

Keywords: mammographic density; breast density; breast cancer; premenopausal

Comparison of fully and semi-automated area-based methods for measuring mammographic density and predicting breast cancer risk

U Sovio^{1,7,8}, J Li^{2,3,7}, Z Aitken^{1,9}, K Humphreys⁴, K Czene⁴, S Moss⁵, P Hall⁴, V McCormack⁶ and I dos-Santos-Silva^{*,1}

¹Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK; ²Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, Singapore; ³Human Genetics Lab, Genome Institute of Singapore, 60 Biopolis Street, 02-01, Singapore 138672, Singapore; ⁴Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, Stockholm 17177, Sweden; ⁵Centre for Cancer Prevention, Queen Mary University of London, Wolfson Institute of Preventive Medicine, Charterhouse Square, London EC1M 6BQ, UK and ⁶Environment and Radiation Section, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon, France

Background: Mammographic density is a strong risk factor for breast cancer but the lack of valid fully automated methods for quantifying it has precluded its use in clinical and screening settings. We compared the performance of a recently developed automated approach, based on the public domain ImageJ programme, to the well-established semi-automated Cumulus method.

Methods: We undertook a case-control study within the intervention arm of the Age Trial, in which ~54 000 British women were offered annual mammography at ages 40–49 years. A total of 299 breast cancer cases diagnosed during follow-up and 422 matched (on screening centre, date of birth and dates of screenings) controls were included. Medio-lateral oblique (MLO) images taken closest to age 41 and at least one year before the index case's diagnosis were digitised for each participant. Cumulus readings were performed in the left MLO and ImageJ-based readings in both left and right MLOs. Conditional logistic regression was used to examine density–breast cancer associations.

Results: The association between density readings taken from one single MLO and breast cancer risk was weaker for the ImageJ-based method than for Cumulus (age–body mass index-adjusted odds ratio (OR) per one s.d. increase in percent density (95% CI): 1.52 (1.24–1.86) and 1.61 (1.33–1.94), respectively). The ImageJ-based density–cancer association strengthened when the mean of left–right MLO readings was used: OR = 1.61 (1.31–1.98).

Conclusions: The mean of left–right MLO readings yielded by the ImageJ-based method was as strong a predictor of risk as Cumulus readings from a single MLO image. The ImageJ-based method, using the mean of two measurements, is a valid automated alternative to Cumulus for measuring density in analogue films.

*Correspondence: Professor I dos-Santos-Silva; E-mail: isabel.silva@lshtm.ac.uk

⁷Joint first authors.

⁸Current address: Department of Obstetrics and Gynaecology, University of Cambridge, Box 223, Level 2, The Rosie Hospital, Robinson Way, Cambridge CB2 0SW, UK.

⁹Current address: Centre for Women's Health, Gender and Society, Melbourne School of Population and Global Health, Level 3, 207 Bouverie Street, The University of Melbourne, VIC 3010, Australia.

Received 11 October 2013; revised 17 January 2014; accepted 20 January 2014; published online 20 February 2014

© 2014 Cancer Research UK. All rights reserved 0007–0920/14

Mammographic density, which represents the amount of radio-dense fibroglandular tissue in the breast that appears as white areas on a mammogram, is of clinico-epidemiological relevance because it is one of the strongest markers of susceptibility to breast cancer (McCormack and dos Santos Silva, 2006) and because it affects the sensitivity of screening mammography (Boyd *et al*, 2007). Several qualitative and quantitative methods to measure mammographic density have been developed with the quantitative computer-assisted interactive thresholding method (Byng *et al*, 1994), as commonly executed by the Cumulus software, being currently regarded as the 'gold' standard for digitised analogue films (McCormack and dos Santos Silva, 2006). This method is a rather time-consuming, labour-intensive and, hence, costly approach that limits its use in large-scale research studies and precludes its incorporation into screening activities. Moreover, although Cumulus has been shown to have high between- and within-reader reliability in controlled research conditions (McCormack *et al*, 2007), it is unlikely that similar high reader reliability values will be observed when Cumulus is used in clinical/screening practice.

A fully automated method, which attempts to mimic Cumulus, was recently developed by some of us (JL, KC, KH and PH) based on a public domain Java image processing programme, ImageJ (Li *et al*, 2012) (referred to as ImageJ-based method hereafter). In essence, this method measures several image parameters (Li *et al*, 2012) that describe features of the breast as they appear on a mammogram and then selects those that, in an image set with known Cumulus-density readings (that is, the training set), are shown to predict Cumulus density. The selected parameters are then used in a regression model to predict density from images with unknown Cumulus values. Although the performance of this ImageJ-based method has been shown to be comparable to that of Cumulus in Swedish postmenopausal women in terms of its percent density measurements (Li *et al*, 2012), its performance has not yet been comprehensively tested and validated in other settings, for example, in premenopausal women who typically have higher mammographic density levels or in a population with mammograms taken under variable conditions (for example, variable X-ray equipment and acquisition settings), or in terms of absolute as well as percent density. The aim of this study is to compare the agreement and reliability of ImageJ-based and Cumulus-based measurements for both absolute and percent density, their associations with known breast cancer risk factors, and their ability to predict breast cancer risk, among predominantly premenopausal women who participated in a British trial of annual mammographic screening at younger ages.

MATERIALS AND METHODS

Study design. A nested case-control study of breast cancer in relation to prediagnostic mammographic density was undertaken within the Age Trial (International Standard Randomised Controlled Trial, number ISRCTN24647151), a British trial of annual mammographic screening at younger ages (Moss, 1999; Moss *et al*, 2006). About 54 000 women aged 40–41 years were randomised to the intervention arm of this trial between 1991 and 1997 and invited to attend annual mammographic screenings from ages 40–41 to 48–49, after which they joined the 3-yearly National Health Service Breast Screening Programme (NHSBSP). A total of 23 NHSBSP centres took part in the study. Screening in the trial was by analogue X-ray mammography. Both cranio-caudal and medio-lateral oblique (MLO) views were taken of each breast at the first trial screening round, but only MLO views in subsequent rounds. Participants have been followed up through NHS health registers for cancer incidence and mortality since their entry into the trial.

Women in the intervention arm who were diagnosed with breast cancer at least 1 year after their first negative screen and before 2006 were eligible cases for the present study ($n = 442$). Up to six eligible controls were initially randomly selected for each case among women in the intervention arm who were still alive and had not been diagnosed with breast cancer at the time of the case's diagnosis matching on screening centre, date of birth (within ± 3 months) and date of each of the case's prediagnostic screens (within ± 3 months). Eligible cases and controls were contacted via their general practitioner, and asked to complete a short questionnaire on anthropometric and reproductive factors and to provide written consent for their mammograms to be accessed. The mammograms of consented participants were retrieved from the relevant NHS screening centres and digitised on an Array 2905 laser digitizer (0–4.0 OD, 50 micron resolution, 12 bit image). The first two control women for whom both questionnaire data and mammograms were obtained were included in the study. 76% of eligible cases and 80% of eligible controls completed the questionnaire and gave consent for access to their mammograms; images could be retrieved for 89% of consented cases and 93% of consented controls.

The study complied with the Helsinki Declaration and was approved by the UK NHS South-East Multi-Centre Research Ethics Committee (05/MREC01/77). All participants provided written informed consent.

Cumulus density readings. Density readings using the Cumulus software (University of Toronto, Toronto, Ontario, Canada) (Byng *et al*, 1994) were performed by three observers (IdSS, VM and ZA). Because the Cumulus readings are labour-intensive, readings were only performed in one single image. The left MLO image taken at age 41 years (that is, first screening round), or closest to age 41 years if the woman missed the first round, was chosen because the MLO views were performed in all screening rounds and because left–right breast density values are very similar (McCormack *et al*, 2007). Images were read in batches, with all images from any given case-control set being included in the same batch and read by the same observer. The observer used the Cumulus software to define the breast area on the digitised image by delimiting the skin edge (masking any non-breast features, for example, pectoral muscle) and to select a grayscale threshold to divide dense and nondense pixels. Cumulus used this information to automatically estimate the breast area (in cm^2), dense area (in cm^2) and percent density for each image. Readers were blinded to the woman's name, age, dates of mammography and case/control status. A random 10% sample of images from all batches were included in each batch as duplicates and read twice independently. Each reader had a high degree of within- and between-batch reliability (for example, intraclass correlation (ICC) (95% CI) for within-batch reliability for percent density varied between 0.94 (0.90, 0.98) and 0.97 (0.95, 0.99) across the various readers, whereas ICC for between-batch variability varied between 0.90 (0.84, 0.97) and 0.97 (0.95, 0.99)).

ImageJ density readings. The fully automated ImageJ-based method of measuring mammographic density has been previously described in detail (Li *et al*, 2012). Briefly, it attempts to mimic Cumulus by using a fully automated thresholding procedure to perform the density readings. It comprises several steps: (i) preprocessing (priming) of DICOM images to remove patient identification tags and automatically delimit the breast region by masking the background; (ii) manual inspection and removal of preprocessed images with unremoved tags and those that did not convert into a readable format (for example, wrong film size, compressed images); (iii) image analysis using 15 global thresholding methods to distinguish dense from nondense areas of the breast; and (iv) conversion of the image analysis output to density readings via a machine learning approach for which an independent set of images with known Cumulus readings were

used for training the model. Principal component analyses (PCs) are used to select the parameters that 'predict' Cumulus values. Each PC is a linear combination of the original variables measured by ImageJ (a list and description of these original variables, and of their weights in the PCs, is given in Additional File 3, Supplementary Table S1, of Li *et al* (2012)). Prediction models were developed using penalised regression procedures as described in the statistical analysis section of Li *et al* (2012). Separate prediction models were fitted to estimate percent density and breast size; absolute dense area was obtained from the product of percent density and breast size.

The digitised images of the current UK study were of variable size and resolution, in contrast to the previously published Swedish study whose images were similar (Li *et al*, 2012). Thus, before using the ImageJ-based approach in the present study, the programme was re-trained in images from two previously conducted British studies with mammogram acquisition contemporaneous to that in the current study, available Cumulus readings (conducted by two of the readers of the present study) and digitised films from a wide range of sizes, resolutions, ages, density levels and image acquisition settings (see Supplementary Data for full details on the methods and results of this training). The same left MLO image selected from each participant for the Cumulus readings was chosen for the ImageJ-based readings to allow direct comparison between the two methods. In addition, the right MLO image from the same screening round was also read automatically using ImageJ (at the cost of having to digitise twice the number of images but of not having to read them individually) to allow estimation of its intra-method reliability and agreement, and also to increase the precision of the measurements by taking the mean of the left and right MLO imageJ-based measurements (referred to as ImageJ-based MLO mean hereafter); if only one MLO image (left or right) was available (38 cases; 72 controls) the value for that image was used instead. After exclusion of five untraced and one compressed file images, left MLO images for preprocessing were available for 768 participants; a further 50 images were subsequently excluded because of unremoved patient identification tags or wrong film size (Figure 1). Right MLO images were available for preprocessing for 766 women (after exclusion of five untraced and one compressed

image), with 110 images being excluded because of unremoved tags or wrong film size (Figure 1). In addition, two clear outliers were excluded from both left and right MLO images due to unreasonable ImageJ values for dense and breast areas (for example, for both outliers, dense area = 939.2 cm² and breast area = 10.2 cm²) indicating a problem in the ImageJ estimation process; Cumulus readings for these images were within normal ranges (dense area = 39.2–94.1 cm²; breast area = 45.2–238.6 cm²).

Statistical methods. Women excluded due to failures in the ImageJ process were compared with women included in the analysis in terms of the distributions of known breast cancer risk factors and Cumulus mammographic density measurements. For the sample included in the analysis, differences in the baseline characteristics of the cases and controls were described, including the distribution of known breast cancer risk factors and mammographic density measurements. Histograms and scatter plots were used to visually examine and compare distributions of Cumulus and ImageJ-based absolute dense area, breast area and percent density. A natural log-transformation was used to normalise the breast area distributions; no transformations were required for absolute dense area or percent density values. Intra-method correlations were assessed by estimating Pearson correlation coefficients (PCC) and level of agreement by estimating mean difference between ImageJ and Cumulus values (with 95% limits of agreement). Concordance correlation coefficients (CCC) were calculated to assess inter-method (ImageJ-based vs Cumulus) reliability and ImageJ-based intra-method reliability (left vs right MLO readings).

Linear regression models were used to assess, in controls, associations of ImageJ-based and Cumulus absolute and percent density values with well-established breast cancer risk factors. Models were adjusted for age and body mass index (BMI) at mammography as continuous variables, except if these two variables were the primary exposures of interest, with models of Cumulus measures being further adjusted for reader. Regression coefficients represent the difference in ImageJ-based or Cumulus density measures associated with a unit change in the explanatory variable.

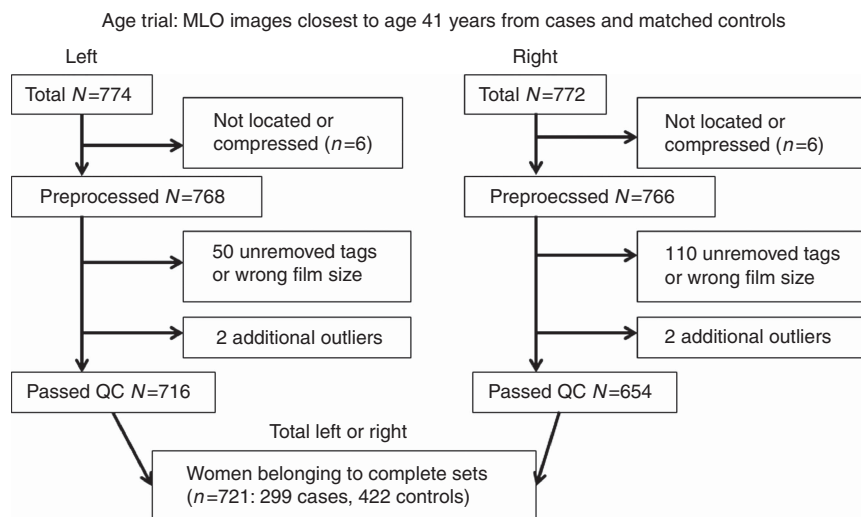


Figure 1. Flow chart detailing case-control selection from the Age Trial for inclusion in this study. ImageJ's built in DICOM reader does not support the reading of compressed DICOM images. The original DICOM images were preprocessed to remove person identification tags and to delimit the breast area from the rest of the image. A higher number of right than left images was found to have 'unremoved tags' because the ImageJ-based method automatically flipped horizontally any image that was annotated as being 'R' in the DICOM header; however, some of these 'R' images were actually scanned like 'L' images and, hence, they did not need to be flipped. Complete sets consist of one case and at least one control, with each having at least one (left or right) MLO image; both left and right MLO images were available for most participants except for 38 cases and 72 controls for whom only one MLO image was available.

Conditional logistic regression models were fitted to examine the association of absolute and percent density, as measured by either ImageJ or Cumulus, with breast cancer risk. Absolute and percent density values were included in the models either as continuous (expressed in s.d. scores) or as categorical variables defined by quartiles of the distributions in controls or by previously published cutoff points (Boyd *et al*, 2007); for the latter, the two highest percent density categories (50–74 and 75+) were combined due to small numbers: <10, 10–24, 25–49 and $\geq 50\%$. These standard cutoffs for percent density corresponded to the 5th, 25th and 79th percentiles of the control distribution for ImageJ-based (MLO mean) measurements and to 14th, 35th and 74th of the control distribution for Cumulus measurements. All models were adjusted for age and BMI at the time of mammography (basic adjustment); reader (for Cumulus) was automatically adjusted for by design as all films belonging to the same case-control set were read by the same observer. Models for selected analyses were further adjusted for number of children, menopausal status and family history of breast cancer (full adjustment).

The area under the curve (AUC) of the receiver operating characteristic curve was used to compare the ability of ImageJ-based and Cumulus percent density, either alone or in combination with other known breast cancer risk factors, to discriminate between women who went on to subsequently develop breast cancer and those who did not. The AUC analyses were performed using frequency (by breaking the matching while adjusting for relevant variables) rather than individual matching as there are no well-established methods for the latter. An AUC of 0.5 indicates that the method has no discriminative value, that is, is no better than a random guess.

All analyses were performed in Stata version 12.1.

RESULTS

A total of 299 complete sets (that is, consisting of one case and at least one control, with each having at least one MLO image) were included in the analysis, corresponding to 299 cases and 422 controls (Figure 1). There was no evidence that women who were excluded from the study, because ImageJ failed to produce valid density values, differed from those included in terms of their age, BMI or reproductive factors but, on average, they had lower Cumulus percent density values: median (interquartile range, IQR) 37.6% (26.6%, 45.9%) for excluded cases vs 42.9% (29.7%, 61.1%) for those included, $P = 0.007$; 31.0% (11.3%, 47.1%) for excluded controls vs 35.7% (18.1%, 51.3%) for those included, $P = 0.02$.

Cases and controls had, as expected, a similar age at mammography, but cases were more likely to be pre/perimenopausal at mammography and to have a positive family history of breast cancer, a lower BMI and a later age at first birth (Table 1). There were no clear case-control differences in Cumulus and ImageJ-based estimates of breast area but cases had, on average, a higher absolute dense area and a higher percent density according to both methods (Table 1).

Among controls the distributions produced using the ImageJ-based approach for dense area and percent density showed similar medians and ranges to the equivalent ones generated by Cumulus (Table 1), but different shapes (Figure 2). Cumulus produced flatter density distributions, that is, with higher proportions of women at the lower and higher ends of the distributions, than the ImageJ-based approach (Figure 2). For example, the percentage of women with dense tissue occupying <5% of the breast was estimated to be 5.5% by Cumulus but only 1.6% by the ImageJ-based approach; similarly, the percentage of those with $\geq 75\%$ dense tissue was 5.6% according to Cumulus but only 2.4% according to the ImageJ-based approach.

Level of agreement between ImageJ-based and Cumulus density estimates. The breast area estimates for the left MLO view yielded by ImageJ were highly correlated to the corresponding values generated by Cumulus ($PCC = 0.91$), although a few outliers were identified; the PCCs for dense area and percent density were lower than those for breast area (0.68 and 0.72, respectively; Figure 3). About 50% of the controls were classified in the same quartile of the percent density distribution by the two methods (that is, 49% if the ImageJ-based left MLO readings were used; 51% if the ImageJ-based mean of left and right MLO readings were used) and about 90% were classified in the same or ± 1 adjacent quartile (that is, 90 and 91%, respectively). There was no evidence of systematic differences between the measurements produced by the two methods except that ImageJ-based estimates tended to slightly under-estimate breast area relative to the reader's evaluation in Cumulus ($P = 0.006$; Table 2). The inter-method reliability for breast area was high ($CCC = 0.92$), but lower for dense area and percent density (0.63 and 0.68, respectively; Table 2). Intra-method reliability for ImageJ-based approach, based on left–right MLO comparisons, was high for breast area ($CCC = 0.90$) but lower for dense area ($CCC = 0.75$) and percent density ($CCC = 0.82$). There were no systematic left–right differences in the estimates of dense area produced by the ImageJ-based approach, but the estimates for breast area were higher, and hence estimates for percent density lower, for the left MLO (Table 2).

Density associations with known breast cancer risk factors.

Overall, associations between well-established breast cancer risk factors and ImageJ-based absolute and percent density were in the direction expected given the effects of these variables on risk, and roughly of similar magnitude to those observed with Cumulus (Supplementary Table 1). Among controls, percent density for both methods was lower in women who were older (although not significantly as the age range was rather narrow) and had a higher BMI at mammography. After adjustment for these two variables (and, for Cumulus measurements, also for reader), percent density was also lower in controls who had ≥ 2 children, but higher among those who were postmenopausal and those who had a later age at menarche; however, none of these associations was statistically significant. Similar patterns were observed for absolute dense area (Supplementary Table 2).

Density and breast cancer risk. Both ImageJ-based and Cumulus percent density were positively associated with breast cancer risk, with no evidence of departure from linearity. For density measurements based on a single MLO image, the magnitude of the age–BMI-adjusted relative risk estimate was weaker for ImageJ-based than for Cumulus (Table 3). However, the magnitude of the association for the mean of the left–right ImageJ-based measurements was remarkably similar to that for a single Cumulus reading – that is, a 61% increase in risk per every s.d. increase in percent density by each approach (Table 3). Further adjustment for parity, menopausal status and family history did not affect these findings (Table 3).

Both methods showed a clear positive trend in the risk of breast cancer across quartiles of density ($P < 0.001$, Table 3). After adjustment for age and BMI, women in the top quartile of Cumulus percent density had a 4.68-fold (95% CI 2.56, 8.55) increase in breast cancer risk compared with those in the lowest quartile; the corresponding associations for ImageJ were slightly weaker – OR of 2.99 (95% CI 1.72, 5.19) for MLO mean and of 3.61 (95% CI 1.99, 6.55) for left MLO (these estimates had wide 95% CIs because of the smaller number of cases in the reference bottom quartile; therefore, we used the third quartile as the referent category in Table 3). Strong positive trends in risk were also observed when published cutoff points were used ($P < 0.001$ for all), albeit less consistently so for the ImageJ-based approach (mainly due to the smaller number of women with density

Table 1. Baseline characteristics, and Cumulus and ImageJ-based density estimates, by case-control status

Baseline characteristics	Cases (N = 299)		Controls (N = 422)	
	N ^a	Mean (s.d.)	N ^a	Mean (s.d.)
Age at mammography (years)	299	41.1 (0.7)	422	41.2 (0.6)
Age at questionnaire (years)	291	53.2 (1.4)	415	51.3 (3.3)
Age at diagnosis (years), cases only	299	46.6 (2.9)	—	—
BMI at mammography (kg m ⁻²) ^b	286	24.9 (4.3)	413	25.5 (4.5)
Age at menarche (years)	288	12.8 (1.7)	327	12.8 (1.5)
Age at first birth (years) ^c	244	26.9 (5.2)	313	26.0 (5.2)
No. of children	292	1.8 (1.1)	418	1.9 (1.0)
	N ^a	% (n ^d)	N ^a	% (n ^d)
Ever breastfed	293	59.0 (173)	198	57.6 (114)
Postmenopausal status at mammography ^e	280	9.6 (27)	403	11.4 (46)
Positive family history of breast cancer	293	17.4 (51)	419	6.9 (29)
Density variables				
	N ^a	Median (IQR)	N ^a	Median (IQR)
Dense area (cm ²)				
Cumulus left MLO	299	50.5 (35.9, 72.1)	422	41.0 (23.5, 59.6)
ImageJ left MLO	292	48.6 (38.3, 64.3)	406	41.7 (30.3, 54.9)
ImageJ left–right MLO mean	299	49.4 (39.4, 62.4)	422	42.0 (30.5, 53.5)
Breast area (cm ²)				
Cumulus left MLO	299	125.1 (95.9, 159.2)	422	124.3 (91.4, 162.4)
ImageJ left MLO	292	122.5 (99.4, 157.8)	406	123.9 (96.7, 155.3)
ImageJ left–right MLO mean	299	125.9 (100.2, 160.4)	422	126.7 (101.3, 159.3)
Percent density				
Cumulus left MLO	299	42.9 (29.7, 61.1)	422	35.7 (18.1, 51.3)
ImageJ left MLO	292	44.1 (31.6, 56.0)	406	38.7 (26.2, 49.3)
ImageJ left–right MLO mean	299	44.4 (31.6, 54.8)	422	37.7 (25.3, 48.5)

Abbreviations: BMI = body mass index; IQR = interquartile range; MLO = medio-lateral oblique view.

^aNumber of women with non-missing data for each of the variables examined.

^bBMI at age 40 years, as ascertained in the questionnaire, was taken as a proxy for BMI at mammography as the two ages are very close. Information on BMI at age 40 was missing for 33 cases and 31 controls; BMI at the time of questionnaire administration was taken as a proxy for BMI at mammography for these women.

^cAmong parous women only.

^dNumber of women who had ever breastfed, were postmenopausal at the time of mammography, or who had a positive family history of the disease.

^eMedian (IQR) age at menopause for participants who were postmenopausal at mammography was 39 (IQR: 35–40) years for cases and 37 (IQR: 34–39) years for controls.

<10%: 72 for Cumulus but only 29 for ImageJ). Similar patterns were observed with dense area, with both methods producing again fully adjusted risk estimates of similar magnitude – that is, a 61–66% increase in risk per every s.d. increase in Cumulus or ImageJ dense area (Supplementary Table 3). It is notable that, for both methods, the magnitude of the risk estimates associated with one s.d. increase in dense area was at least as high, if not higher, than those associated with an equivalent increase in percent density (Table 3; Supplementary Table 3).

Interestingly, when readings from the two approaches were included jointly into the fully adjusted models the OR estimate per one s.d. increase in Cumulus percent density decreased (for example, from 1.58 (Table 3) to 1.43 (95% CI 1.08, 1.88; $P=0.01$) when ImageJ-based readings were based on the left MLO), whereas the corresponding OR estimate for the ImageJ-based readings was no longer significant (OR = 1.14, 95% CI 0.86, 1.52, $P=0.36$), consistent with a single ImageJ-based reading being more affected by measurement error than a single Cumulus measurement.

Cumulus and ImageJ-based (left MLO) percent density showed a similar ability to discriminate between women who subsequently developed breast cancer and those who did not (AUC = 0.61 (95% CI 0.57, 0.65) for each method). Addition of other risk predictor variables (that is, age, BMI, number of children, menopausal status

and family history) to the model improved their discriminatory power only slightly (AUC: 0.65 (95% CI 0.61, 0.69) for Cumulus and 0.64 (95% CI 0.60, 0.69) for ImageJ (left MLO); $P=0.75$ for the difference between the two methods) (Supplementary Figure). The AUC for ImageJ did not improve when mean left–right MLO was used (AUC for the full model: 0.64 (95% CI 0.60, 0.68)). Inclusion of both Cumulus and ImageJ percent density readings jointly into the same model did not improve discrimination (for example, AUC = 0.65 when readings from the two methods were examined in combination with other risk factors).

DISCUSSION

The performance of the fully automated ImageJ-based approach relative to that of Cumulus had so far been assessed only in a previous study (Li *et al*, 2012). This study, which was confined to the evaluation of ImageJ measurements of percent density among a relatively homogeneous sample of Swedish postmenopausal women, showed that ImageJ was as good a predictor of risk of breast cancer as Cumulus, with women whose percent density was >25% having over twice the risk of developing the disease relative to those whose density was <5% (Li *et al*, 2012); no risk estimates based on

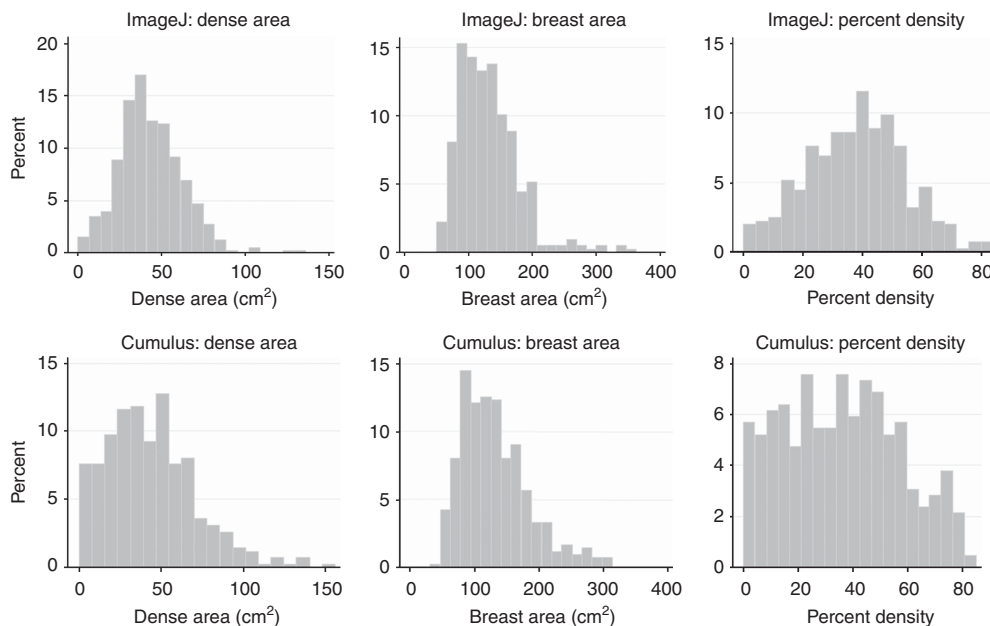


Figure 2. Distributions of Cumulus ($N=422$) and ImageJ-based ($N=406$) breast area and absolute and percent density values in controls, left MLO view.

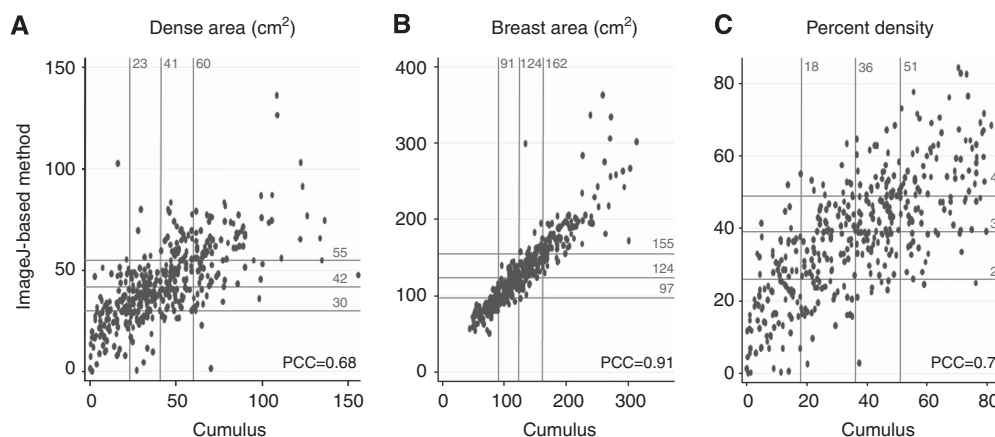


Figure 3. Scatter plots of ImageJ vs Cumulus measurements (left MLO) among controls: (A) dense area; (B) breast area; (C) percent density. The quartile cutoff points for each method are indicated by the vertical and horizontal lines on the graphs.

continuous density measurements were provided in that paper. In the present study, we extended this validation to a population of predominantly premenopausal women who had their mammograms taken in different screening centres using variable X-ray equipment and acquisition settings. Furthermore, we evaluated the performance of the ImageJ-based approach in relation to its measurements of percent as well as absolute dense area.

We have previously shown (McCormack *et al*, 2007) that the average of the left and right MLO Cumulus readings yielded only slightly higher reliability estimates than those produced by a single (left or right) MLO measurement in a study of UK postmenopausal women who underwent mammography as part of routine screening (with Cumulus readings performed by one of the readers in the present study). ICC estimates for percent density were 0.92 for single MLO vs 0.96 for mean left–right MLO; the corresponding figures for dense area were 0.89 vs 0.94 and for breast size 0.96 vs 0.98, respectively. McCormack *et al* (2007) estimated that if the true relative risk for the high density–breast cancer association was 4 then the magnitude of the observed association would be attenuated due to non-differential measurement error to 3.68 if the left–right mean was available and to 3.43 if only one MLO reading

was available. Thus, our study was specifically designed to assess whether one or two ImageJ-based readings per subject performed better than a single Cumulus reading because: (i) it is usually prohibitively expensive (in terms of both labour and time) to perform more than a single Cumulus reading per subject in large-scale studies; and (ii) risk prediction increased only slightly when the mean of two Cumulus readings are used. Our findings showed that Cumulus readings are slightly better predictors of breast cancer risk than those from the ImageJ-based approach when only one single MLO image is read. However, when the ImageJ-based method was based on the average of readings from two MLO images this method performed equally well to Cumulus on a single MLO image in terms of its associations with known breast cancer risk factors and with subsequent breast cancer risk. These findings suggest that twice the number of films would need to be digitised to ensure that the ImageJ-based method achieves the same performance as Cumulus. For some studies, the additional time and resources required to do this may out-weight the time saved by using a fully automated procedure to perform the density readings. Furthermore, often only one image per participant is available in historical collections of digitised images.

Table 2. Inter-method reliability (ImageJ based vs Cumulus), and intra-method reliability for the ImageJ-based approach (left vs right MLO)

Density variable	Inter-method (ImageJ-based-Cumulus)			Intra-method (left-right ImageJ based)			
	Cumulus left MLO, median (IQR)	ImageJ left MLO, median (IQR)	Inter-method CCC ^a	Inter-method mean difference (limits of agreement) ^a	ImageJ right MLO, median (IQR)	Intra-method CCC ^a	Intra-method mean difference (limits of agreement) ^a
Dense area (cm ²)	46 (29, 64)	45 (34, 59)	0.63	1.37 (-39.1, 41.8) P=0.08	44 (32, 57)	0.75	-0.21 (-27.2, 26.8) P=0.80
Breast area (cm ²)	125 (93, 161)	123 (98, 156)	0.92	-1.51 (-34.7, 23.5) P=0.006	126 (102, 157)	0.90	3.87 (-20.2, 35.3) P<0.0001
Percent density	40 (24, 56)	40 (29, 52)	0.68	-0.49 (-30.7, 29.7) P=0.40	39 (27, 51)	0.82	-1.31 (-21.5, 18.9) P=0.002

Abbreviations: CCC = concordance correlation coefficient; IQR = interquartile range; MLO = medio-lateral oblique view.

^aFor log-transformed breast area, the values for CCC, mean difference and 95% limits of agreement have been back-transformed.

Table 3. Association of Cumulus and ImageJ percent density with breast cancer risk

PD measurement (continuous) (per 1 s.d. increase)	Adjusted for age and BMI		Fully adjusted ^a	
	N	OR (95% CI)	N	OR (95% CI)
Cumulus left MLO	672	1.61 (1.33, 1.94)	617	1.58 (1.28, 1.94)
ImageJ left MLO	637	1.52 (1.24, 1.86)	582	1.45 (1.17, 1.80)
ImageJ left-right MLO mean	672	1.61 (1.31, 1.98)	617	1.55 (1.25, 1.93)
PD measurement (categorical)	Quartiles ^b	OR (95% CI) ^c	Cutoffs ^d	OR (95% CI) ^c
Cumulus left MLO (N=672)	1st	0.26 (0.14, 0.48)	<10%	0.30 (0.15, 0.61)
	2nd	0.68 (0.43, 1.10)	10-24%	0.44 (0.26, 0.76)
	3rd (ref)	1.00	25-49% (ref)	1.00
	4th	1.21 (0.80, 1.86)	≥50%	1.36 (0.93, 2.01)
	P _t	<0.001	P _t	<0.001
ImageJ left MLO (N=637)	1st	0.53 (0.29, 0.96)	<10%	0.54 (0.21, 1.38)
	2nd	1.05 (0.66, 1.67)	10-24%	0.52 (0.30, 0.90)
	3rd (ref)	1.00	25-49% (ref)	1.00
	4th	1.91 (1.21, 3.03)	≥50%	1.97 (1.33, 2.94)
	P _t	<0.001	P _t	<0.001
ImageJ left-right MLO mean (N=672)	1st	0.56 (0.33, 0.95)	<10%	0.64 (0.26, 1.59)
	2nd	0.84 (0.53, 1.36)	10-24%	0.60 (0.37, 0.98)
	3rd (ref)	1.00	25-49% (ref)	1.00
	4th	1.67 (1.08, 2.57)	≥50%	2.14 (1.44, 3.16)
	P _t	<0.001	P _t	<0.001

Abbreviations: CI = confidence interval; MLO = medio-lateral oblique view; OR = odds ratio; PD = Percent density; P_t = P-value for linear trend in the ORs across quartiles.

^aAdjusted for age, BMI, number of children, menopausal status and family history of breast cancer; numbers of cases and controls are smaller due to missing values.

^bPercent density values corresponding to these quartile cutoffs in the controls: Cumulus left MLO 18, 36 and 51%; ImageJ-based left MLO 26, 39 and 49%; ImageJ-based MLO mean: 25, 38 and 48%; 3rd quartile taken as reference because of small number of cases in the bottom quartile.

^cAdjusted for age and BMI.

^dPredefined cutoff points (see Materials and Methods section), with the category with the largest number of subjects taken as the reference.

Both the fully automated ImageJ-based method and the semi-automated Cumulus approach attempted to measure the same underlying entity using a thresholding procedure. But although the two methods produced distributions with a similar range of values, their shapes were different, with high inter-method agreement for breast area but only moderate inter-method agreement for absolute and percent density. Relative to the magnitude of the risk estimates associated with the Cumulus readings from one single MLO image, the risk estimates for the ImageJ method were weaker when based on a single image but of remarkably similar magnitude when MLO readings from two breasts were combined. These findings are consistent with the ImageJ-based density measurements being more affected by random errors than the Cumulus density readings. This is further supported by the fact that the intra-method reliability for ImageJ-based density was somewhat lower

than that previously reported for Cumulus (McCormack *et al*, 2007). Interestingly, in this predominantly premenopausal population absolute dense area was as strongly associated with breast cancer risk as percent density for both Cumulus and ImageJ-based methods, adding to the current debate on which measure (absolute or relative) is the best biomarker of risk (Haars *et al*, 2005).

Other methods have been recently developed to measure density in analogue screen-film images (Pawluczyk *et al*, 2003; Heine *et al*, 2008; Aitken *et al*, 2010; Kallenberg *et al*, 2011; Shepherd *et al*, 2011; Heine *et al*, 2012). However, the ImageJ-based approach has several advantages. In contrast to some of the other methods (for example, Shepherd *et al*, 2011), it does not require the use of any special equipment (for example, phantoms, step wedges) during image acquisition and therefore it can be applied to historical collections of images. Thus, the ImageJ-based approach enables the

use of historical collections of analogue films to address clinico-epidemiological questions of current relevance. It is also relatively inexpensive to use as it was developed using an open-source rather than a proprietary software package (Li *et al*, 2012). The ImageJ-based approach requires prior 'training' in independent sets of images with characteristics similar to those from the study population of interest and for which Cumulus density measurements are available; this is necessary to 'train' the model by identifying the image parameter estimates that best predict density. Although this process requires technical expertise in statistical coding, once built, the model can in principle be applied to all images. ImageJ-based method failed to produce valid readings for about 10% of images, a proportion similar to that found in the study by Li *et al* (2012). As the automated thresholding is performed on the whole image, including the region beyond the breast outline, it requires the prior removal of all artefacts, regardless of their position on the film, before thresholding. A large proportion of these images were of poor quality in the original DICOM format (for example, tags superimposed on the breast area, unclear breast edge, non-optimal digitisation). Of particular concern, however, is the fact that the missing values did not occur at random as their frequency was higher for films with lower Cumulus density values. As women with low density values were more likely to be controls the exclusions might have led to an attenuation of the true magnitude of the association between ImageJ-based density and breast cancer risk. This problem was largely overcome when readings from both breasts were combined as almost all participants had at least one valid (left or right) ImageJ-based density value. Further studies are required to assess the extent to which this is a recurring problem of the ImageJ-based approach or a feature specific to our study setting.

The current version of the ImageJ-based method requires visual inspection of the quality of the input images before running the automated thresholding algorithm, as apparently valid readings may be produced for wrong-sized and primed images that do not convert well. Thus, although the ImageJ-based method produces automated density readings its savings in time and labour are somewhat hampered by the need, as with Cumulus, to visually inspect each input image – a rather time-consuming and error-prone process which, similarly to Cumulus, may limit its use in large-scale studies. The incorporation of algorithms that can automatically generate an indicator of the quality of the input images, similar to those generated by other automated methods, should be considered in future refinements of the ImageJ-based method. This would further maximise the benefits of this approach in high-throughput studies by restricting the need for visual inspection to the subset of input images flagged by such in-built algorithms as likely to be of poor quality.

The predictive power of ImageJ percent density to identify, alone or in combination with other risk factors, women at high-risk is low (as demonstrated by its low AUC value), albeit similar to that for Cumulus as observed here and elsewhere (Vachon *et al*, 2007; Li *et al*, 2012). Because of the individual-matched design of the study these AUC values are lower than the corresponding AUC values obtained from an equivalent unmatched study as they cannot take into account the predictive value of the matching variable – that is, age. Nevertheless, the AUC estimates observed here are rather similar to those reported by the previous study of Swedish postmenopausal women whose controls were frequency-matched to cases on age (5-year intervals): 0.60 (95% CI 0.57, 0.63) for Cumulus; 0.59 (95% CI 0.56, 0.62) for the ImageJ-based method (Li *et al*, 2012). The inclusion of both Cumulus and ImageJ into the same model did not improve discrimination – hardly surprisingly as the two methods are attempting to capture the same underlying entity. Thus, the use of either Cumulus or the ImageJ-based method in individual risk prediction is of limited benefit. However, mammographic density, jointly with other genetic and

non-genetic risk factors, can be used to stratify women in the population according to risk for tailored screening (by identifying those who may benefit from more intensive screening as well as those for whom screening may be more harmful than beneficial). This is of particular relevance given the current controversy surrounding the benefits and harms of breast screening and, in particular, the concerns about overdiagnosis (Independent UK Panel on Breast Cancer Screening, 2012). Mammographic density is not currently considered by most screening programmes despite the fact that it affects its sensitivity. BI-RADS (Breast Imaging—Reporting and Data System) is routinely used by US radiologists to visually classify mammograms according to their density and, recently, it became a legal requirement in many states of the USA to inform screen attendees of their breast density. However, BI-RADS only allows classification of women into four rather broad categories and its within- and between-reader reliabilities are far from ideal. Tailored screening strategies would benefit from valid standardised automated methods of measuring breast density.

Limitations of the present study include the fact that the assessment of the ImageJ-based approach was limited to the MLO view, and the lack of detailed information (beyond knowledge that they varied according to screening centre and year of mammography, and films being of different size and resolution) to allow examination of whether its performance differed according to X-ray equipment and image acquisition parameters. The sample size was also relatively small although future re-evaluations on larger sample sizes will be possible as more cases are accrued within the intervention arm of the Age Trial. A small proportion of participants were already postmenopausal at the time of mammography because of a rather early age at menopause (≤ 40 years); however, similar findings were observed when the analysis was restricted to premenopausal women. Data for some potential confounding variables (that is, parity, family history of breast cancer and BMI) were obtained retrospectively at the time of the questionnaire administration but it is unlikely that this might have affected substantially the findings because high agreement between parity and BMI data collected in successive questionnaires were found in a subset of Age Trial women (not included in this study), and between reported family history and cancer registration and death certification data from relatives. Furthermore, any misclassification is unlikely to have been differential as similar associations of these variables with density were observed among cases and controls. Given the on-going switch from analogue to digital mammography, it would be worthwhile to compare the performance of ImageJ (and Cumulus) with the performance of other automated techniques being specifically developed to quantify density on processed or unprocessed digital images. But the availability of historical collections of digitised images, with a wealth of clinical and long-term follow-up data, means that the need for a valid fully automated method to measure density on digitised analogue images will remain well after the switch to digital mammography is completed, and regardless of whether or not such method is found to perform as well on digital images. In short, our findings indicate that among premenopausal women the ImageJ-based approach, using the mean of two measurements, is a valid automated alternative to Cumulus for measuring mammographic density in analogue films and predicting subsequent breast cancer risk.

ACKNOWLEDGEMENTS

We thank the participating National Health Services Breast Screening Programme (NHSBSP) centres for their help with the retrieval of the mammographic films for the study participants. This study was funded by project grants from Breast Cancer

Campaign (2007MayPR23) and Cancer Research UK (G186/11). The work by IdSS was partly undertaken during the tenure of a Senior Visiting Scientist Award by the International Agency for Research on Cancer. The funding bodies had no role in the design of the study; in the collection, analysis and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Aitken Z, McCormack VA, Highnam RP, Martin L, Gunasekara A, Melnichouk O, Mawdsley G, Peressotti C, Yaffe M, Boyd NF, dos Santos Silva I (2010) Screen-film mammographic density and breast cancer risk: a comparison of the volumetric standard mammogram form and the interactive threshold measurement methods. *Cancer Epidemiol Biomarkers Prev* **19**: 418–428.
- Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, Jong RA, Hislop G, Chiarelli A, Minkin S, Yaffe MJ (2007) Mammographic density and the risk and detection of breast cancer. *New Engl J Med* **356**: 227–236.
- Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ (1994) The quantitative analysis of mammographic densities. *Phys Med Biol* **39**: 1629–1638.
- Haars G, van Noord PA, van Gils CH, Grobbee DE, Peeters PH (2005) Measurements of breast density: no ratio for a ratio. *Cancer Epidemiol Biomarkers Prev* **14**: 2634–2640.
- Heine JJ, Carston MJ, Scott CG, Brandt KR, Wu FF, Pankratz VS, Sellers TA, Vachon CM (2008) An automated approach for estimation of breast density. *Cancer Epidemiol Biomarkers Prev* **17**: 3090–3097.
- Heine JJ, Scott CG, Sellers TA, Brandt KR, Serie DJ, Wu FF, Morton MJ, Schueler BA, Couch FJ, Olson JE, Pankratz VS, Vachon CM (2012) A novel automated mammographic density measure and breast cancer risk. *J Natl Cancer Inst* **104**: 1028–1037.
- Independent UK Panel on Breast Cancer Screening (2012) The benefits and harms of breast cancer screening: an independent review. *Lancet* **380**: 1778–1786.
- Kallenberg MG, Lokate M, van Gils CH, Karssemeijer N (2011) Automatic breast density segmentation: an integration of different approaches. *Phys Med Biol* **56**: 2715–2729.
- Li J, Szekeley L, Eriksson L, Hedddson B, Sundbom A, Czene K, Hall P, Humphreys K (2012) High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer. *Breast Cancer Res* **14**: R114.
- McCormack VA, dos Santos Silva I (2006) Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* **15**: 1159–1169.
- McCormack VA, Highnam R, Perry N, dos Santos Silva I (2007) Comparison of a new and existing method of mammographic density measurement: intramethod reliability and associations with known risk factors. *Cancer Epidemiol Biomarkers Prev* **16**: 1148–1154.
- Moss S (1999) A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. *J Med Screening* **6**: 144–148.
- Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L (2006) Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet* **368**: 2053–2060.
- Pawluczyk O, Augustine BJ, Yaffe MJ, Rico D, Yang J, Mawdsley GE, Boyd NF (2003) A volumetric method for estimation of breast density on digitized screen-film mammograms. *Med Phys* **30**: 352–364.
- Shepherd JA, Kerlikowske K, Ma L, Duewer F, Fan B, Wang J, Malkov S, Vittinghoff E, Cummings SR (2011) Volume of mammographic density and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* **20**: 1473–1482.
- Vachon CM, van Gils CH, Sellers TA, Ghosh K, Pruthi S, Brandt KR, Pankratz VS (2007) Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Res* **9**: 217.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)