Assessment of mammographic density changes on plain film mammograms in postmenopausal women on hormone replacement therapy

Sebnem Orguc, Cihan Goktan, Gulgun Yilmaz Ovali, Oznur Karaer* & Semra Oruc*

Departments of Radiology, *Obstetrics & Gynecology, Celal Bayar University Medical Faculty Manisa, Turkey

Received September 29, 2005

Background & objectives: Mammographic screening is an effective tool for the early detection of breast cancer. Hormone replacement therapy (HRT) has been shown to increase mammographic density and thus may hinder early detection of small tumours. We undertook this study to determine and compare the frequency and degree of change in mammographic density in postmenopausal women in HRT using two different methods: the classical Wolfe classification and a new semiquantitative method, which we named as the comparison wheel.

Methods: This study included 285 women, 206 under hormone treatment, and 79 control subjects. All women underwent baseline mammographic study before the beginning of treatment. Mean interval of the follow up mammograms was 16 months. The methods were compared in evaluating the effects of three types of hormone therapies on mammographic density.

Results: The frequency of change was only significant in the combined hormone replacement group when Wolfe classification was used. However, the frequency of increase in density (estrogen group 21%, combined therapy group 42%, tibolone group 28%) was markedly higher when the comparison wheel was used. The inter-rater Kappa value was calculated as 0.977 for the first and 0.957 for the second readings of the two radiologists for the comparison wheel, and 0.973 and 0.968 for the Wolfe classification. The intra-rater Kappa values were determined as 0.972 and 0.957 for the first and 0.963 and 0.926 for the second radiologist for comparison wheel and Wolfe classification respectively.

Interpretation & conclusion: Our findings indicate that the estimated increase of mammographic density depends on the selected hormone regimen, as well as the method of evaluation. The comparison wheel is a semiquantitative method of evaluating changes of mammographic density and is sensitive and reproducible with high inter- and intra-rater Kappa values. This method can be used as an alternative for comparison of digital mammographic applications in the future.

Key words Hormone replacement therapy - mammographic density change - mammography - menopause
An increase in mammographic density has been reported in a significant proportion of postmenopausal women on hormone replacement therapy (HRT)\textsuperscript{1-6}. This increase may compromise the detection of small new tumours, which may appear during the mammographic follow up interval. The epidemiologic studies have found an increased risk for breast cancer as well as a risk of cardiovascular disease after long term HRT\textsuperscript{7-11}. Classically mammographic density is evaluated using the Wolfe classification\textsuperscript{12}. Many investigators used a subjective method of assessment, a modification of percentage scale, with categories based on the amount of dense breast parenchyma in relation to the whole breast volume\textsuperscript{1,4}. Others measured mammographic per cent density with the use of a planimeter (outlining tool) or by using computer-assisted techniques\textsuperscript{3,13,14}.

We compared the effects of tibolone, estrogen alone and estrogen in cyclic or continuous combination with progesteron on the mammographic density in postmenopausal women on HRT. Since visual plain film evaluation of density is subjective, we tried to quantify the changes of density. Therefore, we evaluated the effects of HRT using two different methods: the classical Wolfe classification which has similar categories with American College of Radiology density index from the Breast Imaging Reporting and Data System\textsuperscript{TM} (BI-RADS) and a new method, which we designed and named as the comparison wheel and compared the results with those postmenopausal women who did not receive HRT\textsuperscript{12,15}.

### Material & Methods

This prospective study included 285 women, 206 under hormone treatment (mean age 48.9 ± 4.7 yr), and 79 control subjects (mean age 50.2 ± 5.9 yr). The study was carried out in the Departments of Radiology and Obstetrics and Gynecology of Celal Bayar University Hospital in Manisa, Türkiye during November 2002 and December 2004.

All women underwent baseline mammographic study before the beginning of treatment. Bilateral mediolateral oblique (MLO) and craniocaudal (CC) mammograms were obtained with dedicated mammography equipment (Siemens Mammat 3000, Sweden) using standard compression and automatic exposure mode. The thickness of the breast under a standard compression (18 F) as well as the kvp and mAs values used for baseline mammograms, were noted and the same exposure parameters were repeated for the control mammograms, in order to minimize the intensity variability between successive mammograms of each woman and to estimate actual density change more accurately. The thickness of the breast was not measured, however it was assumed that, standard compression on the same breast yielded to equivalent breast tissue thickness. Minimum interval of the follow up mammograms was 12 months (mean interval: 16 months). Informed consent and ethical committee approval was obtained although no additional examination other than routine mammographic screening was applied. A total of 206 healthy postmenopausal women who were eligible for the study received a HRT regimen, specially chosen by their physicians, according to their gynecological status and personal preferences; 62 women who had hysterectomy and bilateral oopherectomy due to benign lesions received estrogen alone (16 women 0.05 mg/day transdermal estradiol and 46 women 0.625 mg/day conjugated estrogen), 98 received estrogen in cyclic or continuous combination with progesteron (52 women 2 mg/day estradiol valerate on days 1-21 and 1 mg/day cyproterone acetate on days 12-21 and 46 women 0.625 mg/day conjugated estrogen on days 1-25 associated with 10 mg/day medroxyprogesterone acetate) and 46 received 2.5 mg/day tibolone.

Seventy nine postmenopausal women not on HRT were included in the control group. The patient and the control groups consisted of consecutive postmenopausal women who had not received any kind of hormonal therapy previously.
The height and weight of each subject was noted and body mass index (BMI) was calculated as body weight in kilograms divided by the square of height in meters. Only the baseline and follow up cranio-caudal (CC) views of left breast were chosen arbitrarily to assess breast density.

Three types of hormone therapies were evaluated for their effects on mammographic density using both methods. Two radiologists interpreted the mammograms independently. The assessment of density was performed visually on plain films. In case of disagreement, which occurred in 8 cases a joint decision was established with the opinion of the third radiologist. Evaluations were done without the knowledge of HRT status.

Mammographic density of all the films were classified according to Wolfe in four categories: N1, breast parenchyma composed primarily of fat; P1 prominent ductal pattern in up to ¼ of breast volume; P2 prominent ductal pattern in more than one fourth of breast volume; and Dy extremely dense breast parenchyma. N1, P1, P2 and Dy categories of Wolfe corresponded to BI-RADS composition categories 1 through 4 for breast density respectively. An increase of density was recorded as positive when the follow-up mammograms fulfilled the criteria for an upward change in the Wolfe classification in comparison to the baseline mammograms for women under HRT and the control groups.

In addition to Wolfe classification, all mammograms were scaled according to the new scale, which we named as “the comparison wheel”. For this scale a circle on a transparent sheet was divided into 20 equal, wedge shaped sectors, each 18 degrees, forming a wheel pattern. L-CC mammograms for each individual woman were placed on the negatoscope, as mirror images, baseline on the left hand side, and the follow up on the right hand side. Then the wheel was superimposed on mammograms so that, the diameter of the circle bisected each breast through the nipples. The mammographic density of each sector was visually compared with its counterpart. An increase in density was recorded as positive when at least one sector had an increase of density determined by the perceptual judgement of the radiologist corresponding to at least 5 per cent. The total number of sectors, where an increase in density was found, determined the final score. Scoring with this method, which might have ranged between 0-10 also gave a chance of grading the degree of density change.

The same two radiologists re-evaluated the mammograms using Wolfe classification and the comparison wheel methods several months later.

Statistical analysis was performed using SPSS 10.0 (Statistical Package for the Social Sciences) Fisher’s Exact Test, Wilcoxon signed ranks test and Chi square test.

Results

The difference between mean BMI of the HRT group (25.2 kg/m²) and the control group (25.8 kg/m²) was not significant. There was no significant difference between the densities of baseline mammograms according to the Wolfe classification.

Using Wolfe classification, 10 of 206 (4.9%) women using HRT were rated in the upper category in comparison with the baseline mammograms, which was not significantly different from the control group (0%). The frequency of increase in density was most prominent in the estrogen plus progesterone group (8/98) (Table). The degree of density change was not assessed using the Wolfe classification, since all the women who had an increase of density were placed in the category, just above their baseline mammograms.

Using comparison wheel scale 67 of 206 women using HRT showed an increase of density in
comparison with the baseline mammograms, while in control group five women had decrease of density. The frequency of increase in density was most prominent in the estrogen progesteron combination group (41/98), however, no statistically significant difference was observed between continuous and cyclic regimens. A significant number of women under tibolone (13/46) and estrogen (13/62) regimens also showed an increase of density when the new scale was used (Table). There was no degrading from a higher to a lower category of breast density in any of the HRT groups.

Irrespective of type of classification, no increase of mammographic density was recorded in the 79 women, consisting the control group. Five women in the control group shifted to a lower category in the comparison wheel scoring.

The comparison wheel method also gave a chance of grading the degree of density change. The average score of increase in mammographic density (total sum of scores of patients/number of patients) was 0.83 in the HRT group versus -0.06 in the control group (decrease in density). The average scores of increase were 0.76 in tibolone, 0.56 in estrogen and 1.12 in estrogen progesterone combination group (Table).

The comparison wheel method also gave a chance of grading the degree of density change. The average score of increase in mammographic density (total sum of scores of patients/number of patients) was 0.83 in the HRT group versus -0.06 in the control group (decrease in density). The average scores of increase were 0.76 in tibolone, 0.56 in estrogen and 1.12 in estrogen progesterone combination group (Table).

The Fig. illustrates the change in mammographic breast density in 3 individual women, each on a different type of hormone treatment (Fig. A-C). No sign of malignancy or any suspicious mammographic abnormality warranting biopsy were found in the follow up mammograms of both the study and the control groups.

The inter-rater Kappa value for comparison wheel was calculated as 0.977 (Standard error 0.007) for the first and 0.957 (Standard error 0.010) for the second readings of the two radiologists. The intra-rater Kappa value was determined as 0.972 (Standard error 0.008) for the first and and 0.963 (Standard error 0.010) for the second radiologist.

The inter-rater Kappa value for Wolfe classification was calculated as 0.973 (Standard error 0.012) for the first and 0.968 (Standard error 0.013) for the second readings of the two radiologists. The intra-rater Kappa value was determined as 0.957 (Standard error 0.015) for the first and and 0.926 (Standard error 0.019) for the second radiologist.

**Discussion**

HRT with estrogen is widely used in postmenopausal women with its known benefits and risks. The breast is a target organ for sex steroids and there is an increased relative risk of breast cancer with prolonged HRT use. Four large studies have shown a reduction in the sensitivity of screening mammograms between 7 and 21 per cent with HRT use. HRT use is also associated with a reduction of specificity ranging between 12 and almost 50 per cent. Greendale et al. used linear regression analysis to examine the effects of HRT on the change in mammographic breast density in 3 individual women, each on a different type of hormone treatment (Fig. A-C). No sign of malignancy or any suspicious mammographic abnormality warranting biopsy were found in the follow up mammograms of both the study and the control groups.
Fig. Baseline (left) and follow-up (right) mammograms were evaluated for changes in mammographic breast density in 3 individual women, each on a different type of hormone treatment. A. The woman on tibolone treatment with P2 Wolfe pattern on both mammograms had an apparent increase in breast density on two sectors using the comparison wheel. B. The woman receiving estrogen alone upgrading from N1 to P1 on Wolfe classification had an increase in breast density on three sectors according to the comparison wheel. C. The woman on hormonal regimen with combined estrogen and progesterone had an increase of breast density from P1 to P2 on Wolfe classification and a total score of 5 points according to the comparison wheel.
in mammographic per cent density and reported greater mammographic density to be associated with the use of estrogen/progestin combination therapy. Carney et al. found that with increase in breast density the accuracy of screening mammography decreased in their series of 329,495 women. For women with fatty breasts, the sensitivity of mammography was 87 per cent and the specificity was 96.9 per cent. For women with extremely dense breasts, the sensitivity of mammography was 62.9 per cent and the specificity was 89.1 per cent. Chebelowski et al. suggested that estrogen plus progestin may stimulate breast cancer growth and hinder breast cancer diagnosis in their series of 16,608 postmenopausal women. Kerlikowske et al. reported that the use of long duration estrogen and progestin HRT increased the likelihood of developing cancer by 46 per cent compared to non users in their series of 374,465 women. Banks et al. reported that HRT reduced both the specificity and sensitivity of screening mammography in the Million Women Study. Byng et al. used an observer-assisted technique called interactive thresholding that allows reliable quantitative assessment of mammographic density to confirm that mammographic density is one of the strongest risk factors for breast cancer and is present in a large proportion of breast cancer cases. Marias et al. argued that computer assisted radiology could improve the early detection of abnormalities.

Our findings are consistent with the previous studies that reported an increase in density ranging between 17-75 per cent on the mammograms of women undergoing HRT. We determined the frequency of increase in density as 5 per cent using Wolfe classification. A significant percentage of women demonstrated an increase in density when comparison wheel scale was used. The wide range of frequency of mammographic density increase in these studies can be explained due to the differences in methodology as is shown in our study also.

Another factor affecting the frequency and degree of increase in density is the variety of selected HRT regimens. In this study, the highest frequency of density increase was detected in the group using estrogens combined with progesterone. The degree of density increase was only assessed using the comparison wheel and not by using the Wolfe classification. The degree of density increase paralleled the frequency and was higher than the other groups when estrogen plus progesterone was used (average score using comparison wheel: 1.12). The scores of estrogen alone and tibolone were lower as were the frequencies. Combined estrogen and progesterone regimen of HRT seems to be associated with increased mammographic density, and possibly decreased mammographic sensitivity and increased breast cancer incidence more often than the other regimens.

One of the major limitations of our study was not having the chance of digitizing our mammograms and using one of the more sophisticated medical image analysis tools. We are aware that there may be an error in each measurement since visual perception of the interpreter is involved. However, high reproducibility of the comparison wheel method makes it a semiquantitative alternative to computerized techniques in evaluating temporal density changes of baseline and follow up mammograms, especially in women using HRT.

In conclusion, the comparison wheel method was found to be a more sensitive method of scoring increase of mammographic density compared to Wolfe classification, with almost perfect agreement in the inter- and intra-rater Kappa scores which indicate the high reproducibility of the method and can be used in future as an alternative for comparison of digital mammographic applications.

References


26. Harvey JA, Bovbjerg VE, Smolkin ME, Williams MB, Petroni GR. Evaluating hormone therapy-associated...


Reprint requests: Dr Sebnem Orguc, 1811 Sok 29/9 35540, Izmir, Turkey
e-mail: sebnemorguc@superonline.com; sebnem.orguc@bayar.edu.tr