Assessing within-woman changes in mammographic density: a comparison of fully- versus semi-automated area-based approaches

Running title: Assessing within-woman changes in mammographic density

Marta Cecilia Busana[1], Bianca L De Stavola[1], Ulla Sovio[1]*, Jingmei Li[2], Sue Moss[3], Keith Humphreys[2], Isabel dos-Santos-Silva[1]

[1] Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

[2] Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Box 281, 171 77 Stockholm, Sweden

[3] Centre for Cancer Prevention, Queen Mary University of London, Wolfson Institute of Preventive Medicine, Charterhouse Square, London EC1M 6BQ, UK

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

Corresponding author: Isabel dos-Santos-Silva, London School of Hygiene and Tropical Medicine, Department of Non-Communicable Disease Epidemiology, Keppel Street, London, United Kingdom.

Phone: +44 (0)20 7927 2113; Fax: +44 (0)20 7580 6897; Email: isabel.silva@lshtm.ac.uk

List of abbreviations

BMI: body mass index; BC: breast cancer; CC: cranio-caudal; CI: confidence interval; IQR: inter-quartile range; MD: mammographic density; MLO: medio-lateral oblique; PD: percent density; SD: standard deviation

Abstract

Background: Mammographic density (MD) varies throughout a woman's life. We compared the performance of a fully-automated (ImageJ-based) method to the observer-dependent Cumulus approach in the assessment of within-woman changes in MD over time.

Methods: MD was assessed in annual pre-diagnostic films (from age 40 to early 50s) from 313 breast cancer cases and 452 matched controls using Cumulus (left medio-lateral-oblique (MLO) readings) and the ImageJ-based method (mean left-right MLO readings). Linear mixed models were used to compare within-woman changes in MD among controls. Associations between individual-specific MD trajectories and breast cancer were examined using conditional logistic regression.

Results: The age-related trajectories predicted by Cumulus and the ImageJ-based method were similar for all MD measures, except that the ImageJ-based method yielded slightly higher (by 2.54%, 95% CI: 2.07%, 3.00%) estimates for percent MD. For both methods, the yearly rate of change in percent MD was twice faster after menopause than before, and higher BMI was associated with lower mean percent MD, but not associated with rate of change. Both methods yielded similar associations of individual-specific MD trajectories with breast cancer risk.

Conclusions: The ImageJ-based method is a valid fully-automated alternative to Cumulus for measuring within-woman changes in MD in digitised films.

The Age Trial is registered as an International Standard Randomised Controlled Trial, number ISRCTN24647151.

Keywords: mammographic density, breast density, breast cancer, pre-menopausal

Introduction

High mammographic density (MD), which represents a high amount of radio-dense fibroglandular tissue in the breast, is not only associated with an increased breast cancer (BC) risk, independently of other known risk factors [1], but it also affects the sensitivity of mammographic screening [2]. MD varies throughout a woman's life being influenced by factors such as age, parity, menopausal status and hormonal interventions (e.g. hormone therapy (HT) and tamoxifen). Thus, it is conceivable that the rate of change in MD over a woman's lifetime, or at critical periods, may be more relevant to risk than MD measured at any single point. Results from previous studies [3-9] have been conflicting, however, some studies [5,7] showed that within-woman changes in MD over time do not convey any additional risk information beyond that provided by a single measurement, while others [3,4,6,8,9] were consistent with changes being independently associated with risk. But, with the exception of a small study [4], they all relied on qualitative (e.g. Wolfe patterns [3] or BI-RADS [8]) or semi-automated approaches, i.e. Cumulus [5-7,9] to capture within-woman longitudinal changes in MD, all of which are reader-dependent. Reader variability might have introduced noise in the measurement of within-woman changes in MD, thus, leading to an attenuation of their association with risk. Furthermore, small within-woman changes cannot be captured when using the broad Wolfe or BI-RADS categories.

The semi-automated interactive thresholding technique, on which the Cumulus software is based, is currently considered the "gold standard" approach to measure MD [1]. Cumulus measurements of between-women MD differences have been shown to have a high between- and within-reader reliability [10], and to be consistently associated with subsequent BC risk [1]. However, the validity of this method to capture within-woman changes in MD, which are of a smaller magnitude than between-women differences, is unknown. The ImageJ-based method is a fully-automated method which attempts to mimic Cumulus, and whose performance to detect between-women differences in MD has been recently shown to be comparable to that of Cumulus [11,12]. The aim of this study was to compare the performance of this fully-automated method to that of the reader-dependent Cumulus approach in the assessment of within-woman changes in MD. We compared the performance of the ImageJ-based method to that of the "gold-standard" Cumulus in terms of the degree to which they were able to capture: (i) tracking of MD with age; and (ii) within-woman changes in MD among controls and their determinants; and (iii) association of within-woman changes in PD with BC risk.

Materials and Methods

Study design

A nested case-control study of BC in relation to pre-diagnostic MD was undertaken within the Age Trial, a trial of the efficacy of annual mammographic screening at age 40-48 [13,14]. About 54,000 women from the general population aged 40-41 years and resident in the catchment areas of 23 National Health Service (NHS) breast screening centres in Britain were randomized between 1991 and 1996 to the intervention arm of the trial and invited to attend annual screening with analogue mammography. The first screening round included both cranio-caudal (CC) and medio-lateral oblique (MLO) views of each breast; subsequent rounds included only the MLO view. From age 50 years onwards women joined the national breast screening programme and were invited for mammography every 3 years. BC cases among the study population have been ascertained through linkage to the NHS health registers for cancer incidence and mortality since the start of the trial. Women in the intervention arm who were diagnosed with BC between 1993 and 2005, and at least one year after their first negative screen, were eligible for the present study (n=442). For each case, up to six eligible controls were initially randomly selected among women in the intervention arm who had not been diagnosed with BC at the time of the case's diagnosis. Controls were matched to the cases on screening centre, date of birth (± 3 months), date of the first pre-diagnostic screen (± 3 months), and subsequent screens (± 4 months). For 89 (28.4%) cases, no controls were available and therefore the matching was performed only on the first three criteria.

Eligible cases and controls were contacted (for cases after obtaining consent from their general practitioners) and asked to provide written consent for their mammograms to be accessed, and to complete a questionnaire on anthropometric and reproductive variables, and history of breast cancer in first degree relatives. Analogue films were retrieved from relevant NHS screening centres and digitized using a high-quality Array 2959 laser digitiser (Array Corporation Europe, Netherlands). In all, 76% of eligible cases and 80% of eligible controls completed the questionnaire and gave consent for accessing their mammograms; films could be retrieved for 93% of consented cases and 89% of consented controls. The first two matched control women for whom both questionnaire data and films were available were included in the present study.

The study was approved by the UK NHS South-East Multi-Centre Research Ethics Committee (05/MREC01/77).

MD assessment

MD assessment was performed on MLO images (the only view performed on every screening round). For cases, only MLO images taken at least one year prior to BC diagnosis were included. For controls, all MLO images up to the time of diagnosis of the corresponding matched case were read as they were all eligible for control-only analyses; however, only those taken one year prior to the date of diagnosis of the corresponding case were included in case-control analyses (see below).

MD readings using the Cumulus software (version 3, University of Toronto, Toronto, Ontario, Canada) are labour-intensive and were therefore performed only on the left MLO image. This image was preferred because MD values from left and right breasts are highly correlated [10], and the pectoral muscle was less likely to be superimposed on the top of the breast on left images. Images were read in batches of about 250-300 digitized images. All images from a given case-control set were included in the same batch, with images from the same woman ordered randomly. Each batch of images was read by one of three observers, blind to the woman's characteristics and her case-control status. Each observer used Cumulus to delimit the breast area and select a gray-scale threshold to differentiate dense and non-dense tissues, with the software then automatically estimating breast area (in cm^2), dense area (in cm^2), and percent density (PD). Non-dense area was derived by subtracting the dense area from the breast area. Within- and between-reader reliability was estimated by including a random 10% sample of all eligible images as duplicates in each batch; the independent readings provided by these duplicates revealed high within- and between-batch reliability in PD (>0.90 for both for each one of the three observers [12]) Within-observer reliability was >93% for each one of the three readers and between-observer reliability was >0.82 for all pairs of readers.

The ImageJ-based method has been described elsewhere [11]. Briefly, valid digitized images were analysed by the ImageJ software (version 1.46, 26 June 2012) which generated MD readings via an algorithm developed on a training set of independent images with known Cumulus values (further details in Li *et al*. [11] and Sovio *et al.* [12]). Essentially, this approach works by calculating values of various statistical/textural features of the image after applying different thresholding methods to distinguish dense areas from non-dense areas of the breast. The ImageJ-based software includes a machine-learning approach, combining principal component analyses and penalised regression, that develops separate prediction models for estimating PD and breast size; absolute dense area is then estimated from the product of PD and breast size, with non-dense area derived as its complement. ImageJ-based MD measurements were performed in both left and right MLO images of each woman and the left-right mean used in

the analyses. This approach was used because a previous study, based on a subset of the data analysed herein, showed that the mean of ImageJ-based MD readings from the left and right MLO images taken at entry (i.e. closer to age 41 years) performed as well as Cumulus on a single MLO image in terms of the magnitude of the associations of the MD estimates with known BC risk factors, and with subsequent risk of BC [12]. However, we also examined tracking and within-woman changes for the ImageJ-based MD readings taken on the left MLO view only to allow a more direct comparison with similar analyses for Cumulus readings taken on the same view. The ImageJ-based method failed to produce valid readings for about 10% of images (mainly due to poor quality of the digitised images, i.e. tags superimposed on the breast area, unclear breast edge, non-optimal digitisation). Whenever the ImageJ-based estimates were available for only one (left or right) MLO image (for 14.4% cases and 14.1% controls), the value for that image was used instead.

Statistical Methods

A natural log-transformation was used to normalise the distributions of dense and non-dense area values; no transformation was required for PD or breast area values. BMI at each mammography was estimated for each woman by linearly interpolating self-reported BMI at ages 40 and 50 years. For 10.9% cases and 7.3% controls, only BMI at age 40 or 50 years was available and this value was taken to represent their BMI at the time of each screening. Menopausal status was derived from information on age at menopause reported in the questionnaire and was retrospectively determined for each screening appointment. If information on menopausal status was missing (6.4% for cases; 4.2% for controls) the median age at menopause among cases and controls with non-missing information was used. Whenever the values the other variables shown in Table 1 were not know they were treated as missing in the analyses.

MD tracking with age among controls. The readings yielded by each measurement method (i.e. Cumulus, Image-based) and MD measure (i.e. PD, dense area, non-dense area and breast area) combination at a given age were ranked separately, and Spearman rank correlation coefficients (r) between ranks calculated (for Cumulus measures, both overall and stratified by reader).

Within-woman changes in MD among controls. Linear mixed models were fitted by restricted maximum likelihood to estimate individual-specific trajectories [15], separately for each measurement method and MD measure combination. These models, which take into account the correlation among the repeated observations for each

153 woman via one or more random coefficients that capture salient features of each woman trajectory, were specified
154 in terms of linear and quadratic effects of age (centred at age 45 years), number of children (0, 1, 2, ≥3), having ever
155 breastfed (if parous), age-specific BMI (categorised as: <22, 22-24.99 and ≥25kg/m^2), family history of breast cancer,
156 and menopausal status (pre- or post-), and, for Cumulus, also an indicator of reader. Interactions between each
157 variable and current age (on the linear scale) were also included. Different specifications of the random effects
158 component of the models were compared using likelihood ratio tests, with a random intercept and random slope for
159 linear age selected for each combination of measurement method and MD measure. The fixed effects parts of the
160 models were then simplified through backward selection using Wald tests, but always retaining age, BMI and, for
161 Cumulus, reader. Parity was found to be borderline significant for only some of the MD dimensions and was
162 therefore dropped from all models to aid comparison. Model-based estimates of the average trajectories of each
163 MD measure obtained from Cumulus and ImageJ-based values were plotted for different combinations of the
164 predictors to allow a graphical comparison of the performance of the two methods for each MD dimension.

165 To test whether models for data obtained from Cumulus and the ImageJ-based method yielded different predicted
166 trajectories we also fitted a more general model for each MD measure based on readings from both methods and
167 including a "method" binary indicator, as well as interaction terms between this indicator and each of the other
168 selected explanatory variables, while allowing method-specific variances of the residual errors. The model was then
169 simplified by removing terms that were not significant via a backward stepwise procedure using the Wald test as
170 before.

171 Within-woman changes in PD and BC risk. An extension of the models described above was used to evaluate the
172 association between predicted individual trajectories in PD, as yielded by Cumulus or the ImageJ-based method, and
173 BC risk. The models were specified in terms of the variables used above but centred at age 42 (instead of 45) years in
174 order to compare the model predictions from age at entry into the study. First, as before, general mixed effects
175 models were fitted on the cases and fitted again on controls, but using only images taken at least one year prior to
176 diagnosis for each case and the corresponding period for her matched controls. For both groups the models were
177 simplified via backward stepwise selection using the variables mentioned above and always including age, BMI, and,
178 for Cumulus, reader (as before).Woman-specific MD values at age 42 years (the random intercept) and her rate of
179 change from that age (the random slope for age) were predicted (separately by each method) and then included in

conditional logistic regression models to estimate the odds ratios (ORs) of BC. Four selected typical individual PD trajectories were finally plotted relatively to the trajectory of a woman with mean trajectory (i.e. with mean random intercept and mean random slope).

All analyses were performed in Stata version 13.1.

Results

A total of 313 cases and 452 controls were included in the analysis, corresponding to 308 complete sets (i.e. consisting of one case and at least one control, with each having at least one MLO image). Cases and controls had, by design, the same age at first mammographic screen (Table 1). They also had similar ages at menarche, first birth and menopause, and a similar BMI at ages 40 (when available) and 50 years. Cases were more likely to have a positive family history of BC, but less likely to be parous, to have ever breastfed, and to be post-menopausal at the time of their first and last screens (the latter corresponding to the screen taken at least one year prior to BC diagnosis in cases and equivalent time in corresponding controls) (Table 1). Both Cumulus and the ImageJ-based method showed that, at ages 42 years, cases had, on average, higher dense area and PD than controls, but lower non-dense and breast areas (Supplementary Table 1). Cumulus and ImageJ-based PD declined from age 42 to 48 years at a similar rate among controls, i.e. by 1.17% (standard deviation (SD) 1.91%) and 1.07% (SD 1.79%) per year, respectively, reflecting marked increases in non-dense area, as well as smaller decreases in dense area, with increasing age (Supplementary Table 1).

MD tracking among controls. A high degree of tracking in the four MD measures was observed among controls according to both measurement methods (Figure 1 and Supplementary Table 2), with within-woman rank correlations being similar when both methods were based on the left MLO view but slightly higher for the ImageJ-based method when the latter was based on the left-right mean MLO readings (for Cumulus, Figure 1 and Supplementary Table 2 show data for all observers combined as observer-specific correlation coefficients yielded similar values (Supplementary Table 3)). The degree of tracking decreased with increasing time between screens for both methods, but remained high for films taken six years apart, i.e. from age 42 to 48 years (e.g. for PD *r*: 0.66 for Cumulus left MLO readings; 0.66 and 0.77 for ImageJ-based left MLO and left-right MLO mean readings, respectively (Figure 1 and Supplementary Table 2)). Higher degrees of tracking were observed for non-dense and breast areas for both methods (*r* for the latter: ~0.90 for screenings taken 6 years apart; Figure 1 and Supplementary Table 2).

Within-woman changes in MD among controls. The mixed effect models led to broadly similar average trajectories for Cumulus (left MLO) and the ImageJ-based method (left-right MLO mean), as shown in Figure 2 for different combinations of the main predictors (details of the fitted models are given in Supplementary Table 4). Specifically, PD and dense area measured by either method decreased with increasing age, while non-dense area and breast area increased. The linear component of the yearly rate of change in PD was more than twice as fast after the menopausal transition than prior to it for Cumulus (-1.10%; 95% CI: -1.56%, -0.64% vs. -0.50%; 95% CI: -0.81%,-0.18%, respectively). Similarly the yearly rate coefficient in the ImageJ-based model was nearly twice as fast after the menopause (-1.16%; 95% CI: -1.71%, -0.61% vs. 0.67%; 95% CI: 1.15%, -0.18%, respectively). The impact of reaching the menopause (set to be at age 50 years in Figure 2) is -2.10 and highly significant for Cumulus but close to zero (and not significant) for the Image-J based method. BMI was negatively associated with mean levels of PD and dense area, but positively associated with mean levels of non-dense and breast areas, regardless of the measurement method used. However, this variable had no effect on the rate of change of the density measures (Supplementary Table 4). Parity had no effect on mean level and rate of change for any of the MD measures.

For a more direct comparison of the PD trajectories predicted by the two methods we also fitted a common mixed effects model for both sets of measurements, where a binary indicator of method was included as an explanatory variable for both the intercept and the rate of change (Table 2). The results show that the average PD trajectories derived from the ImageJ-based method (left-right MLO mean) were systematically higher than those based on Cumulus (left MLO) (by 2.54, 95% CI: 2.07, 3.00). In contrast, there was no difference in the mean estimates of non-dense area and total breast area at age 45 years yielded by the two methods but the rate of increase with age was less steep for the ImageJ-based method (Table 2). Interestingly, the estimated residual SDs are significantly greater for Cumulus than for the ImageJ-based method for PD (9.10 versus 7.65, $p=0.0002$) and dense area (13.77 versus 10.09, $p<0.0001$) and smaller for non-dense area (0.17 versus 0.23, $p<0.0001$, on a log scale) and total breast area (0.08 versus 1.11, $p<0.0001$, on a log scale) (Table 2).

Analyses based only on readings from the left breast for both Cumulus and the ImageJ-based method show similar patterns. For instance, the PD trajectories derived from the ImageJ-based left MLO readings were also systematically higher than those produced by Cumulus readings on the same view: by 3.26, 95% CI: 2.76, 3.76 (Supplementary Table 5). The estimated residual SDs for the ImageJ-based right-left MLO mean readings were lower than, or similar

to, those for left MLO Cumulus readings (Table 2). However this advantage was lost when only left MLO readings were used (Supplementary Table 5), pointing to the need for averaging the left and right measures when using the automatic ImageJ-based readings in order to reduce the impact of measurement error.

Within-woman trajectories in PD and BC risk. The random coefficients (intercepts and slopes) predicted from the PD models fitted separately on cases and controls (adjusting for age, and BMI) were treated as exposures in conditional logistic regression models for being a case (further adjusting for parity and family history of BC). The parameter estimates were then used to calculate ORs of breast cancer for four women with typical PD trajectories, randomly drawn from the controls with BMI of 22-24.99 kg/m^2, who remained pre-menopausal at the end of the follow-up period, and whose Cumulus readings were performed by the same observer (i.e. observer 1), relatively to a woman with mean trajectory (i.e. with mean random intercept and mean random slope). Their PD predicted trajectories were similar for Cumulus and the ImageJ-based (left-right MLO mean) methods although, consistently with Figure 2, the ImageJ-based PD trajectories tended to be higher than those produced by Cumulus (Figure 3). Their associated ORs for subsequent BC were also similar. Women with a high PD at baseline, which remained high over time, had the highest odds of developing BC according to both methods relative to a woman with mean random intercept and mean slope (OR: 8.10 (95% CI 3.96, 16.6) for Cumulus (left MLO) and 3.42 (2.00, 5.48) for the ImageJ-based method (left-right MLO mean; Figure 3). In contrast, women with the lowest PD at baseline, despite a slight increase in their PD over time, had the lowest odds of developing BC according to both measurement methods (OR: 0.07 (95% CI 0.03, 0.16) for Cumulus (left MLO) and 0.23 (0.12, 0.43) for the ImageJ-based method (left-right MLO mean; Figure 3).

Discussion

The ability of the ImageJ-based method to measure between-women differences in MD at a single point in time has previously been shown to be similar to that of the well-established computer-assisted Cumulus method among post-menopausal Swedish women [11] and pre-menopausal British women [12]. The latter study [12] was conducted on a subset of images included in the present study, i.e. those taken at baseline when the women were close to age 41 years. Herein we extended this comparison to the assessment of within-woman changes in MD over a ~10 year period (from age 41 onwards). The findings showed that the ability of the ImageJ-based method (based on the

average of left and right MLO values) to capture within-woman changes in MD was broadly similar to that of Cumulus (based on a single MLO reading).

Consistently with a previous study [16], both measurement methods showed that MD measures track over time. The degree of tracking was similar for both Cumulus and the ImageJ-based method when both approaches were based on a single (left) image. The degree of tracking for Cumulus was slightly weaker than previously reported reflecting the fact that its MD measurements were based on a single image whereas previous Cumulus work [16] was based on the average of left-right MD measurements. The left–right average MD would have strengthened the degree of tracking for Cumulus by minimising reader measurement error in the assessment of MD. A similar pattern was observed in the present study for the ImageJ-based method, with the degree of tracking being stronger for the left-right average MD than for the left MD only. Nevertheless, the odds of remaining in the top fifth of the PD distribution at later screens for women who were in that category at age 42 years was high according to both methods. Tracking from age 40 onwards implies that between-woman differences in MD are established earlier in life, a finding consistent with recent evidence from women aged 15-30 years [17], and that any within-woman variations after that age are of a much smaller magnitude relative to between-woman differences. These findings explain why MD remains a predictor of breast cancer risk several years after MD assessment and imply that a single MD assessment at young ages (at least as early as age ~40 years) would allow identification of women with high density, and who may benefit the most from risk-lowering interventions or from more intensive screening (e.g. modalities other than mammography, shorter screening intervals).

The age-related MD trajectories predicted by the two methods were broadly similar. On average, PD and dense area declined with age. In contrast, the rate of increase in non-dense and breast areas with age was less pronounced for the ImageJ-based method. The findings from this study are broadly consistent with those reported by previous studies examining longitudinal trends in MD measures and their correlates. *Kelemen et al* [18] and *Boyd et al* [19] also found that PD decreased with age, with a greater decline observed during the menopausal transition. A decline in dense area with increasing age and during the menopausal transition, paralleled with a simultaneous increase in non-dense area, is likely to reflect lobular involution of the breast gland. The amount of fibroglandular tissue decreases with age and the menopausal transition as a result of the decrease in circulating levels of ovarian-

produced oestrogens. Increases in the non-dense area may also result from weight gain and consequent increase in the amount of adipose tissue in the breast.

Reassuringly, both MD measurement methods revealed increased odds of having BC for women with PD trajectories that started at high value and remained high throughout the follow-up, and lower odds for those whose PD trajectories start at a low PD value. A detailed examination of whether between-screen changes in MD convey additional risk information, beyond that provided by a single MD measurement, will be the focus of future analyses within the Age Trial once a larger number of breast cancer cases has been accrued.

The study has several strengths. Most previous studies on changes in MD have focused on screening attendees and therefore mainly on women aged 50 and above. Our study population was unique in that it comprised younger women from the general population who were invited to attend annual routine mammographic screening from age 40/41years. Thus, the young age at recruitment, the availability of multiple screening rounds at short (1-year) intervals, and the relative long follow-up allowed us to map in detail within-woman changes in MD measures over time. The study is also one of the few to have examined not only within-woman changes in PD but also changes in its two components: absolute dense and non-dense areas. For Cumulus, films for a given case-control set were read in a blind way and in a random order (a key methodological feature for the assessment of tracking). Screening mammography equipment might have changed during the follow-up period with more recent systems being based on higher contrast resolutions which make the fibroglandular tissue appear less dense on the films, thus leading to an overestimation of the decline in MD. However, such changes would have had little impact on our ImageJ – Cumulus comparisons as any changes would have affected the two methods similarly as their readings were derived from the same set of images.

The study also had some weaknesses. Only MLO images were available in all screening rounds and, for logistic reasons (time and costs), the Cumulus readings were performed only on the left MLO view. Information on correlates of MD measures at baseline, and their rate of change, was collected retrospectively, hence any misclassification, if present, is likely to have affected similarly the two methods. In particular, screen-specific BMI data were estimated by linearly interpolating the self-reported BMI values at ages 40 and 50 years. Thus, if BMI increased with increasing age to a greater (or lower) degree than our BMI estimates, or if the between-screen

changes in BMI were not linear, residual confounding could have affected the MD trajectories. Information on contraceptive use and hormone therapy was not available, but this would have affected both methods similarly.

Our findings indicate that the ImageJ-based approach, using the mean of two measurements, is a valid fully-automated alternative to Cumulus for measuring within-woman changes in MD. MD is not only a strong BC risk factor but it also affects the sensitivity of screening mammography. Despite its relevance, MD assessment currently has little impact on risk-lowering decisions or screening strategies. Our findings indicate that the ImageJ-based approach, using the mean of two measurements, is a valid fully-automated alternative to Cumulus for measuring within-woman changes in MD. Other fully-automated methods have been developed to measure MD in digitized images (e.g. [20-25]), but the ImageJ-based approach benefits from the fact that it does not require the use of any special equipment (e.g. phantoms) at the time of mammography and hence it can be applied to historical collections of images. In addition, it is relatively inexpensive as it was developed as an open source.

Acknowledgments

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 and C405/A14565). 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. The development of ImageJ was supported by the 2nd Joint Council Office (JCO) Career Development Grant (13302EG065). Jingmei Li is a UNESCO-L'Oréal International Fellow.

Conflict of interest

The authors declare no conflict of interest.

References

1. McCormack VA, dos Santos Silva I (2006) Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for

Cancer Research, cosponsored by the American Society of Preventive Oncology 15 (6):1159-1169. doi:10.1158/1055-9965.EPI-06-0034
2. 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. The New England journal of medicine 356 (3):227-236. doi:10.1056/NEJMoa062790
3. Salminen TM, Saarenmaa IE, Heikkila MM, Hakama M (1998) Risk of breast cancer and changes in mammographic parenchymal patterns over time. Acta oncologica (Stockholm, Sweden) 37 (6):547-551
4. van Gils CH, Hendriks JH, Holland R, Karssemeijer N, Otten JD, Straatman H, Verbeek AL (1999) Changes in mammographic breast density and concomitant changes in breast cancer risk. Eur J Cancer Prev 8 (6):509-515
5. Maskarinec G, Pagano I, Lurie G, Kolonel LN (2006) A longitudinal investigation of mammographic density: the multiethnic cohort. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 15 (4):732-739. doi:10.1158/1055-9965.epi-05-0798
6. Work ME, Reimers LL, Quante AS, Crew KD, Whiffen A, Terry MB (2014) Changes in mammographic density over time in breast cancer cases and women at high risk for breast cancer. International journal of cancer Journal international du cancer. doi:10.1002/ijc.28825
7. Vachon CM, Pankratz VS, Scott CG, Maloney SD, Ghosh K, Brandt KR, Milanese T, Carston MJ, Sellers TA (2007) Longitudinal trends in mammographic percent density and breast cancer risk. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 16 (5):921-928. doi:10.1158/1055-9965.EPI-06-1047
8. Kerlikowske K, Ichikawa L, Miglioretti DL, Buist DS, Vacek PM, Smith-Bindman R, Yankaskas B, Carney PA, Ballard-Barbash R (2007) Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk. Journal of the National Cancer Institute 99 (5):386-395. doi:10.1093/jnci/djk066
9. Lokate M, Stellato RK, Veldhuis WB, Peeters PH, van Gils CH (2013) Age-related changes in mammographic density and breast cancer risk. American journal of epidemiology 178 (1):101-109. doi:10.1093/aje/kws446
10. 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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 16 (6):1148-1154. doi:10.1158/1055-9965.epi-07-0085
11. Li J, Szekely L, Eriksson L, Heddson 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 research : BCR 14 (4):R114. doi:10.1186/bcr3238
12. Sovio U, Li J, Aitken Z, Humphreys K, Czene K, Moss S, Hall P, McCormack V, Dos-Santos-Silva I (2014) Comparison of fully and semi-automated area-based methods for measuring mammographic density and predicting breast cancer risk. British journal of cancer 110 (7):1908-1916. doi:10.1038/bjc.2014.82
13. Moss S (1999) A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. Trial Steering Group. Journal of medical screening 6 (3):144-148
14. 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 (9552):2053-2060. doi:10.1016/S0140-6736(06)69834-6
15. Rabe-Hesketh S, Skrondal A (eds) (2012) Multilevel and Longitudinal Modeling Using Stata, vol Volume I: Continuous Responses. Third edition edn. Stata Press,
16. **McCormack VA**, Perry NM, Vinnicombe SJ, **dos Santos Silva I** (2009) Changes and tracking of mammographic density in relation to Pike's model of breast tissue ageing: a UK longitudinal study. Int J Cancer ([Epub ahead of print])
17. Boyd N, Martin L, Chavez S, Gunasekara A, Salleh A, Melnichouk O, Yaffe M, Friedenreich C, Minkin S, Bronskill M (2009) Breast-tissue composition and other risk factors for breast cancer in young women: a cross-sectional study. Lancet Oncol 10 (6):569-580. doi:S1470-2045(09)70078-6 [pii]

10.1016/S1470-2045(09)70078-6
18. Kelemen LE, Pankratz VS, Sellers TA, Brandt KR, Wang A, Janney C, Fredericksen ZS, Cerhan JR, Vachon CM (2008) Age-specific trends in mammographic density: the Minnesota Breast Cancer Family Study. American journal of epidemiology 167 (9):1027-1036. doi:10.1093/aje/kwn063

19. Boyd N, Martin L, Stone J, Little L, Minkin S, Yaffe M (2002) A longitudinal study of the effects of menopause on mammographic features. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 11 (10 Pt 1):1048-1053
20. 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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 20 (7):1473-1482. doi:10.1158/1055-9965.EPI-10-1150
21. 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. Medical physics 30 (3):352-364
22. 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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 17 (11):3090-3097. doi:10.1158/1055-9965.EPI-08-0170
23. 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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 19 (2):418-428. doi:10.1158/1055-9965.epi-09-1059
24. Kallenberg MG, Lokate M, van Gils CH, Karssemeijer N (2011) Automatic breast density segmentation: an integration of different approaches. Physics in medicine and biology 56 (9):2715-2729. doi:10.1088/0031-9155/56/9/005
25. 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. Journal of the National Cancer Institute 104 (13):1028-1037. doi:10.1093/jnci/djs254