived by the Institutional Review Board. Some study

subjects or cohorts have been previously reported in 'Warren LM, Given-Wilson RM, Wallis MG, et al. (2014). The effect of image processing on the detection of cancers in digital mammography. Am J Roentgenol; 203:387–393'. Methodology: retrospective, experimental study, multicenter study.

Abbreviations and acronyms

CDF	cancer detection fraction
CR	computed radiography
DF	degrees of freedom
DR	digital radiography
FoM	figure of merit
FRF	false recall fraction
JAFROC	jackknife alternative free-response receiver operating characteristics
MGD	mean glandular dose
NIP	needle image plate
PIP	powder image plate
ROI	region of interest
SFM	screen film mammography

References

1. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005; 353:1773–1783. [PubMed: 16169887]
2. Skaane P, Hofvind S, Skjennald A. Randomized trial of screen-film versus full-field digital mammography with soft-copy reading in population-based screening program: follow-up and final results of Oslo II study. Radiology. 2007; 244:708–717. [PubMed: 17709826]
3. Vigeland E, Klaasen H, Klingen TA, Hofvind S, Skaane P. Full-field digital mammography compared to screen film mammography in the prevalent round of a population-based screening programme: the Vestfold County Study. Eur Radiol. 2008; 18:183–191. [PubMed: 17680246]
4. Bluekens AM, Holland R, Karssemeijer N, Broeders MJ, den Heeten GJ. Comparison of digital screening mammography and screen-film mammography in the early detection of clinically relevant cancers: a multicenter study. Radiology. 2012; 265:707–714. [PubMed: 23033499]
5. Van Ongeval C, Van Steen A, Putte GV, et al. Does digital mammography in a decentralized breast cancer screening program lead to screening performance parameters comparable with film-screen mammography? European radiology. 2010; 20:2307–2314. [PubMed: 20455065]
6. Weigel S, Berkemeyer S, Girnus R, Sommer A, Lenzen H, Heindel W. Digital mammography screening with photon-counting technique: can a high diagnostic performance be realized at low mean glandular dose? Radiology. 2014; 271:345–355. [PubMed: 24495234]
7. Keavey E, Phelan N, O'Connell AM, et al. Comparison of the clinical performance of three digital mammography systems in a breast cancer screening programme. Br J Radiol. 2012; 85:1123–1127. [PubMed: 22096222]
8. Bosmans H, De Hauwere A, Lemmens K, et al. Technical and clinical breast cancer screening performance indicators for computed radiography versus direct digital radiography. Eur Radiol. 2013; 23:2891–2898. [PubMed: 23689308]

9. Chiarelli AM, Edwards SA, Prummel MV, et al. Digital compared with screen-film mammography: performance measures in concurrent cohorts within an organized breast screening program. Radiology. 2013; 268:684–693. [PubMed: 23674784]
10. Séradour B, Heid P, Estève J. Comparison of direct digital mammography, computed radiography and screen film in the French national breast screening program. Am J Roentgenol. 2014; 202:229–236. [PubMed: 24370149]
11. Yaffe MJ, Bloomquist AK, Hunter DM, et al. Comparative performance of modern digital mammography systems in a large breast screening program. Med Phys. 2013; 40:12191501-10.
12. Mackenzie A, Dance DR, Workman A, Yip M, Wells K, Young KC. Conversion of mammographic images to appear with the noise and sharpness characteristics of a different detector and x-ray system. Med Phys. 2012; 39:2721–2734. [PubMed: 22559643]
13. Warren LM, Mackenzie A, Cooke J, et al. Effect of image quality on calcification detection in digital mammography. Med Phys. 2012; 39:3202–3213. [PubMed: 22755704]
14. Patel MN, Looney PT, Young KC, Halling-Brown MD. Automated collection of medical images for research from heterogeneous systems: trials and tribulations. Proc SPIE. 2014; 90390C: 90390C-1–7.
15. Warren LM, Green FH, Shrestha L, Mackenzie A, Dance DR, Young KC. Validation of simulation of calcifications for observer studies in digital mammography. Phys Med Biol. 2013; 58:N217–N228. [PubMed: 23880732]
16. Warren LM, Dummott L, Wallis MG, et al. Characterisation of screen detected and simulated calcification clusters in digital mammograms. IWDM 2014; LNCS. 2014; 8539:364–371. [PubMed: 25672453]
17. Mackenzie A, Dance DR, Diaz O, Young KC. Image simulation and a model of noise power spectra across a range of mammographic beam qualities. Med Phys. 2014; 41:1219011-14.
18. Mackenzie A, Warren LM, Dance DR, et al. Using image simulation to test the effect of detector type on breast cancer detection. Proc SPIE. 2014; 9037:90370I-1-14.
19. Dance DR, Skinner CL, Young KC, Beckett JR, Kotre CJ. Additional factors for the estimation of mean glandular breast dose using the UK mammography dosimetry protocol. Phys Med Biol. 2000; 45:3225–3240. [PubMed: 11098900]
20. Oduko JM, Young KC, Burch A. A survey of patient doses from digital Mammography systems in the UK in 2007 to 2009. IWDM 2010, LNCS. 2010:365–370.
21. Looney PT, Mackenzie A, Young KC, Halling-Brown MD. MedXViewer: an extensible web-enabled software package for medical imaging. Proc SPIE. 2014; 9037:90371K-1-7.
22. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: Modelling, analysis and validation. Med Phys. 2004; 31:2313–2330. [PubMed: 15377098]
23. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol. 1992; 27:723–731. [PubMed: 1399456]
24. Soh BP, Lee W, McEntee MF, et al. Screening mammography: test set data can reasonably describe actual clinical reporting. Radiology. 2013; 268:46–53. [PubMed: 23481165]
25. Weigel S, Heindel W, Heidinger O, Berkemeyer S, Hense HW. Digital mammography screening: association between detection rate and nuclear grade of ductal carcinoma in situ. Radiology. 2014; 271:38–44. [PubMed: 24475843]
26. European Commission. European guidelines for quality assurance in breast cancer screening and diagnosis. EUREF 4th edition, European Commission; Brussels, Belgium. 2006.
27. Huda W, Ogden KM, Scalzetti EM, Dance DR, Bertrand EA. How do lesion size and random noise affect detection performance in digital mammography? Acad Radiol. 2006; 13:1355–1366. [PubMed: 17070453]
28. Saunders RS Jr, Baker JA, Delong DM, Johnson JP, Samei E. Does image quality matter? Impact of resolution and noise on mammographic task performance. Med Phys. 2007; 34:3971–3981. [PubMed: 17985642]
29. Evans KK, Birdwell RL, Wolfe JM. If you don't find it often, you often don't find It: why some cancers are missed in breast cancer screening. PloS one. 2013; 8:1–6.

Key points

1. The type of mammography detector can affect the cancer detection rates
2. CR detectors performed worse than DR detectors in mammography
3. Needle phosphor CR performed better than powder phosphor CR.
4. Calcification clusters detection is more sensitive to detector type than other cancers.

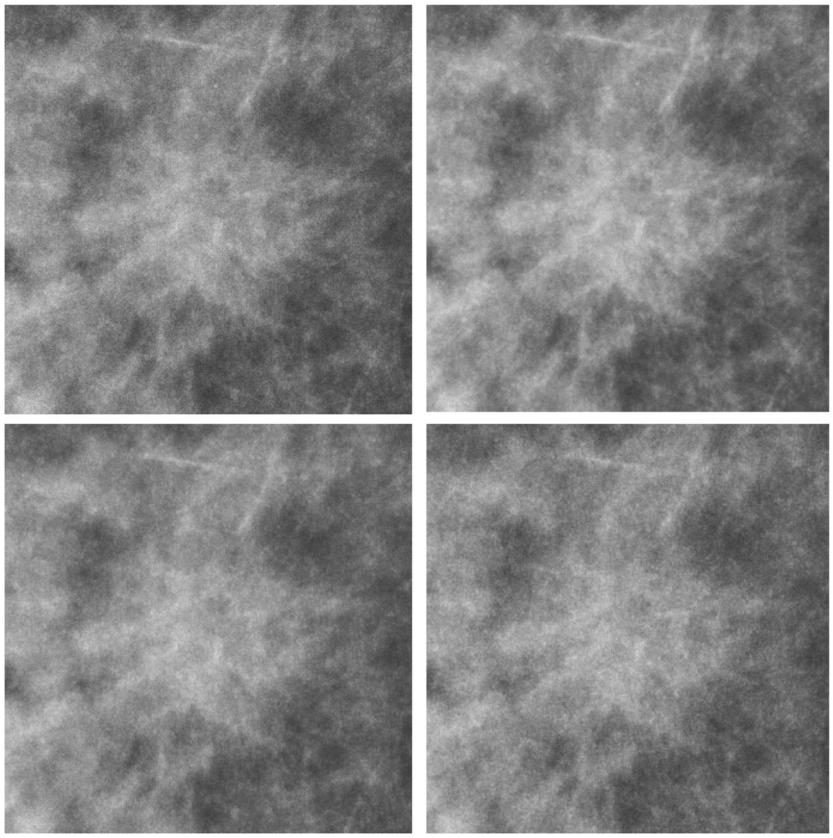

Figure 1.
Example image of a real mass (non-calcification lesion) converted to appear for the 4 arms of the study. Top left: a-Se Arm; Top right: CsI Arm; Bottom left: CR NIP Arm; Bottom right: CR PIP Arm.

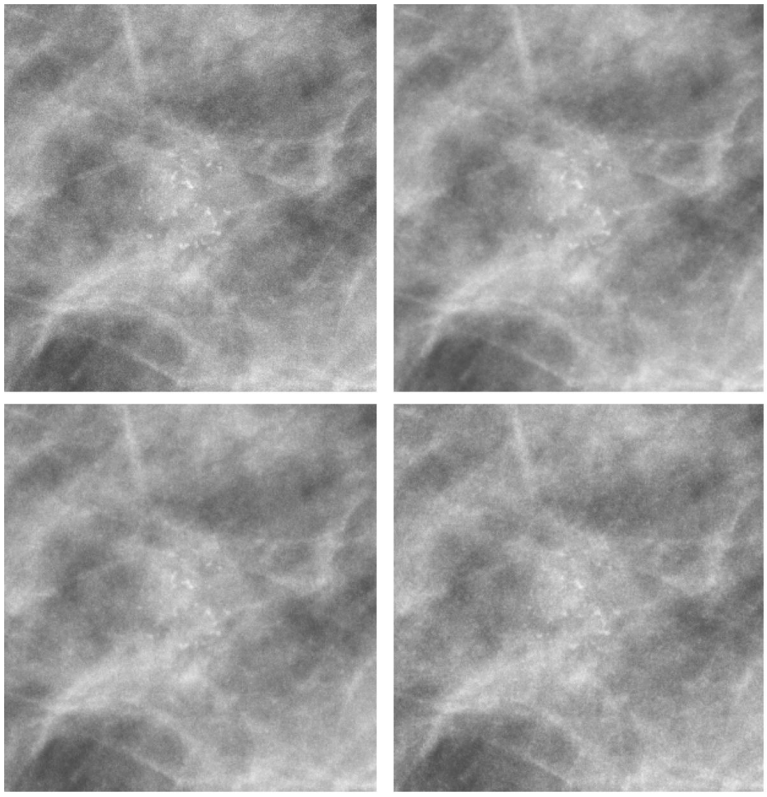

Figure 2.
Example image of an inserted calcification cluster converted to appear for the 4 arms of the study. Top left: a-Se Arm; Top right: CsI Arm; Bottom left: CR NIP Arm; Bottom right: CR PIP Arm.

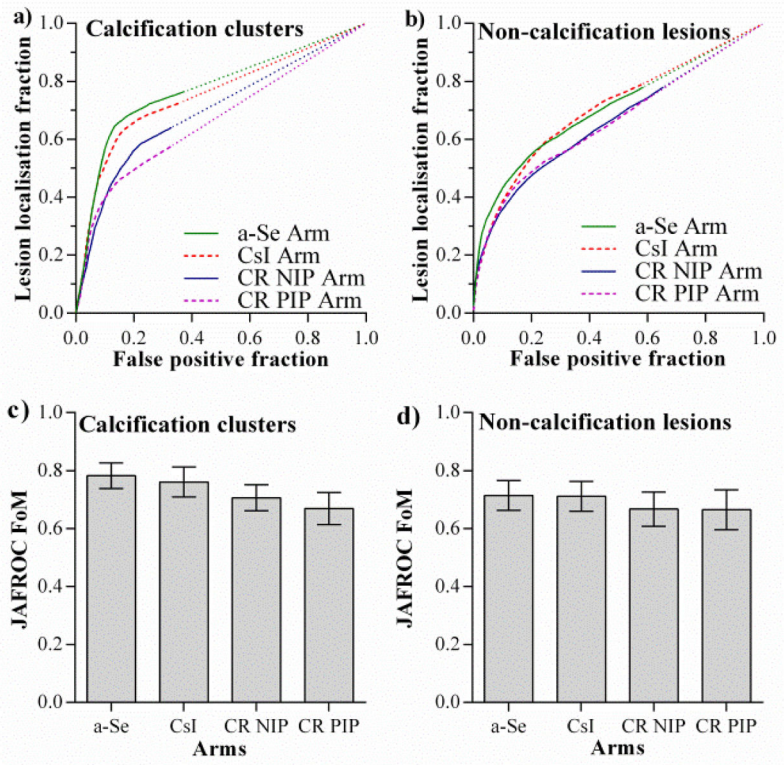

Figure 3.

Reader-average AFROC curves for a) calcification clusters and b) non-calcification lesions. Solid and dashed lines = measured data, dotted lines = straight line extension to (1,1), the area under which is included in the *FoM*. JAFROC *FoM* for calcification clusters and non-calcification lesions. c) & d) JAFROC *FoM*s for each arm combination. Error bars indicate 95% confidence intervals.

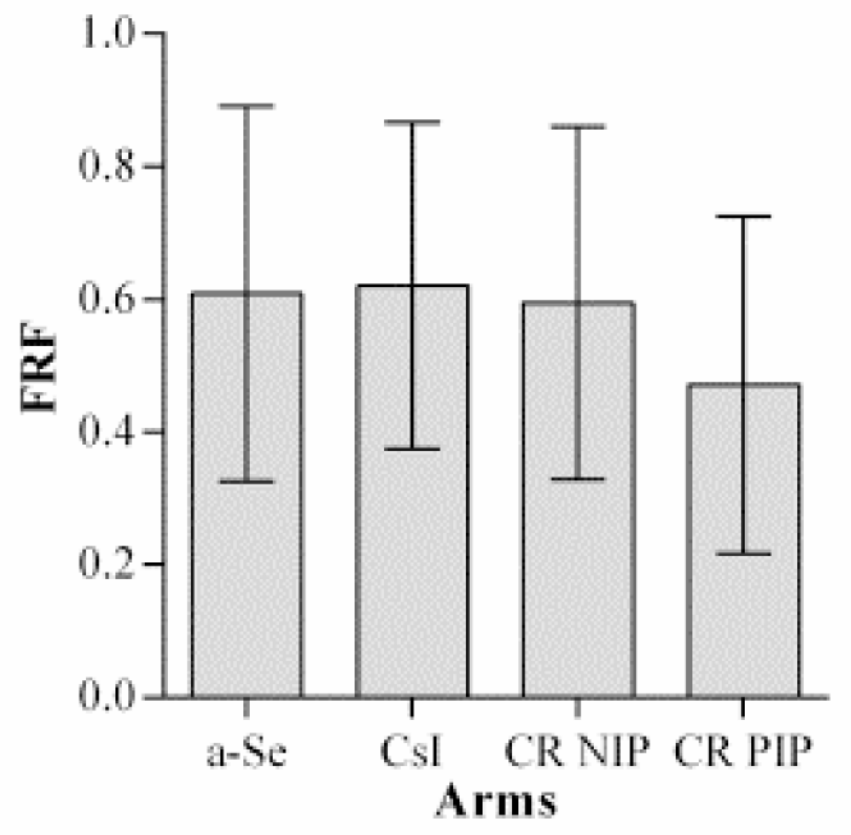

Figure 4.
False recall fraction (FRF) for all non-lesion localisation marks with a rating 4 or greater in cases without cancers. Error bars indicate 95% confidence intervals.

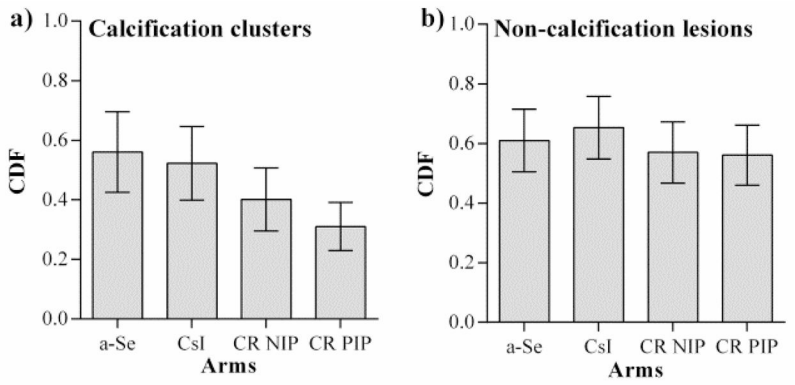

Figure 5.

Cancer detection fraction (CDF) for lesion localisation marks with a rating of 4 or greater for a) calcification clusters and b) non-calcification lesions. Error bars indicate 95% confidence intervals.

Table 1

The four detector arms used for the study and the systems used for characterisation of detector image quality.

Arm name	Technology and detector type	Systems used for detector characterization
a-Se	DR: Amorphous-selenium (a-Se) photoconductor	Hologic Dimensions (Hologic Inc., Bedford, USA) Hologic Selenia (Hologic Inc., Bedford, USA)
CsI	DR: Caesium iodide (CsI) phosphor	GE Essential (GE Healthcare, Milwaukee, USA)
CR NIP	CR: needle image plate	Agfa DX-M with NIP HM5.0 plates (Agfa Healthcare, Mortsel, Belgium) Carestream Direct view Elite CR with SNP plates (Carestream Health Inc., Rochester, USA)
CR PIP	CR: powder image plate	Carestream 900 system with EHR2 image plates (Carestream Health Inc., Rochester, USA)

Table 2

The percentage difference in JAFROC weighted *FoM* relative to second detector in first column with *P*-value in parenthesis for calcification clusters and non-calcification lesions.

Difference between Arms	Calcification clusters	Non-calcification lesions
CsI v a-Se	−2.8% (0.21)	−0.3% (0.86)
CR NIP v a-Se	−9.8% (0.0001*)	−6.6% (0.0017*)
CR PIP v a-Se	−14.5% (<0.0001*)	−6.9% (0.0012*)
CR NIP v CsI	−7.2% (0.0026*)	−6.3% (0.0028*)
CR PIP v CsI	−12.0% (<0.0001*)	−6.6% (0.0019*)
CR PIP v CR NIP	−5.5% (0.036*)	−0.3% (0.90)

*Significant difference $P<0.05$

Table 3

Comparison of false recall fraction (FRF) for CR NIP, CR PIP and DR (a-Se, CsI) for all non-lesion localization marks (rating ≥4). The percentage differences with *P*-value in parenthesis are shown relative to the DR detectors.

DR arm	Difference in FRF of CR NIP and DR	Difference in FRF of CR PIP and DR
DR: a-Se	−2% (0.66)	−23% (0.0003*)
DR: CsI	−4% (0.43)	−24% (0.0001*)
Average difference	−3%	−23%

*Significant difference $P<0.05$

Table 4

Comparison of cancer detection fraction (CDF) for CR NIP, CR PIP and DR (a-Se, CsI) for all lesion localization marks (rating ≥4) for both types of lesions. The percentage differences with *P*-value in parenthesis are shown relative to the DR detectors. The predicted difference between the cancer detection fractions (CDF_{db}) for whole image database for DR and CR.

DR arm	Difference in CDF of CR NIP and DR	Difference in CDF of CR PIP and DR
Calcification clusters		
DR: a-Se	−28% (<0.0001*)	−45% (<0.0001*)
DR: CsI	−23% (0.0002*)	−41% (<0.0001*)
Average difference in CDF	−26%	−43%
Non-calcification lesions		
DR: a-Se	−7% (0.071)	−8% (0.029*)
DR: CsI	−13% (0.0003*)	−14% (0.0001*)
Average difference in CDF	−10%	−11%
Predicted difference in CDF_{db} for whole image database compared to DR	−7% to −15%	−10% to −22%

*Significant difference $P<0.05$

Table 5

Comparison of cancer detection fraction (CDF) for CR NIP and CR PIP for all lesion localization marks (rating ≥4) for both types of lesions. The percentage differences with *P*-value in parenthesis are shown relative to the CR PIP detectors. The predicted difference between the cancer detection fractions (CDF_{db}) for whole image database for CR NIP and CR PIP.

	Difference in CDF of CR NIP compared to CR PIP
Calcification clusters	29% (0.005*)
Non-calcification lesions	1.6% (0.69)
Predicted difference in CDF_{db} for whole image database compared to CR PIP	6% to 11%

* Significant difference $P<0.05$

Table 6

Comparison of published cancer detection rates (CDR) per 1000 women for DR and CR PIP for invasive cancers, in-situ cancers and all cancers. The percentage differences are expressed as the difference relative to DR.

Study	Invasive cancer CDR			In-situ cancers CDR			All cancers CDR			MGD of CR PIP relative to DR
	DR	CR PIP	% difference	DR	CR PIP	% difference	DR	CR PIP	% difference	
Flanders (12)	NA	NA	NA	1.1	0.8	−27%	6.3/5.1[b]	6.4/4.8[b]	+2%/−6%[b]	+60%
Ontario (13)	4.0	2.9	−28%[a]	1.0	0.5	−50%[a]	4.9	3.4	−31%[a]	+17%[c]
Bouches du Rhône (14)	5.0	4.2	−16%	1.6	0.75	−53%[*]	7.1	5.5	−23%[*]	NA

[*] Significantly different ($P<0.05$);

[a] Statistical test not undertaken;

[b] Prevalent/subsequent screening rounds;

[c] Data for 2009 only (11);

NA: Not available