FULL PAPER

Estimation of percentage breast tissue density: comparison between digital mammography (2D full field digital mammography) and digital breast tomosynthesis according to different BI-RADS categories

[1]A S TAGLIAFICO, MD, [2]G TAGLIAFICO, PhD, [3]F CAVAGNETTO, [4]M CALABRESE, MD and [5]N HOUSSAMI, MD

[1]Department of Experimental Medicine, University of Genova, Genova, Italy
[2]Consiglio Nazionale delle Ricerche, Istituto di Matematica Applicata e Tecnologie Informatiche, Genova, Italy
[3]Department of Medical Physics, IST, Istituto Nazionale per la Ricerca sul Cancro, IRCCS Azienda Ospedaliera–Universitaria San Martino, Genoa, Italy
[4]Department of Diagnostic Senology, IST, Istituto Nazionale per la Ricerca sul Cancro, IRCCS Azienda Ospedaliera–Universitaria San Martino, Genova, Italy
[5]Screening and Test Evaluation Program (STEP), School of Public Health, Sydney Medical School, University of Sydney, Sydney, Australia

Address correspondence to: Dr Alberto Stefano Tagliafico
E-mail: alberto.tagliafico@unige.it

This work was presented as a scientific paper in Vienna at the European Congress of Radiology, 2013.

Objective: To compare breast density estimated from two-dimensional full-field digital mammography (2D FFDM) and from digital breast tomosynthesis (DBT) according to different Breast Imaging–Reporting and Data System (BI-RADS) categories, using automated software.

Methods: Institutional review board approval and written informed patient consent were obtained. DBT and 2D FFDM were performed in the same patients to allow within-patient comparison. A total of 160 consecutive patients (mean age: 50±14 years; mean body mass index: 22±3) were included to create paired data sets of 40 patients for each BI-RADS category. Automatic software (MedDensity©, developed by Giulio Tagliafico) was used to compare the percentage breast density between DBT and 2D FFDM. The estimated breast percentage density obtained using DBT and 2D FFDM was examined for correlation with the radiologists' visual BI-RADS density classification.

Results: The 2D FFDM differed from DBT by 16.0% in BI-RADS Category 1, by 11.9% in Category 2, by 3.5% in Category 3 and by 18.1% in Category 4. These differences were highly significant ($p<0.0001$). There was a good correlation between the BI-RADS categories and the density evaluated using 2D FFDM and DBT ($r=0.56$, $p<0.01$ and $r=0.48$, $p<0.01$, respectively).

Conclusion: Using DBT, breast density values were lower than those obtained using 2D FFDM, with a non-linear relationship across the BI-RADS categories. These data are relevant for clinical practice and research studies using density in determining the risk.

Advances in knowledge: On DBT, breast density values were lower than with 2D FFDM, with a non-linear relationship across the classical BI-RADS categories.

To tailor screening and diagnosis protocols, it is important to identify females with an increased risk of breast cancer [1–3]. It has been estimated that females with dense breasts (breast densities of >75%) have 4–6 times higher risk of breast cancer than females with low breast densities [4] and that breast density is increasingly recognised as an independent determinant of breast cancer risk and possibly in prognosis [5]. Assessment of breast density is becoming crucial in epidemiological studies, including the estimation of breast cancer risk and assessing breast density-related risk over time, radiation dose monitoring and monitoring drug-related response [6,7].

Different methods and classifications have been reported to assess breast density: the Tabar classification [8], Wolfe's parenchymal patterns [9], and both semi-quantitative and quantitative computer-aided techniques [10–16]. The Breast Imaging–Reporting and Data System (BI-RADS)

classification, considered as the additional quantitative scheme, is routinely used in the USA and was introduced to standardise reporting. Initially, it was based on four qualitative categories but an additional quantitative scheme was added in 2003, based on the extent of fibroglandular tissue [17]. Mammographic breast density estimation may be limited by the two-dimensional (2D) nature of the imaging technique, whereas a three-dimensional (3D) imaging modality, such as digital breast tomosynthesis (DBT), reduces the appearance of the overlapping parenchymal tissue and may therefore influence or alter density assessments [13,14]. In DBT, high-spatial-resolution tomographic images of the breast are reconstructed from multiple low-dose projection images acquired within a limited range of X-ray tube angles [15]. It has been demonstrated in a few studies that the automated estimation of breast density eliminates subjectivity between comparisons of full-field digital mammography (2D FFDM) and DBT and is more reproducible than a quantitative BI-RADS evaluation [14,16]. However, previous research mainly considered patients with relatively high breast density, with the possibility of the results not being applicable across all density categories and showing whether published percentage breast density differences between 2D FFDM and DBT apply to less dense or non-dense breasts. The purpose of our study was to compare the breast tissue density estimated using 2D FFDM and DBT among patients in a balanced data set of the four BI-RADS categories, using fully automated software.

MATERIALS AND METHODS

The local ethics committee (institutional review board of the National Institute for Cancer Research) approved the study, and written informed consent was obtained from all participating females to evaluate breast percentage densities using DBT and 2D FFDM.

Study population

This was a prospective study of diagnostic females performed from March 2012 to June 2012. Both DBT and 2D FFDM images were obtained in participating females. After DBT and 2D FFDM, further work was carried out as indicated by a standard clinical practice. Females were included prospectively until a preplanned number of 40 patients in each BI-RADS category were enrolled. The categories were determined according to the BI-RADS-defined parenchymal structure (D1: 0–25%, D2: 26–50%, D3: 51–75% and D4: >76%), and all four categories were included because this method is accurate in the evaluation of breast density, especially that of the interquartile range [12,16,18].

Digital mammography

For the 2D FFDM (Dimensions™; Hologic Bedford, MA), a breast was positioned for mediolateral oblique and craniocaudal (CC) views and was compressed using the standard mammographic compression force. After acquisition, digital mammograms were available in an unprocessed format with pixel values linearly proportional to the exposure of the radiograph at the detector and in a processed format, which was used for software analyses.

Breast Imaging–Reporting and Data System density classifications

Breast density was evaluated by two blinded radiologists (MC and AST) with 20 and 5 years of experience in breast examination,

respectively, with >6000 examinations reported every year using the 2D FFDM images and with more than 2 years of experience in DBT each.

Tomosynthesis

DBT was performed by three dedicated technologists with the same devices as used for 2D FFDM. The tomosynthesis images were acquired in the same compression sequence as the 2D images. The device consisted of a custom-designed high-power (mA) tungsten (W) anode X-ray tube and X-ray filters made of rhodium (Rh) and silver (Ag) for 2D FFDM and aluminium (Al) for DBT. These different filters are used in 2D and 3D imaging and produce optimal X-ray spectra based on the breast thickness/composition and imaging modes. With this approach, patient radiation exposure is minimised. The image receptor is a direct-capture 70-μm pixel pitch selenium detector. The tube potential for the two acquisitions was regulated accordingly. The X-ray tube moves over a 15-degree arc while the breast is compressed. The projections are then combined to create a pseudo-3D image set of the breast (a set of 2D images at different heights throughout the breast), with 1-mm slices through the breast. Each DBT data set to be analysed by a software consisted of 15 reconstructed images chosen from CC data set to represent a balanced selection among the lateral and central images. The selection was computerised and assured comparable distances among consecutive analysed images. All DBT images were analysed as reported in the literature [14].

Mammogram density analysis with fully automated software

The data set comprised the 2D FFDM and the DBT reconstructed planes obtained from all 160 patients. All DBT reconstructed planes were analysed and the mean value obtained was used for comparison with 2D FFDM. This method has been previously described in the literature [16]. The 4 BI-RADS density categories (40 patients per category) as classified by the two radiologists were based on 2D FFDM images. For the analysis, a fully automated version of a software previously used on digital analogue mammograms and DBT was used [12,14,16,18]. This software uses an algorithm based on the maximum entropy method, which has been demonstrated to be reliable for breast percentage assessment [12,14]. The difference "delta" between 2D FFDM and DBT was calculated for each BI-RADS category. The CC views were used for both DBT and 2D FFDM images, and the delta calculated consisted in the difference in the means of the values obtained by the data sets of CC views.

Statistical analysis

Statistical analysis was performed using the two-paired Student's t-test to compare breast density obtained on 2D FFDM with DBT. Values were expressed as mean ± standard deviation (SD). To correlate breast density obtained on digital mammograms and DBT, the Pearson's test and linear regression analysis were used. $p < 0.05$ was considered statistically significant. All analyses were performed with the SPSS® software for Windows (v. 10.1.3; SPSS Inc., Chicago, IL) and MedCalc statistical software for Windows (MedCalc, Brussels, Belgium). Intraobserver (re-evaluation of the same data sets after 2 months) and interobserver agreements between the two radiologists evaluating the four BI-RADS categories were calculated using K statistics.

RESULTS

A total of 160 consecutive patients (mean age: 50±14 years; mean body mass index: 22±3) were included. Therefore, this data set relates to a normal body mass index of the population. Pathological findings were three invasive ductal cancers, one invasive lobular cancer and one ductal *in situ* carcinoma. The median tumour size was 11 mm (range, 7–29 mm).

Program performance and time

None of the 2D FFDM or DBT images were excluded for technical problems. The participating radiologists were able to make their evaluations in less than 8 min (mean: 6 min). The mean time to complete the fully automated analysis was 4 min per DBT data set (range, 1–6 min).

2D FFDM *vs* DBT according to BI-RADS

In BI-RADS Category 1 (breast density: 0–25%), 2D FFDM density values were 16% higher than with DBT (Figure 1). The mean values and SD of percentage breast density were 25.8±3.6 for 2D FFDM and 9.7±5.5 for DBT; these differences in percentage density between DBT and 2D FFDM were significant ($p < 0.0001$).

In BI-RADS Category 2 (breast density: 26–50%), 2D FFDM density values were 11.9% higher than with DBT (Figure 2). The mean values and SD of percentage density were 32.6±7.2 for 2D FFDM and 20.7±10.1 for DBT; these differences in percentage density between DBT and 2D FFDM were significant ($p < 0.0001$).

In BI-RADS Category 3 (breast density: 51–75%), 2D FFDM density values were 3.5% higher than with DBT (Figure 3). The mean values and SD of percentage density were 60.7±13.4 for 2D FFDM and 57.2±5.4 for DBT; these differences in percentage density between DBT and 2D FFDM were significant ($p < 0.0001$).

In BI-RADS Category 4 (breast density: >76%), 2D FFDM density values were 18.1% higher than with DBT (Figure 4). The mean values and SD of percentage density were 79.7±7.9 for 2D FFDM and 61.7±6.12 for DBT; these differences in percentage density between DBT and 2D FFDM were significant ($p < 0.0001$).

Figure 1. Comparison of breast density between 2D FFDM and DBT in BI-RADS Category 1 (0–25%). The *y*-axis shows the percentage density. BI-RADS, Breast Imaging-Reporting and Data System; DBT, digital breast tomosynthesis; FFDM, full field digital mammography. The open circle and closed square indicate outliers.
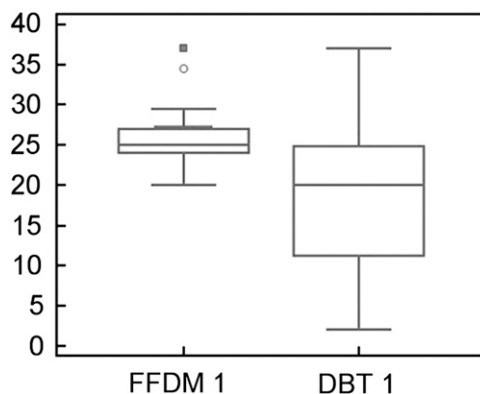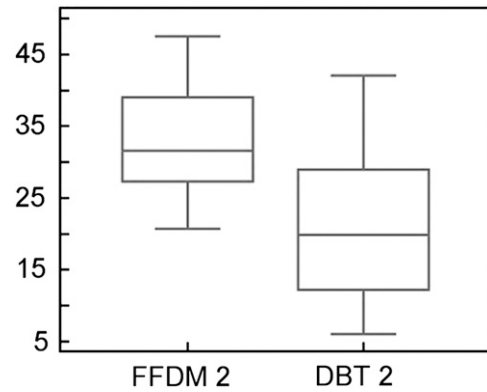

Figure 2. Breast density comparison between 2D FFDM and DBT in BI-RADS Category 2 (26–50%). The *y*-axis shows the percentage density. BI-RADS, Breast Imaging-Reporting and Data System; DBT, digital breast tomosynthesis; FFDM, full field digital mammography.

The differences in delta between 2D FFDM and DBT according to BI-RADS categories are reported in Figure 5. There was a good correlation between BI-RADS categories on a four-grade scale and the density evaluated with DBT and 2D FFDM($r=0.56$, $p < 0.01$ and $r=0.48$, $p < 0.01$). Very good correlation ($r=0.76$, $p < 0.01$ and $r=0.82$, $p < 0.01$) was present with a two-grade scale (D1–2/D3–4).

Overall agreement between DBT, 2D FFDM and radiologists' BI-RADS was good ($k=0.81$ and $k=0.83$). The agreement was better for D1 and D4 categories ($r=84$ and $r=0.93$, respectively), whereas lower levels of agreement were found for D2 and D3 categories ($r=0.57$ and $r=0.55$, respectively).

Using BI-RADS method, intra- and interobserver agreements of the two radiologists in the evaluation of breast density were considered to be very good (intraobserver agreement for Reader 1: $k=0.81$; intraobserver agreement for Reader 2: $k=0.79$; interoberver agreement: Reader 1 *vs* Reader 2: $k=0.89$).

Figure 3. Breast density comparison between 2D FFDM and DBT in BI-RADS Category 3 (51–75%). The *y*-axis shows the percentage density. BI-RADS, Breast Imaging-Reporting and Data System; DBT, digital breast tomosynthesis; FFDM, full field digital mammography.
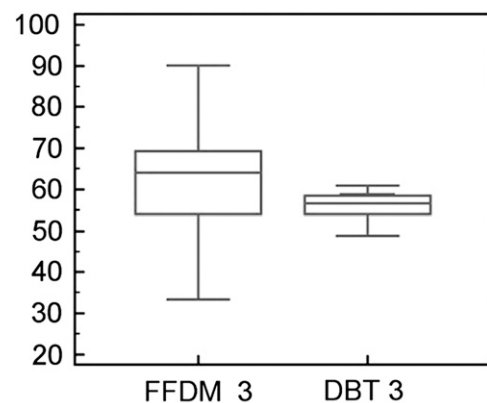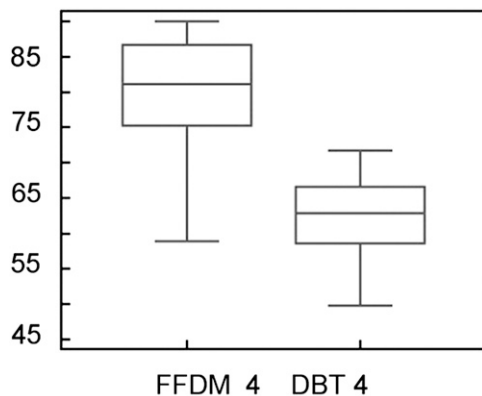
Figure 4. Breast density comparison between 2D FFDM and DBT in BI-RADS Category 4 (>76%). The *y*-axis shows the percentage density. DBT, digital breast tomosynthesis; FFDM, full field digital mammography.


DISCUSSION

This study addresses the existing gap in knowledge about how breast density estimation might differ between 2D FFDM and DBT across all BI-RADS density categories. Previous research in this area has limitations in mostly having patients with dense breasts rather than patients distributed in the classical BI-RADS categories [14,16]. Our data provide a direct comparison of the breast density percentage estimates between patients (meaning the same females who underwent both 2D FFDM and DBT) and also allow comparison of estimates derived from images acquired with 2D FFDM and DBT. The latter is particularly timely, given the potential transition from 2D (standard 2D FFDM) to a pseudo-3D (DBT) mammography, and that, in future, some patients or possibly some screening programmes might shift to DBT for mammography screening or assessment.

Breast density measurements are important for cancer risk predictive models and in both clinical and epidemiological studies, but it is not known if the values of breast density, which are usually lower on DBT [14,16], are consistent across the four BI-RADS categories. Inconsistent breast percentage density differences between DBT and 2D FFDM across BI-RADS categories may hamper the reproducibility and reliability of this measurement, more so given that breast density estimation may be problematic in risk assessment models and in clinical trials because of variability in the approaches used to measure it [4]. In our study, we explore the comparative percentage density across the four BI-RADS categories using DBT and 2D FFDM in the same patients. This design allowed us to assess whether differences between 2D FFDM-based density percentage and DBT-based density percentage were constant across BI-RADS categories. For this purpose, we calculated the delta belonging to each BI-RADS category.

For BI-RADS Categories 2 and 3, there were relatively lower values of delta between 2D FFDM and DBT, whereas for BI-RADS Categories 1 and 4, there were higher values of delta for differences between 2D FFDM and DBT. For BI-RADS Category 1, the mean density of the 2D images was near the upper limit of the category definition. This may suggest that there is a reasonable difference between the radiologists' estimates (based on the 2D images) and the software evaluation for non-dense breasts. In breast cancer prediction models, the delta related to BI-RADS Category 4 may pose a serious concern because of significant underestimation on breast density percentage in patients undergoing DBT. For BI-RADS Category 3, the difference (delta) between 2D FFDM and DBT was lower; therefore, discrepancies in breast density estimation for these patients may be lower. These data are difficult to explain and may require further research. For BI-RADS Categories 1 and 2, the concern related to the lower values of breast density on DBT compared with 2D FFDM is relatively low if the initial evaluation is made on 2D FFDM. If the first evaluation is made on DBT, it is possible that some patients belonging to Categories 3 and 4 may be downgraded to a lower class.

Figure 5. Differences in delta between two-dimensional full-field digital mammography (2D FFDM) and digital breast tomosynthesis (DBT) according to Breast Imaging-Reporting and Data System (BI-RADS) categories. The *y*-axis shows the percentage density.
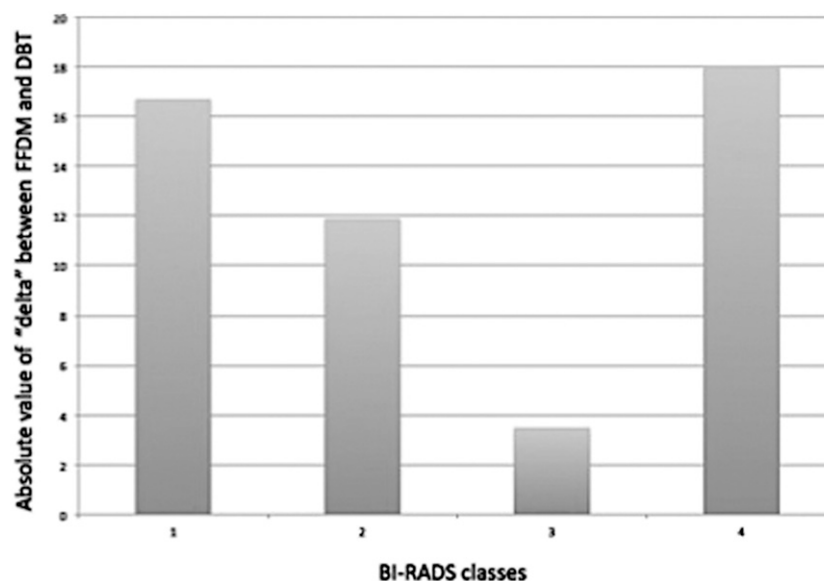
A strength of our study is the fact that the two radiologists who performed the BI-RADS evaluation were dedicated breast radiologists with experience in 2D FFDM and breast tomosynthesis; this may explain the very good correlation for intra- and interobserver agreement. Moreover, the use of the fully automated software eliminates subjectivity in breast density percentage evaluation. The fully automated software can perform image analysis without the need for superiority control; therefore, in large epidemiological studies seeking to integrate density measures, it is possible to analyse large sets of images automatically with efficiency. Recently, researchers have attempted to develop and validate fully automated methods to assess breast density quantitatively [11,12,16,19,20]. This approach could improve breast cancer risk models and may have an impact on diagnostic and therapeutic strategies that factor breast density [11]. Another strength of our study is the within-patient comparative analysis, which essentially "controls" for patient-related factors. In this study, we used reconstructed images; however, it has been shown that the percent density measured using an operator-assisted method from unprocessed and processed 2D FFDM images correlated well and provided similar associations with breast cancer [20]. Previous studies compared visual assessment of breast density evaluated as a continuous variable and found that it correlates better with Cumulus and Quantra™ [11,19].

To our knowledge, this study is the first work comparing breast density percentage between 2D FFDM and DBT with representative subjects across the four BI-RADS density classes. The potential advantage of measuring breast density with DBT is that the entire volume of the breast is considered, overcoming the traditional limitation of the 2D evaluation in 2D FFDM [12]. Also, as DBT is anticipated to be used in future mammography practice, this work highlights the differences between the two modalities in measuring breast density percentage for all density categories. A limitation of this study is that the applied reference standard for breast density was the calibration made on the basis of the radiologists' estimation. This method is not perfect and introduces subjectivity and merely reflects the lack of a "true" reference standard for breast issue density. A recently published study proposed the use of MRI with dedicated sequences as a reference standard [14]; although the MRI approach seems promising and worthy of further research, it was not possible in this study to include MRI examination for patients. Also, even if MRI were used, this would still not allow for an absolute reference standard for breast tissue density. Another potential limitation concerns the generalisability of these results, given that different software for automated density might give slightly different estimates. We acknowledge this possibility while pointing out that this would not substantially affect our relative estimates for density measures (*i.e.* 2D FFDM relative to DBT).

In conclusion, our study compares breast density estimation using four different data sets belonging to the classical BI-RADS categories. The comparison was performed using paired measurements because the same patients had both 2D FFDM and DBT, allowing a reliable and valid comparison of density percentage. Breast density values were significantly lower on DBT than with 2D FFDM, but with a non-linear relationship across BI-RADS categories. Higher differences between breast density estimation for 2D FFDM and DBT were observed for the least dense and the most dense breasts (BI-RADS Categories 1 and 4). These data will assist future clinical application and future research that includes breast density assessment and should be considered where density is measured using 2D FFDM or DBT or combinations thereof. Our findings may be particularly relevant for longitudinal studies planning to measure density as part of a risk assessment and that might cross modalities in measuring breast density over time.

REFERENCES

1. Lehman CD, Blume JD, Weatherall P, Thickman D, Hylton N, Warner E, et al. Screening women at high risk for breast cancer with mammography and magnetic resonance imaging. Cancer 2005;103: 1898–905. doi: 10.1002/cncr.20971

2. Sardanelli F, Podo F, Santoro F, Manoukian S, Bergonzi S, Trecate G, et al. High Breast Cancer Risk Italian 1 (HIBCRIT-1) Study Multicenter surveillance of women at high genetic breast cancer risk using mammography, ultrasonography, and contrast-enhanced magnetic resonance imaging (the high breast cancer risk Italian 1 study): final results. Invest Radiol 2011;46:94–105.

3. Sardanelli F, Podo F, D'Agnolo G, Verdecchia A, Santaquilani M, Musumeci R, et al. Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRIT study): interim results. Radiology 2007;242:698–715. doi: 10.1148/radiol.2423051965

4. Harvey JA, Bovbjerg VE. Quantitative assessment of mammographic breast density: relationship with breast cancer risk. Radiology 2004;230:29–41. doi: 10.1148/radiol.2301020870

5. Houssami N, Kerlikowske K. The impact of breast density on breast cancer risk and breast screening. Curr Breast Cancer Rep 2012. doi:10.1007/s12609-012-0070-z

6. 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. doi: 10.1088/0031-9155/52/19/010

7. 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. doi: 10.1158/1055-9965.EPI-04-0706

8. Gram IT, Funkhouser E, Tabar L. The Tabar classification of mammographic parenchymal patterns. Eur J Radiol 1997;24:131–6.

9. Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E. Percentage density, Wolfe's and Tabar's mammographic patterns: agreement and association with risk factors for breast cancer. Breast Cancer Res 2005;7:R854–861. doi: 10.1186/bcr1308

10. 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.

11. Li J, Szekely L, Eriksson L, Sundbom A, Czene K, Hall P, et al. High-throughput mammographic density measurement: a tool for risk prediction of breast cancer. Breast Cancer Res 2012;14:R114. doi: 10.1186/bcr3238

12. 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. doi: 10.1016/j.breast.2008.09.005

13. Tagliafico A, Astengo D, Cavagnetto F, Rosasco R, Rescinito G, Monetti F, et al. One-to-one comparison between digital spot compression view and digital breast tomosynthesis. Eur Radiol 2012;22:539–44

14. Tagliafico A, Tagliafico G, Astengo D, Airaldi S, Calabrese M, Houssami N. Comparative estimation of percentage breast tissue density for digital mammography, digital breast tomosynthesis, and magnetic resonance imaging. Breast Cancer Res Treat 2013;138: 311–7. doi: 10.1007/s10549-013-2419-z

15. Gur D, Abrams GS, Chough DM, Ganott MA, Hakim CM, Perrin RL, et al. Digital breast tomosynthesis: observer performance study. AJR Am J Roentgenol 2009;193: 586–91. doi: 10.2214/AJR.08.2031

16. Tagliafico A, Tagliafico G, Astengo D, Cavagnetto F, Rosasco R, Rescinito G, et al. Mammographic density estimation: one-to-one comparison of digital mammography and digital breast tomosynthesis using fully automated software. Eur Radiol 2012;22: 1265–70. doi: 10.1007/s00330-012-2380-y

17. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS). Reston, VA: American College of Radiology; 1993.

18. Tagliafico A, Calabrese M, Tagliafico G, Resmini E, Martinoli C, Rebora A, et al. Increased mammographic breast density in acromegaly: quantitative and qualitative assessment. Eur J Endocrinol 2011;164:335–40. doi: 10.1530/EJE-10-0896

19. Ciatto S, Bernardi D, Calabrese M, Durando M, Gentilini MA, Mariscotti G, et al. A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. Breast 2012;21: 503–6. doi: 10.1016/j.breast.2012.01.005

20. Vachon CM, Fowler EE, Tiffenberg G, Scott CG, Pankratz VS, Sellers TA, et al. Comparison of percent density from raw and processed full-field digital mammography data. Breast Cancer Res 2013;15:R1. doi: 10.1186/bcr3372