



This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.

 **creative commons**
C O M M O N S D E E D

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

 **Attribution.** You must attribute the work in the manner specified by the author or licensor.

 **Noncommercial.** You may not use this work for commercial purposes.

 **No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Health professionals' agreement on density judgements and successful abnormality identification within the UK Breast Screening Programme

Iain T. Darker*, Yan Chen & Alastair G. Gale
Applied Vision Research Centre, Loughborough University, Loughborough, UK

ABSTRACT

Higher breast density is associated with a greater chance of developing breast cancer. Additionally, it is well known that higher mammographic breast density is associated with increased difficulty in accurately identifying breast cancer. However, comparatively little is known of the reliability of breast density judgements. All UK breast screeners (primarily radiologists and technologists) annually participate in the PERFORMS self-assessment scheme where they make several judgements about series of challenging recent screening cases of known outcomes. As part of this process, for each case, they provide a radiological assessment of the likelihood of cancer on a confidence scale, alongside an assessment of case density using a three point scale. Analysis of the data from two years of the scheme found that the degree of agreement on case density was significantly greater than no agreement ($p < .001$). However, only a moderate degree of inter-rater reliability was exhibited ($\kappa = .44$) with significant differences between the occupational groups. The reasons for differences between the occupational groups and the relationship between agreement on density rating and case reading ability are explored.

Key words: Breast screening, Breast cancer, Mammography, Mammographic Density, Receiver Operator Characteristic, Sensitivity, Specificity.

1. INTRODUCTION

In the UK the national breast screening programme has run for some 22 years and for which the latest data (2008-2009) show invited 2.6 million women aged 50-70 years for screening with 4.4% of these women recalled for assessment. Some 1,746 benign biopsies were performed and 16,535 cancers were successfully detected¹. Screening is performed nationally at some 110 centres by small groups of screeners who comprise mainly radiologists and technologists (advanced practitioner radiographers who are specially trained to interpret mammograms) and a mixed group of other health professionals with a keen interest in screening (e.g. breast clinicians). Screeners voluntarily take part in an annual self assessment scheme (PERFORMS) which is an educational scheme and runs as part of the UK's screening programme². Currently some 650 screeners take part in this, reading case sets twice a year. Annually these individuals on two occasions (not all individuals read both annual sets for a variety of reasons) read sets of carefully selected difficult exemplars of recent screening cases which are chosen by an expert radiological panel to be challenging. These comprise difficult normal, benign and malignant appearances. Individual screeners inspect cases, reporting their decisions about each case into a computer which then immediately feeds back information about their various responses as compared to the expert panel and to the known case pathology. Furthermore, they subsequently receive detailed anonymised information concerning how they have performed as compared to all the participant screeners in the UK³.

Not surprisingly, variations in participant performances are found and the one key purpose of the scheme is to highlight such variations in an individual's skill back to them as compared anonymously to their peers nationally. This then enables them to consider whether they require further training concerning particular aspects of the mammographic reporting process. Our latest data show that in terms of the basic screening decision of whether a case should or should not be recalled then overall, circa 20% of cases that should be recalled are, in fact, not recalled; whilst circa 13% of cases that should be 'returned to screen' (i.e. in normal screening, that woman should not be recalled for further investigations such as biopsy, ultrasound, etc.) are, in fact recalled. In terms of cancer identification then there is an average missed cancer rate of circa 15%. These errors rates illustrate the complex nature of these particular case sets which are deliberately difficult.

*i.t.darker@lboro.ac.uk; <http://performs.lboro.ac.uk/>

When compiling the self assessment sets of cases a panel of experienced screening radiologists are asked to make a series of decisions about each case concerning the presence of key mammographic features, whether radiologically it should be recalled etc. As part of this process they also rate the breast density of each case on a three point scale of 'fatty', 'mixed' and 'dense'. This rating information is then used to aid in the selection of cases to be included in the case sets of the PERFORMS scheme. When individuals participate in the PERFORMS scheme then, as well as making various decisions about mammographic features and whether to recall the case, they also rate the cases on the same three point density scale. This paper considers various aspects of these density ratings. Firstly, given that perceived density is strongly linked to cancer risk then it could be expected that in a selected case set as here representing difficult exemplars then a predominance of dense cases may be found. In contrast, these are screening cases from a population of women over the age of 50 years and so predominantly fatty breasts could be expected. In the UK, screening has been primarily undertaken by radiologists; more recently this role has also been extended to technologists. Previously screening performance differences have been reported between these two groups primarily related to the experience of the radiologists⁴. Does this also hold for breast density judgements?

Wolfe⁵ first identified the link between breast density and cancer risk and proposed four categories of parenchymal patterns with the highest risk category having 37 times more risk than the lowest risk category. Subsequent work has generally supported the relationship of risk and breast density category but at a much reduced magnitude of relative risk (e.g. commonly four times the risk for the specifically dense category^{6, 7}). Another notable approach is that of Tabar⁸ who proposed five category classifications. Quantitative based scales have been proposed by Boyd⁹ using six categories and more recently BIRADS¹⁰ has become quite widely used. Various studies have reviewed the use of such scales both making comparisons between them^{11,12} and density scales as compared on film and digital images¹³. Additionally various computer imaging based approaches have been developed to estimate breast density based on the digital image data¹⁴. Jaffe¹⁵ has provided a good review of qualitative and quantitative (both 2D [i.e. planimetry and image digitisation] and volumetric [i.e. digital mammography or tomosynthesis]) measurements of breast density. Volumetric assessments of breast density are now clinically available, such as Hologic's Quantra.

2. METHOD

Information was examined from the most recent two complete years of the PERFORMS scheme (2008 and 2009). This comprised data from all those screeners who had read both case sets each year. This encompassed 444 individuals (261 radiologists, 158 technologists, 11 breast clinicians, nine breast physicians, four associate specialists, and one clinical assistant). For the purpose of this study each participant was placed into one of three occupational groups: radiologists, technologists and 'others'.

2.1 Breast screening cases

All participants had examined the same 120 cases each year making a total of 240 cases. For the data analyses here only 209 of these cases are considered, these being either malignant or normal: 87 cases had originally been confirmed malignant at biopsy and 122 cases had been confirmed normal at the subsequent three year follow-up screening round. Some 31 benign cases (as confirmed at biopsy) were excluded from the analysis.

2.2 Individuals' performance and density judgements

In taking part in the scheme each participant inspected four mammographic views of each case (the CC and MLO views of both breasts) and made a screening decision about each breast, noted mammographic features and gave an overall rating of breast density for the whole case using three categories. Three categories have been used in other studies^{e.g.16}. The mammographic feature information is not considered further here. Whilst full screen digital mammography is being rolled out across the UK this research work utilised film mammograms. The screening opinion was collected using a six-point scale, namely: normal; probably benign but return to screen (do not recall); probably benign but recall; indeterminate; suspicious of malignancy; and malignant. This differs from BIRADS ratings but is broadly analogous. The rating for each breast of a case was then collapsed into a single rating to provide an overall opinion for the case by each individual. For malignant cases, the rating associated with the breast containing a malignancy was taken as the rating for the whole case (no cases exhibited bilateral malignancies). For normal cases, the most 'malignant' rating assigned to either breast was taken as the rating for the whole case. Breast density for each case was collected on a

three-point rating scale: 'fatty', 'mixed density', and 'dense'. For each participant, screening performance (i.e. recall or not recall decisions) was measured as accuracy (A_z) calculated using ROCKIT.

2.3 Case difficulty

Case difficulty was assessed by calculating the proportion of correct screening decisions for each case, across radiologists and technologists to gain an overall assessment of case difficulty, as well as within each of these occupational groups to gain an assessment of how the difficulty of each case varied by occupational group. The screening rating decision was collapsed into the binary ratings of 'return to screen' (including normal and probably benign but return to screen [i.e. do not recall] decisions); or 'recall' (including probably benign but recall, indeterminate, suspicious of malignancy, and malignant decisions).

3. RESULTS

3.1 Density judgements

Of the cases studied here, in terms of mode density rating the experts judged 32 to be fatty, 153 to be mixed density and 24 to be dense; comparatively the participants judged 32 to be fatty, 163 to be mixed density, and 14 cases to be dense. There was a significant correlation between the mode density rating given by the experts and radiologists ($r_s = 0.78$, $n = 209$ cases, $p < .0001$) and also that given by the experts and technologists ($r_s = 0.74$, $n = 209$ cases, $p < .0001$). A near perfect correlation was found between the mode density rating given by radiologists and that given by technologists to each case ($r_s = 0.95$, $n = 209$ cases, $p < .0001$). Overall, a moderate measure of agreement was found on density ratings for these malignant and normal cases as measured using Fleiss' Kappa ($\kappa = 0.44$, $SD = 0.004$, $n = 444$, cases = 209). If the benign cases were to be included in the analysis then a similar kappa value ensues; $\kappa = 0.47$ ($SD = 0.003$, $n = 444$, cases = 240).

The nature of the differences between expert ratings and those of radiologists and technologists were examined. Cases in each category of density according to the experts were assigned an ordinal rank (fatty = 1; mixed density = 2; dense = 3). Then for each ordinal rank assigned by the experts, variations in the ordinal rank assigned by radiologists and technologists, and overall participants, were examined. Across cases there was a maximum of one rank difference between the expert and mode participant opinion. Using Wilcoxon signed-rank tests when the expert opinion was that breast density was fatty, there was a significant tendency for both radiologists ($z = -2.00$, $p < .05$, $r = -.35$), and technologists ($z = -2.45$, $p < .05$, $r = -.43$), and overall ($z = -2.00$, $p < .05$, $r = -.35$) to use the mixed density category rather than the fatty category (12.50%, 18.75%, 12.50% of cases respectively). When the expert opinion was that breast density was mixed density, there was no significant tendency for either radiologists ($z = -0.82$, $p = \text{n.s.}$, $r = -.14$), technologists ($z = -0.09$, $p = \text{n.s.}$, $r = -.02$), or overall ($z = -0.82$, $p = \text{n.s.}$, $r = -.14$) to use either the fatty category (2.61%, 1.96%, 2.61% of expert opinion mixed density cases respectively) or dense category (1.31%, 1.96%, 1.31% of expert opinion mixed density cases respectively). When the expert opinion was that breast density was dense, there was a significant tendency for both radiologists ($z = -3.46$, $p < .001$, $r = -.61$), and technologists ($z = -3.74$, $p < .001$, $r = -.66$), and overall ($z = -3.46$, $p < .001$, $r = -.61$) to use the mixed density category rather than the dense category (50.00%, 58.33%, 50.00% of cases respectively). Thus, participants showed a greater tendency towards the mixed density category than experts. To illustrate the range of density appearances figure 1 shows examples of cases with high and low agreements in each density category.

3.2 Occupational Groups

There were striking differences in the levels of agreement in the judgements of breast density exhibited within the occupational groups (figure 2): radiologists, $\kappa = 0.47$, $SD = 0.005$, $n = 261$; technologists, $\kappa = 0.39$, $SD = 0.007$, $n = 158$; others, $\kappa = 0.44$, $SD = 0.019$, $n = 25$. Agreement amongst radiologists was significantly higher than amongst technologists ($p < .0001$) and others ($p < .01$), whilst agreement amongst the others group was significantly higher than amongst technologists ($p < .001$).

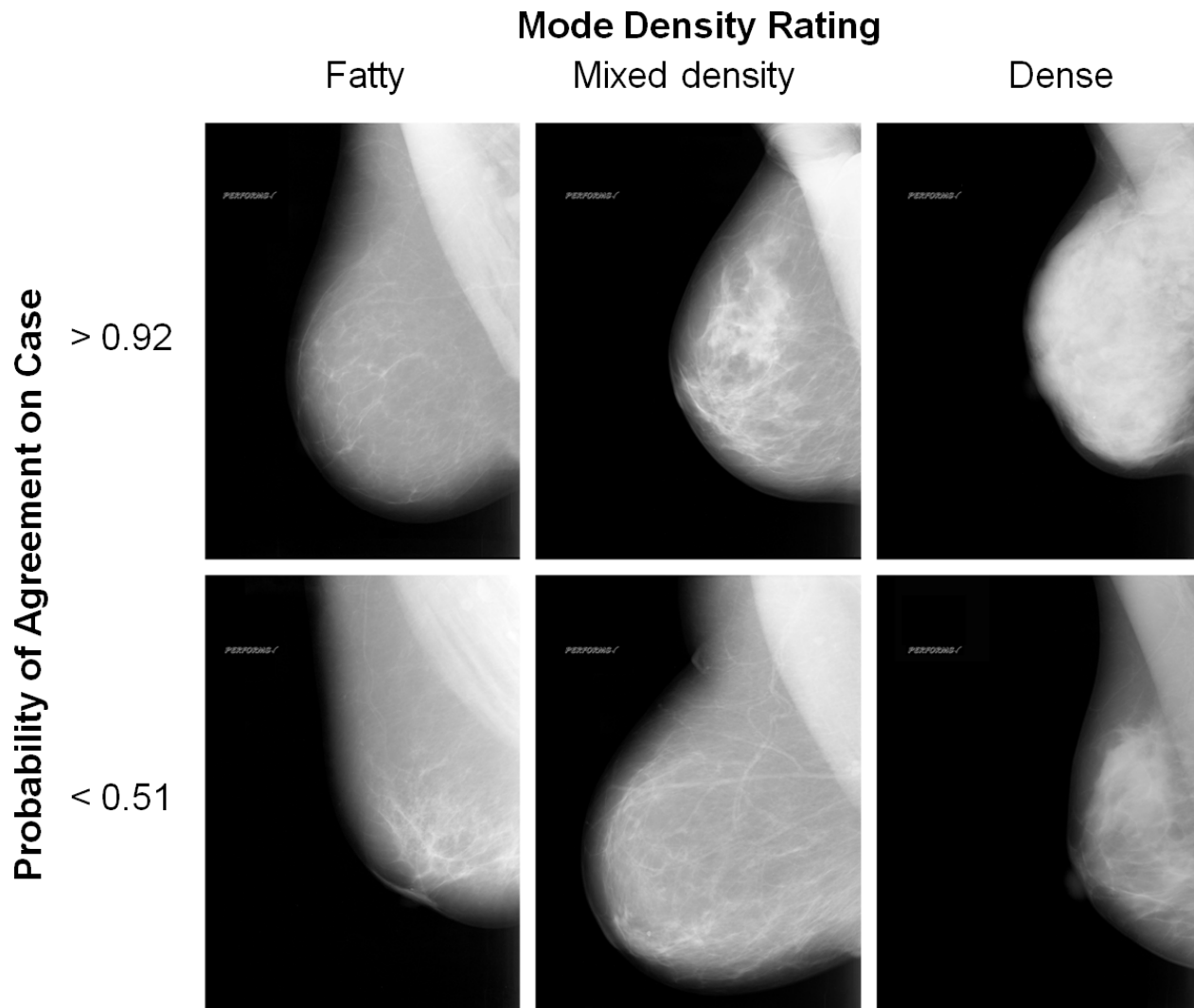


Figure 1: Examples of cases with high and low levels of agreement, and with mode density estimates in the 'Fatty', 'Mixed density', and 'Dense' categories.

3.3 Experience

As would be expected, radiologists were typically more experienced than technologists. Where data on experience were available (for 174 Radiologists and 103 technologists), all technologists had fewer than 16 years experience, whilst 81 radiologists had 16 years or greater experience.

When those radiologists with 16 years or more experience were excluded from the analysis of level of agreement (along with those for whom there were no data on experience), the differences between occupational groups remained largely unchanged with significantly higher levels of agreement amongst radiologists ($p < .0001$): radiologists, $\kappa = 0.48$, $SD = 0.009$, $n = 93$; technologists, $\kappa = 0.38$, $SD = 0.009$, $n = 103$. There were no significant differences between agreement levels amongst radiologists with less than 16 years experience and radiologists of 16 or more years experience ($p = n.s.$): radiologists < 16 years experience, $\kappa = 0.48$, $SD = 0.009$, $n = 93$; radiologists $\Rightarrow 16$ years experience, $\kappa = 0.48$, $SD = 0.010$, $n = 81$ (figure 2.) Thus, degree of experience would not appear to be associated with agreement on breast density rating.

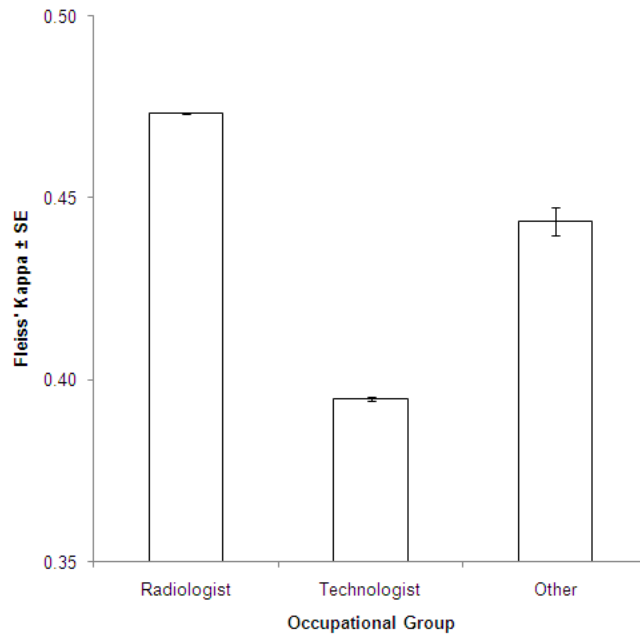


Figure 2: Level of agreement on breast density rating analysed by Occupational Group.

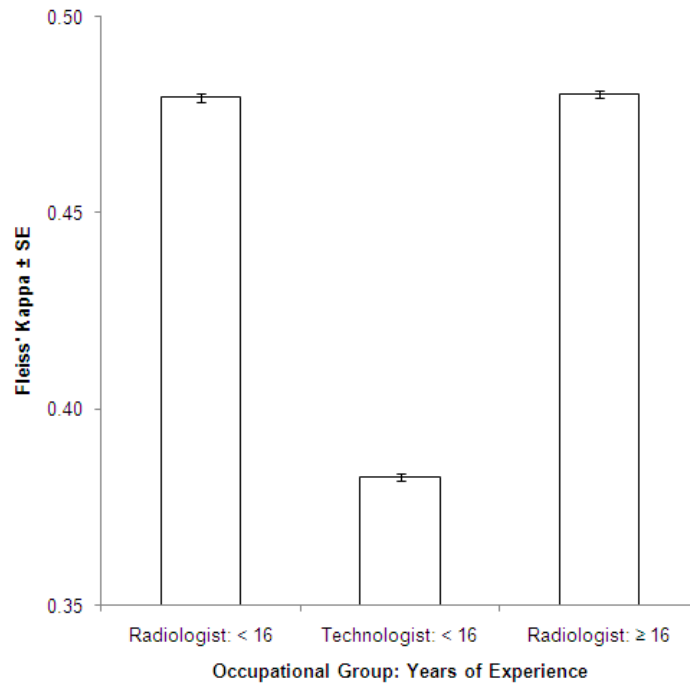


Figure 3: Level of agreement on breast density rating analysed by Occupational Group and Years of Experience.

3.4 Screening performance

A potential predictor of agreement on breast density rating might be overall performance on the PERFORMS scheme. ROCs were generated on a screener by screener basis and A_z scores were derived for each participant. All Radiologists

and Advanced Practitioners were placed into a single pool of participants which was then divided into three similarly sized groups based on a three-way split of A_z score: those with A_z scores between 0.7763 - 0.9028 were assigned to a Low performance group (68 radiologists, 72 technologists); those with A_z scores between 0.9029 - 0.9333 were assigned to a Mid performance group (86 radiologists, 54 technologists); and those with A_z scores between 0.9334 - 0.9986 were assigned to a High performance group (107 radiologists, 32 technologists).

Within each of the performance groups, significant disparity in the levels of agreement between radiologists and technologists remained with higher levels of agreement amongst radiologists (all $ps < .0001$). For radiologists, there were higher levels of agreement in the High and Mid performance groups than in the Low performance group (both $ps < .0001$). Similarly, for technologists there were significantly higher levels of agreement amongst the High performance group than amongst the Low ($p < .01$) or Mid performance groups ($p < .05$).

Thus, those evidencing a higher performance on the PERFORMS scheme here had a greater level of agreement on case density, especially amongst radiologists. However, differences between the occupational groups remained apparent at all levels of ability considered.

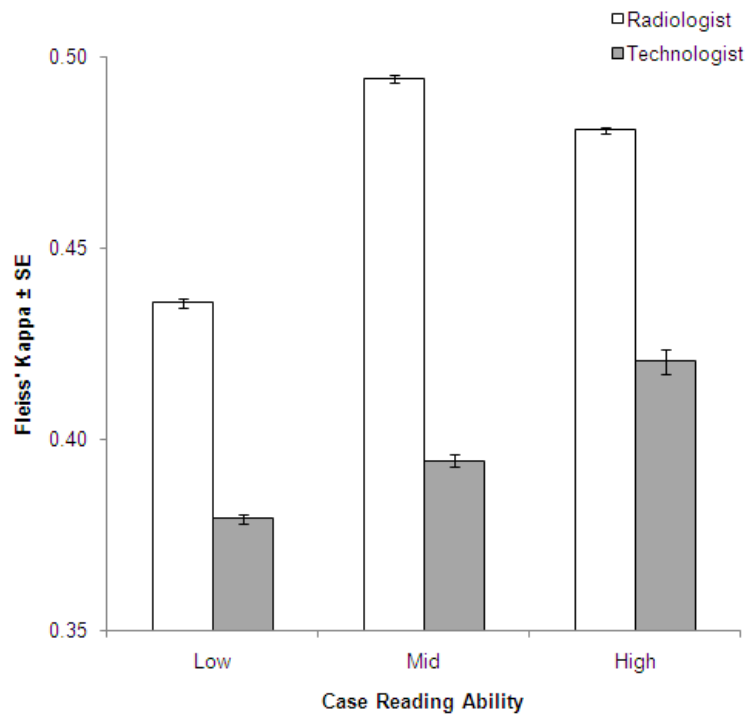


Figure 4: Level of agreement on breast density rating analysed by Occupational Group and Case Reading Ability.

Rate of Correct Recall (figure 5) was analysed by occupation (radiologist, technologists) and breast density. Mauchly's Test of Sphericity for repeated measures factors was failed, $\chi^2(2) = 75.07$, $p < .0001$, and degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.86$). There were main effects of breast density $F(1.72, 715.81) = 199.36$, $MSE = 0.007$, $p < .0001$, $\eta^2_p = 0.3$ and occupation $F(1, 417) = 8.49$, $MSE = 0.015$, $p < .01$, $\eta^2_p = 0.02$. There was no significant interaction between these two factors $F(1.72, 715.81) = 2.27$, $MSE = 0.007$, $p = n.s.$, $\eta^2_p = 0.01$. Pair-wise post-hoc tests were used to resolve the main effect of Breast Density rating and the Bonferroni correction was applied. The proportion of correct recalls was higher in Fatty ($M = 0.80$, $SD = 0.11$) than in Mixed Density ($M = 0.71$, $SD = 0.07$) or Dense ($M = 0.71$, $SD = 0.11$) breasts (both $ps < .05$), but the proportion of correct recalls was equivalent in Mixed Density and Dense breasts ($p = n.s.$). The proportion of correct recalls was higher for radiologists ($M = 0.746$, $SD = 0.07$) than for technologists ($M = 0.725$, $SD = 0.07$) ($p < .01$). Thus, sensitivity was higher

in Fatty breasts. Although the influence of occupation was significant, the size of the effect is extremely small. Its significance may reflect the large sample size.

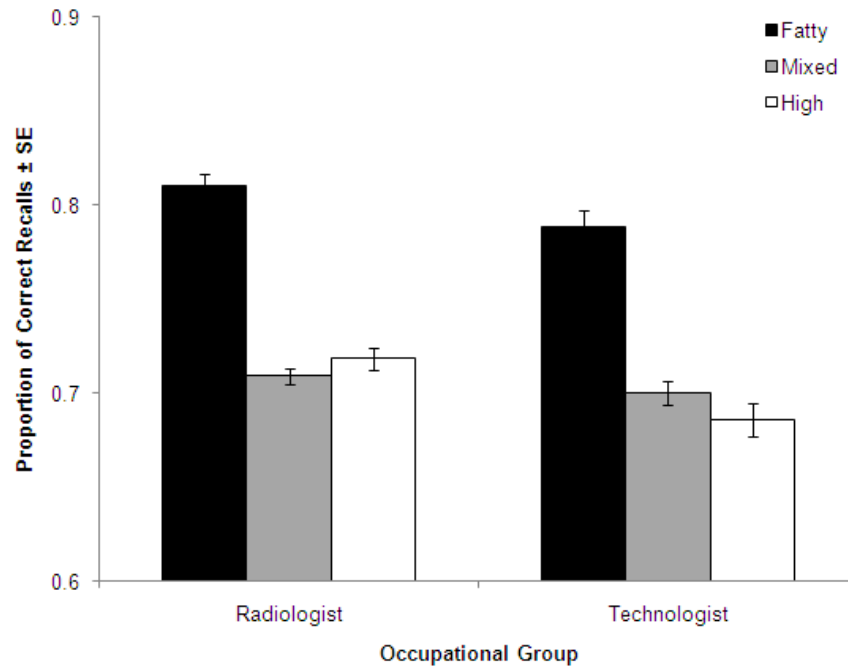


Figure 5: Proportion of Correct Recall decisions analysed by Occupational Group and Expert Density Rating.

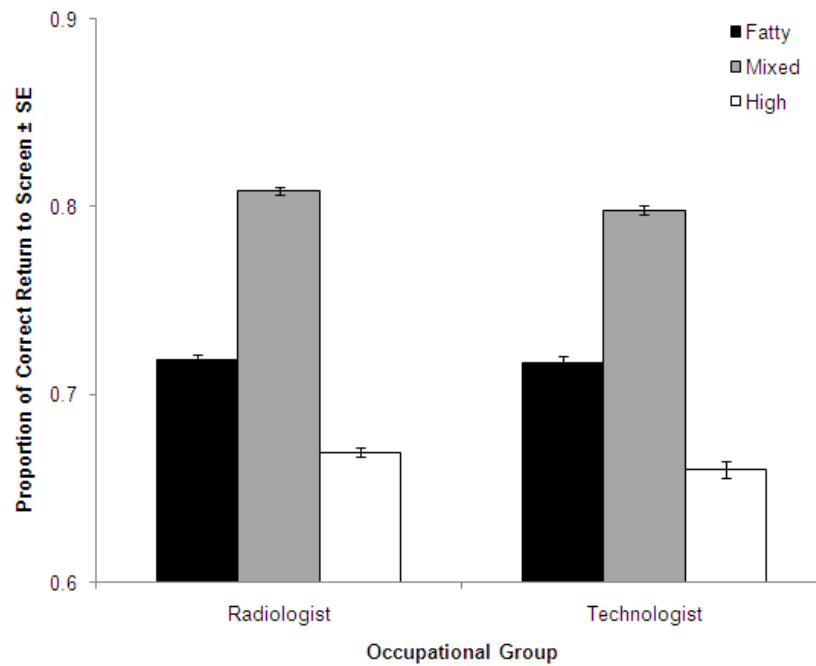


Figure 6: Proportion of Correct Return to Screen decisions analysed by Occupational Group and Expert Density Rating.

Rate of Correct Return to Screen decisions (figure 6) was analysed by occupation (radiologists, technologists) and breast density). Mauchly's Test of Sphericity for repeated measures factors was failed, $\chi^2(2) = 34.55$, $p < .0001$, and degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.93$). There were main effects of breast density rating $F(1.85, 772.44) = 1382.85$, $MSE = 0.002$, $p < .0001$, $\eta^2_p = 0.77$ and occupation $F(1, 417) = 5.52$, $MSE = 0.003$, $p < .05$, $\eta^2_p = 0.01$. There was no significant interaction between these two factors $F(1.85, 772.44) = 1.44$, $MSE = 0.002$, $p = \text{n.s.}$, $\eta^2_p = 0.00$. Pair-wise post-hoc tests were used to resolve the main effect of Breast Density rating and the Bonferroni correction was applied. The proportion of correct return to screen decisions was higher in Mixed Density ($M = 0.80$, $SD = 0.03$) than in Fatty ($M = 0.72$, $SD = 0.05$) breasts and higher in Fatty than in Dense ($M = 0.67$, $SD = 0.05$) breasts (all $ps < .05$). The proportion of correct recalls was higher for Radiologists ($M = 0.732$, $SD = 0.03$) than for technologists ($M = 0.725$, $SD = 0.03$) ($p < .05$). Thus, specificity was higher in Mixed Density breasts. Again, although the influence of occupation was significant, the size of the effect is extremely small. Its significance may reflect the large sample size.

3.5 Case difficulty

Given the association between agreement level on breast density and case reading ability, case difficulty has the potential to predict the level of agreement on a given case. Additionally, the differences between occupational groups in levels of agreement on breast density might, in part, be explained if different cases pose different levels of difficulty for the occupational groups. However, across cases there was no significant correlation between the overall proportion of correct responses for a given case and the overall probability of agreeing on the density rating for that case for either radiologists ($r_s = 0.04$, $n = 209$ cases, $p = \text{n.s.}$) or technologists ($r_s = 0.03$, $n = 209$ cases, $p = \text{n.s.}$), or overall ($r_s = 0.03$, $n = 209$ cases, $p = \text{n.s.}$). Further, radiologists and technologists appeared to find the same cases similarly easy or challenging with a near perfect correlation across cases between the overall proportion of correct responses for a given case for radiologists and for technologists ($r_s = 0.96$, $n = 209$ cases, $p < .0001$).

3.6 Measured observed case density

Although participants examined film mammograms, these were laser printed copies of digitised original film cases. Consequently an initial approach was undertaken to assess image greyscale density as perceived by participants when examining the digital data of these images. An edge detection algorithm was first used to determine the breast edges. To exclude image regions such as the axilla, a circle was fitted to encompass most of the breast. The region between the centre of this circle and either its boundary or the edge of the breast, whichever was closest, was then used in the calculation of estimates of breast density. Mean greyscale pixel value was calculated within the described regions across the four views of each case. Additionally, a simple pixel threshold (greyscale value > 199) was used to define continuous regions of high density. For each case, the number, mean size, and maximum size of these regions was recorded. Also recorded were the mean and maximum sizes of each continuous low density region (greyscale value < 200).

An attempt was then made to delineate the decision on which breast density category to choose by placing these derived measures of breast density into a multiple regression model, across the 209 cases, with mode density rating for each case as the dependent variable. Thus, the amount of variance in mode density rating across cases was described in terms of mean greyscale pixel value, number of high density regions, mean size of high density regions, maximum size of high density region, mean size of low density regions, and maximum size of low density region. (The number of low density regions was not considered because of its necessarily close relationship to the number of high density regions and the associated risk of co-linearity). This was done separately for the mode density ratings of radiologists, technologists, and the initial expert radiologists.

For all three groups mean greyscale pixel value appeared to have the greatest influence regarding breast density rating (highest value of β), when the other factors were held constant. Various weightings were observed for the contributions of the quantifications of high density regions and low density regions. Overall, greater values for high density region quantities exhibited positive associations with density ratings, whilst greater values for low density region quantities exhibited negative associations with density ratings. Thus, when making a decision about breast density, it appears that the overall greyscale value of the breast is influential, alongside the amounts of high and low density regions. The specific contributions of high and low density regions may be dependent upon the method and greyscale pixel level threshold employed here to quantify high and low density regions. However, these multiple regression models serve to

demonstrate that aspects of the individual mammograms were being used in the rating of breast density in a predictable manner. Not considered here are the effects of viewing cases sequentially and comparisons between cases in making breast density decisions.

4. DISCUSSION

In taking part in the PERFORMS scheme participants make several judgements about each case which they examine. Amongst these, they judge the overall density of the breast on a three point scale – this scale being used to simplify this particular judgement and also because this aspect is not of prime importance within the overall objectives of the scheme itself. To some extent then the findings may well be related to the relative degree of importance that participants ascribe to assigning density judgements and whether different results would be found if a larger density rating scale was used is a mute point.

In the data examined here the vast majority of cases were of judged mixed density. This may well reflect the use of three density categories although it was to be expected that more cases would fall into the dense category. The judgements of the expert breast screening radiological panel on case density were closely related to the density judgments of the radiologist participants as well as the technologists. There was very good agreement between radiologists and technologists on the mode density assigned to each case, although overall a moderate agreement was found within each of these groups. The participants used the mixed density category more than did the original expert panel.

Agreement on density category was significantly higher amongst the radiologists than amongst the technologists. One reason for this may be due to experience as the radiologists as a group were the more experienced. However, when this was examined it would appear that degree of experience was not related to agreement on breast density. How well radiologists and technologists did on the PERFORMS scheme was then studied as a possible explanation and it was found that those who did better on the scheme had a higher level of agreement on judged case density. Sensitivity was higher in cases judged fatty and specificity was higher in cases judged of mixed density which would be expected if increasing breast density masks abnormality appearances. In terms of case difficulty both groups found the same cases equally relatively straightforward or difficult; similarly there was no relationship between the number of correct decisions per case and its judged density. In an effort to quantify perceived breast density unsurprisingly it was found that the overall greyscale value of the image is important, together with the amounts of high and low density regions.

5. CONCLUSIONS

Observer variation is rife in all radiological performance studies and here variation was also found in the reported breast density ratings given to these self assessment screening cases. Better agreement amongst participants on density ratings assigned to cases was found for radiologists as compared to technologists and this was not related to the experience of the radiologists. Those who performed better on the self assessment scheme also evidenced better agreement on case density judgements.

Is it important that such differences were found in ratings of breast density? It is if reported breast density is to be used reliably and prospectively to aid or plan personalised treatment and follow-up for a particular individual in a rigorous fashion. Future work will explore further aspects of such reported and measured density ratings in more detail. Currently, in the UK, breast density is not routinely recorded during screening.

6. ACKNOWLEDGEMENTS

This work is supported by the UK NHS Breast Screening Programme.

7. REFERENCES

- [1] Patnick J., [NHS Breast Screening Programme Annual Review 2010: Overcoming Barriers] (2010)
- [2] Gale A.G., "PERFORMS – a self assessment scheme for radiologists in breast screening," *Seminars in Breast Disease*, 6(3), 148-152, (2003)
- [3] Gale A.G., "Maintaining quality in the UK breast screening program", *Proc. SPIE Medical Imaging 2010*, 7627, 1-11 (2010)
- [4] Gale A.G. and Scott H., "Measuring Radiology Performance in Breast Screening," *Contemporary Issues in Cancer Imaging – Breast Cancer* (2010)
- [5] Wolfe J.N., "Breast patterns as an index of risk for developing breast cancer," *AJR* 126, 1130-1139 (1976)
- [6] Harvey J.A. and Bovbjerg V.E., "Quantitative assessment of mammographic breast density: relationship with breast cancer risk," *Radiol.*, 230, 29-41 (2004)
- [7] Sickles E.A., "Wolfe mammographic parenchymal patterns and breast cancer risk", *AJR* 188, 301-303 (2007)
- [8] Gram I.T., Funkhouser E. and Tabar L., "The Tabar classification of mammographic parenchymal patterns," *Eur. J. Radiol.*, 24, 131-136 (1997)
- [9] Boyd N.F., Byng J.W., Jong R.A., et al., "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study," *J. Natl. Cancer Inst.*, 87, 670 -675 (1995)
- [10] American College of Radiology, [ACR breast imaging reporting and data system atlas] (2003)
- [11] Garrido-Esteva M.G., et al., "Evaluation of mammographic density patterns: reproducibility and concordance among scales," *BMC Cancer*, 10, 485-494 (2010)
- [12] Gram I.T., et al., "Percentage density, Wolfe' and Tabar's mammographic patterns: agreement and association with risk factors for breast cancer," *Breast Cancer Res.*, 7(5), 854-861 (2005)
- [13] Pinker K., Perry N., Vinnicombe S., Shiel S. and Weber M., "Conspicuity of breast cancer according to histopathological type and breast density when imaged by full field digital mammography compared with screen film mammography," *Eur. Radiol.*, 21, 18-25, (2011)
- [14] Subashini T.S., Ramalingam V. and Palanivel S., "Automated assessment of breast tissue density in digital mammograms," *Comp. Vision Underst.*, 114, 33-43, (2010)
- [15] Yaffe M.J., "Mammographic density: Measurement of mammographic density," *Breast Cancer Res.*, 10, 209, (2008)
- [16] Gilles R., et al., "Breast cancer in women 35 years old or younger: clinical and mammographic features," *Eur. Radiol.*, 5, 630-632, (1995)