Effect of Tamoxifen on Mammographic Density

Catherine K. Chow,† David Venzon, Elizabeth C. Jones, Ahalya Premkumar, Joyce O’Shaughnessy, and JoAnne Zujewski

Diagnostic Radiology Department, Warren Grant Magnuson Clinical Center [C. K. C., E. C. J., A. P.], Biostatistics and Data Management Section, Division of Clinical Services, National Cancer Institute [D. V.], and Medicine Branch, Division of Clinical Sciences, National Cancer Institute [J. O., J. Z.], NIH, Bethesda Maryland 20892

Abstract

There are strong data showing that increased breast cancer risk is associated with increased mammographic density. Tamoxifen has been shown to decrease the risk of invasive breast cancer and decrease breast density. We sought to demonstrate and calculate the extent of change in mammographic density in women who have taken tamoxifen for up to 2 years. We evaluated mammograms from 28 high-risk women who were taking tamoxifen.

Four different methods of evaluation were used: (a) two qualitative methods (Wolfe criteria and the American College of Radiology Breast Imaging and Reporting Data System criteria); (b) one semiquantitative method (mammograms were assigned one of five semiquantitative scores by visual inspection); and (c) one quantitative method (computer-aided calculation of fibroglandular area from digitized mammograms). The Wolfe criteria showed a 0.03 category decrease per year (P = 0.50). The American College of Radiology Breast Imaging and Reporting Data System criteria showed a 0.1 category decrease per year (P = 0.12). Semiquantitative criteria showed a 0.2 category decrease per year (P = 0.039). Digitized scores showed a 4.3% decrease per year (P = 0.0007).

In conclusion, tamoxifen causes a decrease in mammographic density with use, an effect that is better quantitated with semiquantitative criteria or digitized images. Density change might become useful as a surrogate end point for the effect of tamoxifen and other chemopreventive measures, although our data do not predict an individual’s degree of risk reduction.

Introduction

The association of breast cancer risk and mammographic parenchymal patterns was first proposed in 1976 by John Wolfe (1), who suggested a direct relationship between the density of the breast and the likelihood of developing breast cancer (2).

Subsequently, Wolfe’s classifications fell into disuse, and mammographic parenchymal patterns were recorded for descriptive purposes, providing merely the background in which masses or calcifications were discovered. There has been renewed interest in mammographic density, however, as the result of more recent epidemiological data linking mammographic density to breast cancer risk (3–5). Various methods of quantitative classification of breast density have been described in the literature; they report ORs of between 2.8 and 6.0 for a range in breast density partitions (6). Tamoxifen has been found by several investigators to decrease breast density (7–9) as well as provide a 49% reduction in the risk of invasive breast cancer (10). To confirm and quantify the extent of breast density reduction and to assess the utility of using breast density change as a surrogate marker for invasive breast cancer risk reduction, we have undertaken a study of the mammographic densities of women enrolled in a chemoprevention study who have received a 2-year course of tamoxifen.

Materials and Methods

All 32 women participating in this study were enrolled in a pilot trial of tamoxifen and 4-HPR† in women at high risk for developing breast cancer, which was conducted between August 14, 1994 and February 8, 1998 at the Warren Grant Magnuson Clinical Center, NIH. This protocol was approved by the Institutional Review Board of the National Cancer Institute and was designed to determine the acute and cumulative toxicity of tamoxifen and 4-HPR and to assess the feasibility of obtaining adequate tissue to study potential intermediate biomarkers of proliferative and premalignant disease (11). Patients were selected based on either a histologically documented diagnosis of DCIS, LCIS, or atypical hyperplasia or a risk of invasive breast cancer of 1.7% over the next 5 years, based on the Gail model. Concurrent use of estrogen or progesterone replacement therapy or hormonal contraceptives was not allowed.

Tamoxifen was administered p.o. at 20 mg/day for 23 months; four cycles of 200 mg of 4-HPR were administered every 25 days with a 3-day drug holiday during months 1–4. Standard four-view bilateral mammography was performed before the study and annually.

For those patients with lobular or ductal carcinoma in situ, mammograms of the unaffected side were selected and digitized. For those who had not had surgery, one side was selected based on the amount of dense tissue present (favoring the side with more density) or the presence of mammographically indeterminate findings. Wherever possible, we selected the craniocaudal studies for evaluation and digitization; the mediolateral oblique studies were not preferred for digitization because of the variable appearance of the pectoralis muscle on the

† The abbreviations used are: 4-HPR, 4-(hydroxy)phenyl retinamide; BI-RADS, American College of Radiology Breast Imaging and Reporting Data System; DCIS, ductal carcinoma in situ; LCIS, lobular carcinoma in situ; OR, odds ratio; HRT, hormone replacement therapy.

Received 2/14/00; revised 5/17/00; accepted 6/20/00.

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.

1 To whom requests for reprints should be addressed, at Diagnostic Radiology Department, Warren Grant Magnuson Clinical Center, NIH, Building 10, Room 1C660, 10 Center Drive MSC 1182, Bethesda MD 20892-1182. Phone: (301) 496-7700; Fax: (301) 496-9933; E-mail: cchow@nih.gov.
films. Digitization of the mammograms was performed by first photographing all images on 4 × 5 sheet film using the 4-inch × 5-inch multidigit (LogEtronics, Springfield, VA) and then scanning the negatives with the Dicomed digital camera (Dicomed, Inc., Burnsville, MN) at 267 dots/inch and saving them as TIFF files. Analysis of the digitized mammograms was performed on a Power Macintosh 6100/66 computer using the NIH Image program.3

Using the NIH Image program, a region of interest was selected by manually outlining the entire breast (excluding the pectoralis muscle on the mediolateral oblique images). This region was then interactively thresholded by using the Density Slice tool, which allows segmentation of the image based on gray level and applies a tint over the fibroglandular densities above the specified threshold. Density slicing can be somewhat subjective, in that the reader must judge when the underlying fibroglandular tissue is adequately tinted. The digitized image is converted to a binary image in which all background is assigned a value of 0, and all tinted areas are assigned a value of 255. The percentage of the breast occupied by fibroglandular tissue is calculated as the fraction of the mean pixel value over 255. The digitized images were scored by one reader (C. K. C.) in batches of 20–30 images over the course of the study. We are not aware of any established methods of measuring the fidelity of digitized images. Published studies of analyses of mammographic density using digitized images rely primarily on measurements of inter- and intraobserver variability (6) and strive for agreement. Our study relies on a single reader (thus eliminating interobserver variability) and seeks to measure the change in, rather than the absolute quantification of, mammographic density.

The original study mammograms also were blindly scored by two independent readers using the following criteria: (a) BI-RADS (12); (b) Wolfe’s criteria (1), and (c) visual semi-quantitative criteria following the method of Boyd et al. (4). The BI-RADS method classifies breast parenchyma into four categories: (a) fatty; (b) scattered fibroglandular densities; (c) heterogeneously dense; and (d) extremely dense. The Wolfe criteria were described in his 1976 article and comprise N1 (lowest risk), P1 (low risk), P2 (high risk), and DY (highest risk). In the Boyd et al. (4) classification, six categories of breast density were recognized: (a) density = 0%; (b) density = >0% to <10% of the area of the breast; (c) density = 10% to <25%; (d) density = 25% to <50%; (e) density = 50% to <75%; and (f) density = ≥ 75%. For purposes of calculation, it was assumed that the differences between successive categories are approximately equal in all three methods.

To assess the changes in the scores of digitized images of patients not on hormonal manipulation (tamoxifen), 20 patients from another protocol were selected to form control group. These patients were randomly selected from a cohort undergoing long-term follow-up according to the National Cancer Institute protocol (13). These patients had not been treated with HRT and had been followed for a period of 12–20 years. The NIH Office of Human Subjects Research gave approval for anonymous retrospective review and digitization of these patients’ mammograms. Two annual mammograms from the years 1994–1999 (the same period of time as the chemoprevention trial) were selected randomly from each patient. Changes in the density of digitized images between the two mammograms were calculated and compared with the annual rate of change of the study patients. In addition, to determine operator-dependent variability in the scoring of digitized images, 20 digitized study images from the chemoprevention trial were randomly selected for rescoring.

The rate of change of density for the study patients was calculated as the difference between the density of the last mammogram and that of the baseline mammogram, divided by the interval between them (in years). To compare our results to those of Son and Oh (8), we also calculated the relative change in density (with data from digitized images only) as:

\[
\text{Original density} - \text{final density} \times 100
\]

\[
\text{Original density}
\]

The statistical method used for comparing the distributions of quantitative values between the two groups was the Wilcoxon’s rank-sum test. Paired values and rates of change were tested for differences from a zero using the Wilcoxon signed rank test. Correlations between breast density scores and rates of density change were assessed by the Spearman rank correlation method.

## Results

Thirty-two patients originally enrolled in the chemoprevention trial. The average age of the patients was 49.5 years (range, 36–74 years). Ten patients presented with a history of DCIS, 10 patients presented with LCIS, 1 patient presented with LCIS and atypical ductal hyperplasia, 1 patient presented with DCIS and LCIS, and 6 patients were eligible based on Gail model criteria (1 of these patients also had atypical ductal hyperplasia). Five patients did not complete the trial, three because of drug effects and two because of development of DCIS during treatment. Eighty-two mammograms were digitized, and results from 54 images were used in this study to calculate the rate of change in breast density. In addition, 80 film-screen mammograms were available for assessment with the Wolfe, BI-RADS, and semiquantitative criteria. Individual patients had two to four images available for analysis. Of the two patients who developed DCIS in the contralateral breast, one patient had taken part in the study for 9 months, and the other had taken part in the study for 15 months. The ages of the 20 control patients ranged from 42 to 57 years (average age, 51 years), which is not a significant variation (P = 0.18 by Wilcoxon’s rank-sum test).

The initial digital scores of the study images (calculated from 28 patients, 1 of whom did not complete the protocol) have a mean ± SD of 31.9 ± 19.0%, and range of 1–64.3%. The initial digital scores in the control patients (mean ± SD, 29.7 ± 12.0%; range, 11.9–59.9%) are comparable with those in the study population.

There is a weak trend toward lower initial breast density scores with increasing age at enrollment, amounting to a shift of approximately half a category over a 10–20-year age range in the BI-RADS, Wolfe, and semiquantitative methods and ~6% in the digitized scores; however, it is not statistically significant (Spearman rank correlation, P > 0.22).

### Rate of Change of Density Measurements Using Digitized Mammograms in Study Patients

Baseline mammograms were obtained from 4 days to 9 months (mean, 2.3 months) before the start of tamoxifen therapy. The rate of change of measurements was calculated for 27 patients (1 of the patients had an initial mammogram only). This rate averaged −4.3 ± 6.6%/year (range, −21.5% to 10.1%/year). The changes for

---

3 This public domain program is available at zippy.nimh.nih.gov or from the National Technical Information Service, Springfield, VA (Part No. PB93-504868).
Reproducibility of Scores. Twenty study images chosen randomly and found to be representative of the entire distribution were rescored. The differences between the rescored and original measurements (second set – first set) have a mean of $-2.2 \pm 6.9\%$ (range, $-15.5\%$ to $11.0\%$). These differences are consistent with a mean of zero ($P = 0.17$, Wilcoxon signed rank test).

The two readers using the three other systems had good agreement on the BI-RADS system, with 57 of 81 mammograms given the same score and the remaining 24 of 81 differing by one category. With the Wolfe criteria, the proportion in agreement was again 57 of 81 mammograms. With the semiquantitative system, only 33 of 81 mammograms were in complete agreement, attributable in part to the larger number of categories, but also because reader A tended to score higher across categories 3–6. Because the BI-RADS and Wolfe categories are standardized, we did not seek to measure reproducibility for each reader, instead we used the average of the two readers’ scores.

Changes in Measurements in Control Patients. The changes in scores of the follow-up mammograms for the control patients are effectively centered around zero (mean ± SD, $0.1 \pm 5.4\%$; range, $-9.2\%$ to $10.5\%$; $P = 0.88$ for the test of consistency with a mean of zero). The intervals between the annual mammograms for the control patients ranged between 9 and 19 months, with a median of 12 months. The SD of the changes in control patients is not very different from the SD of the rescored mammograms, which supports the conclusion that there is very little change with time in the scores in the control population.

Association Between Age and Menopausal Status and the Rate of Density Change. There is no evidence of an association between age at enrollment and the rate of density change in the scores of any of the four methods over the 2-year observation period ($P > 0.20$, Spearman method). There are 11 premenopausal and 16 postmenopausal women in the data set, and there is no difference between the two groups in the change in breast density ($P > 0.48$ for each of the four scores; $P = 0.90$ for the digitized system in particular, Wilcoxon’s rank-sum test). For the BI-RADS and Wolfe scores, the median changes are zero in both groups; for the other two scores, the changes are mostly negative, and the ranges of the two groups overlap considerably. The premenopausal women tend to have slightly higher scores at the time of the first measurement for each of the scores, but that tendency is not significant ($P > 0.36$ for each). Because of the numbers of patients in these groups, the power to detect mean differences $<6–8\%$ between them is low.

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|c|}
\hline
Criteria & Average & SD & Range & \(P\) \\
\hline
Wolfe & $-0.03$ & 0.4 & $-1.0$ to $1.5$ & 0.5 \\
BI-RADS & $-0.1$ & 0.4 & $-1.5$ to $0.5$ & 0.12 \\
Semiquantitative & $-0.2$ & 0.5 & $-1.2$ to $0.8$ & 0.039 \\
Digitized & $-4.3\%$ & 6.60\% & $-21.5$ to $10.1\%$ & 0.0007 \\
\hline
\end{tabular}
\caption{Breast density change*}
\end{table}

* The values are calculated per year. The unit for the Wolfe, BI-RADS, and semiquantitative scales is one category.
Discussion

There are strong empirical data suggesting that increased mammographic density is associated with breast cancer risk. Byrne et al. (14) studied a group of 1880 women from the Breast Cancer Detection Demonstration Project and found that dense breasts on mammography were associated with increased risk of breast cancer independent of family history, age at first birth, alcohol consumption, and history of prior biopsy. They also found that the higher relative risk for these patients was present for both premenopausal and postmenopausal women, and it persisted for 10 or more years. Byng et al. (6) estimate that because higher breast density is more common, it may confer a higher attributable risk of breast cancer than other factors, such as the BRCA 1 or BRCA 2 gene. A meta-analysis published by Warner et al. (15) found that the ORs for the risk of increased mammographic density varied according to study design and the method used to classify mammographic pattern; quantitative classification methods yielded higher ORs than four-category systems, which, in turn, yielded higher ORs than three-category systems. Although causality has not been proven, a study by Boyd et al. (3) demonstrated that the risk of hyperplasia (with or without atypia) or carcinoma was related to increasing mammographic density, providing a link to histological data. These authors postulate that collagen in the stroma is responsible for most of the density seen in mammography and that the response of stromal (and epithelial) cells to regulatory factors (endocrine, paracrine, angiogenesis) may be involved in the association with breast cancer risk.

It is well known that hormonal manipulations, such as HRT in postmenopausal women or the use of oral contraceptives in premenopausal women, can lead to changes in mammographic density (16–19), with oral contraceptive regimens decreasing breast density, and HRT increasing it. Tamoxifen has recently been shown to reduce the incidence of breast cancer in women at risk for this disease (10), leading to the Food and Drug Administration’s decision to approve its use as an agent for decreasing the risk for invasive breast cancer. In our chemoprevention protocol, we followed the changes in mammographic density in the treated patients to assess the feasibility of using mammographic density as a surrogate endpoint for tamoxifen effect in chemopreventive studies. Although we found a decrease in density using all four methods of assessment (Wolfe, BI-RADS, semiquantitative, and computer-aided calculation of digitized mammograms), we found that only the calculations performed with the semiquantitative method and the digitized mammograms achieved statistical significance at −0.2 category and −4.3% per year, respectively.

Other investigators have also found a decrease in breast density after tamoxifen therapy (7–9). Ursin et al. (7) reported the results from a pilot trial of 19 patients, of whom 5 received tamoxifen as adjuvant therapy. Based on the evaluation of trabeculations and the appearance of nodular and diffuse densities on the mammograms, Ursin et al. (7) found that the tamoxifen-treated group improved, i.e., showed a statistically significant decrease of 0.38 unit (using a 5-point scale). Their patients were all premenopausal at the start of treatment and were followed for 1 year. Son and Oh (8) followed 152 patients with breast cancer for 1 year. Forty patients received tamoxifen alone, and 62 patients received tamoxifen in conjunction with other therapy (chemotherapy and/or radiation). The control group consisted of 50 patients who had not received tamoxifen and 20 healthy women. The authors reported the degree of change in the contralateral breast by (a) estimating fibroglan-
though we do not believe that 4-HPR played a significant role in the changes in breast density, its effect cannot be entirely excluded. Nevertheless, globally decreased mammographic density may indicate an effective chemopreventive effect and thus may be used as a surrogate end point in the selection and validation of agents to bring into large-scale testing in the future. The clinical significance of a decrease in mammographic density for the individual patient, however, is unclear. The need for validation of surrogate end point markers is underscored by the findings in the two patients who developed new diagnoses of DCIS while on treatment, despite significant decreases in density. The lack of protection from in situ carcinoma may be explained by the hormone independence of DCIS. It is also possible that the decrease in density because of tamoxifen and the decrease in breast cancer risk are coincidental, or that the decrease in breast density attributable to tamoxifen unmasked DCIS that had been mammographically silent. Masking bias has been postulated to explain the association of increased cancer incidence in dense breasts (21). Cancers may be masked initially by the density of the breast but become detectable in later years because of age-related decreases in breast density, which can lead to the conclusion of increased breast cancer incidence in dense breasts.

We believe that data from digitized images should be used to detect and quantitate small changes in breast density, as we attempted in this study. Because of unavoidable differences in film exposure and patient positioning, qualitative assessments of density are likely to vary from interpreter to interpreter. Digitized information allows for the manipulation of gray scale and should provide a more accurate measurement of the fibroglandular area. Multiple publications describing various automated methods of calculating mammographic density are available in the literature (22–25). With the digitization techniques currently available, mammographic data collected in recent large-scale trials could easily be used for the validation testing of breast density as an intermediate marker for the effect of tamoxifen.

References


