

# **Technical evaluation of EIZO RadiForce RX850 8MP monitor**

NCCPM Technical Report 1801

June 2018

Cecilia J Strudley, Lucy M Warren and Kenneth C Young



## Acknowledgements

The authors are grateful to the staff at the Jarvis Breast Screening Centre, Guildford, for their assistance with the evaluation of this equipment.

## Funding

This study was funded for the NHS Breast Screening Programme by Public Health England. One of the authors (LMW) works on the OPTIMAM 2 research programme funded by Cancer Research UK (C30682/A17321).

# Contents

Executive summary	5
1. Introduction	6
1.1 Testing procedures and performance standards for image display	6
1.2 Objectives	6
2. Methods	7
2.1 System tested	7
2.2 Physics QC tests	8
2.3 Observer study	8
3. Results	11
3.1 Physics QC tests	11
3.2 Observer study	12
4. Discussion	17
4.1 Physics QC tests	17
4.2 Comparison of observer performance between monitors	17
4.3 The effect of using quadrant view in addition to full field view	17
4.4 Study design	18
4.5 Summary	18
5. Conclusions	19
References	20
Appendix 1: Manufacturer's comments	21
Appendix 2: Jack-knife alternative free-response receiver operating characteristic	22
Appendix 3: Results for individual observers	23

## Executive summary

The objective of this evaluation was to determine whether use of the EIZO 8MP monitor would make any difference to cancer detection in breast screening if used as an alternative to a pair of conventional 5MP monitors.

A study was carried out where seven observers assessed three hundred cases using a pair of 5MP monitors and a single 8MP monitor. No significant difference was found between the monitors when quadrant view was used in addition to full field view. When only the full field view was used, observers perform marginally better in detecting malignant calcifications using the pair of 5MP monitors. When readers are generalised to the population, this difference becomes non-significant.

Therefore, if only the full field view is used to view subtle or very subtle calcification clusters, without zooming, then the 8MP monitor may not be a suitable replacement for the pair of 5MP monitors. However, provided that all images are magnified at least to the extent that image pixels are mapped 1:1 to the monitor pixels, no significant disadvantage in using the 8MP monitor has been demonstrated. The 8MP monitor is also able to display colour images, which were not assessed during this evaluation.

Performance in detecting calcifications was considerably improved by using quadrant view in addition to full field view when using either the pair of 5MP monitors or the 8MP monitor.

# 1. Introduction

It is current practice in the NHSBSP that a pair of 5MP monitors is used to display mammograms when reporting screening mammograms. Recent innovations in display technology have made available larger format monitors allowing display of a pair of mammograms on a single monitor at a similar resolution. This evaluation is of an EIZO 8MP colour monitor, which approximates to a pair of 4MP displays on a single screen.

## 1.1 Testing procedures and performance standards for image display

Current guidelines for the testing of pairs of 5MP mammography displays are given in NHSBSP Equipment Report 0604<sup>1</sup>, and include measurements of the dynamic range of the display, the matching and uniformity of luminance of the pair of monitors, a test of the resolution of the display and conformance to the DICOM greyscale standard. These tests were carried out on the 8MP monitor. However, additional tests were needed to determine the effect, if any, of using a single 8MP display instead of the conventional pair of 5MP monitors. A study was carried out to compare the performance of observers using a conventional pair of 5MP monochrome EIZO monitors to that when using the EIZO 8MP colour monitor to display mammograms. There was no assessment of the colour performance.

## 1.2 Objectives

The objectives of this evaluation were:

- to evaluate the performance of the EIZO 8MP monitor
- to determine whether use of the EIZO 8MP monitor would make any difference to cancer detection in breast screening, if used as an alternative to a pair of conventional 5MP monitors.

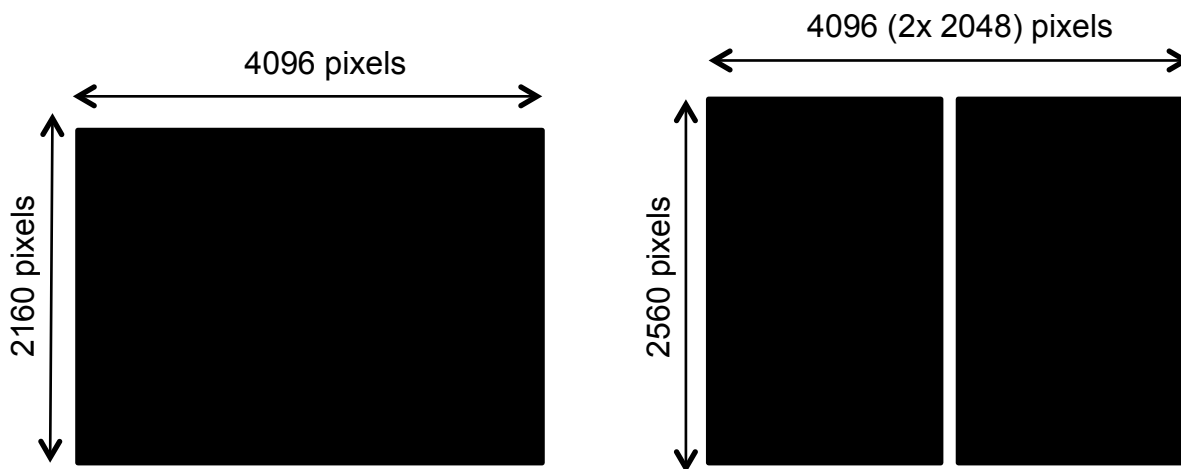
## 2. Methods

### 2.1 System tested

The equipment tested was an EIZO RadiForce RX850 8MP colour monitor which is an LED backlit LCD display. For comparison, tests were also carried out on a pair of EIZO 5MP monochrome monitors. Details of the monitors used in this evaluation are given in Table 1 and Figure 1.

**Table 1. Details of the monitors used in this evaluation**

	8MP monitor	Pair of 5MP monitors
Model	RadiForce RX 850	GX540-CL
Serial number(s)	20241114	20280033 and 20257033
Date of manufacture	November 2014	September 2009
Format	4096 pixels x 2160 pixels	Each: 2048 pixels x 2560 pixels
Pixel pitch	0.1704 mm x 0.1704 mm	0.165 mm x 0.165 mm
Viewable image size	697.9 mm x 368.0 mm	Each: 337.9 mm x 422.4 mm



**Figure 1. Dimensions of the two monitor types used in this evaluation. The 8MP monitor is shown on the left and the pair of 5MP monitors on the right**

The width of the 8MP monitor in pixels is equivalent to the combined width of the two 5MP monitors (4096 pixels versus 2 x 2048 pixels), though the physical width of the viewable image is slightly greater (by 3%) for the 8MP monitor (698 mm versus 2 x 338 mm) due to the increased pixel pitch. The height of the 8MP monitor in pixels is significantly less than the height of the 5MP monitors (2160 pixels versus 2560 pixels) and the physical height of the viewable image is 87% of the viewable image height for the 5MP (368 mm versus 422 mm).

## 2.2 Physics QC tests

Physics tests were carried out on the 8MP monitor and on the pair of 5MP monitors used for comparison in the observer study. Measurements were made using an external light-meter.

Ambient lighting for the observer study was measured at the centre of each screen with the sensor pointing outwards under various conditions.

As recommended in NHSBSP 0604, test patterns designed by the American Association of Physics in Medicine (AAPM) were used in this evaluation. These test patterns are available at [euref.com](http://euref.com). AAPM TG18-QC was displayed on each monitor and the visibility of the resolution gratings, difference between adjacent greyscale squares and corner patches within each greyscale square and sharpness of high and low contrast lines and borders were assessed. Further assessment of the resolution was made using the TG18-LPV-50 and LPH-50 line pair patterns.

To measure luminance uniformity and to compare central luminance of the two halves of the 8MP display and of the two 5MP displays, use was made of the AAPM TG18-UN80 test pattern.

Greyscale DICOM conformance was measured using the AAPM TG18 greyscale steps. The ambient luminance was measured with the sensor facing each monitor at a distance of 30cm with the monitors off.

## 2.3 Observer study

Seven observers (4 radiologists and 3 radiographers), with an average of 8.8 years of experience (range one to twenty four years) were enrolled for the observer study. They had read an average of 6700 mammograms per annum (range 3200 to 12000), 300 clinical cases, imaged using Hologic Selenia mammography systems, were selected from the pool of mammograms in the OPTIMAM image database<sup>2,3</sup>. The cases comprised 100 normal cases and 200 cases with biopsy proven cancers of subtle or very subtle appearance. Of these, 100 cases included calcifications and 100 cases included non-calcified lesions. The women whose cases were normal had since returned for the next screening episode and the screening images were found to be still normal. The cases containing calcifications included 111 malignant clusters and 8 benign clusters. The cases containing non-calcified lesions included 101 lesions, of which 45 appeared as a mass, as opposed to a focal asymmetry or distortion. Seven of these were spiculated and 38 were ill-defined. The locations and conspicuity of the cancers were recorded by experienced radiologists in advance of the study and the diagnosis for each cancer was confirmed by checking the pathology data records. Although each case included two views of each breast, only one view per breast was used for the study. For cases including a cancer, the view where the cancer was least obvious was selected, provided that it was judged to be visible. For normal cases the view to be used was randomly selected. Thus, for the study, each case consisted of a pair of CC or MLO views. Two thirds of the cases included a cancer of subtle or very subtle appearance, and some cases included more than one cancer in which case one of



the cancers may have been of obvious appearance. The least conspicuous cancers were used for this study in order to maximise the sensitivity of the study to any differences in cancer detection.

The study was conducted using MedXViewer software<sup>4</sup>, which allows the observer to mark suspected cancers on the images and record answers to questions about each mark.

It was expected that any difference in the visibility of cancers between the two types of monitor would be greatest when entire images were viewed, due to the smaller screen size and reduced number of pixels for those displayed on the 8MP monitor. However in clinical practice readers normally use quadrant view and zoom as well, which may overcome any effect due to these differences between the two display types. It was therefore decided to make the assessment of each case a two stage process. Each case was first displayed full field and the observers were asked to mark any cancers and give their overall assessment of the case. Next the same case was made available in quadrant view (1:1 pixel in image to pixel on monitor), and the observers were able to cycle through the quadrant views and the full field view, change, remove or add marks made in the first stage, and then reassess the case.

Cases were divided into five sets of 60 cases and each set was read by each observer on each monitor type, alternating between the two types of monitor. The sets were sequenced differently for each observer and the order in which cases within each set were presented was randomised for each observer. Restrictions were placed on the pace of the study, such that there was a minimum of two weeks before observers saw any case for the second time, to reduce any risk of cases being remembered. Observers were also restricted to a maximum of two sets of cases per day to avoid fatigue.

Prior to the main study a set of 15 cases containing more obvious cancers were used for training purposes to familiarise the observers with the software and to test the study setup. Observers were instructed to work in ambient light conditions conforming to the NHSBSP standard, as described in section 3.1.

Observers were first presented with a pair of MLO or CC images for each case, displayed full field, with quadrant view and magnification controls disabled, and asked to mark any possible cancers they could see. Cancers were marked electronically using the mouse pointer and a right click to indicate the position of the possible cancer. A box was then displayed on the screen asking the following three questions about that mark. The responses available for selection are shown in brackets after each question below:

1. What type of cancer is this? (Calcification cluster / Calcified lesion / Non-calcified lesion)
2. How confident are you that this is a cancerous calcification cluster / calcified lesion / non-calcified lesion? (Slider bar to indicate 1% to 100%)
3. Would you recall on the basis of this lesion? (Yes / No)

When the observer had finished marking any cancers they could see they were instructed to press “enter” on the keyboard.

For the second stage of the assessment of each case the views retained any marks made in the first stage. Quadrant view was enabled allowing the observers to step through the four quadrant views as well as the full field view. Observers were asked to review any marks already made and to look for any further cancers. Options were available to change responses given to questions regarding each mark made in the first stage, to delete marks and to add further marks. When the observers had finished the second stage they were instructed to again press “enter” in order to move on to the next case.

Reader marks for both stages of each case were recorded in a database and retrieved for later analysis. The data recorded included date and time information so that case reading speeds could be compared for the two monitor types.

Each mark made was compared to the ground truth data stored for each image. If the position of the mark fell within the boundary of a recorded cancer the mark was designated a true positive. Where more than one mark was made within a single recorded cancer then the mark with the highest confidence rating was used and the others discarded. Marks, on the basis of which observers had indicated that they would recall a case for further assessment, were classified as either true positives or false positives. These, together with data on all the cases in the study, were fed into JAFROC analysis software.

Analysis was carried out using Chakraborty’s JAFROC analysis software (version 4)<sup>5</sup>. A brief explanation of JAFROC analysis is found at Appendix 2.

## 3. Results

### 3.1 Physics QC tests

The results of the Physics QC tests are summarised in Table 2 below.

Measurements of ambient illuminance were found to be within the 10 lux limit providing the overhead lights were dimmed to the minimum setting and the light-box turned off.

The visual assessment of greyscale contrast steps indicated satisfactory contrast across the greyscale range. For both the 8MP and the pair of 5MP monitors, the steps in contrast between adjacent squares were noticeable, contrasts of 5% on 0% and 95% on 100% luminance were visible and +/-1.6% corner patches on each of the contrast squares were just discernible, with the exception of the black for both monitor types.

High contrast borders appeared sharp with no smearing or artefacts for either monitor type.

When viewing the TG18-QC test pattern, the high contrast resolution gratings of 2 pixel and 1 pixel line width were resolved with the naked eye at all positions on both monitor types, as were the 2 pixel low contrast lines. Using a magnifying glass, low contrast single pixel line pairs were seen on the 5MP monitors, but not on the 8MP monitor. The assessment was repeated using the TG18-LPV-50 and LPH-50 test patterns to display the narrow low contrast resolution grating over a larger area. It was possible to resolve these test patterns on both monitor types.

When a uniform 10% luminance was displayed on the two monitor types, a slight difference in the shade of grey was noticed. The 5MP monitor had a slightly pinkish hue with some angular dependence, whereas the 8MP monitor displayed a colder blue shade of grey. Both monitor types conformed to the uniformity standard, which stipulates a maximum deviation of 30% from the luminance at the centre of the screen. The matching of luminance between the two 5MP monitors conformed to the standard which stipulates a maximum difference of 5%. This limit was also satisfied when applied to the left and right halves of the 8MP monitor.

Measurements showed that both monitor types conformed to within 10% of the DICOM greyscale standard.

**Table 2. Summary of physics QC test results**

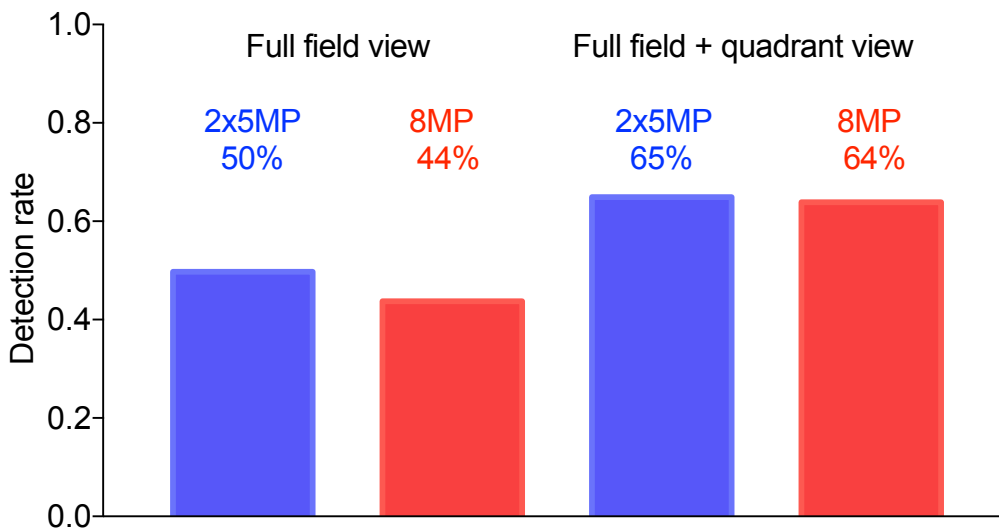
Test	Limiting value	Left 5MP	Right 5MP	8MP
Lmin	<1.5 cd/m <sup>2</sup>	0.48	0.55	0.41
Lmax	>450 cd/m <sup>2</sup>	497.8	485.5	471
Luminance ratio (Lmax/Lmin)	>300	1037	883	1149
Ambient luminance	<0.4*Lmin	0.27*Lmin	0.24*Lmin	0.17*Lmin
Visual assessment of greyscale steps	Noticeable difference between 0% and 5% and between 95% and 100% luminance	Yes	Yes	Yes
Sharpness	No artefacts, smearing	OK	OK	OK
Resolution	All contrast line pairs seen	OK	OK	Partially OK
Luminance uniformity	Max 30% difference from centre	3%	2%	3%
Matching	Max 5% difference in white luminance between centres of left & right displays	2.5%	3%	(3% & 2%)*
DICOM greyscale standard	Max deviation of 10%	3.1%	3.9%	5.6%

\*For the 8MP monitor results are shown treating it both as a single display and in brackets as two adjacent displays, left and right

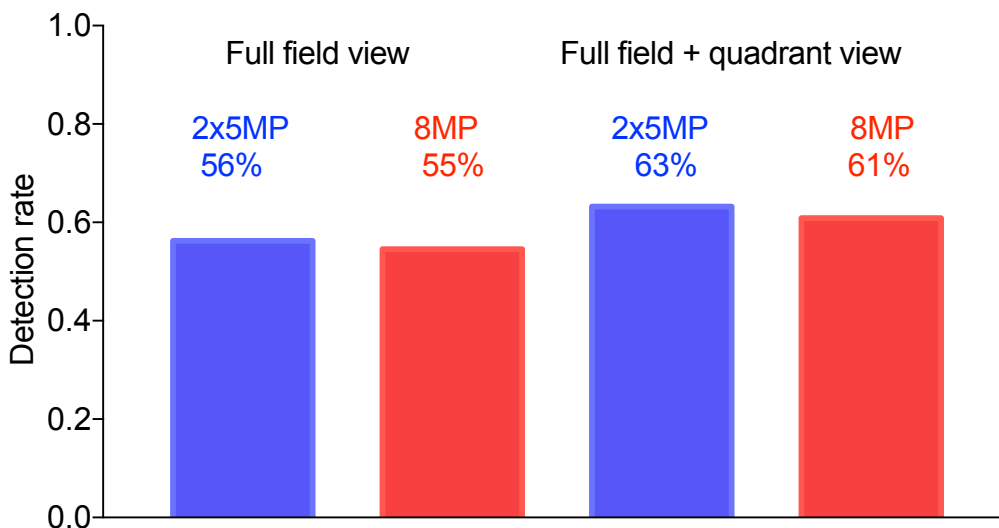
## 3.2 Observer study

### 3.2.1 Comparison of two monitor types: Detection rates

Detection rates were calculated in terms of the proportion of confirmed cancers that each reader correctly located and would recall. The mean detection rates for viewing calcification clusters or non-calcified lesions on each monitor type using full field view only or full field plus quadrant views are shown in Figures 2 and 3 below.



**Figure 2. Mean cancer detection rates for calcification clusters viewed on each monitor type using full field view only or full field view plus quadrant views**



**Figure 3. Mean cancer detection rates for non-calcified lesions viewed on each monitor type using full field view only or full field view plus quadrant views**

### 3.2.2 Comparison of two monitor types: JAFROC analysis

The JAFROC analysis used has greater statistical power than ROC analysis because true positives are only generated when the location of a cancer is correctly identified. JAFROC analysis results are in terms of a figure of merit, which is the area under the curve in the plot of lesion localisation fraction versus false positive fraction.

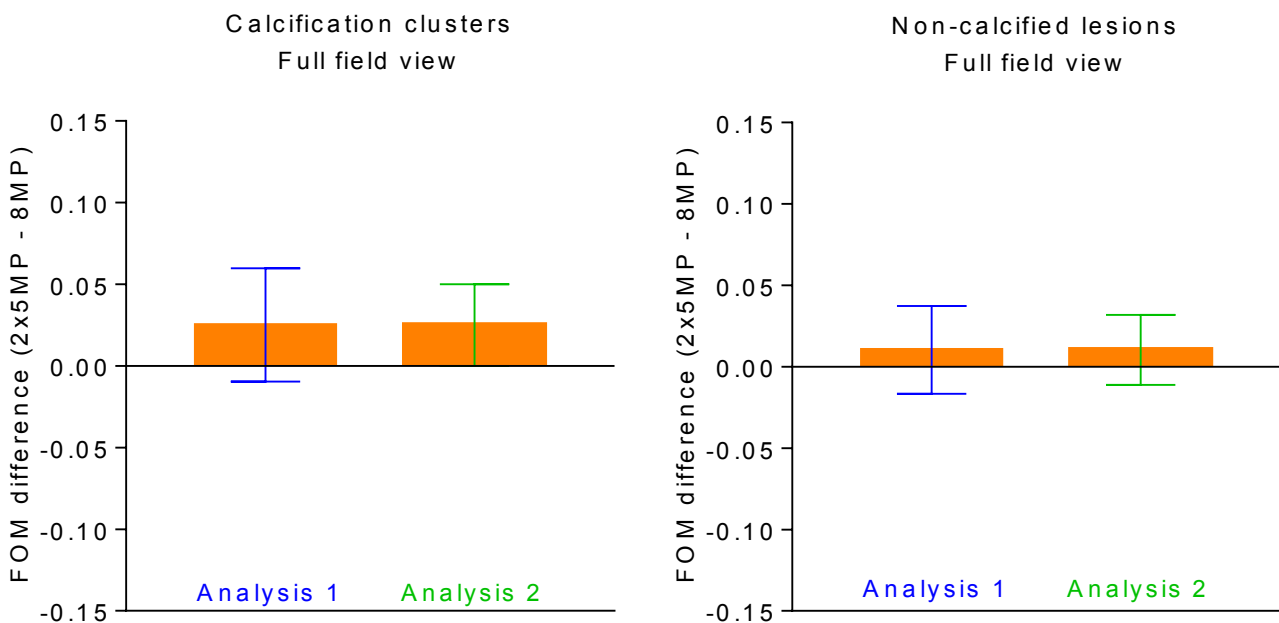
The mean figures of merit for the seven readers were obtained for calcification clusters and for non-calcified lesions. These figures of merit are presented in Table 3.

**Table 3. Mean figures of merit for calcification clusters and non-calcified lesions when cases are viewed full field, and full field plus quadrant view**

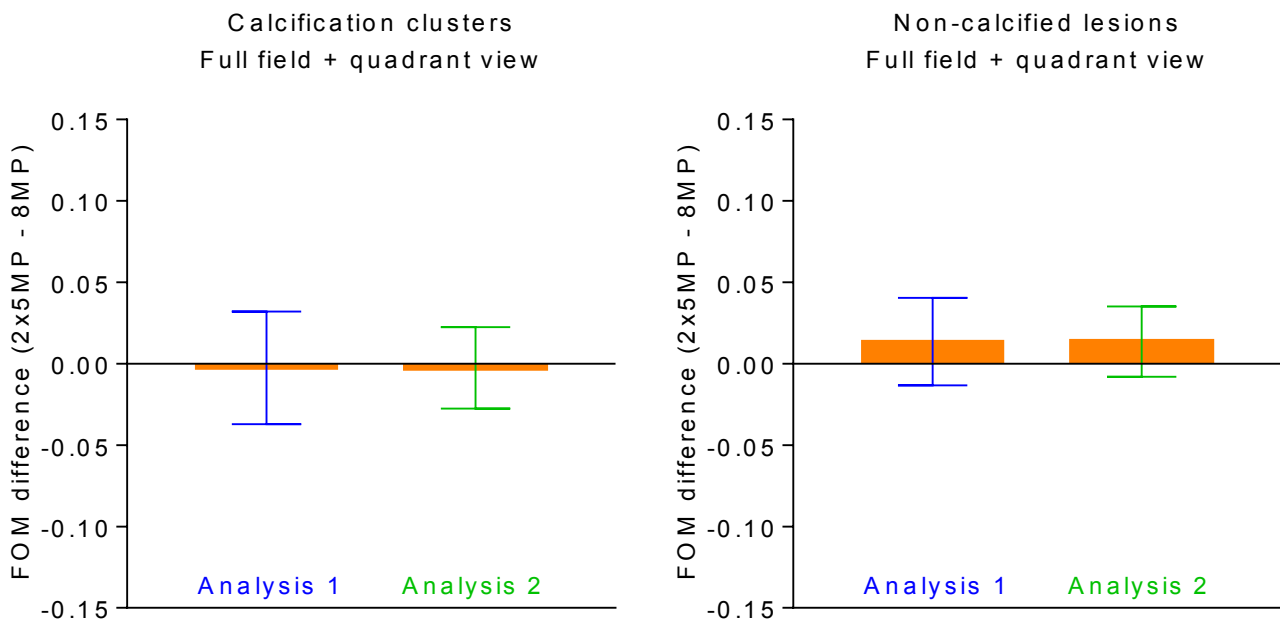
	FOM (full field view)		FOM (full field plus quadrant view)	
	2x 5MP	8MP	2x 5MP	8MP
Calcifications	0.665	0.640	0.731	0.734
Non-calcified lesions	0.715	0.704	0.732	0.719

\* Result indicating statistically significant difference between the two monitor types

The differences between the two monitor types and the *p* values for these differences resulting from two analyses: Analysis 1 (random reader and random case) and Analysis 2 (fixed reader and random case) are shown in Figures 4 and 5. In each graph the differences in the figures of merit are presented with error bars showing the 95% confidence limits on these differences. In each case the figure of merit for the 8MP monitor is subtracted from that for the pair of 5MP monitors such that a positive difference indicates a better performance for the pair of 5MP monitors and a negative difference indicates better performance for the 8MP monitor. Where the error bars straddle the x axis, the difference calculated is not statistically significant.



**Figure 4. Difference in figure of merit between pair of 5MP monitors and 8MP monitor when calcification clusters or non-calcified lesions are viewed full field. The error bars indicate the 95% confidence limits for Analysis 1 (random reader and random case) and Analysis 2 (fixed reader and random case).**



**Figure 5. Difference in figure of merit between pair of 5MP monitors and 8MP monitor when calcification clusters or non-calcified lesions are viewed using quadrant view in addition to full field. The error bars indicate the 95% confidence limits for Analysis 1 (random reader and random case) and Analysis 2 (fixed reader and random case).**

### 3.2.3 Effect of using quadrant view in addition to full field view

An improvement in cancer detection was noted when quadrant view was used in addition to full field view. Tables 4 and 5 show the differences in figure of merit between full field view and full field plus quadrant view for the pair of 5MP monitors and for the 8MP monitor, together with the *p* values for Analysis 1 and Analysis 2.

**Table 4. Mean figures of merit for pair of 5MP monitors, with *p* values calculated for Analysis 1 (random reader and random case) and Analysis 2 (fixed reader and random case)**

	FOM (2 x 5MP)		Difference	<i>p</i> value	
	Full field view	Full field + quadrant view		Analysis 1	Analysis 2
Calcifications	0.665	0.731	0.067	0.001*	0.000*
Non-calcified lesions	0.715	0.732	0.018	0.187	0.108

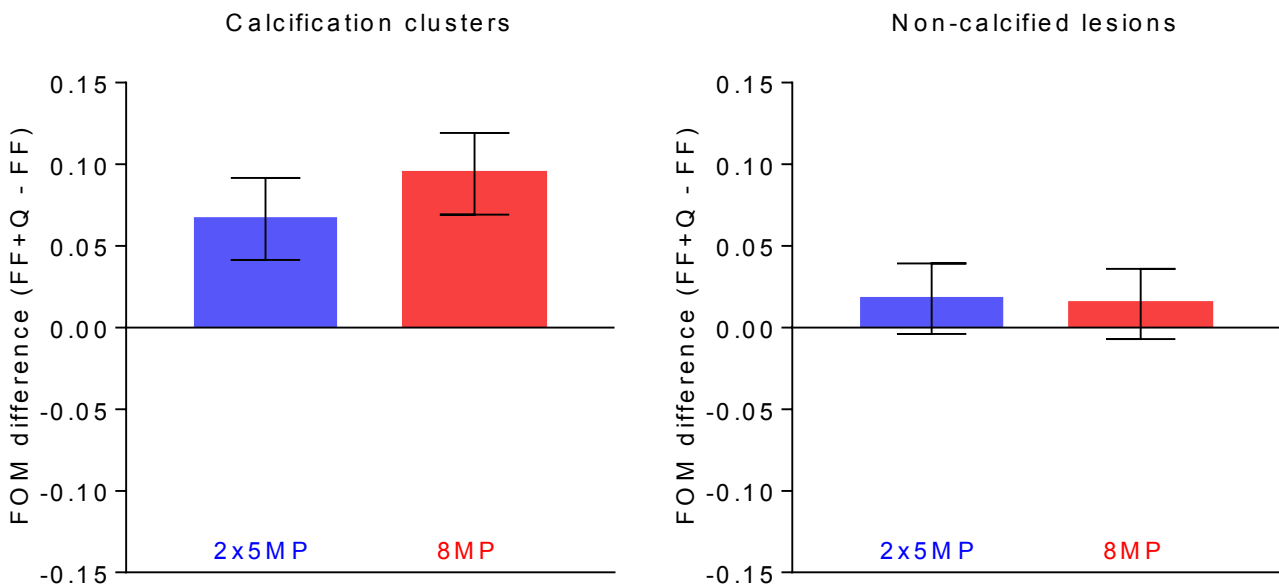
\* Result indicating statistically significant difference between the two monitor types

**Table 5. Mean figures of merit for 8MP monitor, with *p* values calculated for Analysis 1 (random reader and random case) and Analysis 2 (fixed reader and random case)**

	FOM (8MP)		Difference	<i>p</i> value	
	Full field view	Full field + quadrant view		Analysis 1	Analysis 2
Calcifications	0.640	0.734	0.094	0.000*	0.000*
Non-calcified lesions	0.704	0.719	0.014	0.277	0.188

\* Result indicating statistically significant difference between the two monitor types

The results shown in Tables 4 and 5 (Analysis 2) is also presented graphically in Figure 6. A difference greater than zero indicates better performance when using quadrant view in addition to full field view.



**Figure 6. Difference in figure of merit between full field view only and full field plus quadrant views for the pair of 5MP and for 8MP monitors when viewing calcification clusters or non-calcified lesions. The error bars indicate the 95% confidence limits for Analysis 2 (fixed reader and random case).**



## 4. Discussion

### 4.1 Physics QC tests

Tests carried out on the pair of 5MP monitors and on the 8MP monitor showed that they mostly met the NHSBSP standards for digital mammography. It was more difficult to see the low contrast narrow (single pixel) line pairs on the 8MP monitor than on the pair of 5MP monitors. They were only visible when displayed full-screen on the 8MP monitor.

### 4.2 Comparison of observer performance between monitors

The greatest difference in detection (recall) rate between the two monitor types was found when viewing calcification clusters using full field view only. (50% for the pair of 5MP monitors and 44% for the 8MP monitor). When using quadrant view in addition to full field view, or when viewing non-calcified lesions, the differences in detection rates for the two monitor types were reduced to one or two percent

JAFROC analysis software was used to generate figures of merit indicating the mean performance of the observers, taking account of false positives as well as correct lesion localisations. This statistical analysis includes an estimate of the significance of differences in figures of merit between modalities.

When mammograms were displayed using full field view only, the mean JAFROC figure of merit for the pair of 5MP monitors was significantly better ( $p=0.05$ ) when viewing calcification clusters and using Analysis 2 (fixed reader, random case). Differences when viewing non-calcified lesions, or using Analysis 1 (random reader, random case) for any lesion type, were not statistically significant.

When quadrant view was used in addition to full field view, JAFROC analysis showed no statistically significant differences between the two types of display.

### 4.3 The effect of using quadrant view in addition to full field view

Detection rates when using quadrant view in addition to full field view were in all cases greater than when using full field view only, and the observed differences were greatest when viewing calcification clusters. For example, when using a pair of 5MP monitors to view calcification clusters, the use of quadrant view improved the detection rate from 50% to 65%.

JAFROC figures of merit were used to test the effect of using quadrant view in addition to full field view. The additional use of quadrant view gave statistically significant improvements in performance ( $p<0.001$ ) when viewing calcification clusters on either type of display, using either

Analysis 1 or Analysis 2. Differences in performance when viewing non-calcified lesions were not statistically significant.

It should be noted that the observer study was designed to compare the two types of display rather than the effect of using quadrant view in addition to full field view. As each case was assessed using full field view, immediately followed by additional use of the quadrant views, the two modalities were not independent and the validity of the statistical analysis in this regard is therefore questionable. The magnitude of the effect of using quadrant views may be actually be less than that measured in this study, but this study does illustrate the considerable advantage that one may expect from using magnified views in addition to full field views for the detection of calcification clusters.

#### 4.4 Study design

In this study all the cancers used were subtle or very subtle, assuming that the detection of more obvious cancers would be less sensitive to the monitor used. In the screening environment, a much wider range in visibility of cancer is present. Calculating the proportion of malignant cancers detected through screening which have the same characteristics as the cancers used in this study would allow us to calculate the expected change in cancer detection in a screening environment, which corresponds to the change in cancer detection found in this study.

#### 4.5 Summary

This study has demonstrated the clear advantage of using magnified views in addition to the full field view in detecting calcification clusters, regardless of whether one is using a pair of 5MP monitors or the single 8MP monitor. It has also shown that if only the full field view is used to view subtle or very subtle calcification clusters, without zooming, then the 8MP monitor may not be a suitable replacement for the pair of 5MP monitors. However, provided that all images are magnified at least to the extent that image pixels are mapped 1:1 to the monitor pixels, no significant disadvantage in using the 8MP monitor has been demonstrated. The 8MP monitor is also able to display colour images, which were not assessed during this evaluation.

## 5. Conclusions

Physics QC tests conducted on the 8MP mammography display demonstrated compliance with all the requirements of the NHSBSP standards for primary displays, except one, which failed (visibility of low-contrast line pairs on the TG-18 QC pattern). The low contrast line pairs were only visible when displayed full screen on the 8MP monitor.

When quadrant view was used in addition to full field view there was no significant difference between using the pair of 5MP monitors and using the 8MP monitor to detect malignant calcifications or non-calcified lesions. When only the full field view was used, observers perform marginally better ( $p=0.050$ ) in detecting malignant calcifications using the pair of 5MP monitors. When readers are generalised to the population, this difference becomes non-significant. There was no significant difference between the monitors for the detection of malignant non-calcified lesions when only full field view was used.

Performance in detecting calcifications was improved by using quadrant view in addition to full field view when using either the pair of 5MP monitors or the 8MP monitor.

## References

1. Kulama E, Burch A, Castellano I et al. Commissioning and routine testing of full field digital mammography systems (NHSBSP Equipment Report 0604, Version 3). Sheffield: NHS Cancer Screening Programmes, 2009
2. Halling-Brown MD, Looney PT, Patela MN et al. The Oncology Medical Image Database (OMI-DB). Proceedings of SPIE 9039 (2014)
3. Halling-Brown MD, Looney PT, Patela MN et al. Mammographic image database (MIDB) and associated web-enabled software for research. In: Fujita H, Hara T, Muramatsu C, editors. Breast imaging, 12<sup>th</sup> international workshop, IWDM 2014, vol. 8539. LNCS; 2014. p. 514–9.
4. Looney P T, Young K C, Mackenzie A et al. MedXViewer: An extensible web-enabled software package for medical imaging. Proceedings of SPIE 9037 (2014)
5. Chakraborty D P and Berbaum K. S. Observer studies involving detection and localization: Modelling, analysis, and validation. Med. Phys. 31, 2313–2330 (2004).

## Appendix 1: Manufacturer's comments

EIZO are pleased with the finding that an 8MP does not reduce the ability of a reader to diagnose breast images in quadrant view (see section 5). The goal of the product is to make a workstation more flexible by having colour while not reducing the viewer's ability to read breast images. As colour becomes a greater part of radiological imaging in the future, the ability to show colour in an image will become more important. The viewer should also find that they experience less eye fatigue using a single display instead of two separate units as they eye does not have to sweep across the dark section in the centre of the field of view due to the two separate displays having wide bezels. Any reduction in bezel width should allow the user to see more of the important radiological images and less of the unimportant plastic of the bezel. EIZO felt that the testing was very well thought out and conducted to give fair results.

## Appendix 2: Jack-knife alternative free-response receiver operating characteristic

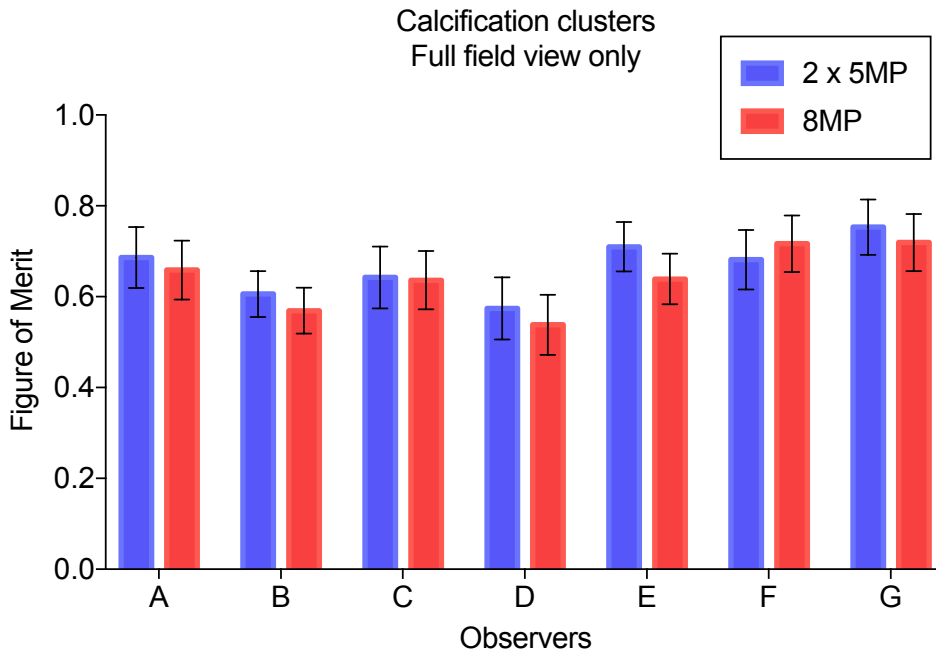
In Receiver Operating Characteristic (ROC) analysis, cases are assessed for disease probability and rated according to the risk of malignancy. Comparing assessments against the truth for each case allows each assessment to be categorised as a true positive (TP) or true negative (TN). The true positive fraction (TPF) is obtained by dividing the number of TPs by the total number of malignant cases. The false positive fraction (FPF) is obtained by dividing the number of FPs by the total number of normal cases. TPF is plotted against FPF, with multiple points corresponding to different operating points as denoted by the confidence rating level selected. The area under the curve is used as a figure of merit. The maximum area under the curve is one, corresponding to a perfect diagnostic system. Guessing the diagnosis, where half of the cases are malignant, would give an area under the curve of 0.5.

ROC analysis takes no account of the location of any detected lesion, whereas in free-response ROC (FROC) the location of each suspected lesion is recorded and is only counted as a TP if there is actually a lesion at that location, regardless of the presence of any further lesions elsewhere in the images for that case. The number of correct lesion localisations is divided by the total number of lesions to obtain the lesion localisation fraction (LLF). In alternative FROC (AFROC) analysis the figure of merit is the area under the curve when LLF is plotted against FPF. Jack-knife AFROC (JAFROC) analysis is a method of calculating the area under the curve for FROC data obtained from multiple readers assessing multiple cases. The figure of merit (area under the curve) is the probability that a malignant lesion is rated as more suspicious than a false positive on a normal image.

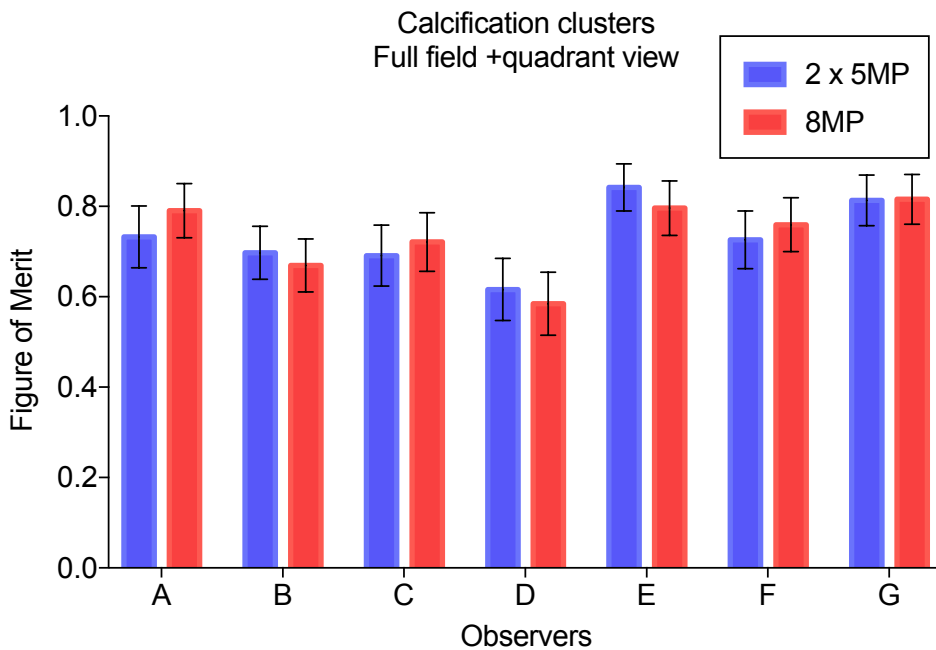
Chakraborty's JAFROC software, which was used to analyse the data for this study, generates  $p$  values relating to differences in figures of merit for pairs of modalities. A  $p$  value of 0.05 or less indicates a confidence of 95% or more that there is a real difference between the two modalities. For each pair of modalities two analyses are carried out: Analysis 1 (random reader, random case) and Analysis 2 (fixed reader, random case). The fixed reader is an average of the observers employed in the study. The random readers and random cases are restricted to those used in the study, and may or may not be representative of the general population of readers.

## Appendix 3: Results for individual observers

JAFROC figures of merit for individual observers detecting calcification clusters using full field view only and using full field and quadrant views are shown in Figures A1 and A2. There is a variation in results between observers, which is typical for these types of observer study.



**Figure A1. Figures of merit for individual observers when viewing calcifications on the two monitor types, using full field view only**



**Figure A2. Figures of merit for individual observers when viewing calcifications on the two monitor types, using full field plus quadrant views**