ORIGINAL ARTICLE

Reliability of the Percent Density in Digital Mammography with a Semi-Automated Thresholding Method

Guiyun Sohn, Jong Won Lee, Sung Won Park[1], Jihoon Park, Jiyoung Woo, Hwa Jung Kim[2], Hee Jung Shin[3], Hak Hee Kim[3], Kyung Hae Jung[4], Joohon Sung[5], Seung Wook Lee[6], Byung Ho Son, Sei-Hyun Ahn

Department of Surgery, [1]Department of Radiology, Health Promotion Center, [2]Department of Biostatistics and Clinical Epidemiology, [3]Department of Radiology and Research Institute of Radiology, and [4]Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul; [5]Department of Epidemiology, School of Public Health and Institution of Health and Environment, and [6]Graduate School of Public Health, Seoul National University, Seoul, Korea

Purpose: The reliability of the quantitative measurement of breast density with a semi-automated thresholding method (Cumulus™) has mainly been investigated with film mammograms. This study aimed to evaluate the intrarater reproducibility of percent density (PD) by Cumulus™ with digital mammograms. **Methods:** This study included 1,496 craniocaudal digital mammograms from the unaffected breast of breast cancer patients. One rater reviewed each mammogram and estimated the PD using the Cumulus™ method. All images were reassessed by the same rater 1 month later without reference to the previously assigned values. The repeatability of the PD was evaluated by an intraclass correlation coefficient (ICC). All patients were grouped based on their body mass index (BMI), age, family history of breast cancer, breastfeeding history and breast area (calculated with Cumulus™), and subgroup analysis for the ICC of each group was performed. All patients were categorized by their Breast Imaging

Reporting and Data System (BI-RADS) density pattern, and the mean and standard deviation of the PD by each BI-RADS categories were compared. **Results:** The ICC for the PD was 0.94, indicating excellent repeatability. The discrepancy between the paired PD values ranged from 0 to 23.93, with an average of 3.90 (standard deviation = 3.39). The subgroup ICCs for the PD ranged from 0.88 to 0.96, indicating excellent reliability in all subgroups regardless of patient variables. The ICCs of the PD for the high-risk (BI-RADS 3 and 4) and low-risk (BI-RADS 1 and 2) groups were 0.90 and 0.88, respectively. **Conclusion:** This study suggests that PD calculated with digital mammograms has an acceptable reliability regardless of patient age, BMI, family history of breast cancer, breastfeeding history, breast size, and BI-RADS density pattern.

Key Words: *Breast, Mammography, Observer variation*

INTRODUCTION

Breast density is considered to be an independent risk factor for breast cancer. It is estimated that women with an increased breast density have 4 to 6 times higher risk of breast cancer than women with less dense breasts [1-3]. The relative risk of cancer related to breast density is greater than most

traditional risk factors such as nulliparity and early menarche [4,5]. Recently, mammographic density has also been investigated as a surrogate marker of breast cancer treatment outcomes [6-8]. Therefore, assessment of breast density is gaining importance, not only in terms of breast cancer risk, but also for predicting the responsiveness of adjuvant radiotherapy or systemic therapy [9,10].

Traditionally, mammographic density was described through visual evaluation methods such as Wolfe's scale [11], the six-categorical assessment [12], and the four-category tissue composition description of the American College of Radiologists' Breast Imaging Reporting and Data System (BI-RADS) [13,14]. Of these subjective methods, the BI-RADS four-category tissue composition description is still the most commonly used, and it shows consistent association with breast cancer risk [2]. However, it is based on the subjective

Correspondence to: Jong Won Lee
Division of Breast and Endocrine Surgery, Department of Surgery, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro, 43-gil, Songpa-gu, Seoul 138-736, Korea
Tel: +82-2-3010-5603, Fax: +82-2-474-9027
E-mail: asanbreastcenter@gmail.com

estimates of the clinician and may not be comparable across different studies. It also subjectively categorizes density patterns instead of quantifying density with continuous variables. A study conducted by Grove et al. [15] found that misclassification of density patterns may lead to a significant underestimation of breast cancer risk. Accordingly, potential objective and quantitative methods for density assessments have become popular.

Currently, the computer-assisted thresholding method (Cumulus™; Sunnybrook Health Sciences Centre, Toronto, Canada) (Figure 1) [12], which measures the area of dense tissue and total breast area to calculate the percent density (PD; the dense area expressed as a percentage of the total breast area) is frequently being used as a reliable tool for quantitatively measuring mammographic density, with several studies reporting a significant correlation between PD and breast cancer risk [16-19]. Despite many advantages, the computer-assisted methods are limited by the fact that, though they are relatively objective and fully quantitative, they are calculated semi-automatically, causing an issue of variability and reproducibility. In addition, previous validation studies on PD used analog film mammograms, which need to be digitized through scanning [20]. Digital mammograms are now more widespread, and density measurement does not require further digitization. However, it has been demonstrated that the PD and absolute dense area tend to be lower when measured in digital mam-

Figure 1. Computer assisted semi-automated thresholding method, Cumulus™. The green area indicates dense area, and the red line indicates total breast area.

mograms than in analog films [16], which means that the reliability of the PD performed with processed digital mammograms might need additional validation.

The purpose of the present study was to assess the intra-observer reliability of PD performed with digital mammograms alone [18,20-22]. In a subgroup analysis, we also sought to determine whether the reliability is consistent regardless of patients' characteristics and the BI-RADS density pattern.

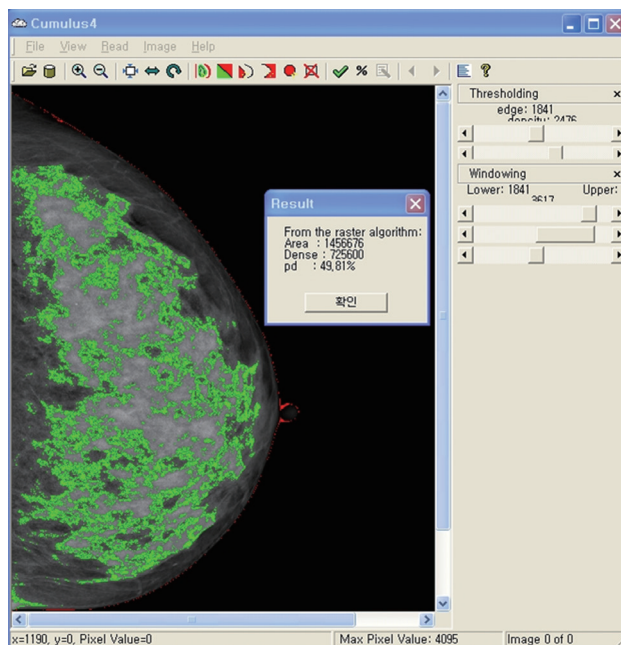## METHODS

Ethics statement

Since this study is a retrospective analysis of follow-up mammography of breast cancer patients, Institutional Review Board of Asan Medical Center approved this study didn't require written or verbal consent.

Subjects and mammograms

This study included 1,496 women with breast cancer who had a breast cancer operation from 2001 to 2003 and were followed for at least 5 years after operation in the Breast and Endocrine Division, Department of Surgery of Asan Medical Center. All mammograms were craniocaudal digital mammograms obtained from unaffected breast 5 years after surgery as a routine surveillance examination. All mammograms were performed with either a Senographe DS or Senographe Essential unit (GE Healthcare, Milwaukee, USA). Mammograms not eligible for density assessment due to multiple granulomas after foreign body injections were excluded.

Measurements and statistical methods

All digital mammograms were reviewed by a single technician (W.J., clinical nurse specialist) who had previous training in density assessment with Cumulus™. She reviewed each mammogram and estimated the PD using the Cumulus™, version 4.2. All images were reassessed by W.J. 1 month later without reference to the previously assigned values and without any patient information. The intrarater agreement of PD was evaluated by the intraclass correlation coefficient (ICC). All patients were grouped based on their clinical variables, which were body mass index (BMI), age, family history of breast cancer, breastfeeding history, and breast area (calculated with Cumulus™), and the ICC of the PD was calculated and compared for each group. BI-RADS density patterns were evaluated by multiple experienced radiologists at the Radiology Unit of Asan Medical Center. The mean and standard deviation (SD) of the PD was compared for each BI-RADS group. The ICC and Student t-test was used for the analysis. For the ICC values, 0.8–1.0 was considered to indicate excel-

lent repeatability, and all tests were two-sided with significance level of $p < 0.05$. All statistical analysis was performed with IBM SPSS statistical software version 19 (SPSS Inc., Chicago, USA).

RESULTS

To assess the agreement of the paired PD values, 1,496 digi-
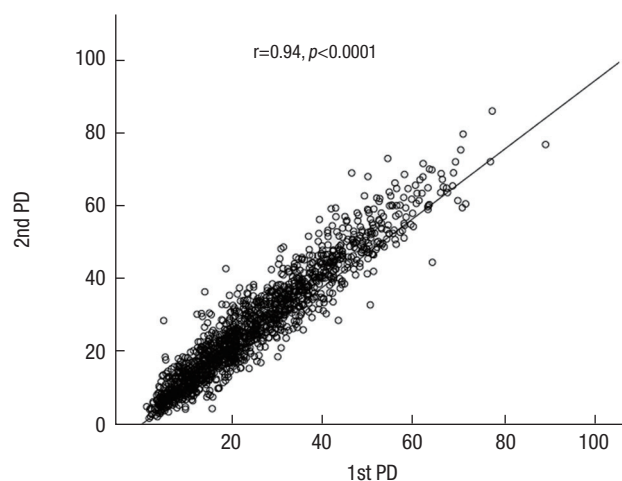
Figure 2. Intrarater reproducibility for two percent density value by one rater calculated with Cumulus™. Intraclass correlation coefficient between two PD values was 0.94, which indicates excellent agreement between two values.
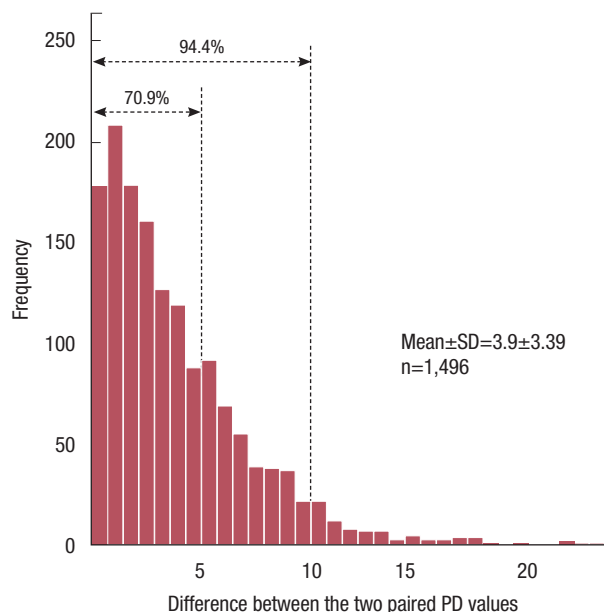
Figure 3. This histogram represents the distribution of the difference between two percent density (PD) values by one rater. 70.9% fell in within difference of 5, and 94.4% within difference of 10 between two paired PD values.

tal mammograms were included in the analysis. The correlation between the two independent PD measurements is shown in Figure 2. The ICC was 0.94, which indicates an excellent agreement between the first and second measurements. The discrepancy between the two paired PD measurements is shown in Figure 3. It ranged from 0 to 23.93, and the average was 3.90 (SD = 3.39). As seen in the figure, 70.9% of all the discrepancy values belonged to a range of 0 to 5, and 94.4% belonged to a range of 0 to 10.

The subgroup analysis of the ICC between the paired PD values is shown in Table 1. All patients were grouped according to their BMI, age, family history of breast cancer, breast-feeding history, breast area, and BI-RADS parenchymal pattern and the ICC was calculated for the PD pairs for each group. The ICC for each group ranged from 0.88 to 0.96, which indicates excellent repeatability in all groups regardless of their characteristics.

The relationship between the BI-RADS pattern and the PD calculated with Cumulus™ is shown in Figure 4. Of the 1,496 patients, 64.2% were estimated to be BI-RADS 2 (n = 961), whereas only 4.7% of patients belonged to BI-RADS 1 (n = 70). The mean PD values for each BI-RADS group increased as the density patterns indicated denser breasts.

Further analysis was done to identify the characteristic of

Table 1. Subgroup analysis of intrarater repeatibility

Characteristic	No. (%)	ICC
BMI		
<20	182 (12.2)	0.94
20–25	901 (60.2)	0.93
>25	413 (27.6)	0.92
Age (yr)		
<40	275 (18.4)	0.93
40–49	712 (47.6)	0.94
50–59	363 (24.3)	0.90
≥60	146 (9.8)	0.89
Breast area (pixels)*		
1st quartile (≤954,735)	374 (25.0)	0.91
2nd quartile (954,736–1,194,622)	374 (25.0)	0.94
3rd quartile (1,194,623–1,501,439)	374 (25.0)	0.96
4th quartile (≥1,501,440)	374 (25.0)	0.93
Family history of breast cancer		
Yes	112 (7.5)	0.95
No	1,369 (91.5)	0.94
Breastfeeding history		
Yes	1,034 (69.1)	0.93
No	435 (9.0)	0.94
History of endocrine therapy		
Yes	993 (66.4)	0.94
No	490 (32.8)	0.96
Unknown	13 (0.9)	

ICC = intraclass correlation coefficient; BMI = body mass index.
*Calculated with Cumulus™.

r=0.94, *p*<0.0001
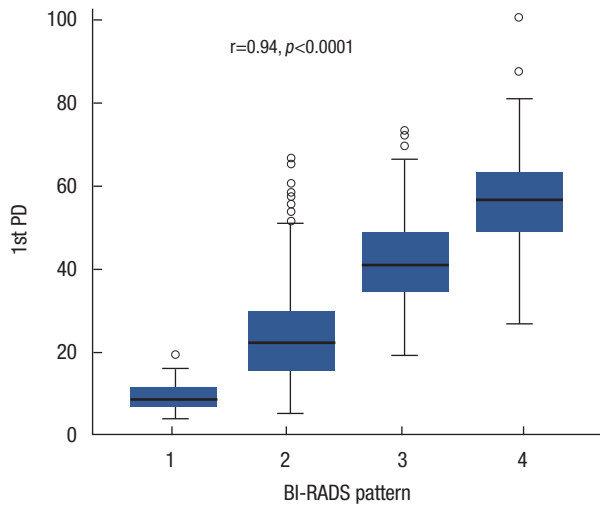
BI-RADS parenchymal pattern	PD[1]		ICC
	Mean ± SD (%)	Range (%)	
BI-RADS 1 (n=70, 4.7%)	0.42±3.10	1.33-15.77	0.88
BI-RADS 2 (n=901, 64.2%)	19.07±10.02	1.86-18.56	
BI-RADS 3 (n=284, 19.0%)	35.48±9.70	2.71-68.81	0.90
BI-RADS 4 (n=181, 12.0%)	48.40±12.18	14.71-64.01	

Figure 4. This distribution plot represents the distribution of first measured percent density (PD) value in each Breast Imaging Reporting and Data System (BI-RADS) density pattern group. The attached table indicates intraclass correlation coefficient (ICC) between two PD values by one rater in fatty breast group (BI-RADS 1 and 2), and in dense breast group (BI-RADS 3 and 4) separately.

population with extreme discrepancy between two paired PD values (≥ 10, n = 84). There was no demographic difference between two groups, but group with large discrepancy had a high average PD value with statistical significance (24.7 vs. 31.3, $p < 0.0001$). Figure 5 presents the Bland-Altman plot of PD value and its discrepancy between two paired PD values.

DISCUSSION

In the present study, we demonstrated high intrarater agreement of PD by Cumulus[TM], particularly when ICC values were used as indicators. The repeatability was consistent, irrespective of various patient variables, such as the BI-RADS parenchymal patterns. The various traditional methods have shown significant correlations with breast cancer risk in many studies, but they have limitations due to their subjective and categorical measurements, which are not optimal for statistical analysis and clinical application. Several studies present the superiority of quantitative methods over qualitative methods for the estimation of breast density and breast cancer risk assessment [17,18,23,24]. Tagliafico et al. [23] reported that automated or semi-automated estimation of breast density
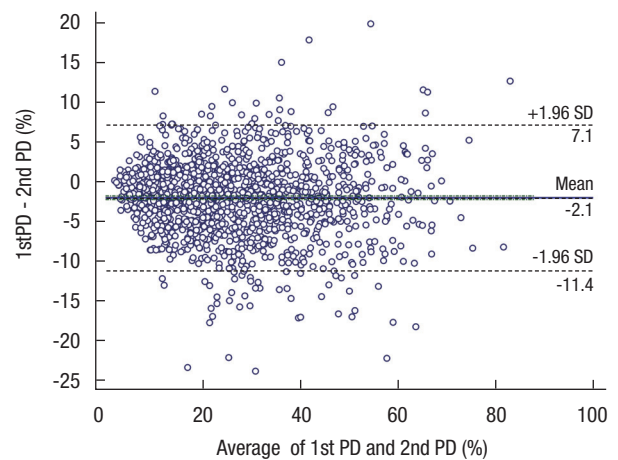
Figure 5. Bland-Altman plot of discrepancy between the two pared percent density (PD) values. X-axis indicates the average of two PD values and Y-axis indicates the discrepancy between two PD values, and in overall this plot represents the distribution of the discrepancy between two PD values, and it doesn't show any specific pattern in its distribution.
SD=standard deviation.

eliminates subjectivity, and is more accurate than the BI-RADS quantitative evaluation, and Gram et al. [24] reported that quantitative methods convey additional information on breast cancer risk over the qualitative methods in the classification of mammograms into high- and low-risk groups. In addition, in studies comparing the quantitative assessment of breast density and the qualitative Wolfe classification, the quantitative analysis was proven to have a stronger association with breast cancer [17,18].

Since Cumulus[TM], which, of several quantitative methods, has recently been gaining popularity, is a semi-automated method that requires a trained specialist, its reliability must be guaranteed regardless of the clinical characteristics and ethnicity of the patients. The present study suggests that Cumulus[TM], a computer-assisted thresholding method, shows high overall intrarater repeatability, as reported elsewhere [18,20-22,25,26] In our subgroup analysis that stratified patients by their variables, such as BMI, age, family history of breast cancer, breastfeeding history, and breast area, PD values showed high intrarater repeatability in all subgroups. Accordingly, Cumulus[TM], which is not currently routinely used in radiology units, may, given its reliability, be a promising candidate for density assessment.

We hypothesized that the parenchymal density pattern might affect the reliability of Cumulus[TM], and evaluated its repeatability in both dense and fatty BI-RADS parenchymal pattern groups. We grouped all mammograms into four BI-RADS density patterns, and calculated the mean PD and its variability for each group. As the BI-RADS density pattern in-

dicated more dense breasts, the mean PD value increased, which shows an acceptable correlation between the BI-RADS density pattern and the PD calculated with Cumulus™. Subgroup analysis for reliability after dichotomizing the four BI-RADS patterns into two subgroups demonstrated that the PD values also had a high intrarater repeatability that was irrespective of the BI-RADS parenchymal patterns (Figure 3).

Also there were group of patients with extreme variability between two paired PD values (≥ 10). Two groups showed no demographic difference, and showed difference only in PD itself. Cumulus™ is semi-automated method which requires a technician to outline each breast parenchyme in green and total breast area in red as shown in Figure 1. From my personal experience with Cumulus™, heterogenous parenchymal pattern might make it difficult to outline its breast parenchyme correctly, which further cause relative variance in its PD value. Therefore to identify which group of patients has a risk of under- or overestimation of its breast density, we might need to concentrate on its parenchymal pattern or characteristics which might cause parenchymal heterogeneity in future study.

This study has valuable clinical implications in several aspects. As mentioned above, all mammograms in the present study, unlike most previous studies performed with scanned analog mammogram images for the evaluation of breast density [20,25,26], were full digital mammograms. To our knowledge, this is the first series to exclusively address the reliability of PD from full digital mammogram images. When breast density is estimated based on Cumulus™ using digital mammograms, the dense area percentage may be lower and more variable that that of scanned film mammograms [16,27], in part due to a better delineation of the breast edge on digital mammograms and an increased nondense area from the image processing algorithms that improve visualization of the skin line and subcutaneous tissues. Therefore, the shift to digital mammography brings new concerns about the reliability of the computer-assisted thresholding method. Our study, with a relatively large sample size, demonstrated that the intrarater repeatability of PD using digital mammograms is highly agreeable and consistent, regardless of patient variables or parenchymal density pattern. Although density measured with Cumulus™ has been considered relatively objective and accurate compared to conventional subjective methods, its use was limited by the need to digitize the film mammograms before they could be read, adding to the resources and time required. Our results suggest that PD calculated with digital mammograms has as high a reliability as scanned mammograms, which has been shown in many previous studies [20,25,26], and both time and resources could be saved.

Interestingly, we identified a small number of populations

with extreme discrepancies in the breast density assessment between the two distinct methods: those with high PD values in BI-RADS category 1 and those with extremely low PD values in BI-RADS categories 3 or 4. Fundamentally, the BI-RADS density pattern is a parenchymal pattern analysis, different from the PD, which is a quantification of the absolute amount of dense area and the whole breast area. Nonetheless, it would be interesting to investigate further which subgroup shows the most extreme discrepancy between the two different methods and to determine which method is more accurate in density assessment for certain subgroups.

For the distribution of breast density, the results from this study should be interpreted with caution. The present study does not exactly reflect healthy unaffected Korean female populations since all mammograms in the study were collected from the unaffected breast of breast cancer patients. More than two-thirds of the patients belonged to BI-RADS pattern 1 or 2, indicating a larger proportion of nondense breast compared to that seen in a previous study of Korean populations [28]. This discrepancy and proportionate decrease in dense breast might have been caused by adjuvant hormonal therapy after breast cancer operation, as suggested by other studies [29,30]. This study was not intended to evaluate the distribution of breast density of Korean populations, and thus the results regarding the distribution of breast density in the Korean populations should be interpreted with caution.

In conclusion, our study demonstrated that mammographic PD, based on a semi-automated thresholding method, Cumulus™, and performed with full digital mammography, shows a highly acceptable intrarater agreement, which suggests that PD is a highly reliable quantitative value for breast density assessment regardless of patient variables or parenchymal density patterns.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med 2007;356:227-36.
2. 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.
3. Harvey JA, Bovbjerg VE. Quantitative assessment of mammographic breast density: relationship with breast cancer risk. Radiology 2004; 230:29-41.
4. Cummings SR, Tice JA, Bauer S, Browner WS, Cuzick J, Ziv E, et al.

Prevention of breast cancer in postmenopausal women: approaches to estimating and reducing risk. J Natl Cancer Inst 2009;101:384-98.

5. Boyd NF, Rommens JM, Vogt K, Lee V, Hopper JL, Yaffe MJ, et al. Mammographic breast density as an intermediate phenotype for breast cancer. Lancet Oncol 2005;6:798-808.

6. Morishita M, Ohtsuru A, Hayashi T, Isomoto I, Itoyanagi N, Maeda S, et al. Clinical significance of categorisation of mammographic density for breast cancer prognosis. Int J Oncol 2005;26:1307-12.

7. Kim J, Han W, Moon HG, Ahn SK, Shin HC, You JM, et al. Breast density change as a predictive surrogate for response to adjuvant endocrine therapy in hormone receptor positive breast cancer. Breast Cancer Res 2012;14:R102.

8. Park CC, Rembert J, Chew K, Moore D, Kerlikowske K. High mammographic breast density is independent predictor of local but not distant recurrence after lumpectomy and radiotherapy for invasive breast cancer. Int J Radiat Oncol Biol Phys 2009;73:75-9.

9. Highnam R, Jeffreys M, McCormack V, Warren R, Davey Smith G, Brady M. Comparing measurements of breast density. Phys Med Biol 2007;52:5881-95.

10. Diorio C, Pollak M, Byrne C, Mâsse B, Hébert-Croteau N, Yaffe M, et al. Insulin-like growth factor-I, IGF-binding protein-3, and mammographic breast density. Cancer Epidemiol Biomarkers Prev 2005;14:1065-73.

11. Wolfe JN, Saftlas AF, Salane M. Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a case-control study. AJR Am J Roentgenol 1987;148:1087-92.

12. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. Phys Med Biol 1994;39:1629-38.

13. American College of Radiology. ACR Breast Imaging Reporting and Data System. Reston: American College of Radiology; 1993.

14. American College of Radiology. ACR Breast Imaging Reporting and Data System Atlas. Reston: American College of Radiology; 2003.

15. Grove JS, Goodman MJ, Gilbert FI Jr, Russell H. Wolfe's mammographic classification and breast cancer risk: the effect of misclassification on apparent risk ratios. Br J Radiol 1985;58:15-9.

16. Byrne C, Schairer C, Wolfe J, Parekh N, Salane M, Brinton LA, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. J Natl Cancer Inst 1995;87:1622-9.

17. Saftlas AF, Hoover RN, Brinton LA, Szklo M, Olson DR, Salane M, et al. Mammographic densities and risk of breast cancer. Cancer 1991;67:2833-8.

18. Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, Miller AB, 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.

19. Kato I, Beinart C, Bleich A, Su S, Kim M, Toniolo PG. A nested case-control study of mammographic patterns, breast volume, and breast cancer (New York City, NY, United States). Cancer Causes Control 1995;6:431-8.

20. Gao J, Warren R, Warren-Forward H, Forbes JF. Reproducibility of visual assessment on mammographic density. Breast Cancer Res Treat 2008;108:121-7.

21. Boyd NF, Lockwood GA, Martin LJ, Knight JA, Jong RA, Fishell E, et al. Mammographic densities and risk of breast cancer among subjects with a family history of this disease. J Natl Cancer Inst 1999;91:1404-8.

22. Martin LJ, Melnichouk O, Guo H, Chiarelli AM, Hislop TG, Yaffe MJ, et al. Family history, mammographic density, and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2010;19:456-63.

23. Tagliafico A, Tagliafico G, Tosto S, Chiesa F, Martinoli C, Derchi LE, et al. Mammographic density estimation: comparison among BI-RADS categories, a semi-automated software and a fully automated one. Breast 2009;18:35-40.

24. Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E. Percentage density, Wolfe's and Tabár's mammographic patterns: agreement and association with risk factors for breast cancer. Breast Cancer Res 2005;7:R854-61.

25. Kataoka M, Atkinson C, Warren R, Sala E, Day NE, Highnam R, et al. Mammographic density using two computer-based methods in an isoflavone trial. Maturitas 2008;59:350-7.

26. Heine JJ, Carston MJ, Scott CG, Brandt KR, Wu FF, Pankratz VS, et al. An automated approach for estimation of breast density. Cancer Epidemiol Biomarkers Prev 2008;17:3090-7.

27. Harvey JA. Quantitative assessment of percent breast density: analog versus digital acquisition. Technol Cancer Res Treat 2004;3:611-6.

28. Jeon JH, Kang JH, Kim Y, Lee HY, Choi KS, Jun JK, et al. Reproductive and hormonal factors associated with fatty or dense breast patterns among Korean women. Cancer Res Treat 2011;43:42-8.

29. Boyd NF. Tamoxifen, mammographic density, and breast cancer prevention. J Natl Cancer Inst 2011;103:704-5.

30. Cuzick J, Warwick J, Pinney E, Duffy SW, Cawthorn S, Howell A, et al. Tamoxifen-induced reduction in mammographic density and breast cancer risk reduction: a nested case-control study. J Natl Cancer Inst 2011;103:744-52.