Comparison of a New and Existing Method of Mammographic Density Measurement: Intramethod Reliability and Associations with Known Risk Factors

Valerie A. McCormack,[1] Ralph Highnam,[2] Nicholas Perry,[3] and Isabel dos Santos Silva[1]

[1]Department of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine; [2]Highnam Associates Limited, Wellington, New Zealand; and [3]Breast Assessment Centre, St. Bartholomew's Hospital, London, United Kingdom

Abstract

Background: **Mammographic density is one of the strongest risk factors for breast cancer. It is commonly measured by an interactive threshold method that does not fully use information contained in a mammogram. An alternative fully automated standard mammogram form (SMF) method measures density using a volumetric approach.**

Methods: **We examined between-breast and between-view agreement, reliability, and associations of breast cancer risk factors with the threshold and SMF measures of breast density on the same set of 1,000 digitized films from 250 women who attended routine breast cancer screening by two-view mammography in 2004 at a London population-based screening center. Data were analyzed using random-effects models on transformed percent density.**

Results: **Median (interquartile range) percent densities were 12.8% (5.0-22.3) and 21.8% (18.4-26.6) in the threshold and SMF methods, respectively. There was no evidence of systematic differences between left-right breasts or between views in either method. Reliability of a single measurement was lower in the SMF than in the threshold method (0.77 versus 0.92 for craniocaudal and 0.68 versus 0.89 for mediolateral oblique views). Increasing body mass index and parity were associated with reduced density in both methods; however, an increase in density with hormone replacement therapy use was found only with the threshold method.**

Conclusion: **Established properties of mammographic density were observed for SMF percent density; however, this method had poorer left-right reliability than the threshold method and has yet to be shown to be a predictor of breast cancer risk.** (Cancer Epidemiol Biomarkers Prev 2007; 16(6):1148–54)

Introduction

Mammographic density, the proportion of radiodense fibroglandular tissue in the breast, has repeatedly been shown to be one of the strongest predictors of breast cancer risk. A recent meta-analysis found that women with high mammographic density (>75%) had a 4.6-fold increased breast cancer risk [95% confidence interval (95% CI), 3.6-5.9] compared with women with little (<5%) dense tissue (1). This biomarker has great potential to be used as an intermediate end point in studies of breast cancer etiology, to gain understanding of disease pathways, and in clinical settings as a prognostic indicator or as a tool to monitor risk reduction strategies (2). Despite its major potential, there are a variety of definitions and ways of estimating mammographic density and there is currently no worldwide standardized tool to measure it.

Existing established methods of density classification and measurement using screen-film mammography are area based; that is, they are derived from projected areas of dense tissue as they appear on a mammogram. These include the Wolfe grade (3), Breast Imaging Reporting and Data System classification (4), Tabar grade (5), six category classification, and a computer-assisted threshold method (6, 7). The latter produces a quantitative continuous measure of density and is thus considered here as the current standard; however, it is limited by its oversimplicity in representing a volume by a projected area, by the subjectivity in defining a threshold for dense breast tissue, and by its time-intensive nature.

Development of methods of breast density estimation using volume-based approaches is under way. Proposed methods include incorporation of a calibration step-wedge (8) or using dual-energy approaches similar to those used in bone densitometry (9). Both of these approaches require alterations to be made at the time of mammography and hence can only be used prospectively. Highnam and Brady, (10, 11) have developed a volume-based method of breast density measurement, known as the standard mammogram form (SMF), which can be applied retrospectively to films already taken. SMF is a fully automated method of breast density measurement based on a complex model of the breast volume and of the imaging process. It estimates the volume of dense and nondense tissue while standardizing the image for variations arising from differential X-ray imaging conditions. This method ideally incorporates imaging data (e.g., voltage and exposure variables) but has also been adapted for application to films for which such calibration data are unavailable (10). Similar methods are being applied to digital mammography (12) where calibration data are easily available and images are much more reliably segmented.

Studies examining the relationship between SMF percent density and subsequent breast cancer risk are on-going and will provide the ultimate and essential test of its use as a predictor of breast cancer risk. In the meantime, internal validity can be assessed and comparison with other methods of measurement can be made (13). Here, we evaluate SMF based on established properties of mammographic density, including agreement by breast laterality and projection (14), and increased density with younger age, lower parity, premenopausal status, and use of

Received 1/30/07; revised 3/27/07; accepted 4/6/07.

Grant support: Cancer Research UK 5-year programme grant and a Cancer Research UK Fellowship (V. McCormack).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Valerie McCormack, Noncommunicable Disease Epidemiology Unit, Department of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E HT, United Kingdom. Phone: 44-20-7927-2301; Fax: 44-20-7436-4230. E-mail: valerie.mccormack@lshtm.ac.uk

doi:10.1158/1055-9965.EPI-07-0085

hormonal therapy (15). These properties are assessed and compared for both the SMF and threshold measures of density made on the same set of films.

Materials and Methods

Study Population. A subset of women who had participated in a study of determinants of breast density were included in this analysis. In the larger study (not yet published), women ages 53 to 65 years who underwent routine screening by mammography in 2004 as part of the United Kingdom National Health Service Breast Screening Programme at the Central and East London Breast Screening Service were randomly selected and invited to participate. For this analysis, we included a random sample of 250 of these women for whom both craniocaudal and mediolateral oblique views (as per normal screening practice) were taken on both breasts at the same screening visit. For all 1,000 films, voltage (kVp) and exposure (mAs) data were extracted from screening records. All screen films were retrieved and digitized in 2006 using an Array 2905 laser digitizer with optical density range 0 to 4.0, 12-bit depth, and pixel size of 75 μm.

Each woman self-completed a questionnaire on known breast cancer risk factors, including self-reported height, weight, menstrual status, age at menopause, total number of pregnancies, date of each pregnancy, and use of hormone replacement therapy. Body mass index (BMI) was calculated as weight/height^2 (kg/m^2). Signed informed consent was provided by each woman and ethical approval was obtained from the East London and the City Local Research Ethics Committee.

Threshold Method. The interactive-threshold method was implemented with done using the Cumulus software (6) by a single trained observer (V.M.). Films were randomly ordered and first masked to ensure blinding to all patient and image identifiers. The breast area was defined with the aid of a thresholding technique to identify the skin edge. The pectoral muscle and non–breast tissue were masked out in mediolateral oblique views. A grayscale threshold was then interactively selected to dichotomize the total breast area into dense and nondense areas. Percent density was calculated as the total dense area divided by the total breast area (6). A random 10% sample of films was read twice (blindly) and intrarater reliability was 0.90 for a single measurement.

SMF Method. SMF percent densities were obtained using the programme GenerateSMF version 2.2β from Siemens Molecular Imaging for optical densities 0 to 4.0. Full technical details of the SMF method are provided elsewhere (10, 11). In brief, the algorithm identifies the breast area, removing the pectoral muscle in the mediolateral oblique view, and creates a standardised image (SMF), which represents the image corrected for variations in imaging conditions that are not due to the breast tissue itself. It accounts for the anode-heel effect (i.e., less radiation reaching the breast the further away from the chest wall), compression plate slant (by assuming a 0.5 cm linear variation across the image), and glare caused by the intensifying screen, and it removes the effects of both scattered and extrafocal radiation. The algorithm incorporates imaging acquisition variables (calibration data) for voltage, exposure, and anode-filter combination. These variables affect the intensity and total energy emitted by the X-ray source and the degree to which X-rays are absorbed by the filter before exposing the film. Knowledge of these variables improves the accuracy of standardising the image. When these variables are partially known or unavailable, the program assumes a voltage of 28 kVp and a molybdenum-molybdenum anode-filter combination, and estimates the exposure from the image itself. The distance between the plates (breast thickness) is always estimated from the image itself in SMF version 2.2β, by identifying an area in the breast that is almost entirely fat and then using the image information to estimate the thickness of fat (i.e., breast thickness) that would have led to the degree of attenuation observed.

SMF models the difference in the X-ray energy entering and leaving the breast as a function of breast thickness and X-ray attenuation properties of the materials in the intervening cross-section. The material is predominantly either fibroglandular (referred to as ''interesting'') or fatty tissue, and these have different X-ray attenuation coefficients. Thus, given the total height of the breast (i.e., the distance between the compression plates), the height of dense interesting tissue at each pixel, denoted by h_{int}, can be estimated. SMF percent density is then calculated as the total volume of dense tissue (sum of heights h_{int}) divided by the total breast volume (either including or excluding the edge region at the periphery of the breast where the breast thickness is less than the distance between the plates but cannot be precisely estimated).

Output from SMF version 2.2β accessible to the user includes an indication of the quality of the SMF image (excellent/poor) and an image showing the segmentation regions. Poor-quality images are those for which SMF produced negative estimated heights of dense tissue for >5% of the breast region, indicating probable errors. These SMF values were considered unreliable and thus removed from the analysis. The SMF algorithm was run with and without incorporation of the calibration data (data for the latter are not presented in full here).

Statistical Methods. Neither the SMF nor threshold percent density values were normally distributed, so a natural logarithmic transformation of SMF percent density and a square-root transformation of threshold values were taken in all analyses. Distributions are therefore described in terms of percentiles rather than means/SD, for ease of interpretation. Random-effects normal models were used to model correlated measures of density from the four films for each woman, (on the transformed scale). Such models were used to assess whether there were systematic differences between breasts or between views, in addition to estimating view-specific between-women (σ_b^2) and between-breast variance (σ_w^2). Repeatability of a density measurement made on the same film is not assessed as the SMF method is 100% repeatable; that is, it always estimates the same value for a given film. Reliability [intraclass correlation coefficient (ICC), %] of a single density value obtained from either the left or right craniocaudal or mediolateral oblique view, as a measure of the average density from that view, was estimated as the percentage of the total variance that is due to between-subject variance [for a single image $\sigma_b^2 / (\sigma_b^2 + \sigma_w^2)$]. Bland-Altman plots (not shown) of the mean left-right difference against the average value were examined to check that mean differences were constant across densities. To investigate whether missing values (where the SMF image was of ''poor'' quality) occurred at random, or were related to the underlying breast density, a logistic regression model for the odds of being missing was fitted with threshold density and breast cancer risk factors as the explanatory variables. Analyses were conducted in Stata 9.

Nondifferential measurement error in breast density will lead to an attenuation of relative risks when this variable is included as an explanatory variable in regression models to predict breast cancer risk. Using the estimated ICC and assuming a true relative risk (RR) for breast cancer of 4 between the extreme categories of density, the attenuated relative risk was calculated as 4^{ICC}.

Results

The mean age of the 250 women was 60.1 years (range 55-65 years) and 98% were postmenopausal at mammography in 2004.

Percentage density values were obtained for all 1,000 images for the threshold method and for 976 images in the SMF method (2.4% were considered to be of poor quality; 15 craniocaudal and 9 mediolateral oblique images distributed among 17 women). There was evidence that the distribution of these missing values was not at random, as the percentage missing was 1.1% (95% CI, 0.4-1.9) among images with threshold densities of <20%, 4.8% (1.5-8.1) for threshold densities between 20% and 29% and 6.0% (1.9-10.0) in films with threshold densities 30% or greater (test for linear trend $P < 0.001$). Relative to those without missing SMF values ($n = 233$), women with at least one excluded image ($n = 17$) were slightly younger (mean age 58.9 versus 60.2 years; $P = 0.05$), had lower BMI (24.0 versus 27.1 kg/m^2; $P = 0.02$), and were more likely to be nulliparous (25% versus 13%, $P = 0.18$).

To carry out all further comparisons on the same sample of images, the remaining results are restricted to the 976 images with nonmissing data for both threshold and SMF densities.

Overall Distributions and Agreement of Ranks. Table 1 shows the percentiles of measures of breast size, dense tissue, nondense tissue, and percent mammographic density. In both measurement methods, and for both views, similar patterns of differences were observed between left and right breasts. There was no evidence of laterality differences for percent density measures or for the area/volume of dense tissue; however, for both threshold and SMF methods, the left breast was larger than the right (mean area of 178.1 cm^2 in the left versus 175.5 cm^2 in the right breast; SMF breast volume of 605.0 cm^3 in the left and 595.1 cm^3 in the right), corresponding to a larger nondense area/volume in the left breast. Although the breast size was larger in the mediolateral oblique than in the craniocaudal view, particularly so in the SMF method, in both methods there was no evidence that the overall percent density measures were different between views in either method.

The shape of the distribution of mammographic density as defined by the two methods is shown by the histogram of mean left-right density in the craniocaudal views (Fig. 1). As the mean age at mammography in this sample is 60.1 years (SD 2.6), the distribution of threshold density is as expected with a majority of densities under 20% (also see percentiles in Table 1). However, several key characteristics differ greatly between the distributions of SMF and threshold percent densities. First, the absolute values lie in very different ranges. Five percent of the threshold densities lie below 1.2%, whereas the corresponding 5th percentile is 14.8% in the SMF method including the edge (19.7% excluding it). None of the SMF percent densities were very low (minimum 9.1% for SMF including the edge). Second, the range of SMF values is much narrower. The interquartile range spans 19% in the threshold method but only 7.6% and 10.5% in the SMF method when the edge is included and excluded, respectively (Table 1).

Mean SMF percent densities were higher when calibration data were included in the SMF algorithm than when they were not (statistics not shown). For craniocaudal and mediolateral oblique views, respectively, mean SMF percent density (including the edge) were 0.87 (95% CI, 0.52-1.22) and 2.30 (95% CI, 2.02-2.56) percentage points higher when calibration data were included. Similar but larger differences were observed for SMF percent density excluding the edge.

Although the threshold and SMF methods are not expected to agree, they both attempt to measure the same underlying entity; thus, their determinants should be correlated. As shown in Fig. 2, SMF density increased with increasing threshold density. A 10% increase in threshold percent density corresponded to a 3.2% (95% CI, 2.9-3.6) increase in SMF percent

Table 1. Mammographic features as assessed by the threshold and SMF methods: percentiles and agreement between breasts and by view

		Breast size	Dense tissue	Nondense tissue	Percent density including edge	Percent density excluding edge
	Threshold measure (units):	Area (cm^2)	Area (cm^2)	Area (cm^2)	Of projected area (%)	NA
	SMF measure (units):	Volume (cm^3)	Volume (cm^3)	Volume (cm^3)	Of breast volume (%)	Of breast volume excluding edge (%)
CC views (244 women, 485 films)						
Threshold	Percentiles: 50th (25, 75)					
	Left	171 (126, 225)	20.8 (9.5, 33.0)	146 (99, 207)	12.8 (4.8, 23.6)	—
	Right	164 (118, 221)	19.6 (8.9, 34.6)	143 (98, 197)	13.5 (4.6, 23.6)	
	Test for left-right difference	$P = 0.04$	$P = 0.98$	$P = 0.04$	$P = 0.62$	
	Limits of agreement*	−34, +38	−13.2, +20.2	−34.1, 38.6	−7.6, + 11.1	
SMF	Percentiles: 50th (25, 75)					
	Left	569 (385, 784)	120 (89, 170)	440 (287, 607)	21.4 (18.6, 26.4)	27.4 (23.6, 34.4)
	Right	547 (369, 773)	119 (83, 172)	433 (280, 593)	22.1 (18.9, 26.3)	29.0 (24.7, 34.6)
	Test for left-right difference	$P = 0.21$	$P = 0.94$	$P = 0.18$	$P = 0.33$	$P = 0.11$
	Limits of agreement*	−138, 186	−42, +63	−116, +161	−6.5, +9.1	−8.1, +11.3
MLO views (246 women, 491 films)						
Threshold	Percentiles: 50th (25, 75)					
	Left	171 (135, 217)	18.3 (9.5, 33.0)	148 (109, 200)	12.1 (5.7, 21.7)	—
	Right	170 (131, 213)	20.3 (8.6, 32.9)	146 (105, 191)	12.9 (5.0, 20.4)	
	Limits of agreement*	−32, +36	−13.7, +21.4	−33.0, +37.1	−7.9, +11.9	
	Test for left-right difference	$P = 0.02$	$P = 0.71$	$P = 0.03$	$P = 0.37$	
SMF	Percentiles: 50th (25, 75)					
	Left	630 (460, 872)	133 (95, 189)	480 (341, 673)	21.6 (18.1, 26.5)	28.5 (23.8, 34.2)
	Right	625 (443, 867)	134 (97, 191)	483 (339, 667)	22.0 (18.0, 26.9)	28.9 (23.9, 35.5)
	Limits of agreement*	−167, +230	−56, +97	−143, +206	−8.0, +12.6	−10.4, +16.1
	Test for left-right difference	$P = 0.12$	$P = 0.88$	$P = 0.06$	$P = 0.20$	$P = 0.20$
CC-MLO differences						
Threshold	Test for CC-MLO difference	$P = 0.03$	$P = 0.84$	$P = 0.01$	$P = 0.19$	—
SMF		$P < 0.001$	$P < 0.001$	$P < 0.001$	$P = 0.58$	$P = 0.40$

Abbreviations: CC, craniocaudal; MLO, mediolateral oblique; NA, not applicable.
*Limits of agreement refer to the range in which 95% of left-right differences are expected to lie, calculated as mean difference ± 2 SD (difference).

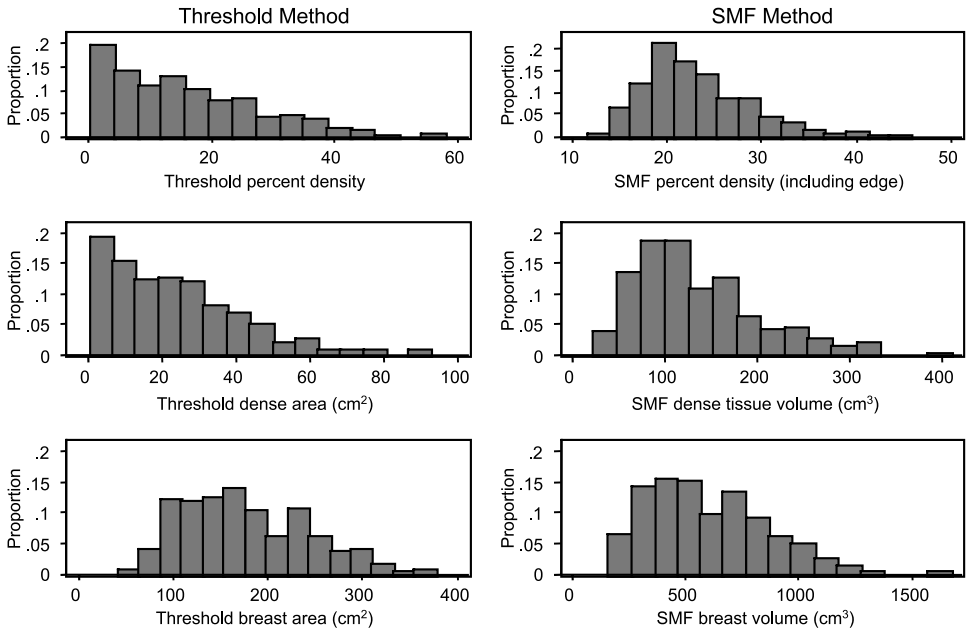

Figure 1. Distributions of SMF and threshold values of density in craniocaudal view (average left-right values, $n = 250$ women).

density including the edge, or a 4.4% increase (95% CI, 3.9-4.8) in SMF percent density excluding the edge. The correlations (Spearman) of density rankings between the threshold and SMF method (including the edge) were 0.71 for the average of left-right craniocaudal densities and 0.64 for average left-right mediolateral oblique densities and were higher (0.77 and 0.68, respectively) for SMF percent density excluding the edge. Correlations between ranks were very similar in magnitude for SMF densities estimated without calibration data. Measures of breast size (area in the threshold and volume in the SMF method) were highly correlated (Spearman correlation coefficient of 0.99 for both craniocaudal and mediolateral oblique views; Fig. 2); thus, differences in rankings of percent density in the threshold and SMF methods are predominantly due to differences between rankings of the area and volume of dense tissue (Spearman correlation coefficient of 0.32 and 0.29 for craniocaudal and mediolateral oblique views; Fig. 2).

Reliability. The correlation structure of density measurements by laterality and view revealed similar patterns in both measurement methods. A single density was more closely correlated with the density value (independently and blindly measured) not for the alternative view for the same breast but for the same view for the contralateral breast; for example, in the threshold method, the correlation coefficient of left craniocaudal was 0.90 with right craniocaudal and 0.87 with left mediolateral oblique (corresponding values were 0.76 and 0.59 for SMF percent density with edge). Thus, reliability of a single or left-right mean density as representative of the density from that view is assessed (Table 2). High reliability (>0.90) was found for both the threshold and SMF methods for measures of total breast area/volume, dense area/volume, and nondense area/volume in craniocaudal views. Reliabilities for percent densities (including edge) for the threshold method were much higher than those generated by SMF (0.92 versus 0.77 for a single craniocaudal view). However, the reliability of SMF was considerably improved by taking the left-right average of craniocaudal views (0.87 and 0.88 including and excluding the edge, respectively). In both methods, reliability coefficients were lower for the mediolateral oblique than for the craniocaudal view. When calibration data were not included in SMF estimation, reliability coefficients were slightly higher, resulting from greater between-subject variance (data not shown).

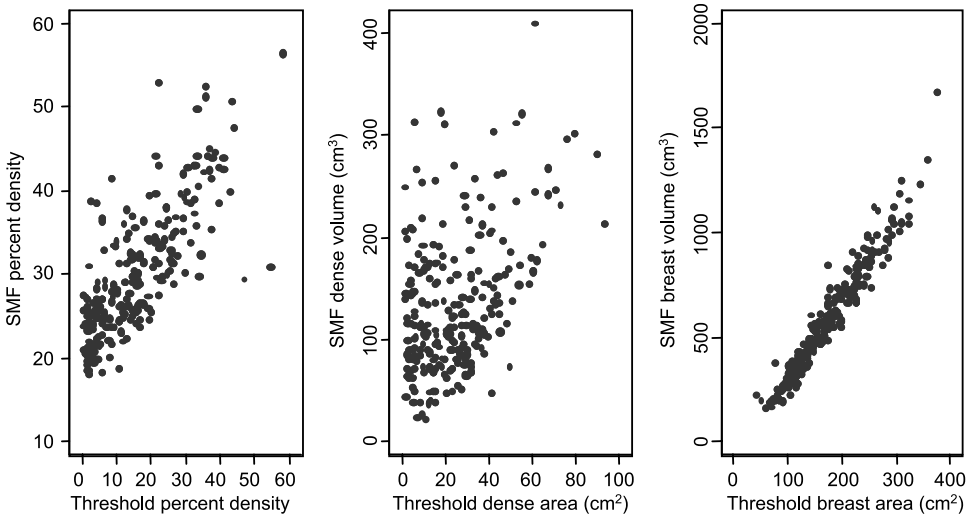

Figure 2. Scatter plots of breast density measures in SMF versus threshold method, craniocaudal views.

If associations between percent breast density and breast cancer risk were estimated for both density measurement methods, attenuation of relative risks would be lower when density is measured from the craniocaudal view (due to higher reliability; Table 2). Relative risks would be considerably reduced in the SMF method, especially if only a single mediolateral oblique view was used to assess density (e.g., from a true relative risk of 4 to an observed relative risk of 2.6; Table 2).

Associations with Breast Cancer Risk Factors. As the absolute values and distributions of percent density in the two methods are not comparable, associations of breast cancer risk factors with density are assessed on standardized density values so that the magnitude could be compared directly. Table 3 thus reports the number of SDs change in density associated with each breast cancer risk factor (all mutually adjusted). In both methods, crude analyses revealed that older women had lower density; however, this difference became nonsignificant after adjusting for BMI. A difference in BMI of 2 kg/m^2 corresponded to lower percent densities by 0.15 SDs (95% CI, 0.10-0.20) and 0.05 SDs (−0.03-0.10) in the threshold and SMF methods, respectively. However, contrasting effects were observed for the area/volume of dense tissue, where higher BMI was associated (borderline significant) with a lower dense area in the threshold method but with a larger volume of dense tissue in the SMF method. Higher parity was associated with lower percent density in both methods, the magnitude of which was greater for the SMF than the threshold method. Differential results were found for the effect of hormone replacement therapy on density. As expected, women who were taking hormone replacement therapy at the time of mammography had a higher threshold percent density (by 0.44 SDs; 95% CI, 0.09-0.79); however, the corresponding difference for SMF was much reduced and not statistically significant (0.25 SDs; 95% CI, −0.12-0.63).

Discussion

This study is the first to directly compare the SMF with the interactive threshold method. Expected properties of mammographic density were observed for the SMF method of measurement, such as laterality agreement and associations with breast cancer risk factors. However, the SMF method was less reliable than the threshold method. Some of these limitations may be overcome by altered study design, which we outline below.

Theoretically, volumetric-based measures of mammographic density are expected to be better markers of breast cancer risk than the currently used area-based methods. The SMF method is such a volumetric method, which, in contrast to other volumetric approaches, can be applied to retrospectively collected films because it does not require phantoms or wedges to be placed on the X-ray plate at the time of mammography (9).

The first drawback of the SMF method in this study was the presence of missing values. Although the overall percentage of missing values was low (2.4%), it was higher than the two previously reported figures of 0.7% (10, 13). Critically, these missing values did not appear to occur at random, but were more prevalent in women with more dense breasts. Thus, in studies using SMF percent density, exclusion of women based on a missing value may lead to biases. This problem may be exacerbated in populations of younger women where density is on average higher; however, as our study sample only included women ages 55 to 65 years, we were unable to assess this. On the contrary, missing values never or rarely occur in the threshold method. We note, however, that SMF version 2.2 was tested and developed only on Canon digitized images (an Array digitizer was used here) and it is likely that better performance can be achieved by further improving the segmentation algorithms to work for all digitizers, or in the future by only using directly digital mammograms.

For both the threshold and SMF methods, on average the left breast was slightly larger than the right, a finding that has previously been reported (16). Despite this, for the SMF method, as for the threshold method, we found no evidence of systematic differences in percent density between left-right breasts or between craniocaudal and mediolateral oblique views, suggesting that these measures can be used interchangeably. However, as between-breast variation in densities was smaller in the craniocaudal than the mediolateral oblique view, the former should be selected if a restricted number of measurements are made. We found that a single SMF density was less reliable than a single threshold density (e.g., reliability 0.91 versus 0.77 for craniocaudal views); however, by taking the average of SMF densities for two or more views, reliability

Table 2. Reliability of measures of mammographic features in the threshold and SMF methods and resulting attenuations of relative risk estimates

Mammographic feature (definition and units as per Table 1)	Reliability coefficient				Attenuated RR (original RR, 4)*			
	Threshold		SMF		Threshold		SMF	
	Single film	Left-right mean	Single film	Left-right mean	Single film	Left-right mean	Single film	Left-right mean
CC views								
Breast size	0.96	0.98	0.95	0.97				
Dense tissue	0.89	0.94	0.91	0.95				
Nondense tissue	0.96	0.98	0.94	0.97				
Percent density including edge	0.92	0.96	0.77	0.87	3.58	3.78	2.91	3.34
Percent density excluding edge	—	—	0.79	0.88	—	—	2.99	3.39
MLO views								
Breast size	0.96	0.98	0.94	0.97				
Dense tissue	0.86	0.93	0.87	0.93				
Nondense tissue	0.96	0.98	0.93	0.97				
Percent density including edge	0.89	0.94	0.68	0.81	3.43	3.68	2.57	3.07
Percent density excluding edge	—	—	0.68	0.81	—	—	2.57	3.07
CC and MLO (average of four films)								
Percent density including edge	0.97		0.87		3.84		3.34	
Percent density excluding edge	—		0.88		—		3.39	

Abbreviation: RR, relative risk.
*Attenuated relative risk (true relative risk, 4) of associations with density as an exposure, assuming nondifferential measurement error.

Table 3. Mutually adjusted regression coefficients for established risk factors for mammographic density measured by the threshold and SMF methods in 250 women

Explanatory variable	Increment/ category	Mean (SD) or percentage (%)	Dense tissue		Percent density including edge		Percent density excluding edge
			Threshold (area)	SMF (volume)	Threshold	SMF	SMF
Age	Per 1 y	60.1 (2.61)	−0.005 (−0.05, 0.04) $P = 0.84$	−0.02 (−0.06, 0.02) $P = 0.35$	−0.007 (−0.05, −0.04) $P = 0.76$	−0.03 (−0.07, 0.02) $P = 0.26$	−0.02 (−0.07, 0.02) $P = 0.31$
BMI	Per 2 kg/m^2 increase	26.9 (5.1)	−0.05 (−0.10, 0.003) $P = 0.06$	0.20 (0.16, 0.25) $P < 0.001$	−0.15 (−0.20, −0.10) $P < 0.001$	−0.07 (−0.12, −0.02) $P = 0.006$	−0.10 (−0.15, −0.05) $P < 0.001$
Parity	1-2 (reference)	30.6%	0	0	0	0	0
	3-5	46.5%	−0.16 (−0.43, 0.12) $P = 0.27$	−0.25 (−0.50, −0.00) $P = 0.05$	−0.19 (−0.45, 0.07) $P = 0.16$	−0.45 (−0.73, −0.18) $P = 0.001$	−0.42 (−0.69, −0.15) $P = 0.002$
	6+	9.0%	−0.42 (−0.89, 0.06) $P = 0.08$	−0.63 (−1.05, −0.21) $P = 0.004$	−0.40 (−0.84, 0.04) $P = 0.08$	−0.83 (−1.30, −0.36) $P = 0.001$	−0.66 (−1.12, −0.20) $P = 0.005$
	Nulliparous	13.9%	0.30 (−0.09, 0.68) $P = 0.13$	−0.10 (−0.44, 0.25) $P = 0.59$	0.37 (0.01, 0.73) $P = 0.04$	0.09 (−0.29, 0.48) $P = 0.63$	0.25 (−0.12, 0.63) $P = 0.18$
HRT	No (reference)	88.2%	0	0	0	0	0
	Yes	11.8%	0.42 (0.04, 0.79) $P = 0.03$	0.08 (−0.25, 0.42) $P = 0.63$	0.44 (0.09, 0.79) $P = 0.02$	0.25 (−0.12, 0.63) $P = 0.19$	0.26 (−0.10, 0.63) $P = 0.16$

NOTE: Effect of the explanatory variable on the standardized density measure (mean across the four values for each woman). Coefficients are the number of SDs change in mammographic density (or dense area/volume) associated with each explanatory variable.
Abbreviation: HRT, hormone replacement therapy.

of an SMF average density was improved considerably especially for the craniocaudal view. The automated nature of SMF means that it is possible in practice to obtain SMF averages easily as this would incur greater digitization time but no extra reading time. Reduced reliability may be overcome in epidemiologic studies through increased sample size, but would have more serious consequences if SMF was to be used clinically for individual breast density or risk assessment. Reliability was not compromised by the nonavailability of calibration data, further supporting the robustness of the calibration parameter compensation method; that is, the method by which SMF percent density is estimated when imaging acquisition data are not known (10). However, there were systematic differences between SMF percent densities generated with and without knowledge of calibration data, so such densities are not directly comparable and should not be combined.

Greater reliability in the threshold method may partly be attributed to total control of segmentation for each individual film with its particular intricacies (e.g., skin-folds, labels). In the threshold method, the user has the flexibility to exclude any parts of the image and can use their judgment to do this. Automated segmentation in the SMF method was not 100% perfect. Visual inspection of segmentation images for the films in this study revealed that even among films that were not considered as poor quality (which were excluded from these analyses), there were a few images where segmentation failed to identify the correct area on the film that corresponded to the breast or failed to completely remove the pectoral muscle in mediolateral oblique views. It is thus advisable to visibly examine a sample of both craniocaudal and mediolateral oblique projections to gauge whether there are common segmentation errors not classified as poor-quality images. Such errors are rare though.

The above observations have important implications when SMF is examined as a predictor of breast cancer risk. Women with high density who are more likely to become breast cancer cases would have a greater probability of being excluded and thus any real differences in SMF percent density between cases and noncases would be artificially reduced. This potential bias, coupled with the lower reliability of SMF percent density, may attenuate estimated SMF-breast cancer risk associations. To avoid these biases and to maintain maximum statistical power, where possible, researchers should ensure that several films are digitized per woman and the algorithm developers should pay special attention to why the performance of segmentation algorithms is worse on denser breasts.

The absolute values of SMF percent density did not agree with the corresponding threshold densities and the range of values in the former was greatly reduced (by over a half). This observation concurs with an evaluation of SMF in a Glasgow cohort (13) in which the differences in SMF percent density between the two extreme categories (corresponding to a difference of 75% in the threshold method) was only 18% on the SMF scale. Absolute differences in density from the two methods should therefore not be directly compared. The theoretical definitions of the two methods account for some of these differences. In a predominantly fatty breast that only has thin regions of dense tissue that are not perceived to exceed the density threshold, the threshold percent density would be low. For the same breast in the SMF method, all dense tissue (no matter how thin) contributes to the volume of dense tissue, so very low values are rarely achieved. Similarly, in a very dense breast, high values are realized in the threshold method (say 80-90%) as only the fatty edge would not be considered as dense. In the SMF method, however, such a film would have a lower density, as effectively the threshold method considers the cross-section above dense areas to be 100% dense, whereas in the SMF method it could not be so high due to the layer of subcutaneous fat at the periphery.

Although comparisons between the SMF and threshold methods were made, differences between them should be interpreted with caution. The threshold method, although the most commonly used, is not a gold standard. This method was examined here to check that the sample of films revealed associations that we would expect based on previous studies. Imperfect agreements of ranks between the two density

measurement methods may equally result from errors in the threshold method as from errors in the SMF method. However, when both methods produce similar associations with breast cancer risk factors, this consistency is consider to be evidence that they are both measuring the same underlying variable, breast density, albeit on different scales. Further promising indications of SMF as a measure of mammographic density are revealed in the similarity of the direction, size, and magnitude of its standardized associations with age, BMI, and parity. For both measures, density decreased with increasing BMI, with the magnitude of the effect being remarkably similar (in standardized terms). This association was present for SMF whether or not the edge was included; thus, it remains important to adjust for BMI when analyzing SMF density, as is the case for the threshold method (17).

We have shown here that, based on the rarity of missing or invalid values, reliability, correlations with the threshold method, and associations with known risk factors, the SMF method may be a viable method of density measurement, albeit not without caveats. Studies investigating whether it is a marker of breast cancer risk will provide the definitive test of its use as a predictor of breast cancer risk.

Acknowledgments

We thank the women from the Central and East London Breast Screening Service who agreed to participate in this study of mammographic features; Akua Asare who digitized the films; Siemens Molecular Imaging Limited, suppliers of the SMF software; and Professor Mike Brady for providing access to the Array digitizer.

References

1. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2006;15:1159–69.
2. Boyd NF, Rommens JM, Vogt K, et al. Mammographic breast density as an intermediate phenotype for breast cancer. Lancet Oncol 2005;6: 798–808.
3. Wolfe J. Breast patterns as an index of risk for developing breast cancer. AJR Am J Roentgenol 1976;126:1130–9.
4. BI-RADS. Breast imaging reporting and data system. 3rd ed. Reston, VA: American College of Radiology; 1998.
5. Tabar L, Dean PB. Mammographic parenchymal patterns. Risk indicator for breast cancer? JAMA 1982;247:185–9.
6. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. Phys Med Biol 1994;39:1629–38.
7. Boyd NF, Byng JW, Jong RA, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. J Natl Cancer Inst 1995;87:670–5.
8. Pawluczyk O, Augustine BJ, Yaffe MJ, et al. A volumetric method for estimation of breast density on digitized screen-film mammograms. Med Phys 2003;30:352–64.
9. Shepherd JA, Herve L, Landau J, Fan B, Kerlikowske K, Cummings SR. Novel use of single X-ray absorptiometry for measuring breast density. Technol Cancer Res Treat 2005;4:173–82.
10. Highnam R, Pan X, Warren R, Jeffreys M, Davey Smith G, Brady M. Breast composition measurements using retrospective standard mammogram form (SMF). Phys Med Biol 2006;51:2695–713.
11. Highnam R, Brady M. Mammographic image analysis. London: Kluwer Academic Publishers; 1999.
12. van Engeland S, Snoeren PR, Huisman H, Boetes C, Karssemeijer N. Volumetric breast density estimation from full-field digital mammograms. IEEE Trans Med Imaging 2006;25:273–82.
13. Jeffreys M, Warren R, Highnam R, Smith GD. Initial experiences of using an automated volumetric measure of breast density: the standard mammogram form. Br J Radiol 2006;79:378–82.
14. Byng JW, Boyd NF, Little L, et al. Symmetry of projection in the quantitative analysis of mammographic images. Eur J Cancer Prev 1996;5: 319–27.
15. Byrne C, Schairer C, Wolfe J, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. J Natl Cancer Inst 1995;87:1622–9.
16. Senie RT, Saftlas AF, Brinton LA, Hoover RN. Is breast size a predictor of breast cancer risk or the laterality of the tumor? Cancer Causes Control 1993; 4:203–8.
17. Boyd NF, Martin LJ, Sun L, et al. Body size, mammographic density, and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2006;15:2086–92.

Comparison of a New and Existing Method of Mammographic Density Measurement: Intramethod Reliability and Associations with Known Risk Factors

Valerie A. McCormack, Ralph Highnam, Nicholas Perry, et al.

Cancer Epidemiol Biomarkers Prev 2007;16:1148-1154.