JOURNAL OF CLINICAL ONCOLOGY

Estrogen Alone in Postmenopausal Women and Breast Cancer Detection by Means of Mammography and Breast Biopsy

Rowan T. Chlebowski, Garnet Anderson, JoAnn E. Manson, Mary Pettinger, Shagufta Yasmeen, Dorothy Lane, Robert D. Langer, F. Allan Hubbell, Anne McTiernan, Susan Hendrix, Robert Schenken, and Marcia L. Stefanick

From the Los Angeles Biomedical Research Institute at Harbor, University of California, Los Angeles Medical Center, Torrance; University of California Davis, Sacramento; University of California, Irvine; Stanford University, Palo Alto, CA; The State University of New York, Stony Brook, NY; Fred Hutchinson Cancer Research Center, Seattle, WA; Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; Outcomes Research Institute, Geisinger Health System, Danville, PA; Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI; and Department of Obstetrics and Gynecology, The University of Texas Health Science Center, San Antonio, TX.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Rowan T. Chlebowski, MD, PhD, Los Angeles Biomedical Research Institute at Harbor, University of California Los Angeles Medical Center, 1124 W Carson St, Torrance, CA, 90502; e-mail: rchlebowski@gmail.com.

A B S T R A C T

Purpose
As the influence of estrogen alone on breast cancer detection is not established, we examined this issue in the Women's Health Initiative trial, which randomly assigned 10,739 postmenopausal women with prior hysterectomy to conjugated equine estrogen (CEE; 0.625 mg/d) or placebo.

Methods
Screening mammography and breast exams were performed at baseline and annually. Breast biopsies were based on clinical findings. Effects of CEE alone on breast cancer detection were determined by using receiver operating characteristic (ROC) analyses of mammogram performance.

Results
After a 7.1-year mean follow-up, fewer invasive breast cancers were diagnosed in the CEE than in the placebo group, but the difference was not statistically significant. Use of CEE alone increased mammograms with short-interval follow-up recommendations (cumulative, 39.2% v 29.6.3%; $P < .001$) but not abnormal mammograms (ie, those suggestive of or highly suggestive of malignancy; cumulative, 7.3% v 7.0%; $P = .41$). Breast biopsies were more frequent in the CEE group (cumulative, 12.5% v 10.7%; $P = .004$) and less commonly diagnosed as cancer (8.9% v 15.8%, respectively, with positive biopsies; $P = .04$). Mammographic breast cancer detection in the CEE group was significantly compromised only in the early years of use.

Conclusion
CEE alone use for 5 years results in approximately one in 11 and one in 50 women having otherwise avoidable mammograms with short-interval follow-up recommendations or breast biopsies, respectively. Although the breast biopsies on CEE were less commonly diagnosed as cancer, breast cancer detection was not substantially compromised. These findings differ from estrogen-plus-progestin use, for which significantly increased abnormal mammograms and a compromise in breast cancer detection are seen.

J Clin Oncol 28:2690-2697. © 2010 by American Society of Clinical Oncology

INTRODUCTION

After initial reports from the Women's Health Initiative (WHI) trial of combined hormone therapy,[1,2] use of menopausal hormonal therapy substantially decreased[3,4] but continues in wide use. In particular, estrogen alone, indicated for women with prior hysterectomy primarily for climacteric symptoms, currently is still used by millions of women in the United States.[5]

In the Women's Health Initiative (WHI) clinical trial for postmenopausal women without prior hysterectomy, combined hormone therapy with conjugated equine estrogen (CEE) plus medroxy-progesterone acetate (MPA) increased breast cancer incidence[2,6] and compromised breast cancer detec-

tion, contributing to diagnostic delay.[7] In contrast, in the WHI randomized trial evaluating CEE alone in women with prior hysterectomy, breast cancer incidence was not increased.[8] Although mammograms with short interval follow-up recommendations were increased,[8] details of CEE effects on breast cancer detection were not reported. Therefore, we assessed CEE influence on breast cancer detection by means of screening mammography and breast biopsy during the trial.

METHODS

A total of 10,738 postmenopausal women enrolled on the WHI trial evaluating CEE alone at 40 clinical centers between 1993 and 1998. Detailed eligibility criteria and

recruitment procedures have been described.[9,10,11] Eligibility included age between 50 to 79 years, postmenopausal status, and requirement for written informed consent. Major exclusions included prior breast cancer, other prior cancer within 10 years except nonmelanoma skin cancer, or medical conditions likely to result in death in 3 years. Women using menopausal hormones were eligible after a 3-month washout. A baseline mammogram and breast clinical exam not suggestive of cancer were eligibility requirements. Women in the CEE-alone trial also could participate in the WHI dietary modification (DM) trial and/or the WHI trial of calcium and vitamin D (CaD). Approximately 30% and 60% joined the DM and CaD trials, respectively. The study was approved by human subjects committees at each institution, and all participants provided written informed consent.

Participants were randomly assigned to CEE (Premarin 0.625 mg; Wyeth Ayerst, Collegeville, PA) or an identical-appearing placebo. A computerized randomization procedure was developed by the WHI Clinical Coordinating Center (CCC) and was implemented at local clinical centers. Both participants and staff were blinded to study medication allocation.

Data Collection

Baseline information was collected by using standardized self-report instruments and a brief physical exam. Interviewer-administered questionnaires were used to collect information on past hormone therapy use.[9]

Follow-up included a contact 6 weeks after random assignment to assess adherence, at 6-month intervals to assess clinical outcomes, and annually for clinic visits. Mammography was performed at WHI clinical centers as well as at more than 3,000 community sites.[12] Mammogram reports were obtained, reviewed at the local clinical centers, and coded to reflect radiologist recommendations. Screening mammograms and clinical breast exams were required annually. Dispensing of study medications required their completion and clearance of findings suggestive of breast cancer. Therefore, mammograms with recommendations for additional imaging evaluation (equivalent to BIRADS category 0)[13] were considered incomplete studies and were considered missing mammograms for these analyses. Work-up of breast findings, including recommendations for biopsy, were directed primarily by community physicians (Fig 1, CONSORT diagram).

Breast cancer self reports were verified by centrally trained WHI physician adjudicators who reviewed medical records and pathology reports (available in 98.2% of participants).[14] Breast cancers required histologic confirmation. Final adjudication and coding were performed at the WHI Clinical Coordinating Center by using the Surveillance Epidemiology and End Results coding system.[15] The intervention was stopped by the National Insti-

tutes of Health after a mean intervention period of 7.1 years on the basis of increased stroke risk and assessment that benefit for coronary heart disease was unlikely.[16]

Statistical Analyses

Participant characteristics were compared in the two randomized groups by using χ^2 statistics or t tests. Breast cancer results by randomly assigned group were assessed with time-to-event methods and were based on the intent-to-treat principle. Breast cancer incidence was compared by using hazard ratio (HRs) and corresponding 95% CIs estimated from Cox proportional hazard models stratified by age and dietary modification trial random assignment. Subgroup analyses were determined with Cox proportional hazard models, for which P values were determined with Wald statistics.

The observations in this report are based on mammogram recommendations after screening mammograms performed in protocol-defined time windows. Information on performance of diagnostic mammograms ie, those driven by clinical findings between screenings) was not captured and is not a component of these results.

Estimates of mammogram sensitivity, specificity, and positive and negative predictive values were compared by randomly assigned group. CIs were calculated by using the efficient-score method with continuity correction. Breast cancer occurrences were defined as invasive breast cancer or ductal carcinoma in situ. Mammograms with findings suggestive or highly suggestive of malignancy were considered abnormal or positive. All other completed mammograms were considered negative. A positive mammogram was considered a true positive if breast cancer was diagnosed within 1 year of the mammogram. A negative mammogram was considered a true negative if breast cancer was not diagnosed within 1 year. Specificity was defined as the percentage of true-negative exams among women without a breast cancer diagnosis within 1 year. Sensitivity was defined as the percentage of true-positive exams among women with a breast cancer diagnosis made within 1 year. Positive predictive value was defined as the percentage of true-positive exams among women with positive exams. Similarly, negative predictive value was the percentage of true-negative exams among women with negative exams.

Receiver operating curves (ROCs) and the corresponding area under the curve (AUC) were used to assess the diagnostic accuracy of mammograms.[17] The ROCs plot the true-positive rate (ie, sensitivity) versus the false positive rate (ie, specificity) for a mammogram result, as the cutoff for defining an abnormal test is allowed to vary across the range of the five possible mammogram recommendation categories (ie, normal, benign, short-interval follow-up recommended, suggestive of cancer, and highly suggestive of cancer). The AUC is a general measure of the diagnostic accuracy of the mammography. An ROC that corresponds to a fair coin toss classifier (ie, a nonpredictive model) is a straight line connecting the coordinates (0,0) to (1,1) and has an AUC of 0.50. An ROC that correspond to a perfect classifier is a pair of vertical and horizontal lines connecting the coordinates (0,0) to (0,1) to (1,1) and has an AUC of 1.00. CIs for the AUC were computed by using the bootstrap method.

ROCs are presented for three distinct time periods (ie, 1 to 2, 3 to 4, and ≥ 5 years from entry) and are compared across random assignment groups. A generalized estimating equation (GEE) approach was used to take into account the correlation between multiple mammogram results for a participant within each time period. Testing was done within the GEE model to assess the interaction between random assignment group and the mammography result. Statistical significance was determined for each period after categorizing the mammogram into abnormal versus normal, for which abnormal is defined as mammograms with a recommendation for short-interval follow-up or with findings suggestive or highly suggestive of malignancy.

Biopsy frequency was evaluated by using the semi-annual reports of breast biopsies, and the time to each woman's first biopsy report was determined and compared between random assignment groups using a log-rank statistic. Biopsy dates were collected for women with breast cancer diagnoses. Because the dates of biopsies with benign findings were not collected, a breast biopsy was considered to be a true positive if an invasive breast cancer or ductal carcinoma in situ was diagnosed during the 6-month
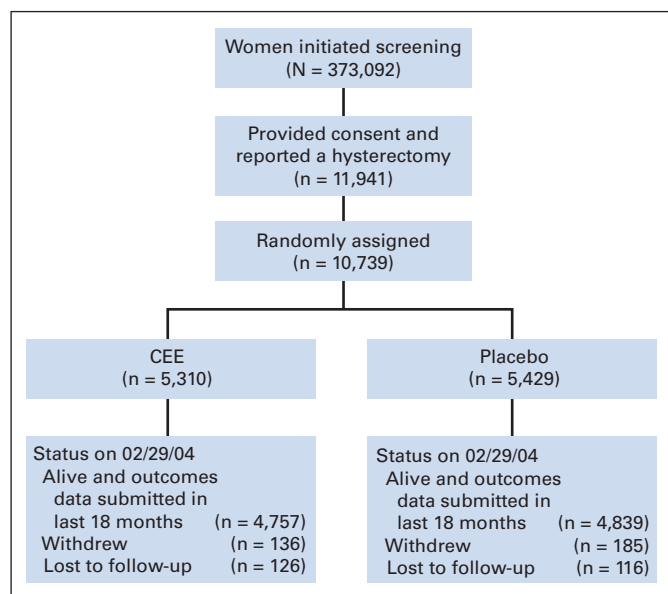

Fig 1. CONSORT diagram. CEE, conjugated equine estrogen.

Women initiated screening
(N = 373,092)

Provided consent and
reported a hysterectomy
(n = 11,941)

Randomly assigned
(n = 10,739)

CEE
(n = 5,310)

Placebo
(n = 5,429)

Status on 02/29/04
Alive and outcomes
data submitted in
last 18 months   (n = 4,757)
Withdrew   (n = 136)
Lost to follow-up   (n = 126)

Status on 02/29/04
Alive and outcomes
data submitted in
last 18 months   (n = 4,839)
Withdrew   (n = 185)
Lost to follow-up   (n = 116)

Table 1. Patient Demographics and Clinical Characteristics in the CEE Trial by Random Assignment Group

| | Group | | | |
| | CEE | | Placebo | |
Variable	No.	%	No.	%
Age at screening, years	63.71	7.25	63.73	7.33
Mean				
SD				
Age group at screening, years				
50-59	1,637	30.8	1,673	30.8
60-69	2,387	45.0	2,465	45.4
70-79	1,286	24.2	1,291	23.8
Ethnicity				
White	4,007	75.5	4,075	75.1
Black	782	14.7	835	15.4
Hispanic	322	6.1	333	6.1
American Indian	41	0.8	34	0.6
Asian/Pacific Islander	86	1.6	78	1.4
Unknown	72	1.4	74	1.4
Education				
0-8 years	181	3.4	148	2.8
Some high school	354	6.7	370	6.9
High school/GED diploma	1,233	23.5	1,188	22.1
School after high school	2,271	43.2	2,350	43.7
College degree or higher	1,216	23.1	1,327	24.7
Age at menarche, years				
≤ 11	1,215	23.0	1,280	23.7
12-13	2,805	53.1	2,853	52.8
≥ 14	1,259	23.9	1,274	23.6
Body mass index, kg/m^2	30.09	6.14	30.07	6.21
Mean				
SD				
< 25	1,110	21.0	1,096	20.3
25 to < 30	1,795	34.0	1,912	35.5
≥ 30	2,376	45.0	2,383	44.2
Physical activity, kcal/wk/kg				
None	1,081	22.2	1,043	21.3
> 0-3.5	887	18.2	930	19.0
> 3.5-8.0	983	20.1	983	20.0
> 8.0-16.5	981	20.1	945	19.3
> 16.5	948	19.4	1,003	20.5
Alcohol intake				
Nondrinker	718	13.7	737	13.7
Past drinker	1,277	24.3	1,270	23.6
< 1 drink per month	767	14.6	766	14.2
< 1 drink per week	1,001	19.1	1,049	19.5
1 to < 7 drinks per week	1,027	19.6	1,091	20.3
≥ 7 drinks per week	457	8.7	475	8.8
Smoking				
Never smoked	2,723	51.9	2,705	50.4
Past smoker	1,986	37.8	2,089	38.9
Current smoker	542	10.3	571	10.6
5-year Gail risk				
< 1.25	2,129	40.1	2,149	39.6
1.25-1.74	1,620	30.5	1,688	31.1
≥ 1.75	1,561	29.4	1,592	29.3
Age at first birth, years				
Never pregnant/no term pregnancy	491	10.4	463	9.5
< 20	1,193	25.2	1,234	25.3
20-29	2,846	60.0	2,914	59.8
≥ 30	210	4.4	260	5.3

(continued in next column)

Table 1. Patient Demographics and Clinical Characteristics in the CEE Trial by Random Assignment Group (continued)

| | Group | | | |
| | CEE | | Placebo | |
Variable	No.	%	No.	%
No. of children breastfed				
None	2,468	47.2	2,491	46.7
1-2	1,621	31.0	1,732	32.4
≥ 3	1,138	21.8	1,116	20.9
No. first-degree relatives with breast cancer				
None	4,202	85.8	4,352	86.4
1	634	12.9	597	11.9
≥ 2	63	1.3	88	1.8
Benign breast disease				
No	3,894	80.7	3,787	78.3
Yes, 1 biopsy	683	14.2	756	15.6
Yes, ≥ 2 biopsies	250	5.2	295	6.1
Weight, kg	78.57	17.21	78.55	17.45
Mean				
SD				
Oophorectomy				
No	2,717	44.2	2,090	41.6
Yes				
Partial	802	16.3	827	16.4
Bilateral	1,938	39.5	2,111	42.0
Age at oophorectomy, years				
< 40	946	34.9	1,021	35.2
40-49	1,230	45.4	1,304	44.9
≥ 50	533	19.7	579	19.9
Menopausal hormone therapy usage status				
Never used	2,769	52.2	2,770	51.1
Past user	1,871	35.2	1,948	35.9
Current user*	669	12.6	708	13.1
Prior estrogen only use, years				
No	2,872	54.1	2,891	533
Yes				
< 2	738	13.9	792	14.6
2-5	579	10.9	576	10.6
> 5	1,121	21.1	1,170	21.6
Prior estrogen plus progestin use, years				
No	5,093	95.9	5,178	95.4
Yes				
< 2	88	1.7	95	1.8
2-5	56	1.1	63	1.2
> 5	73	1.4	93	1.7
Regency of hormone use, years				
Current at screening*	669	26.3	708	26.7
< 2	248	9.8	272	10.2
2-5	318	12.5	314	11.8
> 5	1,305	51.4	1,362	51.3

Abbreviations: CEE, conjugated equine estrogen; SD, standard deviation; GED, general education development.
*Using at time of screening, this required 3 months wash out prior to entry of study.

interval when a biopsy was reported or within 2 months after the interval when the biopsy was reported.

Analyses regarding mammogram performance by random assignment were based on an intent-to-treat principle. Additional analyses adjusted for study medication adherence were conducted by censoring follow-up for
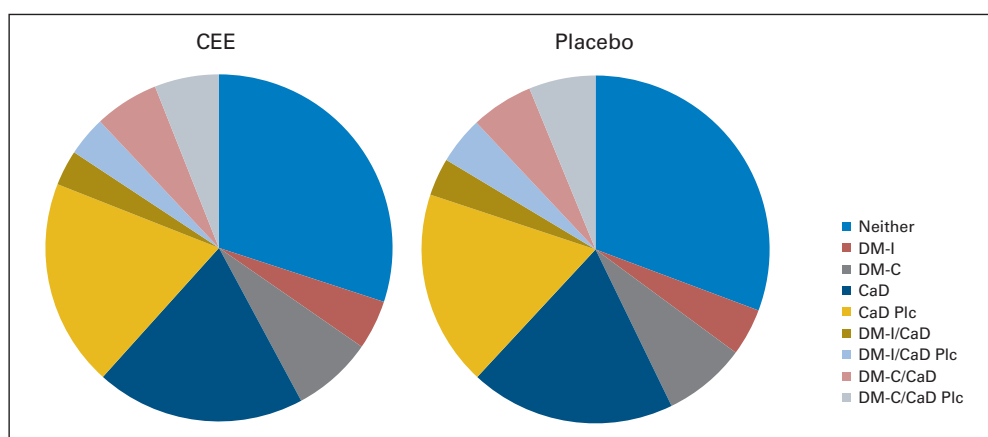
Fig 2. Random assignment in the Women's Health Initiative (WHI) dietary modification (DM) and calcium/vitamin D (CaD) clinical trials by random assignment group in the WHI conjugated equine estrogen (CEE) clinical trial. As women could participate in all three trials and could be randomly assigned to an intervention (I) or control (C) group in the DM trial or in the active drug (CaD) or placebo (CaD Plc) groups in the CaD trial, a total of nine categories are described.

women 6 months after they became nonadherent (ie, consuming < 80% of study pills or starting nonstudy hormone therapy during the most recent study interval).

RESULTS

All women were postmenopausal and had prior hysterectomy. As reported, 41% had prior bilateral oophorectomy, and the median age was 63 years; 31% entered onto the study between 50 and 59 years of age. Most factors associated with breast cancer risk and with abnormal mammograms were balanced between the hormone and placebo groups (Table 1). Participation in other WHI clinical trial components was also balanced between random assignment groups (Fig 2). A full CONSORT diagram is available (Fig 1).[11]

As previously reported,[8] fewer invasive breast cancers were diagnosed in the CEE alone than in the placebo group, but the difference

Table 2. Clinical Breast Cancer Outcomes by Treatment Group							
	Group				Outcome Analysis		
	CEE (n = 5,310)		Placebo (n = 5,429)				
Breast Cancer Outcome	No.	%	No.	%	HR	95% CI	P*
Total breast cancer†	129	0.34	161	0.42	0.82	0.65 to 1.04	.10
Invasive breast cancer	104	0.28	133	0.34	0.80	0.62 to 1.04	.09
In situ breast cancer	25	0.07	30	0.08	0.86	0.51 to 1.46	.58
Tumor size, cm*‡							.09
Mean	1.75		1.40				
SD	1.44		0.94				
Tumor size, categorized, cm§							
≤ 2	65	0.17	102	0.26	0.66	0.48 to 0.90	.008
> 2	26	0.07	23	0.06	1.15	0.65 to 2.01	.63
Missing	8	0.02	5	0.01	1.68	0.55 to 5.14	.36
Lymph nodes examined							
No	11	0.03	11	0.03	1.03	0.45 to 2.38	.94
Yes	92	0.24	118	0.31	0.80	0.61 to 1.05	.11
Missing	1	< 0.01	4	0.01	0.25	0.03 to 2.27	.18
No. of lymph nodes examined‖							.63
Mean	10.6		10.1				
SD	9.7		7.4				
No. of positive lymph nodes¶							.11
Mean	1.4		1.0				
SD	3.6		3.2				
Positive lymph nodes§							
No	60	0.16	92	0.24	0.67	0.49 to 0.93	.02
Yes	33	0.09	28	0.07	1.19	0.72 to 1.97	.49
Missing	11	0.03	13	0.03	0.87	0.39 to 1.94	.73

Abbreviations: CEE, conjugated equine estrogen; HR, hazard ratio; SD, standard deviation.
*The P values, hazard ratios, and 95% CIs were estimated from a Cox proportional hazards model stratified by age and dietary modification trial random assignment.
†Data on total, invasive, and in situ breast cancer had been previously reported.
‡Mean and SD only apply to those with a known tumor size for invasive cancers.
§P value from a t test using log-transformed values.
‖Mean and SD only apply to those with a known number of lymph nodes examined, including those with zero nodes examined.
¶Mean and SD only apply to those with a known number of positive lymph nodes, including those with zero positive nodes.

was not statistically significant (104 cancers in 5,310 hormone-group participants v 133 cancers in 5,429 placebo-group participants; hazard ratio, 0.80; 95% CI, 0.65 to 1.04; $P = .09$; Table 2). Staging information new to this report now includes results from an additional 47 occurrences, previously reported as having missing data.[8] Significantly fewer invasive breast cancers were diagnosed with tumors ≤ 2 cm in women in the CEE group (65 and 102 cancers for CEE alone v placebo, respectively; $P = .008$; Table 2), but there was no effect on tumors greater than 2 cm (26 and 23 cancers for CEE alone v placebo, respectively; $P = .63$). As a result of having fewer small cancers in the hormone group, the average cancer size was somewhat greater than in the placebo group. Although there were five more cancers with lymph node–positive involvement in the hormone group (33 v 28, respectively; $P = .49$), there were 32 fewer cancers with lymph node–negative disease in the hormone group (60 v 92 cancers, respectively; $P = .02$).

The cumulative frequency of mammograms with a short-interval follow-up recommendation by year and random assignment group are outlined in Figure 3A. Women in the hormone group had approximately 4% greater risk of having such a mammogram after 1 year and approximately a 9% greater risk after 5 years (30.7% v 22.0%; $P < .001$; Fig 3A). However, in the same period, there was no increase in mammograms with more serious findings either suggestive or highly suggestive of breast cancer (cumulative, 5.4% v 5.1% for CEE v placebo; $P = .53$; Fig 3A).

The frequency of clinically indicated breast biopsies by year and random assignment group are outlined in Figure 3B. The cumulative percent of women with a biopsy through year 7 was significantly greater in the CEE-alone group than in the placebo group (12.5% v 10.7%; $P = .004$), and the time to first biopsy report was significantly shorter ($P = .006$). Of 1,838 breast biopsies performed, breast cancer was diagnosed in 112 (8.9%) of 1,007 biopsies in the CEE group and in 127 (15.8%) of 831 biopsies in the placebo group ($P = .04$).

The year-by-year performance characteristics of mammograms by random assignment group are listed in Table 3. As seen, the sensitivity and positive predictive value of mammograms were compromised by CEE-alone use, whereas the specificity and the negative predictive value of mammograms were similar in the CEE-alone and placebo groups.

ROCs and AUC statistics were used to compare mammogram diagnostic performance by time on study and by random assignment group (Fig 4). Overall, performance was significantly inferior in the CEE-alone group in the first 2 years but not in subsequent periods. Looking across time intervals for women in the placebo group, the ROC AUCs that assessed mammogram performance were 0.89, 0.85, and 0.92 for mammograms in years 1 to 2, 3 to 4, and ≥ 5, respectively. For women in the CEE-alone group, the ROC AUCs were 0.85 ($P = .03$), 0.88 ($P = .79$), and 0.91 ($P = .58$) for mammograms performed in years 1 to 2, 3 to 4, and ≥ 5, respectively, with P values evaluating the effect of random assignment on mammogram diagnostic performance (Fig 4A). In adherence-adjusted analyses that censored follow-up 6 months after nonadherence to study medication use, mammogram diagnostic performance for women in the CEE-alone group was not significantly different from those in the placebo group for any period examined ($P = .38, .49,$ and $.06$ in years 1 to 2, 3 to 4, and ≥ 5, respectively; Fig 4B).

Information regarding cumulative adherence is provided by year and random assignment group in (Appendix Table A1, online only). Drop-outs represent those patients discontinuing study medication
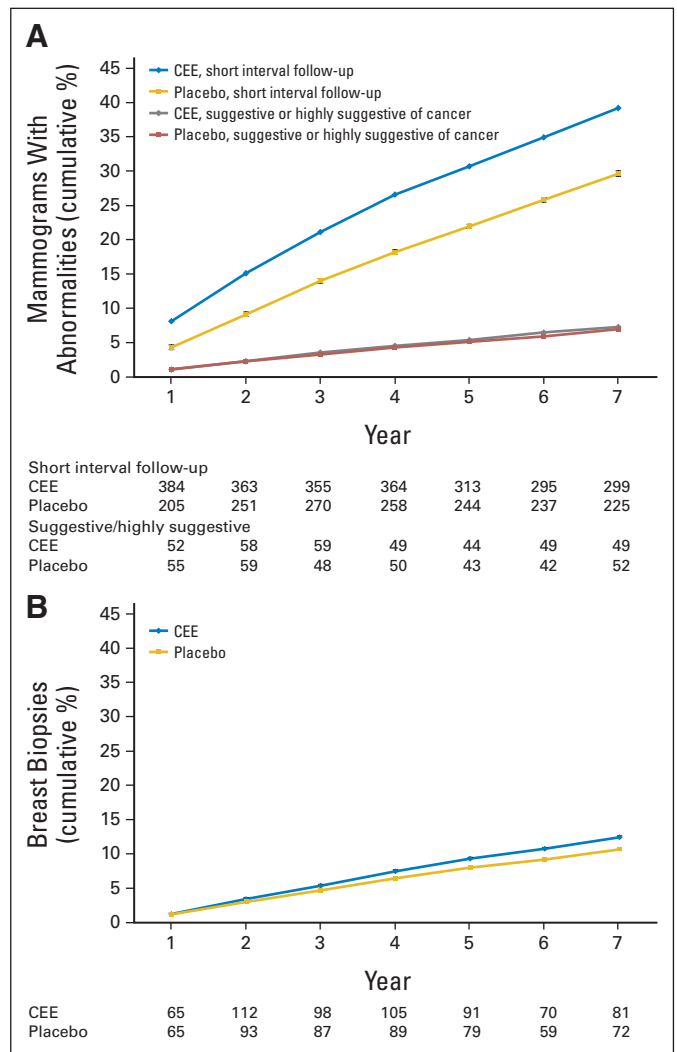

Fig 3. Initial report of a mammogram with (A) recommendation for short-interval follow-up or suggestive or highly suggestive of cancer and (B) initial report of breast biopsy for participants by year on study and random assignment group. Women in the conjugated equine estrogen (CEE) group had about a 4% greater risk of a mammogram with recommendation for short-interval follow-up after 1 year and 9% greater risk after 5 years of use ($P < .001$). Women in the CEE group had about a 2% greater risk of having a biopsy after 5 years of use ($P = .004$). Standard error (SE) bars were plotted at each data point but may not be visible because of their small sizes relative to the data point.

use. Drop-ins represent those patients initiating nonprotocol hormone therapy.

DISCUSSION

Use of CEE alone significantly increased mammograms with short-interval follow-up recommendations but not those with more serious findings. Significantly more breast biopsies were performed for clinical indications in the hormone group, yet they less frequently diagnosed cancer. However, mammogram diagnostic performance differences between random assignment groups were seen only in the early years of exposure. Thus, mammograms in women in the CEE-alone group were generally able to diagnose breast cancer in a timely

Table 3. Performance Characteristics of Mammography by Year on Study in CEE v Placebo Groups

Characteristic	Year 1		Year 2		Year 3		Year 4	
	CEE	Placebo	CEE	Placebo	CEE	Placebo	CEE	Placebo
No. of mammograms	4,703	4,755	4,488	4,608	4,408	4,463	4,272	4,338
No. of breast cancers	15	22	19	26	22	24	19	31
Sensitivity	33.3	45.5	47.4	65.4	50.0	58.3	68.4	58.1
95% CI	13.0 to 61.3	25.1 to 67.3	25.2 to 70.5	44.4 to 82.1	28.8 to 71.2	36.9 to 77.2	43.5 to 86.4	39.3 to 74.9
Specificity	99.0	99.1	98.9	99.1	98.9	99.3	99.2	99.3
95% CI	98.7 to 99.3	98.7 to 99.3	98.5 to 99.2	98.8 to 99.3	98.5 to 99.2	98.9 to 99.5	98.8 to 99.4	99.0 to 99.5
PPV	9.8	18.5	15.5	29.3	18.6	29.8	26.5	36.7
95% CI	3.7 to 22.2	9.7 to 31.9	7.8 to 27.9	18.5 to 42.9	10.1 to 31.3	17.8 to 45.1	15.4 to 41.3	23.8 to 51.7
NPV	99.8	99.7	99.8	99.8	99.7	99.8	99.9	99.7
95% CI	99.6 to 99.9	99.5 to 99.9	99.6 to 99.9	99.6 to 99.9	99.5 to 99.9	99.6 to 99.9	99.7 to 99.9	99.5 to 99.8
				(continued on following page)				

Abbreviations: CEE, conjugated equine estrogen; PPV, positive predictive value; NPV, negative predictive value.
*Total is less than 290 because in five women breast cancer was diagnosed more than a year after the last mammogram.

manner at a cost of an increase in both mammograms with short-interval follow-up recommendation and breast biopsies, which were more likely to be false positive.

The influence of CEE alone on breast cancer detection in women with prior hysterectomy can be compared with that of CEE plus MPA in women with no prior hysterectomy in the parallel WHI randomized trial.[1,2] There were similarities, as both increased mammograms with short-interval follow-up recommendation and both significantly increased breast biopsies that less reliably detected cancer. However, combined hormone therapy also increased abnormal mammograms with more serious findings and more commonly compromised mammogram performance.[2,7] During 5.6 years of combined use of CEE plus MPA, invasive breast cancers were increased, and they were diagnosed at higher stage,[2] consistent with combined hormone therapy effects to both stimulate breast cancer growth and delay breast cancer diagnosis.[2,6,7] However, in the CEE-alone trial with a longer intervention duration of 7.1 years, there were fewer invasive breast cancers in the CEE group, and mammographic cancer detection was not substantially compromised.[8] Breast cancers were diagnosed without substantial delay for women using CEE alone, but more breast biopsies were needed to find the tumors.

Post hoc analyses in the CEE-alone group found significantly ($P = .008$) fewer small invasive breast cancers ≤ 2 cm (65 v 102, respectively), and significantly fewer ($P = .02$) cancers with negative lymph nodes (60 v 92, respectively), whereas there was no significant increase in larger invasive breast cancers or those with positive nodes. These findings may indicate CEE reduces the incidence of smaller, early-stage cancers. Ongoing postintervention follow-up will provide additional information regarding the long-term effects of CEE exposure on breast cancer incidence.

To our knowledge, this is the first comprehensive description of the time course of the influence of CEE alone on the diagnostic performance of mammography and breast biopsy in a randomized clinical trial. The preponderance of observational studies have combined results for estrogen alone with those for estrogen plus progestin use[18,19,20] or have reported no difference in performance between regimens.[21,22,23] In the Million Women Study, conjugated estrogen either alone or combined with progestin had closely comparable adverse influence on the false-positive recall rate.[24] However, as this end point includes both abnormal mammograms and those given a short

follow-up recommendation, it did not address the difference seen in this report, for which combined hormone therapy only increased mammogram with a short-interval follow-up recommendation.

As an increase or even preservation of breast density could potentially adversely affect mammogram diagnostic performance,[21,25] breast density was assessed in ancillary studies in the WHI hormone therapy trials. In 445 randomly assigned women, CEE alone use for 2 years modestly increased breast density compared with placebo (absolute difference in percent breast density, 2.9%; $P < .001$).[26] The effect of combined CEE plus MPA, in a parallel study of similar design and size, was substantially greater, with an absolute difference in percent breast density for combined hormone therapy compared with placebo after 2 years of 6.9% ($P < .001$).[27] Although additional study is needed to correlate breast density change with mammogram recommendations in individuals, breast density change may contribute to mammographic performance differences seen.

The mammography performance regarding sensitivity and specificity reported in this randomized trial reflects that of a large number of mammography centers and interpreting radiologists throughout the nation. Direct comparison to national benchmarks[28] for performance is not appropriate, given substantial differences in study population and protocol directives. For example, when considering the performance of the year 1 mammogram in the WHI trial, all women were required to have a clear mammogram in the past year, information on only completed mammograms was recorded, and no reliable evidence on estrogen effects on performance was available. Nonetheless, comparisons of current results with the Breast Cancer Surveillance Consortium (BCSC) mammogram screening program are favorable. With respect to tumor size, the BCSC program found tumors less than 1 cm in 37% and tumors of 1 to 2 cm in 42%, and the mean tumor size was 1.6 cm. In the placebo arm of the WHI study, tumors less than 1 cm were found in 40%, tumors of 1 to 2 cm were found in 40%, and the mean tumor size was 1.4 cm, which is a performance similar to that of the BCSC program.[28,29]

Study strengths include the randomized design, large sample size, requirement for annual mammography, and the central adjudication of breast cancers by reviewers blinded to random assignment. The trial evaluated CEE at 0.625 mg/d, and the results may not apply to other oral or transdermal hormonal therapies.

Table 3. Performance Characteristics of Mammography by Year on Study in CEE v Placebo Groups (continued)

Characteristic	Year 5		Year 6		Year 7		Total*	
	CEE	Placebo	CEE	Placebo	CEE	Placebo	CEE	Placebo
No. of mammograms	4,123	4,134	3,604	3,618	2,198	2,218	29,184	29,606
No. of breast cancers	22	24	12	16	7	13	125	160
Sensitivity	59.1	62.5	75.0	68.8	71.4	69.2	54.4	60.0
95% CI	36.7 to 78.5	40.8 to 80.5	42.8 to 93.3	41.5 to 87.9	30.3 to 94.9	38.9 to 89.6	45.3 to 63.3	51.9 to 67.6
Specificity	99.3	99.3	99.0	99.3	99.3	99.4	99.0	99.2
95% CI	98.9 to 99.5	99.0 to 99.6	98.6 to 99.3	98.9 to 99.5	98.8 to 99.6	98.9 to 99.6	98.9 to 99.2	99.1 to 99.3
PPV	30.2	35.7	20.0	29.7	23.8	39.1	19.7	29.7
95% CI	17.7 to 46.3	22.0 to 52.0	10.1 to 35.1	16.4 to 47.2	9.1 to 47.6	20.5 to 61.2	15.7 to 24.3	24.9 to 35.1
NPV	99.8	99.8	99.9	99.9	99.9	99.8	99.8	99.8
95% CI	99.6 to 99.9	99.6 to 99.9	99.7 to 100	99.7 to 99.9	99.6 to 100	99.5 to 99.9	99.7 to 99.8	99.7 to 99.8

These findings have clinical implications. Women using estrogen alone for climacteric symptoms for durations comparable to those in the study can be reassured regarding breast cancer risk and detection. However, they will experience more mammograms with short-interval follow-up recommendations and more false-positive breast biopsies. For mammographers, identification of the nature of the findings leading to short-interval follow-up recommendations in women on estrogen should be a priority. Given the suggestion of early
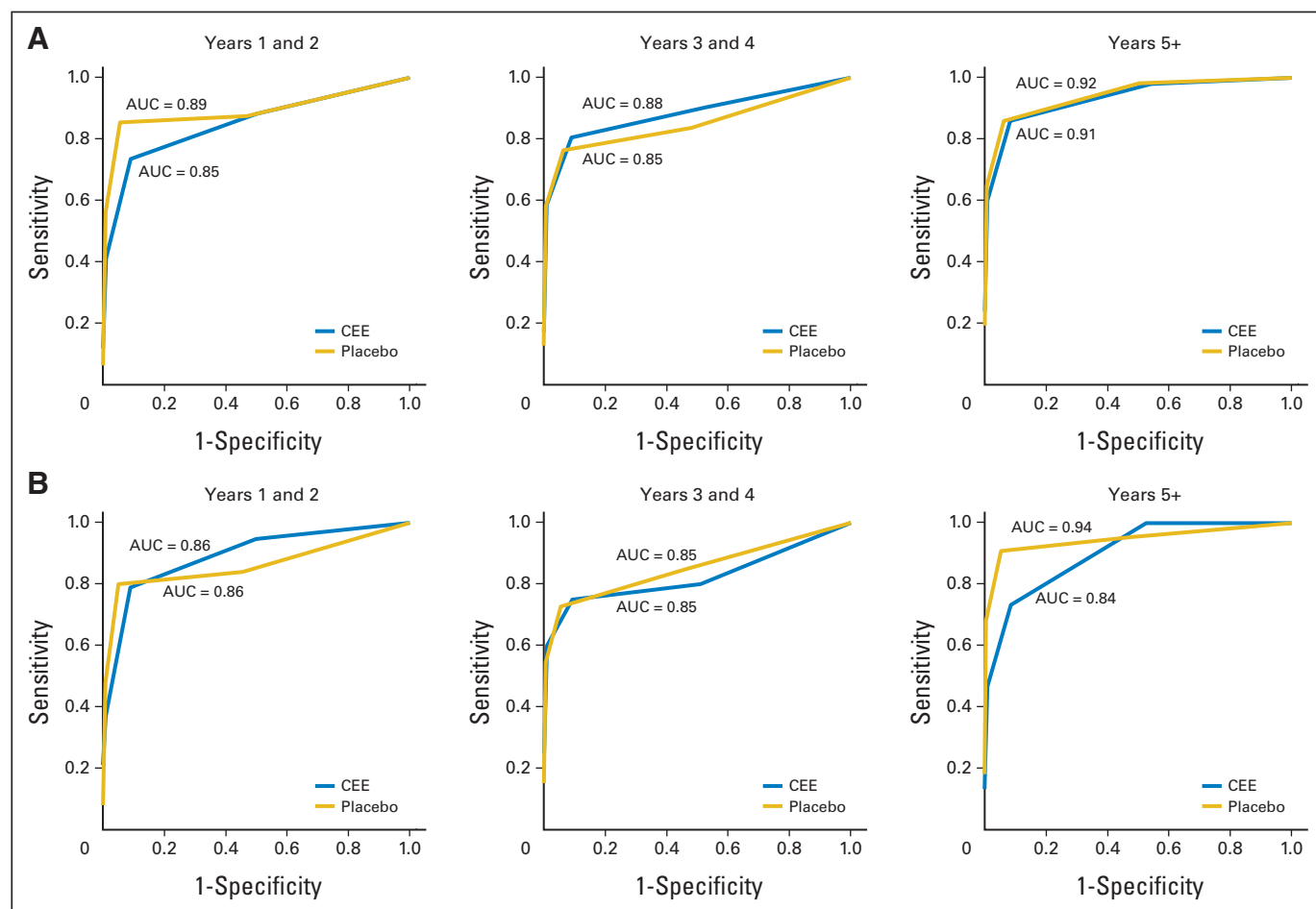
Fig 4. Diagnostic performance of mammograms by study period and randomization group. (A) Intent-to-treat analysis conducted using receiver operating curves (ROC) and nested area under the curve (AUC) statistics. For placebo-group women, the ROC AUCs were 0.89, 0.85, and 0.92 for mammograms performed in years 1 to 2, 3 to 4, and ≥ 5, respectively. For women in the conjugated equine estrogen (CEE) group, the ROC AUCs were 0.85 (P = .027), 0.88 (P = .792), and 0.91 (P = .580) for mammograms performed in year 1 to 2, 3 to 4, and ≥ 5, respectively. P values compared abnormal versus normal mammograms for hormone-group versus placebo-group women. (B) Sensitivity analysis, excluding mammogram findings 6 months after a woman became nonadherent to study medication use. In these analyses, mammogram diagnostic performance for women in the CEE group was not significantly different from those in the placebo group for any period examined (P = .38, .49, and .06 in years 1 to 2, 3 to 4, and ≥ 5, respectively).

interference with mammographic performance after estrogen-alone initiation, extra diligence in review of mammograms obtained in this setting also could be recommended.

In conclusion, use of CEE alone for about 5 years results in approximately one in 11 and one in 50 women who had an otherwise avoidable mammogram with recommendation for short-interval follow-up or a breast biopsy, respectively. Clinically indicated breast biopsies were significantly more frequent among CEE users but less commonly led to breast cancer diagnoses. These findings differ from those with combined CEE plus MPA use, for which abnormal mammograms were increased and for which there was strong evidence of diagnostic delay.

REFERENCES

1. Rossouw JE, Anderson GL, Prentice RL, et al: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the Women's Health Initiative randomized controlled trial. JAMA 288:321-333, 2002

2. Chlebowski RT, Hendrix SL, Langer RD, et al: Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: The Women's Health Initiative randomized trial. JAMA 289:3243-3253, 2003

3. Hersh AL, Stefanick ML, Stafford RS: National use of postmenopausal hormone therapy. JAMA 291:47-53, 2004

4. Ravdin PM, Cronin KA, Howlander B, et al: The decrease in breast cancer incidence in 2003 in the United States. N Engl J Med 356:1670-1674, 2007

5. Kim N, Gross C, Curtis J, et al: The impact of clinical trials on the use of hormone replacement therapy: A population based study. J Gen Intern Med 20:1026-1031, 2005

6. Chlebowski RT, Kuller LH, Prentice RL, et al: Breast cancer after use of estrogen plus progestin in postmenopausal women. N Engl J Med 360:573-587, 2009

7. Chlebowski RT, Anderson G, Pettinger M, et al: Estrogen plus progestin and breast cancer detection with mammography and breast biopsy. Arch Intern Med 168:370-377, 2008

8. Stefanick ML, Anderson GL, Margolis K, et al: Effects of conjugated equine estrogens on breast cancer and mammography in postmenopausal women with hysterectomy. JAMA 295:1647-1657, 2006

9. Anderson GL, Manson J, Wallace R, et al: Implementation of the Women's Health Initiative study design. Ann Epidemiol 13:S5-S17, 2003 (suppl)

10. Hays J, Hunt JR, Hubbel A, et al: The Women's Health Initiative recruitment methods and results. Ann Epidemiol 13:S18-S77, 2003 (suppl)

11. Anderson GL, Limacher M, Assaf AR, et al: Effects of conjugated equine estrogen on postmenopausal women with hysterectomy: The Women's Health Initiative randomized controlled trial. JAMA 291:1701-1712, 2004

12. Yasmeen S, Romano PS, Pettinger M, et al: Frequency and predictive value of mammographic recommendation for short-interval follow-up. J Natl Cancer Inst 95:429-436, 2003

13. American College of Radiology: BI-RADS Mammography (ed 4). http://www.acr.org/SecondaryMainMenuCategories/quality_safety/BIRADSAtlas/BIRAD

14. Curb JD, McTiernan A, Heckberr SR, et al: Outcomes ascertainment and adjudication methods in the Women's Health Initiative. Ann Epidemiol 13:S122-S128, 2003 (suppl)

15. National Cancer Institute: Surveillance Epidemiology and End Results, 2006. http://seer.cancer.gov/

16. Anderson GL, Kooperberg C, Geller N, et al: Monitoring and reporting of the Women's Health Initiative randomized hormone therapy trials. Clinical Trials 4:207-217, 2007

17. Pepe MS: The Statistical Evaluation of Medical Tests for Classification and Prediction. New York, NY, Oxford University Press, 2003

18. Kavanagh AM, Cawson J, Byrnes GB, et al: Hormone replacement therapy, percent mammographic density, and sensitivity of mammography. Cancer Epidemiol Biomarkers Prev 14:1060-1064, 2005

19. Chiarelli AM, Kirsh VA, Klar NS, et al: Influence of patterns of hormone replacement therapy use and mammographic density on breast cancer detection. Cancer Epidemiol Biomarkers Prev 15:1856-1862, 2006

20. Banks E: Hormone replacement therapy and the sensitivity and specificity of breast cancer screening: A review. J Med Screen 8:29-34, 2001

21. 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 138:168-175, 2003

22. Laya MB, Larson EB, Taplin SH, et al: Effect of estrogen replacement therapy on the specificity and sensitivity of screening mammography. J Natl Cancer Inst 88:643-649, 1996

23. Thurfjell E, Holmberg L, Persson I: Screening mammography: Sensitivity and specificity in relation to hormone replacement therapy. Radiology 203:339-341, 1997

24. Banks E, Reeves G, Beral V, et al: Hormone replacement therapy and false positive recall in the Million Women Study: Patterns of use, hormonal constituents and consistency of effect. Breast Cancer Res 8:R8, 2006

25. Ginsburg OM, Martin LJ, Boyd NF: Mammographic density, lobular involution, and risk of breast cancer. Br J Cancer 99:1369-1374, 2008

26. McTiernan A, Chlebowski RT, Martin C, et al: Conjugated equine estrogen influence on mammographic density in postmenopausal women in the Women's Health Initiative Randomized Trial. J Clin Oncol 27:6135-6143, 2009

27. McTiernan A, Martin CF, Peck JD, et al: Estrogen plus progestin use and mammographic density in postmenopausal women: Women's Health Initiative Randomized Trial. J Natl Cancer Inst 97:1366-1376, 2005

28. Feig SA: Auditing and benchmarks in screening and diagnostic mammography. Radiol Clin North Am 45:791-800, vi, 2007

29. Breast Cancer Surveillance Consortium: NCI-funded Breast Cancer Surveillance Consortium cooperative agreement (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA-69976, U01CA63731, U01CA70040). http://breastscreening.cancer.gov/data/benchmarks/diagnostic/table6.html