NIH Public Access
Author Manuscript
*Ann Intern Med*. Author manuscript; available in PMC 2010 January 7.

Published in final edited form as:
*Ann Intern Med*. 2009 June 2; 150(11): 752–765.

# Short-term hormone therapy suspension and mammography recall: the Radiological Evaluation and Breast Density (READ) Randomized Trial

**Diana S.M. Buist, PhD, MPH**[1], **Melissa L. Anderson, MS**[1], **Susan D. Reed, MD, MS, MPH**[2], **Erin J. Aiello Bowles, MPH**[1], **E. Dawn Fitzgibbons, MPH**[1], **Juleann C. Gandara, MD**[3], **Deborah Seger**[1], and **Katherine M. Newton, PhD**[1]

[1]Group Health Center for Health Studies, Seattle, WA 98101

[2]University of Washington, Department of Obstetrics and Gynecology, Seattle, WA 98104

[3]Group Health Permanente, Seattle, WA 98112

## Abstract

**Background—**Without population-based evidence, some clinicians recommend short-term hormone therapy (HT) suspension to improve mammography performance. HT raises breast density and abnormal screening mammograms are more common among women with greater breast density and among women using HT.

**Objective—**Test whether brief (1–2 month) HT suspension before screening mammography decreases additional mammographic imaging (recall) in women aged 45–80.

**Design—**Three-group randomized controlled trial.

**Setting—**Integrated health plan in Washington state, US between 2004–2007.

**Patients—**1704 women aged 44–80 who used HT at their most recent screening (index) mammogram, were due for a screening (study) mammogram, and were still using HT.

**Address for reprint requests:** Diana S.M. Buist, PhD, MPH, buist.d@ghc.org, 206-287-2931; 1730 Minor Ave, Suite 1600; Group Health Center for Health Studies, Seattle, WA 98101-1448.
**Current mailing addresses for all authors:**
Diana S.M. Buist, PhD, MPH; Melissa L. Anderson, MS; Erin J. Aiello Bowles, MPH; E. Dawn Fitzgibbons, MPH; Deborah Seger; Katherine M. Newton, PhD
Group Health Center for Health Studies; 1730 Minor Ave, Suite 1600; Seattle, WA 98101-1448
Susan D. Reed, MD, MS, MPH
Obstetrics and Gynecology; Harborview Medical Center; Box 359865; Seattle, WA 98104-2499
Juleann C. Gandara, MD
Group Health Family Health Center; 125 16th Ave. E. Seattle, WA 98112

**Publisher's Disclaimer:** "This is the prepublication, author-produced version of a manuscript accepted for publication in Annals of Internal Medicine. This version does not include post-acceptance editing and formatting. The American College of Physicians, the publisher of Annals of Internal Medicine, is not responsible for the content or presentation of the author-produced accepted version of the manuscript or any version that a third party derives from it. Readers who wish to access the definitive published version of this manuscript and any ancillary material related to this manuscript (e.g., correspondence, corrections, editorials, linked articles) should go to www.annals.org or to the print issue in which the article appears. Those who cite this manuscript should cite the published version, as it is the official version of record."

**Trial Registration:** clinicaltrials.gov identifier: NCT00117663

**Protocol: Available to interested readers by contacting Diana Buist, PhD at buist.d@ghc.org**
**Statistical Code: Available to interested readers by contacting Melissa Anderson, MS at Anderson.Melissa@ghc.org**
Data: not available

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

**Intervention—**Block randomization (by breast density and HT type) to: no HT suspension (N=567) or suspension for 1 (N=570) or 2 (N=567) months before the study mammogram. One blinded expert radiologist interpreted all mammograms.

**Measurements—**Recall was the primary outcome and change in mammographic breast density (percentage and dense area) between index and study mammograms was the secondary outcome.

**Results—**Mammography recall rates were 11.3% (61/542 no-suspension), 12.3% (50/478 1-month), and 9.8% (44/451 2-month). We identified no subgroups where brief suspension resulted in decreased mammography recall. With suspension, decreases in percentage breast density were orderly and statistically significant: 0.1% (no-suspension), −0.9% (1-month), and −1.5% (2-month). Comparable ordered declines were observed for dense area. Suspension groups experienced increased menopause symptoms.

**Limitations—**Our results can only be generalized to women aged 45–80 years who have used HT for at least 1 year and will consider short-term suspension; the majority (61%) of eligible women refused participation. Mammography recall was determined by one expert radiologist.

**Conclusions—**Brief HT suspension was associated with small changes in breast density and did not affect recall rates. There is no evidence to support short-term HT suspension before mammography.

### Keywords

Randomized controlled trial; mammographic breast density; mammography recall; mammography performance; hormone therapy; hormone therapy suspension; hormone therapy suspension

## INTRODUCTION

Each year, breast cancer is diagnosed in over 182,000 U.S. women.(1) Mammography is the only available screening method proven to reduce breast cancer mortality,(2) yet it is imperfect. Over 10% of U.S. women screened for breast cancer receive a recommendation for additional imaging at the time of their mammogram and only 4.4% of U.S. women who receive additional imaging after a screening mammogram will be diagnosed with breast cancer within 1 year.(3,4) False-positive mammograms have an enormous impact on women (e.g., increased patient anxiety) and subsequent healthcare use and cost.(5) False-positive mammograms raise annual U.S. healthcare costs by approximately $200 million ($500 per workup).(6)

As mammographic breast density increases, both sensitivity and specificity of mammography are decreased.(7–13) Breast density is not static and is affected by exogenous and endogenous reproductive hormones.(14–17) We have previously reported clinically and statistically significant higher rates of mammography recall among HT users(18,19) and have demonstrated HT discontinuation between mammography exams (average time between exams=2 years) can reduce density and recall to levels comparable to non-users of HT.(18,19) Several reports encourage women to consider short-term HT suspension before receiving a mammogram, with the idea that cancers are more easily detected.(18,20–22) Some providers already recommend short-term HT suspension in the hopes it will improve mammography performance; however, population-based evidence has been lacking for widespread or ad-hoc adoption in clinical practice.

This study was designed to test whether brief HT suspension before a screening mammogram improves mammography performance. To assess the impact of HT suspension on mammography sensitivity and specificity, a three-group randomized trial would need to enroll 700,000 women, making it unfeasible. The vast literature on screening mammography performance supports that recall rate is an important measure of the interpretive performance

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of screening tests(23). There is evidence that US radiologists could lower their recall rates without negatively impacting their cancer detection rates(24–27). Screening mammography aims to maximize performance, which involves weighing the sensitivity against the specificity of a test. As recall rates increase, specificity decreases. The recall rate includes both false-and true-positive mammograms, but the majority of positive screening mammograms are false-positive; therefore, lowering recall rate in a screening population mostly translates into lowering the false positive rate. We therefore chose mammography recall rates as our primary outcome and change in mammographic breast density as our secondary outcome. This trial specifically tests whether brief (1–2 month) HT suspension before screening mammography decreases mammographic recall in women aged 45–80.

## METHODS

### Design Overview

This study was a three-group randomized controlled trial to test whether brief HT suspension before screening mammography decreases mammography recall in women aged 45–80. We identified and attempted to recruit 5861 potentially eligible women over 35 months from November 2004 through September 2007 and enrolled a total of 1704 women. Participants were considered enrolled only after we verified their eligibility, they returned a consent form and Health Insurance Portability and Accountability Act authorization form, and randomization was completed. Women were followed for 12 months following their study screening mammogram (through December 2008) for adverse events. The institutional review boards (IRBs) at Group Health and the U.S. Department of Defense reviewed and approved this study.

### Setting and Participants

The Radiological Evaluation and Breast Density (READ) randomized controlled trial was conducted within Group Health, an integrated health plan based in western Washington State (US), with approximately 550,000 members. Group Health has a population-based Breast Cancer Screening Program (BCSP) that women are invited to join when they turn 40 or join Group Health.(28,29) Demographic and breast cancer risk factor information is gathered by a self-administered questionnaire completed when women enter the BCSP and at each mammogram.(29)

We used automated administrative databases to identify and recruit women aged 45–80 years who were due for a screening (“study”) mammogram. Women had to have a previous mammogram (“index”) at Group Health within the past 2 years where they self-reported using HT, and evidence of continued HT dispensing at the time of recruitment. Women were ineligible if they had a history of myocardial infarction, angina treated with medication, coronary revascularization surgery, stroke, blood clots, breast cancer, mastectomy, or breast implants. We further excluded women with any dispensings of tamoxifen or raloxifene because of these medications’ potential influence on breast density. Women who had a breast density rating of “entirely fat” [Breast Imaging-Reporting and Data System® (30) (BI-RADS) breast density 1] at their index mammogram were also excluded since their breast density was unlikely to influence mammography recall.

### Randomization and Interventions

Women were randomized to one of three study groups (no-suspension or 1-or 2-month suspension) using a permuted block algorithm. Once a week, the study programmer identified women eligible for randomization, and each was assigned to one of six stratification cells, based on BI-RADS breast density at index mammogram and type of HT used in the 6 months before recruitment based on electronic pharmacy dispensing records [estrogen plus progestin

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(EPT) versus estrogen alone (ET)]. Within each stratification cell, a random number was generated for each woman using the ranuni function in SAS (version 9.0 SAS Institute Inc., Cary, NC), and women were assigned to study groups in order based on the rank of their random number within the stratification cell. Randomization group assignment was maintained in the study database, available to the study programmer, statistician, and staff responsible for mailing follow-up questionnaires, but the study radiologist and investigators were blinded to randomization group.

All women were mailed a letter with a baseline questionnaire and instructions personalized to their randomization group. The instructions included the participant's mammography appointment date and, for women in the suspension groups, the date they should stop HT. Participants were given the study nurse's toll-free number to discuss any study-related concerns and report any symptoms or adverse events. The study nurse called all women participants to answer questions, and confirm the baseline questionnaire had been completed and returned. These calls occurred two weeks before their scheduled HT suspension date (suspension groups) and one month before their scheduled mammogram (no-suspension).

**Data Collection—**We collected participants' self-reported information from four sources: BCSP questionnaires from the index and study mammograms; baseline study questionnaire at enrollment; and follow-up study questionnaire at study mammogram. Body mass index was calculated from self-reported height and weight collected at the index and study mammograms.

The baseline study questionnaire collected information on medical history, physical activity, tobacco and alcohol use. Both study questionnaires gathered information about HT use, sleep and menopausal symptoms. Participants rated the importance of getting a good night's sleep (1=not important to 10=very important) and reported the frequency (days/week) of nine common sleep disturbances selected from the General Sleep Disturbance Scale.(31) We used a modified Wiklund Menopause Symptom Checklist(32,33) to ascertain the presence and severity (0=not present to 10=severe) of menopausal symptoms (sleep, mood swings, vaginal bleeding, vaginal dryness, hot flashes and night sweats).

**Hormone therapy groups—**We used automated pharmacy dispensing data to identify potentially eligible women and to classify type of HT used for stratification. Women had to have two estrogen dispensings in the 6 months before recruitment; HT type was distinguished by the presence of any progestin dispensing in those 6 months (EPT).(34) We further classified HT dose using information on the highest daily conjugated equine estrogen equivalent dose and categorized these into: low (<0.625mg), medium (0.625mg), and high (>0.625mg).(34–36) Progestin dose was defined as: low (2.5mg medroxyprogesterone acetate or 100mg micronized progesterone) and standard dose (all other progestins).

## Outcomes and Measurements

**Mammography recall (primary outcome)—**Women completed their study mammogram as part of routine clinical care. All mammograms were interpreted in accordance with Group Health standard clinical practice during the trial, and clinical care was continued as usual. Within 1–2 months of the screening mammogram, the expert study radiologist (JCG) independently interpreted each study mammogram using the index mammogram for comparison and provided assessments and recommendations blinded to: randomization group, clinical assessment, and any clinical recommendations. Mammography recall was defined as any assessment requiring additional imaging or evaluation of either breast (BI-RADS assessment of 0 for either or both breasts)(30) determined exclusively from the study radiologists' interpretation. All women in the study had ≥1 screening mammogram at Group Health before being recruited; so all women had comparison films that were used at the time

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of the study screening mammogram. Our primary outcome was defined *a priori* as recommended recall of any imaging (for technical reasons or otherwise). Technical reasons for recall were indicated by our expert radiologist primarily for the reason of not having enough tissue in one or more of the mammographic views to have complete confidence that there were no changes in the breast from the prior views. We believe it is important to include mammograms that were recalled for technical reasons because there is no reason to expect recall for technical reasons to differ across randomization group unless cessation had an effect on the ability to get a technically adequate image including ability to sufficiently compress the breast or to get clear images. Time between mammograms was determined in part by a woman's recommended screening interval (determined by her risk profile)(37) and her choice (median time between index and study mammograms=720 days).

**Breast density (secondary outcome)—**We digitized the left breast craniocaudal projection from the index and study mammograms using a Kodak Lumisys 85 scanner. We used Cumulus software(38) (University of Toronto) to measure dense area and total breast area measured in pixels. Small films were scanned at 87 microns/pixel and large films at 116 microns/pixel and were converted to $cm^2$ for dense area and breast area as follows: $7.554\times10^{-5}cm^2$/pixel (small films) and $1.350\times10^{-4}cm^2$/pixel (large films). Percent breast density was computed as the ratio of dense area to total breast area. All breast density measures were interpreted by the same trained reader (EAB) in batches of 50 films each. Participant's index and study films were read in random order within the same batch of films, with the reader blinded to participant identifiers, randomization group, and the timing of the exam. To evaluate quality assurance, we included four intra-batch repeats within each batch. The mean absolute difference (standard error, SE) in quality control samples was 3.3% (0.2%) for percent breast density, 4.7 (0.4) $cm^2$ for dense area; and 2.5 (0.2) $cm^2$ for breast area. The concordance correlation between intra-batch repeats was 0.9 for percent breast density, 0.95 for dense area, and 1.0 for breast area.

## Follow-up Procedures and Monitoring

The self-administered follow-up study questionnaire was completed before the scheduled mammogram. The questionnaire asked about tolerance of and adherence to HT suspension, menopausal symptoms, and specific adverse events, but did not address any questions related to anxiety related to mammographic procedures, possible recall or mammographic findings. Adherence was based on self-report from the follow-up questionnaire. Women in the 1-and 2-month suspension groups were asked "Were you able to stay off your HT as instructed for this study?" Women who responded "yes" were considered adherent. Adherence among the no-suspension group was determined from a "yes" response to the question "Did you take any estrogen in the last month?" We could not assign adherence status for the 3.8% of women who did not complete these questions.

**Adverse experiences—**We collected adverse experience information, from recruitment to 12 months following study mammogram, for serious adverse events (myocardial infarction, angina, stroke, coronary bypass surgery, coronary angioplasty/stent, pulmonary embolism, deep vein thrombosis, or death), anticipated adverse events (depression, dizziness or faint spells, palpitations, or vaginal bleeding), and unexpected adverse events (any event not previously listed). Adverse experiences were collected from: the study hotline, the follow-up questionnaire, and monthly reviews of Group Health automated data using International Classification of Diseases, 9th Revision and Current Procedural Terminology codes. The study physician (SDR) and a designated medical monitor adjudicated all potential events identified. All serious adverse events were also submitted to the Data Safety and Monitoring Board and Group Health's and the U.S. Department of Defense's IRBs. The Data Safety and Monitoring Board met four times and monitored serious adverse events, anticipated adverse events, and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

unexpected adverse events; interim analyses were examined at two meetings to determine if stopping guidelines had been met.

### Statistical analysis

In designing our study, we estimated study power based on preliminary data demonstrating a 13% mammography recall rate among women using HT.(19) Enrolling 1500 women to three equal size groups provided 85% power to detect a 5% decrease in mammography recall rate (8% mammography recall in the combined suspension groups).

Chi-square tests, Kruskall-Wallis tests, and analysis of variance were used to compare demographic and potential confounding variables among the three groups. We used generalized linear models with a log link to estimate the relative risk (RR) and 95% confidence intervals (CI) of mammography recall for women in the suspension groups relative to the no-suspension group. We used linear regression to compare the mean change from index to study mammograms in breast density across study groups. All statistical analyses were completed using STATA, v.10.0 (StataCorp. College Station, TX).

All main analyses were based on a modified intention to treat analysis, where group differences were estimated based on randomized group assignment, regardless of adherence. Since mammography recall and breast density (the primary and secondary outcomes) could only be assessed among women completing their mammogram, all analyses were limited to women who did not withdraw from the study before their study mammogram, making a strict intention to treat analysis impossible. We report unadjusted results, because sensitivity analyses adjusted for covariates (body mass index, age, and HT type) did not alter the findings. We repeated the main analysis limited to women who adhered to the study protocol.

To investigate effect modification, we specified *a priori* four participant characteristics for subgroup analyses that we thought could alter the association between HT suspension and the study outcomes: age at study enrollment; duration of HT use at baseline; and estrogen and progestin strength. The associations between mammography recall and brief suspension in these subgroups were estimated by including an interaction term between randomization group and the subgroup variables in the multivariable regression models.

## RESULTS

We approached 5861 women to assess their eligibility and to invite them to participate in the trial. We excluded 4157 women: 2999 (72.1%) refused; 977 (23.5%) were determined ineligible after initial contact (e.g., no longer taking HT, heart disease history, had recent mammogram); 179 (4%) could not be contacted; and 2 had unknown reasons for exclusion. Most women who refused participation (58.4%) were unwilling to stop HT, even briefly (Figure 1). Among the 4884 potentially eligible women invited, 65.1% declined participation. There were a few small differences between women who refused compared with those who participated. Women who refused were older than participants (mean 61.3 vs. 60.3 years, $p<.001$), leaner (mean=27.1 vs. 27.6 kg/m$^2$, $p<.002$), less likely to be Caucasian (89.6% vs. 91.3%, $p=<.001$), and less educated (≤high school 19.2% vs. 14.0%, $p<.001$).

We randomly assigned 1704 women to the three study arms: 567 no suspension, 570 1-month suspension, and 567 2-month suspension. A total of 228 (13.4%) randomly assigned women withdrew from the study. Withdrawal differed across groups: 24/567 (4.2%) no suspension; 90/570 (15.8%) 1-month suspension; and 114/567 (20.1%) 2-month suspension. Reasons for withdrawal varied, but the majority of women in the suspension groups withdrew because they refused to discontinue HT after receiving their randomization assignment. Nearly one third of the withdrawals in the no-suspension group (7/24) withdrew because they discontinued HT

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

before their mammogram. Despite differential withdrawal by randomization group, there were no statistically significant characteristics of women who withdrew after randomization compared with women who completed participation: age ($p$=.81), HT type ($p$=.50), body mass index ($p$=.181), race ($p$=.29), education ($p$=.21), breast cancer risk ($p$=.32), family history ($p$=.25), HT duration ($p$=.71) or breast density ($p$=.25). We had complete study data including mammogram assessments and both questionnaires for 95.7% of the 1476 enrolled women who remained in the study. An additional five women (one no-suspension, two 1-month suspension, and two 2-month suspension) were excluded from the primary analysis because they did not complete their study mammogram or their film was not available for assessment by the study radiologist. This resulted in a final analytic sample size of 1471.

The mean age at recruitment among enrolled women was 60.3 years, mean body mass index 27.6 kg/m$^2$, and 91.5% of the study population was Caucasian (Table 1). There were no statistically significant differences in age, body mass index, race, ethnicity, education, or breast cancer risk across randomization groups. Most of the randomly assigned women (62.0%) used unopposed estrogen, and nearly half had previously tried to quit HT. Approximately one third of the women self-reported having hypertension, high cholesterol, and/or depression. Health status and behaviors did not differ across randomization groups. Self-reported health status was high across all groups, with <10% reporting fair or poor health (data not shown). Most women were non-smokers (56.5% never, 37.2% former) and were not heavy alcohol drinkers (28.9% never-drinkers and 14.1% drinking daily, with 70.9% of drinkers reporting having 1 drink/day on days they drank).

Adherence to random assignment was high across randomization arms (99.0% no suspension, 92.7% 1-month, and 87.3% 2-month); but was statistically significantly lower ($p$<.006) in the 2-month suspension group compared to the 1-month group. There were no crossovers designed in this study. Women in the suspension groups were asked to report all reasons for nonadherence, and their most frequent responses included hot flashes (70.5%), night sweats (65.9%), sleep disturbances (65.9%), and “just feeling better” taking HT (64.8%).

## Mammography recall

Just over one in 10 women (11.2%) were recalled for additional mammography imaging with the largest proportion of mammography recall occurring in the women randomly assigned to 1-month suspension (12.3%) followed by no suspension (11.3%) and 2-month suspension (9.8%) ($p$=.45) (Table 2). Recall rates did not change when limited to women adherent with randomization ($p$=.39). Reason for recall included technical (7.3%), work-up of a probably benign lesion (84.1%), work-up of a possible malignant lesion (6.1%) and other (reason not specified, 3.1%). Technical reasons were indicated as the reason for recall for 3.3% in the no suspension, 11.9% in 1-month suspension and 6.8% in 2-months suspension. After excluding the women who had a recall for technical reasons from the numerator and denominator, the pattern of recall remained the same by randomization group: 1-month suspension (11.0%) followed by no suspension (10.9%) and 2-month suspension (9.2%) ($p$=.57).

Recall patterns varied by HT strata. ET users had a statistically non-significant ($p$=.30) lower recall (9.7%) than EPT users (13.5%). There were no statistically significant differences in recall across randomization groups for ET ($p$=.42) or EPT users ($p$=.80). Among ET users, there was a suggestion that women in the 1-month suspension group had a higher recall rate compared to the no-suspension group. Among EPT users, recall rates had statistically non-significant drops from 14.5% among continued users to 12.1% for women who stopped EPT for 2 months.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Pre-specified effect modification analyses by age at recruitment, HT use duration, and HT strength identified no subgroup with statistically significant decreases in mammography recall (data not shown).

## Mammographic breast density

Recall rates increased with increasing amounts of dense area on mammograms but were affected little by percent density (Table 3). After adjusting for body mass index and age, there was an increase in the relative risk of recall with increasing mammographic breast density (percent and dense area), but there was no relation between change in density and recall rates.

There were ordered and statistically significant decreases in percent breast density and dense area with 1 and 2 months of HT suspension (Table 4); decreases in density were greater with longer suspension. Percent breast density decreased in both suspension groups, with a mean (SE) change of -0.9% (0.3%) in the 1-month suspension group and -1.5% (0.4%) in the 2-month group, with no change [0.1% (0.3%)] in the no-suspension group. Comparable ordered declines were observed for dense area. Declines in percent breast density and dense area were larger after excluding non-adherent women.

When stratified by HT type, changes in percent breast density and dense area shared the same pattern as for overall users, but statistically significant differences were limited to EPT users with short-term suspension with mean decreases in percent breast density of −1.3% and −2.2% for 1-and 2-month suspension, compared to no change (0.6%, p=.009 1 month vs. no suspension, p<.0001 2 months vs. no suspension) in the no-suspension group.

In our sub-analyses by age, duration of HT use before index mammogram, and estrogen and progestin strength, we observed the greatest declines in percent breast density and dense area among women <60 years and in women using standard progestin dose (Figure 2). Virtually every subgroup analysis showed ordered decreases in breast density with increasing HT suspension. We observed a nonsignificant increase in mammography recall for women on progestin doses of medroxyprogesterone acetate 2.5 mg compared to higher doses. Although we did not assess continuous vs. cyclical patterns, medroxyprogesterone acetate 2.5 mg is almost never given in a cyclical pattern and is the most common progestin dose for a continuous regimen.

## Adverse events and symptoms

Eight serious adverse events occurred among randomized participants (through 7/18/08): two myocardial infarctions (no-suspension), two strokes (one no-suspension, one 2-month), and four deaths (two no-suspension, one 1-month, one 2-month) (data not shown).The no-suspension group had the most unanticipated events: three diagnoses of breast cancer, two women with atypical chest pain, and eight other events (e.g., fracture/fall, laceration). Six unanticipated events occurred among women in the 1-month group, but none occurred in the 2-month group. All women diagnosed with breast cancer adhered to randomization. Of the women diagnosed with breast cancer, all three in the no-suspension group used EPT, and one in the 1-month suspension group used ET until 1 month before her mammogram.

Vaginal bleeding was the most common anticipated adverse event, occurring in seven women in 1-month, six in 2-month, and three in no-suspension. Hot flashes, sleep disturbances and night sweats were the symptoms that increased the most in the suspension groups, with the greatest increase in symptoms in the 2-month suspension group (Table 5). None of our measured menopausal symptoms improved in either suspension group.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

## DISCUSSION

We have demonstrated that short-term suspension of ET or EPT for 1 or 2-months is associated with small changes in breast density and that suspension does affect mammography recall rates. Although most women who agreed to participate in our study tolerated short-term HT suspension, suspension was associated with statistically significant increases in menopausal symptoms for women using ET or EPT.

No large-scale randomized trials have measured how quickly density changes with HT suspension. Moreover, no studies have quantified the amount of density reduction needed to improve mammography performance (recall, sensitivity, specificity or overall accuracy). Theoretically, even small decreases in breast density and recall have the potential to have a clinically meaningful impact on screening performance in larger groups. However, no studies have quantified the amount of decrease in continuous mammographic breast density (percent or area) needed to reduce mammography recall.

We designed our intervention to test 1-and 2-month HT suspension because we believed it was impractical for women to suspend HT for longer periods between annual and biennial screening mammograms. Previous studies(11,18,19) have demonstrated clinically and statistically significant improvements in mammography recall with HT suspension, but those studies were not randomized and were based on longer periods of HT suspension. The majority (61%) of eligible women approached to participate in this trial refused because they were unwilling to consider short-term HT suspension, even if short-term suspension might improve mammography quality. There were no clinical or statistical differences in breast density or HT type or duration for women who refused compared to women who participated. Among women who agreed to participate in the trial, 13% withdrew after randomization; there were also no clinical or statistical differences between women who agreed to participate and withdrew after randomization to women who remained in the study. Our study was designed to determine the effectiveness of short-term HT suspension, making the high proportion of refusals an important finding of this study.

Linear trends of increased recall were seen with increasing percent density and increasing dense area at the study mammogram, with slightly stronger effects for dense area. However, these trends were not observed in relation to change in percent density or change in dense area and recall. These findings emphasize that density is an important factor influencing recall rates, but that it is likely that change in density needs to be large to be able to have any clinical affect on recall rates.

Percent density changes in our study are similar in magnitude to changes observed during a menstrual cycle(14) and the menopause transition.(39) They are also comparable to other smaller studies measuring change in breast density over longer (12-month) periods. The exception is the Postmenopausal Estrogen/Progestin Intervention (PEPI) Trial, a randomized trial including 571 women aged 45–64(40), designed to examine differences in heart disease risk factors by ET and EPT vs. placebo. PEPI reported some of the largest mean changes in percent breast density over a 12-month period, with breast density increasing from 3.1% to 4.8% for women in the EPT groups and decreasing by −0.1% in the placebo group.(40) Additionally, one small (n=48) observational study found a reduction in density from 2 weeks of HT suspension, and suggested brief suspension could improve mammography performance.(20) However, this study was limited to women with new or enlarging circumscribed breast masses, which excluded abnormalities that were unlikely to be related to HT use. Our current study was not designed to examine pattern changes (e.g., masses/calcifications) in density in response to HT suspension, but this may be an important area for future examination.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Assessing clinically meaningful changes in breast density is challenging for continuous and BI-RADS density measures.(41) Large changes in density are needed to shift BI-RADS categories if radiologists use the recommended distribution of 25% density in each BI-RADS category.(30) Moving one BI-RADS breast density category amounts to a 14%-18% change in breast density.(40) Boyd(42) previously estimated the relative risk of breast cancer increases by 2% for each 1% increase in percent breast density when measured continuously. However, the clinical significance of changing breast density remains unknown, particularly in the subgroup of women for whom HT increases breast density. Understanding characteristics of women whose breast density changes with HT and other exogenous exposures, including which breast-level changes result, may be critical to understanding hormone exposures in breast cancer development and detection.

Several small randomized trials—and one large study nested in a randomized trial—have assessed the effects on density of various HT formulations, including dose, drug type, application, and duration of use.(43–48) Together, these studies suggest estrogens and progestins increase density in some women, with a differential effect by route and dose. Our study was insufficiently large to estimate recall and density changes by HT dose, mode (continuous sequential vs. cyclical EPT), route, or type (synthetic vs. natural. Interestingly, we observed a suggestion of higher mammography recall for women using medroxyprogesterone acetate 2.5 mg (commonly used in continuous regimens) than for women using higher progestin doses (used cyclically), suggesting that suspension of continuous EPT may increase recall rates, although full interpretation of these findings is limited by a small sample size. In vitro studies of breast tissue suggest continuous progestin inhibits cell growth.(49,50) If continuous progestins have an anti-proliferative effect, it is possible that upon withdrawal, particularly in the month following suspension, breast tissue may experience rebound proliferation leading to an increase in recall.

We recruited a population of women who were due for a screening mammogram for whom we had access to index and study mammogram images within our clinical system. Electronic pharmacy records enabled targeted recruitment to women who were still receiving HT dispensings and to block-randomize women by HT type (ET vs. EPT). Automated pharmacy data allowed evaluation of outcome differences by HT type, but we were unable to examine differences by HT mode, although inferences based on progestin dose can be made.

We designed this study to minimize measurement error including having one breast density interpreter; using one computer, monitor, and scanner; ensuring distribution of women across randomization groups in each breast density batch; and including intra-and inter-rater quality control readings in each batch. Despite all these controls, the variability in quantitative density assessments is important, and can limit the ability to detect change.(41) The standard error of the mean change in percent breast density ranged from 0.3–0.4%. This resembles other studies that have reported measurement error in mammographic breast density.

There is a tremendous literature on variability in mammography interpretive performance (23,51–54). We purposefully used an expert radiologist to interpret all study mammograms in the same setting to decrease recall variability that could arise from including multiple radiologists.(23,51–54) Having recall defined by one expert is a strength of the study, which increases internal validity by reducing variation in recall due to inter-reader variation, but it could also decrease the generalizability of the findings if other radiologists have different patterns of recall.(55)

Our results are generalizable to women aged 45–80 years who have used HT for at least 1 year; nearly half our study population had used HT for ≥15 years. With no evidence of an interaction between duration of EPT or ET use and mammography recall, we have no reason to believe

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

that results from this trial cannot be generalized to women who have shorter-term HT use than the women in our study. We found no evidence of a benefit for short-term HT suspension in any age groups, including women over 70 years; an age group been previously demonstrated to have the greatest reduction in recall from HT suspension.(18,19)

Any intervention that can decrease mammography recall has important public health consequences regardless of the intervention effect size; false-positive mammograms account for 25% of the overall costs of US mammography.(56) In the US population in 2005, Medicare reimbursed 11.4 million screening mammograms among non-HMO women between ages 50 and 79 years.(57) Finding interventions to decrease recall for additional imaging without compromising sensitivity is critical to the sustainability of breast cancer detection in the US. Results from this large, population-based randomized trial provide evidence that brief HT suspension before screening mammography does not decrease mammography recall for any subgroups of women using HT, despite decreases in breast density. Our results also demonstrate a substantial negative impact on women from increased menopausal symptoms. For breast cancer detection and prevention, we need to understand whether subsets of women whose breast density changes in response to endogenous exposures (e.g., parity, cyclical changes) and exogenous exposures (e.g., various HT types and doses, tamoxifen, aromatase inhibitors) may benefit differentially from interventions designed to decrease mammographic breast density.

## Acknowledgments

**Role of funding source**

The READ study was funded by the Department of Defense (PI, D Buist; DAMD17-03-1-0447). Registered clinical trial number: NCT00117663. Study participants were recruited from the Group Health Breast Cancer Screening Program funded by the National Cancer Institute (PI: D Buist, U01CA63731). The study sponsors had no involvement in the: study design; collection, analysis, and interpretation of the data; in the writing of the report; and in the decision to submit the paper for publication.

The study team had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. This study could not have been completed without the assistance of Tammy Dodd, Linda Palmer, RN, BS or Melissa Rabelhofer. We would further like to thank members of our advisory board: Hermien Watkins, Paula Hoffman, Deb Schiro, and Margrit Schubiger; members of the Data Safety and Monitoring Board: Susan Heckbert, MD, PhD, Chair, University of Washington Department of Epidemiology; Ben Anderson, MD, University of Washington; Mary Anne Rossing, DVM, PhD, Fred Hutchinson Cancer Research Center; Robert D. Rosenberg, MD, University of New Mexico; Thomas Lumley, PhD, University of Washington Department of Biostatistics; and Elizabeth Lin, MD, Group Health Permanente medical monitor. We would further like to thank Robert Karl, MD, Donna White, MD and Jo Ellen Callahan for their support for implementing this trial at Group Health. We also acknowledge Stephen Taplin, MD, MPH for his collaboration in getting this study funded when he was an investigator at Group Health Cooperative. We thank Rebecca Hughes for her editorial assistance.

Drs. Reed and Newton receive grant support from Pfizer.

**Grant support:**

A population-based trial to assess the effects of brief hormone therapy (HT) suspension on mammography assessments and breast density". The READ study was funded by the Department of Defense (PI, D Buist; DAMD17-03-1-0447). Registered clinical trial number: NCT00117663. Study participants were recruited from the Group Health Breast Cancer Screening Program funded by the National Cancer Institute (PI: D Buist, U01CA63731).

## References

1. American Cancer Society. Cancer Facts and Figures 2008. 2008
2. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;137(5 Part 1):347–360. [PubMed: 12204020]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript


3. National Cancer Institute BCSC. Abnormal Interpretations for 4,032,556 Screening Mammography Examinations from 1996–2005. 2007
4. Brown ML, Houn F, Sickles EA, Kessler LG. Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. Am J Roentgenol 1995;165(6):1373–1377. [PubMed: 7484568]
5. Barton MB, Moore S, Polk S, Shtatland E, Elmore JG, Fletcher SW. Increased patient concern after false-positive mammograms: clinician documentation and subsequent ambulatory visits. J Gen Intern Med 2001;16(3):150–156. [PubMed: 11318909]
6. Randal J. For breast cancer screening, there is room for improvement, report concludes. J Natl Cancer Inst 2004;96(16):1200–1201. [PubMed: 15316051]
7. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003;138(3):168–175. [PubMed: 12558355]
8. Ma L, Fishell E, Wright B, Hanna W, Allan S, Boyd NF. Case-control study of factors associated with failure to detect breast cancer by mammography. J Natl Cancer Inst 1992;84(10):781–785. [PubMed: 1573665]
9. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval-and screen-detected cancers. J Natl Cancer Inst 2000;92(13):1081–1087. [PubMed: 10880551]
10. Byrne C, Schairer C, Wolfe J, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. J Natl Cancer Inst 1995;87(21):1622–1629. [PubMed: 7563205]
11. Laya MB, Gallagher JC, Schreiman JS, Larson EB, Watson P, Weinstein L. Effect of postmenopausal hormonal replacement therapy on mammographic density and parenchymal pattern. Radiology 1995;196(2):433–437. [PubMed: 7617857]
12. Rosenberg RD, Hunt WC, Williamson MR, et al. Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183,134 screening mammograms in Albuquerque, New Mexico. Radiology 1998;209(2):511–518. [PubMed: 9807581]
13. Laya MB, Larson EB, Taplin SH, White E. Effect of estrogen replacement therapy on the specificity and sensitivity of screening mammography. J Natl Cancer Inst 1996;88(10):643–649. [PubMed: 8627640]
14. Buist DS, Aiello EJ, Miglioretti DL, White E. Mammographic breast density, dense area, and breast area differences by phase in the menstrual cycle. Cancer Epidemiol Biomarkers Prev 2006;15(11):2303–2306. [PubMed: 17119062]
15. White E, Velentgas P, Mandelson M, et al. Variation in mammographic breast density by time in menstual cycle among women aged 40–49 years. J Natl Cancer Inst 1998;90(12):906–910. [PubMed: 9637139]
16. Ursin G, Parisky YR, Pike MC, Spicer DV. Mammographic density changes during the menstrual cycle. Cancer Epidemiol Biomarkers Prev 2001;10(2):141–142. [PubMed: 11219771]
17. Baines CJ, Vidmar M, McKeown-Eyssen G, Tibshirani R. Impact of menstrual phase on false-negative mammograms in the Canadian National Breast Screening Study. Cancer 1997;80(4):720–724. [PubMed: 9264355]
18. Boudreau DM, Buist DS, Rutter CM, Fishman PA, Beverly KR, Taplin S. Impact of hormone therapy on false-positive recall and costs among women undergoing screening mammography. Med Care 2006;44(1):62–69. [PubMed: 16365614]
19. Rutter CM, Mandelson MT, Laya MB, Seger DJ, Taplin S. Changes in breast density associated with initiation, discontinuation, and continuing use of hormone replacement therapy. JAMA 2001;285(2):171–176. [PubMed: 11176809]
20. Harvey JA, Pinkerton JV, Herman CR. Short-term cessation of hormone replacement therapy and improvement of mammographic specificity. J Natl Cancer Inst 1997;89(21):1623–1625. [PubMed: 9362162]
21. Speroff L. The impact of the Women's Health Initiative on clinical practice. J Soc Gynecol Investig 2002;9(5):251–253.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript


22. Burnside ES, Trentham-Dietz A, Kelcz F, Collins J. An example of breast cancer regression on imaging. Radiology Case Reports 2006;1:27–37.
23. Institute of Medicine. Improving Breast Imaging Quality Standards. Washington, D.C: The National Academies Press; 2005.
24. Smith-Bindman R, Chu PW, Miglioretti DL, et al. Comparison of screening mammography in the United States and the United kingdom. JAMA 2003;290(16):2129–2137. [PubMed: 14570948]
25. Elmore JG, Nakano CY, Koepsell TD, Desnick LM, D'Orsi CJ, Ransohoff DF. International variation in screening mammography interpretations in community-based programs. J Natl Cancer Inst 2003;95(18):1384–1393. [PubMed: 13130114]
26. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of mammographic accuracy. J Natl Cancer Inst 2005;97(5):358–367. [PubMed: 15741572]
27. Hofvind S, Vacek PM, Skelly J, Weaver DL, Geller BM. Comparing screening mammography for early breast cancer detection in Vermont and Norway. J Natl Cancer Inst 2008;100(15):1082–1091. [PubMed: 18664650]
28. Taplin SH, Thompson RS, Schnitzer F, Anderman C, Immanuel V. Revisions in the risk-based Breast Cancer Screening Program at Group Health Cooperative. Cancer 1990;66(4):812–818. [PubMed: 2386908]
29. Taplin SH, Ichikawa L, Buist DS, Seger D, White E. Evaluating organized breast cancer screening implementation: the prevention of late-stage disease? Cancer Epidemiol Biomarkers Prev 2004;13 (2):225–234. [PubMed: 14973097]
30. American College of Radiology. Breast imaging reporting and data system (BIRADS™). Vol. 3rd ed.. Reston, VA: American College of Radiology; 1998.
31. Lee HP, Gourley L, Duffy SW, Esteve J, Lee J, Day NE. Risk factors for breast cancer by age and menopausal status: a case-control study in Singapore. Cancer Causes Control 1992;3(4):313–322. [PubMed: 1617118]
32. Wiklund I, Holst J, Karlberg J, et al. A new methodological approach to the evaluation of quality of life in postmenopausal women. Maturitas 1992;14(3):211–224. [PubMed: 1508061]
33. Wiklund I, Karlberg J. Evaluation of quality of life in clinical trials. Selecting quality-of-life measures. Control Clin Trials 1991;12(4 Suppl):204S–216S. [PubMed: 1663856]
34. Buist DS, Newton KM, Miglioretti DL, et al. Hormone therapy prescribing patterns in the United States. Obstet Gynecol 2004;104(5 Pt 1):1042–1050. [PubMed: 15516400]
35. North American Menopause Society. Menopause Core Curriculum Study Guide. Vol. 2nd ed.. Cleveland, OH: North American Menopause Society; 2002.
36. A new conjugated estrogen. Med Lett Drugs Ther 1999;41(1058):67–68. [PubMed: 10445019]
37. Taplin SH, Ichikawa L, Yood MU, et al. Reason for late-stage breast cancer: absence of screening or detection, or breakdown in follow-up? J Natl Cancer Inst 2004;96(20):1518–1527. [PubMed: 15494602]
38. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. Automated analysis of mammographic densities. Phys Med Biol 1996;41(5):909–923. [PubMed: 8735257]
39. Kelemen LE, Pankratz VS, Sellers TA, et al. Age-specific trends in mammographic density: the Minnesota Breast Cancer Family Study. Am J Epidemiol 2008;167(9):1027–1036. [PubMed: 18385204]
40. Greendale GA, Reboussin BA, Slone S, Wasilauskas C, Pike MC, Ursin G. Postmenopausal hormone therapy and change in mammographic density. J Natl Cancer Inst 2003;95(1):30–37. [PubMed: 12509398]
41. Byrne C. Invited commentary: assessing breast density change--lessons for future studies. Am J Epidemiol 2008;167(9):1037–1040. [PubMed: 18385205]
42. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ. Mammographic densities and breast cancer risk. Cancer Epidemiol Biomarkers Prev 1998;7(12):1133–1144. [PubMed: 9865433]
43. Marchesoni D, Driul L, Ianni A, et al. Postmenopausal hormone therapy and mammographic breast density. Maturitas 2006;53(1):59–64. [PubMed: 15939560]
44. Christodoulakos GE, Lambrinoudaki IV, Vourtsi AD, et al. The effect of low dose hormone therapy on mammographic breast density. Maturitas 2006;54(1):78–85. [PubMed: 16198519]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript


45. Erel CT, Esen G, Seyisoglu H, et al. Mammographic density increase in women receiving different hormone replacement regimens. Maturitas 2001;40(2):151–157. [PubMed: 11716993]
46. Junkermann H, von Holst T, Lang E, Rakov V. Influence of different HRT regimens on mammographic density. Maturitas 2005;50(2):105–110. [PubMed: 15653007]
47. Lundstrom E, Wilczek B, von Palffy Z, Soderqvist G, von Schoultz B. Mammographic breast density during hormone replacement therapy: differences according to treatment. Am J Obstet Gynecol 1999;181(2):348–352. [PubMed: 10454681]
48. Harvey J, Scheurer C, Kawakami FT, Quebe-Fehling E, de Palacios PI, Ragavan VV. Hormone replacement therapy and breast density changes. Climacteric 2005;8(2):185–192. [PubMed: 16096175]
49. Lange CA, Richer JK, Horwitz KB. Hypothesis: Progesterone primes breast cancer cells for cross-talk with proliferative or antiproliferative signals. Mol Endocrinol 1999;13(6):829–836. [PubMed: 10379882]
50. Groshong SD, Owen GI, Grimison B, et al. Biphasic regulation of breast cancer cell growth by progesterone: role of the cyclin-dependent kinase inhibitors, p21 and p27(Kip1). Mol Endocrinol 1997;11(11):1593–1607. [PubMed: 9328342]
51. Beam CA, Guse C, Sullivan D. A sequential chart for the audit-based evaluation of screening mammogram interpretation. Acad Radiol 1999;6:216–223. [PubMed: 10894079]
52. Beam CA, Layde P, Sullivan D. Variability in the interpretation of screening mammograms by US radiologists. Arch Intern Med 1996;156:209–213. [PubMed: 8546556]
53. Elmore JG, Wells C, Lee C, Howard D, Feinstein A. Variability in radiologists' interpretations of mammograms. N Engl J Med 1994;331:1493–1499. [PubMed: 7969300]
54. Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology breast imaging reporting and data system. J Natl Cancer Inst 1999;90(23):1801–1809. [PubMed: 9839520]
55. Taplin SH, Rutter CM, Lehman CD. Testing the effect of computer-assisted detection on interpretive performance in screening mammography. AJR Am J Roentgenol 2006;187(6):1475–1482. [PubMed: 17114540]
56. Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 1998;338(16):1089–1096. [PubMed: 9545356]
57. Centers for Medicare & Medicaid Services. In the Medicare population in the year 2005, 11.4 million non-HMO women between the ages of 50 to 79 years of age received a reimbursed screening mammogram. Centers for Medicare & Medicaid Services 2005;Vol. 2008
58. Reed SD, Newton KM, LaCroix AZ, Grothaus LC, Ehrlich K. Night sweats, sleep disturbance, and depression associated with diminished libido in late menopausal transition and early postmenopause: baseline data from the Herbal Alternatives for Menopause Trial (HALT). Am J Obstet Gynecol 2007;196(6):593 e1–593 e7. discussion 593 e7. [PubMed: 17547910]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



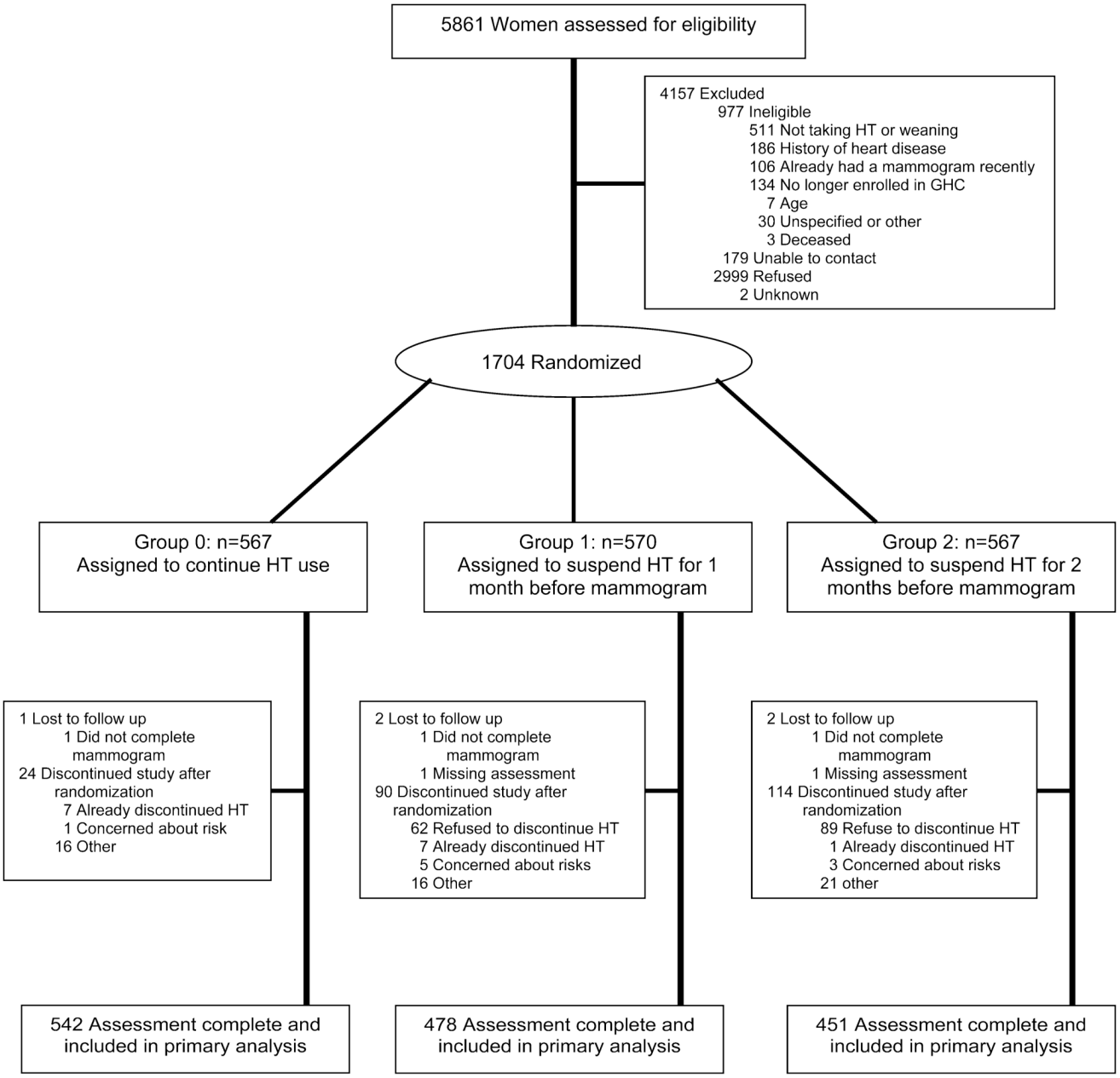


**Figure 1.**
Recruitment for READ Study

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

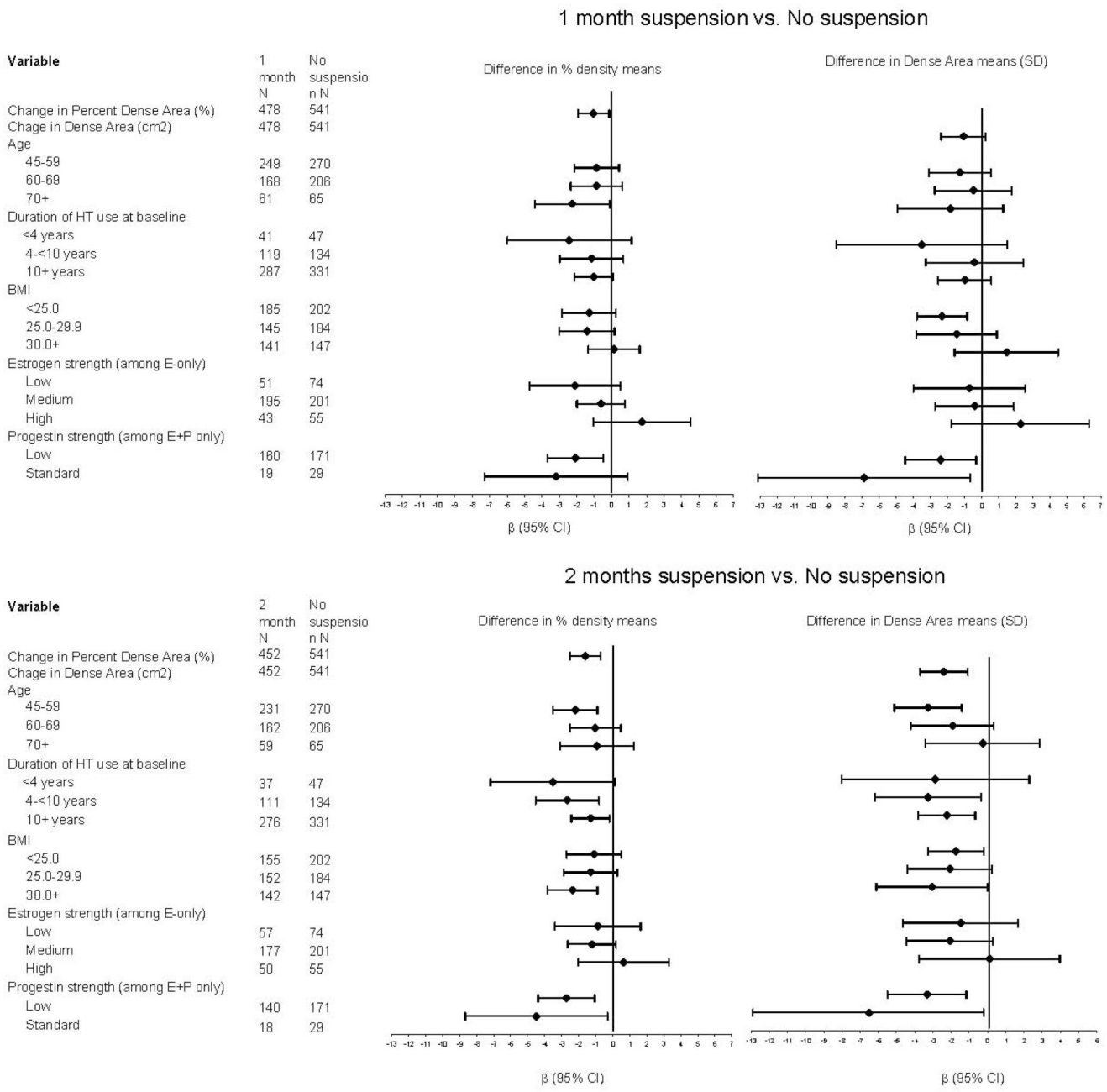


**Figure 2.**

Difference Between Index and Study Mammogram in Means of Percent Breast density and Dense Area for 1-Month Suspension vs. No-Suspension and 2-Month Suspension vs. No-Suspension, by Age, Duration of Hormone Therapy Use, Body Mass Index, and Estrogen and Progestin strength

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



## Table 1

Baseline Characteristics of READ Study Participants by Randomization Assignment

| Characteristics | No-suspension (n = 567) | 1-month suspension (n = 570) | 2-month suspension (n = 567) |
|---|---|---|---|
| Age at study recruitment, y | | | |
| 44–50 | 34 (6.0) | 41 (7.2) | 35 (6.2) |
| 51–55 | 118 (20.8) | 111 (19.5) | 121 (21.3) |
| 56–60 | 151 (26.6) | 178 (31.2) | 175 (30.9) |
| 61–65 | 125 (22.1) | 113 (19.8) | 117 (20.6) |
| 66–70 | 82 (14.5) | 68 (11.9) | 59 (10.4) |
| 71–81 | 57 (10.1) | 59 (10.4) | 60 (10.6) |
| Mammographic breast density | | | |
| Scattered fibroglandular | 194 (34.2) | 196 (34.4) | 194 (34.2) |
| Hetereogeneously dense | 315 (55.6) | 315 (55.3) | 315 (55.6) |
| Extremely dense | 58 (10.2) | 59 (10.4) | 58 (10.2) |
| Type of hormone therapy | | | |
| Estrogen + progestin | 215 (37.9) | 217 (38.1) | 215 (37.9) |
| Unopposed estrogen | 352 (62.1) | 353 (61.9) | 352 (62.1) |
| Body mass index kg/$m^2$, index mammogram | | | |
| Mean (SD) | 27.5 (5.7) | 27.7 (6.0) | 27.6 (5.6) |
| Range | 17.4, 51.1 | 17.5, 49.5 | 17.7, 51.9 |
| Race | | | |
| Caucasian | 525 (92.8) | 516 (90.9) | 511 (90.8) |
| African American | 12 (2.1) | 12 (2.1) | 13 (2.3) |
| Asian | 8 (1.4) | 11 (1.9) | 14 (2.5) |
| Other | 21 (3.7) | 29 (5.1) | 25 (4.4) |
| Hispanic | | | |
| No | 544 (98.6) | 540 (97.0) | 536 (96.8) |
| Yes | 8 (1.5) | 17 (3.1) | 18 (3.3) |
| Time between index and study mammogram (days) (n=1471) | | | |
| Mean (SD) | 620.4 (165) | 618.4 (158) | 618.6 (163) |
| Range | 292, 1232 | 339, 875 | 349, 1234 |
| Education | | | |
| ≤High school | 80 (14.3) | 74 (13.2) | 82 (14.6) |
| Some college | 243 (43.4) | 219 (39.0) | 223 (39.7) |
| College graduate | 99 (17.7) | 127 (22.6) | 112 (19.9) |
| ≥Graduate school | 138 (24.6) | 142 (25.3) | 145 (25.8) |
| Breast cancer risk§ | | | |
| Low | 486 (85.7) | 484 (84.9) | 468 (82.5) |
| Higher | 81 (14.3) | 86 (15.1) | 99 (17.5) |
| **Information from baseline questionnaire** | | | |
| Women completing baseline questionnaire | 540 (95.2) | 485 (85.1) | 477 (84.1) |

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

| Characteristics | No-suspension (n = 567) | 1-month suspension (n = 570) | 2-month suspension (n = 567) |
|---|---|---|---|
| Duration of prior HT use at baseline | | | |
| Mean (SD), years | 14.1 (9.4) | 14.1 (9.2) | 13.6 (8.9) |
| Categorical | | | |
| <5 years | 70 (13.4) | 62 (13.3) | 59 (12.9) |
| 5–9 years | 113 (21.6) | 103 (22.1) | 102 (22.3) |
| 10–14 years | 113 (21.6) | 107 (22.9) | 110 (24.1) |
| 15+ years | 228 (43.5) | 195 (41.8) | 186 (40.7) |
| Days per month taking estrogen at baseline | | | |
| <15 days | 12 (2.3) | 10 (2.1) | 16 (3.4) |
| 15–29 days | 58 (10.9) | 55 (11.5) | 46 (9.8) |
| Daily | 464 (86.9) | 414 (86.4) | 408 (86.8) |
| Ever tried to quit HT previously | | | |
| No | 275 (51.8) | 268 (56.2) | 244 (51.9) |
| Yes | 256 (48.2) | 209 (43.8) | 226 (48.1) |
| Health history | | | |
| Hypertension | 204 (38.4) | 185 (38.8) | 176 (37.9) |
| High cholesterol | 160 (31.1) | 146 (31.3) | 147 (32.4) |
| Diabetes | 34 (6.5) | 46 (9.6) | 20 (4.3) |
| Bone fracture after age 44 or kyphosis | 74 (14.0) | 55 (11.7) | 66 (14.3) |
| Depression | 192 (37.0) | 170 (36.0) | 157 (34.1) |
| Last episode of vaginal bleeding | | | |
| ≤12 months ago | 57 (10.6) | 57 (13.0) | 46 (10.0) |
| 1–5 years ago | 78 (14.8) | 68 (14.4) | 58 (12.6) |
| >5 years ago | 393 (74.6) | 348 (73.6) | 357 (77.4) |

*Data are presented as a number (percentage) of women unless otherwise noted.

‡Women with entirely fatty breasts were ineligible to participate

§Low risk = recommended 2-year screening interval, High risk = recommended annual screening

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript



**Table 2**

Mammography Recall Rates by Randomization Group, Stratified by Type of Hormone Therapy

| Characteristics | Mammography recall by Study Group | | | | Relative Risk of Mammography recall | |
|---|---|---|---|---|---|---|
| | No-suspension n = 542 | 1-month suspension n = 478 | 2-months suspension n = 451 | *p*-Value | 1-month vs. No-suspension | 2-month vs. No-suspension |
| **All study participants** | | | | | **RR (95% CI)** | **RR (95% CI)** |
| # recalled | 61 | 59 | 44 | | | |
| Recall rate (95% CI) | 11.3 (8.7, 14.2) | 12.3 (9.5,15.6) | 9.8 (7.2, 12.9) | .45 | 1.10 (0.78, 1.53) | 0.87 (0.60, 1.25) |
| **Estrogen only at baseline** | n = 335 | n = 295 | n = 286 | | | |
| # recalled | 31 | 34 | 24 | | | |
| Recall rate (95% CI) | 9.3 (6.1, 12.4) | 11.5 (7.9, 15.2) | 8.4 (5.2, 11.6) | .42 | 1.25 (0.79, 1.97) | 0.91 (0.55, 1.51) |
| **Estrogen + progestin at baseline** | n = 207 | n = 183 | n = 165 | | | |
| # recalled | 30 | 25 | 20 | | | |
| Recall rate (95% CI) | 14.5 (9.7, 19.3) | 13.7 (8.6, 18.7) | 12.1 (7.1, 17.2) | .80 | 0.94 (0.58, 1.54) | 0.84 (0.49, 1.42) |

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript



**Table 3**

Recall rate and the adjusted* relative risk of recall by mammographic breast density (percent and dense area) at study mammogram and change in breast density from the index to study mammogram

| | # Women | # Recalled | Recall Rate (95% CI) | RR* (95% CI) | P-value for linear trend |
|---|---|---|---|---|---|
| **Breast Density at Study Mammogram** | | | | | |
| % Breast Density | | | | | |
| <10% | 307 | 31 | 10.1 (6.7, 13.5) | 1.00 | 0.046 |
| 10 to <20% | 395 | 46 | 11.6 (8.5, 14.8) | 1.40 (0.90, 2.18) | |
| 20 to <30% | 338 | 38 | 11.2 (7.9, 14.6) | 1.48 (0.92, 2.38) | |
| 30 to <40% | 213 | 25 | 11.7 (7.4, 16.1) | 1.58 (0.93, 2.70) | |
| ≥ 40% | 217 | 24 | 11.1 (6.9, 15.2) | 1.71 (0.98, 3.00) | |
| Dense Area | | | | | |
| <15 $cm^2$ | 341 | 28 | 8.2 (5.3, 11.1) | 1.00 | 0.004 |
| 15 to <25 $cm^2$ | 405 | 43 | 10.6 (7.6, 13.6) | 1.41 (0.89, 2.24) | |
| 25 to <35 $cm^2$ | 287 | 34 | 11.8 (8.1, 15.6) | 1.57 (0.97, 2.55) | |
| 35 to <50 $cm^2$ | 260 | 34 | 13.1 (9.0, 17.2) | 1.84 (1.13, 2.97) | |
| ≥ 50 $cm^2$ | 177 | 25 | 14.1 (9.0, 19.3) | 1.87 (1.12, 3.11) | |
| **Change in breast density** | | | | | |
| Change in % Breast Density | | | | | |
| Decrease > 8% | 200 | 27 | 13.5 (8.7, 18.3) | 1.10 (0.71, 1.70) | 0.52 |
| Decrease, 2 to 8% | 207 | 35 | 8.6 (5.9, 11.3) | 0.69 (0.46, 1.04) | |
| No change (±2%) | 402 | 53 | 13.2 (9.9, 16.5) | 1.00 | |
| Increase ≥ 2% | 461 | 49 | 10.6 (7.8, 13.4) | 0.89 (0.61, 1.28) | |
| Change in Dense Area | | | | | |
| Decrease ≥ 10 $cm^2$ | 223 | 34 | 15.2 (10.5, 20.0) | 1.64 (1.02, 2.62) | 0.109 |
| Decrease, 2 to <10 $cm^2$ | 425 | 35 | 8.2 (5.6, 10.9) | 0.92 (0.57, 1.47) | |
| No change (±2 $cm^2$) | 336 | 29 | 8.6 (5.6, 11.6) | 1.00 | |

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



| | # Women | # Recalled | Recall Rate (95% CI) | RR* (95% CI) | P-value for linear trend |
|---|---|---|---|---|---|
| Increase ≥ 2 $cm^2$ | 486 | 66 | 13.6 (10.5, 16.6) | 1.52 (1.00, 2.30) | |

*adjusted for body mass index and age

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript



**Table 4**

Breast density (Mean, SE) and Change in Breast density by Randomization Group

| | Mean (SE) | | | Difference in means (No suspension is reference group) | |
|---|---|---|---|---|---|
| | **No-suspension** | **1-month suspension** | **2-months suspension** | **1-month suspension vs. No-suspension** | **2-months suspension vs. No-suspension** |
| **# Mammograms completed** | 541 | 478 | 452 | **β (95% CI)** | **β (95% CI)** |
| **Percent Dense Area (%)** | | | | | |
| *All study participants* | | | | | |
| Index mammogram | 24.4 (0.6) | 24.1 (0.7) | 24.1 (0.7) | −0.30 (−2.15, 1.55) | −0.38 (−2.26, 1.50) |
| Study mammogram | 24.5 (0.6) | 23.2 (0.7) | 22.6 (0.7) | −1.34 (−3.19, 0.51) | **−1.98 (−3.85, −0.10)** |
| Change in % dense | 0.1 (0.3) | −0.9 (0.3) | −1.5 (0.4) | **−1.04 (−1.93, −0.15)** | **−1.60 (−2.50, −0.70)** |
| *Estrogen only at baseline* | | | | | |
| Index mammogram | 22.9 (0.8) | 23.0 (0.9) | 22.7 (0.9) | 0.14 (−2.14, 2.42) | −0.22 (−2.52, 2.07) |
| Study mammogram | 22.7 (0.8) | 22.3 (0.9) | 21.6 (0.9) | −0.33 (−2.61, 1.95) | −1.12 (−3.42, 1.19) |
| Change in % dense | −0.2 (0.4) | −0.7 (0.4) | −1.1 (0.4) | −0.47 (−1.57, 0.63) | −0.89 (−2.00, 0.22) |
| *Estrogen + progestin at baseline* | | | | | |
| Index mammogram | 27.0 (1.1) | 25.9 (1.2) | 26.5 (1.2) | −1.03 (−4.15, 2.08) | −0.49 (−3.69, 2.71) |
| Study mammogram | 27.6 (1.1) | 24.6 (1.1) | 24.3 (1.2) | −3.00 (−6.09, 0.09) | **−3.28 (−6.46, −0.11)** |
| Change in % dense | 0.6 (0.5) | −1.3 (0.6) | −2.2 (0.6) | **−1.96 (−3.45, −0.48)** | **−2.79 (−4.31, −1.27)** |
| **Dense Area ($cm^2$)** | | | | | |
| *All study participants* | | | | | |
| Index mammogram | 30.5 (0.9) | 31.9 (1.0) | 30.3 (1.0) | 1.38 (−1.26, 4.02) | −0.17 (−2.85, 2.51) |
| Study mammogram | 30.1 (0.8) | 30.4 (1.0) | 27.5 (0.9) | 0.30 (−2.13, 2.74) | **−2.57 (−5.04, −0.10)** |

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



| | Mean (SE) | | | Difference in means (No suspension is reference group) | |
|---|---|---|---|---|---|
| | **No-suspension** | **1-month suspension** | **2-months suspension** | **1-month suspension vs. No-suspension** | **2-months suspension vs. No-suspension** |
| Change in area | −0.5 (0.5) | −1.5 (0.5) | −2.9 (0.5) | **−1.07 (−2.37, 0.22)** | **−2.40 (−3.71, −1.08)** |
| *Estrogen only at baseline* | | | | | |
| Index mammogram | 30.2 (1.1) | 31.0 (1.2) | 30.8 (1.3) | 0.86 (−2.47, 4.20) | 0.58 (−2.78, 3.94) |
| Study mammogram | 29.0 (1.0) | 29.8 (1.1) | 27.9 (1.2) | 0.80 (−2.23, 3.82) | −1.07 (−4.13, 1.98) |
| Change in area | −1.2 (0.6) | −1.3 (0.6) | −2.9 (0.7) | −0.07 (−1.78, 1.65) | −1.65 (−3.38, 0.07) |
| *Estrogen + progestin at baseline* | | | | | |
| Index mammogram | 31.1 (1.5) | 33.3 (1.8) | 29.6 (1.4) | 2.20 (−2.14, 6.54) | −1.43 (−5.88, 3.03) |
| Study mammogram | 31.8 (29.0) | 31.3 (1.7) | 26.8 (1.3) | −0.51 (−4.59, 3.58) | **−5.04 (−9.24, −0.85)** |
| Change in area | 0.8 (0.7) | −1.9 (0.7) | −2.8 (0.8) | **−2.71 (−4.66, −0.75)** | **−3.61 (−5.62, −1.61)** |

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript



**Table 5**

Mean of Symptom Severity Scores by Intervention Group at Baseline and Follow-up and Change in Severity of Symptoms

| | Baseline Questionnaire | | | | Follow-up Questionnaire | | | |
|---|---|---|---|---|---|---|---|---|
| | No Suspension (n = 529)* | 1-month Suspension (n = 465) | 2-month Suspension (n = 444) | *p*-Value | No Suspension (n = 527) | 1-month Suspension (n = 462) | 2-month Suspension (n = 443) | *p*-Value |
| Hot flashes/night sweats, % | | | | | | | | |
| Do you have hot flashes or night sweats? | | | | .40 | | | | <.001 |
| No | 48.7 | 51.7 | 52.8 | | 49.9 | 15.0 | 11.6 | |
| Yes | 51.3 | 48.3 | 47.2 | | 50.1 | 85.0 | 88.4 | |
| Among women who have hot flashes or night sweats: | n = 268 | n = 223 | n = 208 | | n = 262 | n = 390 | n = 388 | |
| How many hot flashes plus night sweats per day? | | | | 0.145 | | | | <.001 |
| < 1 per day | 56.3 | 52.0 | 44.2 | | 48.9 | 17.7 | 11.1 | |
| 1–2 per day | 23.1 | 24.7 | 36.5 | | 30.5 | 19.2 | 17.3 | |
| 3–4 per day | 13.4 | 14.8 | 13.0 | | 13.4 | 22.6 | 20.9 | |
| 5–6 per day | 3.7 | 3.1 | 4.3 | | 5.0 | 16.9 | 20.1 | |
| > 6 per day | 2.6 | 4.5 | 1.9 | | 1.9 | 23.1 | 29.9 | |
| missing | 0.8 | 0.9 | 0.0 | | 0.4 | 0.5 | 0.8 | |
| Intensity of hot flashes | | | | .76 | | | | <.001 |
| Mild | 47.0 | 48.9 | 52.9 | | 54.6 | 29.7 | 21.4 | |
| Moderate | 38.1 | 38.6 | 34.6 | | 34.4 | 53.6 | 56.2 | |
| Severe | 4.9 | 5.4 | 6.3 | | 5.3 | 14.4 | 20.6 | |
| missing | 10.1 | 7.2 | 6.3 | | 5.7 | 2.3 | 1.8 | |
| Intensity of night sweats | | | | .31 | | | | <.001 |
| Mild | 26.9 | 18.8 | 22.1 | | 29.0 | 18.7 | 12.6 | |
| Moderate | 32.8 | 36.8 | 36.1 | | 34.0 | 25.9 | 24.2 | |
| Severe | 37.7 | 40.8 | 39.9 | | 35.1 | 53.9 | 61.1 | |
| Missing | 2.6 | 3.6 | 1.9 | | 1.9 | 1.5 | 2.1 | |
| How much do night sweats interfere with sleep? | | | | .007 | | | | <.001 |
| Do not interfere | 20.9 | 13.5 | 16.4 | | 19.5 | 14.6 | 9.0 | |
| Woke up, right back to sleep | 55.6 | 52.0 | 63.5 | | 57.6 | 43.9 | 47.2 | |
| Kept me awake | 20.2 | 29.6 | 17.3 | | 19.9 | 37.7 | 39.2 | |
| Ruin my night | 0.8 | 0.9 | 1.4 | | 1.2 | 2.3 | 2.8 | |

*Ann Intern Med*. Author manuscript; available in PMC 2010 January 7.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



| | Baseline Questionnaire | | | | Follow-up Questionnaire | | | |
|---|---|---|---|---|---|---|---|---|
| | **No Suspension (n = 529)*** | **1-month Suspension (n = 465)** | **2-month Suspension (n = 444)** | ***p*-Value** | **No Suspension (n = 527)** | **1-month Suspension (n = 462)** | **2-month Suspension (n = 443)** | ***p*-Value** |
| Missing | 2.6 | 4.0 | 1.4 | | 1.9 | 1.5 | 1.8 | |
| Sleep, mean (SD) | | | | | | | | |
| Importance of good night's sleep | 9.0 (1.5) | 8.9 (1.6) | 9.0 (1.5) | .47 | 8.9 (1.7) | 8.7 (1.9) | 8.7 (1.7) | .074 |
| Sleep Disturbance Scale‡ | 2.60 (1.42) | 2.54 (1.42) | 2.57 (1.41) | .77 | 2.55 (1.45) | 2.98 (1.51) | 3.13 (1.46) | <.001 |
| Menopausal symptoms, mean (SD) | | | | | | | | |
| Menopausal symptom scale | 1.92 (1.33) | 1.99 (1.48) | 2.01 (1.47) | .57 | 1.81 (1.33) | 2.48 (1.73) | 2.89 (1.76) | <.001 |
| Vasomotor symptom sub-scale† | 1.71 (2.20) | 1.85 (2.44) | 1.93 (2.43) | .35 | 1.64 (2.12) | 4.31 (2.99) | 5.30 (2.91) | <.001 |
| Forgetfulness | 2.6 (2.4) | 2.5 (2.4) | 2.6 (2.4) | .82 | 2.4 (2.4) | 2.2 (2.7) | 2.6 (2.8) | .024 |
| Mood swings | 1.3 (1.8) | 1.5 (2.1) | 1.5 (1.9) | .31 | 1.2 (1.8) | 1.9 (2.7) | 2.2 (2.7) | <.001 |
| Trouble concentrating | 2.0 (2.4) | 1.9 (2.3) | 2.1 (2.3) | .52 | 1.8 (2.3) | 2.2 (2.8) | 2.6 (2.9) | <.001 |
| Vaginal bleeding | 0.1 (0.7) | 0.1 (0.7) | 0.1 (0.7) | .99 | 0.1 (0.8) | 0.2 (0.8) | 0.1 (0.8) | .81 |

*Number returning questionnaire

†Subscale of the Wiklund(28, 29) symptom checklist, which includes hot flashes and night sweats

‡General Sleep Disturbance Scale(27, 58) average score across sleep difficulty questions