Supplementary Materials

Supplementary Table 1. Derivation of the data used for the Breast Cancer Risk Assessment Tool (BCRAT).

Model components	Whites	Blacks	Hispanics	Asians
Population Incidence SEER breast cancer incidence rates used in the online BCRAT	From whites from 1983 to 1987	From Blacks from 1994-1998	From Hispanics from 1990-1996	From Asians from 1998-2002
Relative risks	Data from Breast Cancer Detection Demonstration Project (BCDDP)[1,2] 2,852 US white women (0.3% ≥75 years) diagnosed with invasive or non-invasive breast cancer between 1973-1980 and 3,146 matched controls Unconditional logistic regression, odds ratios used to approximate relative risks	Data from Women's Contraceptive and Reproductive Experiences (CARE) study[4] Data from 1,622 African American women ages 35-64 diagnosed with invasive breast cancer between 1994-1998 and 1,661 controls Unconditional logistic regression; since the estimated effect of age at first live birth was nearly 0 and the interaction with family history was not significant, these terms were omitted	Same as for white women	Data from Asian American Breast Cancer Study (AABCS)[5] Data from 589 women 20-55 years diagnosed with invasive breast cancer between 1983-1987and 589 controls Unconditional logistic regression; interactions between age at first live birth and family history and between age and number of breast biopsies were omitted and the number of affected first-degree relatives was dichotomized (0 vs ≥1).
Attributable risk (1-attributable risk used in the model)	Estimated at 0.5788 for women ≥50 years Derived from the observed exposure distribution of cases (ages 20-54 diagnosed between 1980-1982) in the Cancer and Steroid Hormone Study.[3]	Estimated at 0.7440 for women ≥50 years Derived from the CARE study[4]	Same as for white women	Estimated at 0.5032 for women ≥50 years Derived from the AABSC study[5]
Competing risks (data on deaths from the National Center from Health Statistics)	Data from whites from 1995-2003	Data from Blacks from 1996-2000	Data from Hispanics from 1990-1996	Data from Asians from 1998-2002

References for Table 1:
1. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81:1879-86.
2. Baker LH. Breast Cancer Detection Demonstration Project: five-year summary report. CA Cancer J Clin. 1982;32:194-225.
3. Wingo PA, Ory HW, Layde PM, Lee NC. The evaluation of the data collection process for a multicenter, population-based, case-control design. Am J Epidemiol. 1988;128:206-17.
4. Gail MH, Costantino JP, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in African American women. J Natl Cancer Inst. 2007;99:1782-92.

5. Matsuno RK, Costantino JP, Ziegler RG, et al. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. J Natl Cancer Inst. 2011;103:951-61.

Supplementary Table 2. Breast Cancer Risk Assessment Tool (BCRAT)'s Calibration (Expected/Observed [E/O] breast cancer cases and 95% Confidence Intervals) among Nurses' Health Study (n=71,293) and the Women's Health Initiative-Extension study (n=79,611) Participants by Deciles of Risk.

BCRAT Deciles of Risk	n	Expected breast cancer cases (E)	Observed breast cancer cases (O)	E/O ratios (95% CI)
WHI-ES				
Age 55-74 years				
1	5700	61.6	84	0.73 (0.59-0.91)
2	5701	75.9	67	1.13 (0.89-1.44)
3	5701	83.4	98	0.85 (0.70-1.04)
4	5702	88.8	94	0.94 (0.77-1.16)
5	5700	95.1	97	0.98 (0.80-1.20)
6	5701	102.4	97	1.06 (0.86-1.29)
7	5701	111.6	110	1.01 (0.84-1.22)
8	5701	127.4	140	0.91 (0.77-1.07)
9	5701	158.2	160	0.99 (0.85-1.15)
10	5701	241.3	166	1.45 (1.25-1.69)
Age 75+ years				
1	2235	25.1	25	1.00 (0.68-1.48)
2	2285	31.7	33	0.96 (0.68-1.35)
3	2260	34.3	43	0.80 (0.59-1.07)
4	2261	36.5	33	1.11 (0.79-1.55)
5	2260	39.6	40	0.99 (0.73-1.35)
6	2260	42.9	40	1.07 (0.79-1.46)
7	2261	46.7	41	1.14 (0.84-1.55)
8	2260	53.9	52	1.04 (0.79-1.36)
9	2260	66.9	59	1.13 (0.88-1.46)
10	2260	101.4	69	1.47 (1.16-1.86)
NHS				
Age 57-74 years				
1	5237	65.6	61	1.08 (0.84-1.38)
2	5193	74.9	89	0.84 (0.68-1.04)
3	5197	80.0	72	1.11 (0.88-1.40)
4	5235	85.4	87	0.98 (0.80-1.21)
5	5190	91.0	75	1.21 (0.97-1.52)
6	5209	98.7	93	1.06 (0.87-1.30)
7	5212	105.9	107	0.99 (0.82-1.20)
8	5216	121.9	119	1.02 (0.86-1.23)
9	5211	154.4	116	1.33 (1.11-1.60)
10	5211	229.8	138	1.66 (1.41-1.97)
Age 75+ years				
1	1917	25.6	21	1.22 (0.79-1.87)
2	1914	29.2	25	1.17 (0.79-1.73)
3	1925	31.5	38	0.83 (0.60-1.14)
4	1915	33.8	20	1.69 (1.09-2.62)
5	1912	36.1	39	0.93 (0.68-1.27)
6	1922	38.3	32	1.20 (0.85-1.69)
7	1927	42.7	28	1.52 (1.05-2.21)
8	1912	49.6	35	1.42 (1.02-1.97)
9	1920	61.5	44	1.40 (1.04-1.88)
10	1918	92.2	55	1.68 (1.29-2.18)

Supplementary Table 3. Sample characteristics of Women's Health Initiative-Extension Study participants by whether they were originally enrolled in the clinical trials or the observational study.

	WHI-Clinical Trials Participants			WHI-Observational Study Participants			CT-OS p-value
	Overall	57-74 years	75+ years	Overall	55-74 years	75+ years	Overall
N	46,128	33,610	12,518	50,953	35,131	15,822	
Breast Cancer Risk Assessment Tool Risk Factors							
Age, mean (SD)	71 (6.6)	67 (4.4)	79 (3.1)	71 (7.0)	67 (4.6)	79 (3.2)	
Race							<0.0001
Non-Hispanic White,%	82.3	80.3	87.8	86.9	85.7	89.5	
Non-Hispanic Black,%	9.2	10.2	6.4	5.7	6.4	4.1	
Hispanic,%	4.3	5.1	2.3	3.4	3.9	2.5	
Asian,%	2.1	2.2	1.7	2.2	2.2	2.1	
Hawaiian/Pacific Islander,%	0.3	0.3	0.1	0.1	0.2	0.1	
Native American,%	1.1	1.2	0.9	0.8	0.9	0.6	
Unknown,%	0.7	0.7	0.8	0.9	0.7	1.1	
Age at menarche (yrs)							0.01
≤11,%	21.9	23.5	17.7	22.2	23.9	18.6	
12-13,%	55.0	54.5	56.2	55.5	55.4	55.7	
≥14,%	22.8	21.7	25.7	21.9	20.4	25.3	
Unknown	0.3	0.3	0.4	0.4	0.3	0.5	
Age at first birth (yrs)							<0.0001
≤19, %	14.1	16.1	8.8	10.4	12.3	6.3	
20-24,%	39.7	40.7	36.9	38.2	40.1	34.1	
25-29,%	20.4	18.6	25.1	23.2	21.3	27.3	
≥30, %	7.0	6.1	9.4	7.6	6.5	9.9	
Nulliparous	2.4	2.6	1.9	2.6	2.9	2.0	
Unknown	16.4	15.8	18.0	18.0	16.9	20.4	
Number of biopsies							<0.0001
0, %	73.7	73.7	73.6	70.4	70.2	70.8	
1, %	17.0	17.0	16.9	18.0	18.1	17.7	
2+, %	9.4	9.3	9.5	11.6	11.7	11.5	
First-degree relatives with history of breast cancer							<0.0001
0, %	87.4	88.0	85.8	86.0	86.7	84.5	
1, %	11.5	11.0	12.6	12.7	12.3	13.8	
2+, %	1.1	1.0	1.5	1.3	1.1	1.7	
History of atypia on breast biopsy							NA
Yes	0.5	0.6	0.4	NA[b]	NA	NA	
Other Covariates of Interest							
Mammogram in past 2 years							<0.0001
Yes, %	82.4	83.2	80.3	86.5	89.2	80.6	
No, %	17.5	16.7	19.7	13.2	10.5	19.1	
Unknown, %	0.1	0.1	0.1	0.3	0.3	0.3	
Body Mass Index (kg/m^2), %[†]							<0.0001
<20	2.4	2.1	3.2	4.0	3.9	4.4	
20-24	24.5	23.2	28.0	36.0	35.7	36.7	
25-29	35.0	34.0	37.9	34.7	33.8	36.7	
30+	37.0	39.7	29.6	24.5	25.9	21.4	
unknown	1.1	1.0	1.3	0.8	0.7	0.8	
Oophorectomy, %							<0.0001
At least 1 ovary intact	81.2	81.4	80.6	80.7	81.2	79.6	
Bilateral oophorectomy	17.9	17.9	18.0	18.7	18.3	19.5	
At least 1 ovary removed	0.9	0.7	1.4	0.6	0.5	1.0	

Significant Illness							
Diabetes, %	11.9	11.8	12.4	8.0	7.6	8.7	<0.0001
Myocardial infarction, %	3.5	2.5	6.2	3.3	2.2	5.6	0.04
Stroke, %	2.3	1.7	4.0	2.1	1.5	3.5	0.04
Peripheral artery disease, %	1.7	1.4	2.6	1.8	1.4	2.7	0.30
Congestive heart failure, %	2.2	1.6	4.0	2.1	1.3	4.0	0.23
Emphysema, %*	2.7	2.4	3.2	8.6	7.8	10.3	<0.0001
Number of significant illnesses (from list above)							<0.0001
0, %	79.9	81.9	74.4	79.1	81.9	73.1	
1, %	16.6	15.3	20.2	16.8	15.1	20.7	
2+, %	3.5	2.8	5.4	4.1	3.1	6.3	
Outcomes							
Breast cancer diagnosed during study period, %	2.0	2.0	1.9	2.0	2.0	1.9	0.81
Died during study period, %	5.3	3.2	10.9	5.5	2.9	11.1	0.20

* Emphysema and diabetes were based on self-report but congestive heart failure, myocardial infarction, peripheral artery disease, and stroke were physician-adjudicated with medical records by WHI. Emphysema was not reassessed before the WHI-ES for WHI-CT participants. Emphysema was reassessed before the WHI-ES for WHI-OS participants.

† NA=Not available

Supplementary Table 4. Calibration and Discrimination of the Breast Cancer Risk Assessment Tool (BCRAT) among subsets of Women's Health Initiative (WHI) extension study participants*

CALIBRATION	WHI-Clinical Trials Participants				WHI-Observational Study Participants			
	Overall	55-74 years	75+ years	p value	Overall	55-74 years	75+ years	p value
N	38,218	28,054	10,164	age comparison	41,393	28,955	12,438	age comparison
Primary analyses (Expected/observed ratios) †	**E/O 95% CI**	**E/O 95% CI**	**E/O 95% CI**		**E/O 95% CI**	**E/O 95% CI**	**E/O 95% CI**	
BCRAT expected/observed ratios	1.02 (0.95-1.09)	0.99 (0.91-1.07)	1.10 (0.95-1.26)	0.19	1.08 (1.01-1.16)	1.07 (0.99-1.16)	1.10 (0.97-1.25)	0.67
Sensitivity analyses								
Including women with information on atypia	1.01 (0.94-1.09)	0.98 (0.91-1.07)	1.10 (0.95-1.26)	0.19	Not available	Not available	Not available	
DISCRIMINATION								
BCRAT c-statistics	0.577 (0.556-0.597)	0.575 (0.551-0.599)	0.586 (0.545-0.626)	0.64	0.579 (0.559-0.599)	0.581 (0.557-0.605)	0.573 (0.537-0.609)	0.72
Sensitivity analyses								
Including women with information on atypia‡	0.575 (0.554-0.595)	0.572 (0.548-0.596)	0.584 (0.544-0.625)	0.61				

* These are the WHI observational study and clinical trial participants who chose to participate in the extension study in 2005.

† Presence or absence of atypical hyperplasia was captured for the 38,112 WHI-CT participants during the main trial (but not for observational study participants) and we used these data in these analyses.

Supplementary Table 5. Breast cancer incidence among SEER, NHS and WHI.

	Breast Cancer Incidence in Non-Hispanic Whites								
	SEER 1983-1987	SEER 1995-2003	SEER 2006-2010	NHS 2004-2009			WHI EXTENSION STUDY I		
Age	Incidence*	Incidence*	Incidence*	Incidence*	Cases	n	Incidence*	Cases	n
55-59	272.1	334.0	273.8	310.2	81	5,222	278.8	42	3,013
60-64	334.8	397.4	359.2	366.2	289	15,784	423.7	322	15,200
65-69	392.3	448.8	430.7	366.5	291	15,881	389.4	395	20,290
70-74	417.8	489.5	445.4	381.7	275	14,411	426.9	410	19,209
75-79	443.9	546.1	462.3	360.7	223	12,364	420.2	336	15,992
80-84	442.1	482.7	436.3	330.4	109	6,598	352.1	139	7,896
85+	410.9	404.1	365.0	0.0	0	23	375.3	28	1,492

*Average breast cancer incidence per year per 100,000 non-Hispanic White women.

* Abbreviations: SEER: NCI's Surveillance, Epidemiology, and End Results Program; NHS=Nurses' Health Study beginning in 2004; WHI=Women's Health Initiative Extension Study 2005

Supplementary Table 6A. Recent mammography use by history of significant illness among participants that reported on mammography use in NHS and WHI-Extension Study

	NHS No. of significant illnesses (n=60,383)			p value*	WHI No. of significant illnesses (n=96,909)			p value*
Mammography Screening	0 (n=42,240)	1 (n=13,883)	2+ (n=4,260)	<0.0001	0 (n=77,169)	1 (n=16,242)	2+ (n=3,670)	<0.0001
Yes	89.5%	87.2%	83.3%		85.5%	82.8%	77.4%	
No	10.5%	12.8%	16.7%		14.5%	17.2%	22.6%	

* We used the Mantel-Haenszel Test of Trend.

Supplementary Table 6B. Breast cancer tumor size at diagnosis by history of significant illness among NHS and WHI-ES participants with known tumor size

	NHS No. of significant illnesses (n=1,117)			p value	WHI No. of significant illnesses (n=1,790)			p value*
Tumor Size	0 (n=881)	1 (n=193)	2+ (n=43)	0.076	0 (n=1,413)	1 (n=290)	2+ (n=87)	0.20
≤2.0 cm	76.4%	72.5%	67.4%		76.8%	73.1%	75.9%	
2.1 to 4.0 cm	19.3%	20.2%	27.9%		18.9%	22.4%	13.8%	
4.1+ cm	4.3%	7.3%	4.7%		4.3%	4.5%	10.3%	

* We used the Mantel-Haenszel Test of Trend.

Supplementary Table 7. Baseline characteristics, overall and by age, among Women's Health Initiative-Extension Study participants without a history of cancer (n=97,081) and with a history of cancer (n=106,893, excludes women with a history of breast cancer). [*,†]

	WHI-ES no cancer[*]			WHI-ES with history of cancer included[*]		
	Overall	**55-74 years**	**75+ years**	**Overall**	**55-74 years**	**75+ years**
N	97,081	68,741	28,340	106,893	74,871	32,022
Breast Cancer Risk Assessment Tool Risk Factors						
Age, mean (SD)	71 (6.8)	67 (4.5)	79 (3.1)	71 (6.9)	67 (4.5)	79 (3.1)
Race						
Non-Hispanic White,%	84.7	83.0	88.8	85.1	83.3	89.1
Non-Hispanic Black,%	7.4	8.3	5.1	7.2	8.2	4.9
Hispanic,%	3.8	4.5	2.4	3.7	4.3	2.3
Asian,%	2.1	2.2	1.9	2.1	2.2	1.9
Hawaiian/Pacific Islander,%	0.2	0.2	0.1	0.2	0.2	0.1
Native American,%	1.0	1.1	0.7	0.9	1.1	0.7
Unknown,%	0.8	0.7	1.0	0.8	0.7	1.0
Age at menarche (yrs)						
≤11,%	22.1	23.7	18.2	22.1	23.8	18.3
12-13,%	55.2	54.9	55.9	55.3	54.9	56.0
≥14,%	22.4	21.1	25.5	22.3	21.0	25.3
Unknown	0.3	0.3	0.4	0.3	0.3	0.4
Age at first birth (yrs)						
≤19, %	12.2	14.2	7.4	12.2	14.3	7.4
20-24,%	38.9	40.4	35.4	38.9	40.5	35.4
25-29,%	21.9	20.0	26.4	21.7	19.8	26.2
≥30, %	7.3	6.3	9.7	7.3	6.3	9.6
Nulliparous	2.5	2.7	1.9	2.5	2.7	2.0
Unknown	17.2	16.4	19.3	17.3	16.4	19.5
Number of biopsies						
0, %	71.9	71.9	72.0	71.7	71.7	71.8
1, %	17.5	17.5	17.4	17.5	17.6	17.4
2+, %	10.6	10.6	10.6	10.7	10.7	10.8
First-degree relatives with history of breast cancer						
0, %	86.7	87.3	85.1	86.5	87.2	85.0
1, %	12.1	11.7	13.3	12.3	11.8	13.3
2+, %	1.2	1.0	1.6	1.2	1.0	1.7
Outcomes[†]						
Breast cancer diagnosed during study, %	2.0	2.0	1.9	2.0	2.0	1.9
Died during study, %	5.4	3.1	11.0	6.0	3.5	11.8

* WHI=Women’s Health Initiative Extension Study which began in 2005. Women with a history of cancer (excluding breast cancer) were included in the second group.

† Participants were followed for 5 years.

Supplementary Table 8. Calibration and Discrimination of the Breast Cancer Risk Assessment Tool among WHI-ES (n=87,569) participants including women with prior cancer history (excluding a history of breast cancer).

CALIBRATION	WHI-ES*			
	Overall	55-74 years	75+ years	p value
N	87,569	62,080	25,489	age comparison
Primary analyses (Expected/observed ratios) *, †	E/O 95% CI	E/O 95% CI	E/O 95% CI	
BCRAT expected/observed ratios	1.06 (1.01-1.11)	1.03 (0.98-1.09)	1.11 (1.02-1.22)	0.14
DISCRIMINATION				
BCRAT c-statistics	0.577 (0.563-0.590)	0.578 (0.561-0.594)	0.575 (0.550-0.601)	0.88

* These are the WHI observational study and clinical trial participants who chose to participate in the extension study in 2005 who did not have a history of breast cancer.

† We used SEER breast cancer incidence rates for non-Hispanic Whites, Hispanics, and non-Hispanic Blacks from 2006-2010 and from 1998-2002 for Asians.

Supplementary Table 9. Calibration and discrimination of the Breast Cancer Risk Assessment Tool among NHS (n=73,070) and WHI-ES (n=96,059) participants including women with missing data*

CALIBRATION	NHS*				WHI-ES*			
	Overall	57-74 years†	75+ years	p value	Overall	55-74 years	75+ years	p value
N	73,070	53,355	19,715	age comparison	96,317	68,251	28,066	age comparison
Primary analyses (Expected/observed ratios) ‡,§	E/O 95% CI	E/O 95% CI	E/O 95% CI		E/O 95% CI	E/O 95% CI	E/O 95% CI	
BCRAT expected/observed ratios (95% Confidence intervals)	1.20 (1.14-1.27)	1.16 (1.09-1.24)	1.30 (1.17-1.45)	0.03	1.01 (0.96-1.05)	0.98 (0.93-1.04)	1.06 (0.97-1.15)	0.15
DISCRIMINATION‡								
BCRAT c-statistics	0.565 (0.549-0.581)	0.566 (0.548-0.584)	0.566 (0.535-0.596)	0.99	0.573 (0.560-0.586)	0.571 (0.556-0.587)	0.577 (0.552-0.601)	0.72

* NHS=Nurses' Health Study included participants alive in 2004; WHI-ES=Women's Health Initiative Extension Study began in 2005

† The youngest women in NHS in 2004 were 57 years.

‡ We compared the expected (E) number of breast cancers based on BCRAT estimates (calculated using the BCRAT SAS macro) to the observed number (O) in each cohort stratified by age (55-74, 75+).[7] To determine the 95% CI of the E/O ratios, we used the Poisson variance of the logarithm of the observed number of cases.[19] To test whether E/O ratio estimates differed by age within cohort, we used the normal approximation z-test.

Supplementary Table 10. Sample characteristics of Women's Health Initiative-Extension Study participants by whether or not they had missing data on Breast Cancer Risk Assessment Tool Risk Factors

	WHI Ext I		
	Women not missing data on BCRAT risk factors	**Women having missing data on BCRAT risk factors**	**p-value**
N	79,611	16,706	
Age, mean (SD)	71 (6.8)	71 (7.2)	
Race			<0.0001
Non-Hispanic White,%	86.4	80.5	
Non-Hispanic Black,%	7.0	9.6	
Hispanic,%	3.4	6.2	
Asian,%	2.1	2.6	
Hawaiian/Pacific Islander,%	0.2	0.2	
Native American,%	1.0	0.9	
Age at menarche (yrs)			0.5
≤11,%	22.2	22.1	
12-13,%	55.5	55.2	
≥14,%	22.3	22.7	
Age at first birth (yrs)			0.4
≤19, %	14.7	14.2	
20-24,%	47.0	49.7	
25-29,%	26.4	23.4	
≥30, %	8.8	9.5	
Nulliparous	3.0	3.2	
Number of biopsies			0.001
0, %	72.1	71.2	
1, %	17.5	17.5	
2+, %	10.4	11.3	
First-degree relatives with history of breast cancer			0.011
0, %	86.5	87.3	
1, %	12.3	11.5	
2+, %	1.2	1.2	
Breast cancer diagnosed during study period, %	1.9	2.0	0.5
Died during study period, %	5.3	5.8	0.009

* WHI=Women’s Health Initiative Extension Study which began in 2005.

† Participants were followed for 5 years.

DATA DICTIONARY

Below we describe each variable used in the current analyses and how it was defined using data from the Nurses' Health Study (NHS) and Women's Health Initiative (WHI).

Age

NHS –We measured age by calculating the number of months between a participant's birth month and year and month and year of return of each questionnaire. The number of months was divided by 12 to determine age in years.

WHI – We measured age by calculating days since randomization at the time of the event, dividing by 365, and adding to age in years at randomization (CT) or enrollment (OS).

Race/ethnicity

NHS – Race/ethnicity was assessed in the 1992 questionnaire: white, African American, Native American, Asian, Hawaiian/PI and Hispanic or non-Hispanic. The NHS did not capture specific Asian ethnicities (Japanese, Chinese, Filipino, other Asians), used in the Gail model. Therefore, for the 0.83% of Asians in the population, we assigned Asian ethnicity randomly using proportions found in 1970 census (the census before the NHS began, 31% Japanese, 20% Chinese, 15% Filipino, 34% other Asians,1970 census: Characteristics of Populations, Volume I, Chapter B, Table 48.)

WHI – Race/ethnicity (non-Hispanic white, African American, Asian [Japanese, Chinese, Filipino, other Asians] or Pacific Islander, Hispanic, Native American) were assessed at baseline screening. Participants could then indicate multiple ethnicities or no specific ethnicities on a race addendum. If a participant answered that they were Hispanic then they were identified as Hispanic regardless of race. If a participant indicated multiple ethnicities, then we assigned that participant the ethnicity with the highest baseline breast cancer risk according to SEER 1983-1987 used in the BCRAT (Hawaiian, then Native American, then Non-Hispanic Blacks, then Non-Hispanic Whites, then Japanese, then other Pacific Islanders, then Filipino, then Chinese, then other Asians). However, we categorized participants that reported being Non-Hispanic Black and Non-Hispanic White, as being Non-Hispanic Black. When a participant indicated that she was Asian Indian, Korean, Vietnamese or Other Asian, without indicating that she was Hawaiian, Japanese, Filipino or Chinese, we categorized this woman as Other Asian.

Breast cancer

NHS –We included confirmed cases of invasive breast cancer diagnosed within 60 months of return of their 2004 questionnaire or within 60 months of June 2004 for women that did not return their 2004 questionnaire. Eighty-seven percent of self-reported breast cancers in our sample were confirmed by review of pathology reports and 13% were confirmed through follow-up interviews with the nurses themselves.

WHI – We included confirmed cases of invasive breast cancer diagnosed between when a woman gave consent to participate in the Extension study in 2005 and five year follow-up. Self-reported breast cancers were first verified at each local clinical center using medical records, and then reviewed and adjudicated centrally at the WHI Clinical Coordinating Center (CCC) by trained WHI physicians.

Death

NHS –Most deaths are reported by participants' next of kin or by postal authorities. These reports are supplemented by searches of the National Death Index (NDI); Using these methods >98% of deaths have been identified.

WHI – Data on death were obtained from next of kin, postal authorities, and review of the National Death Index, Medicare/Medicaid databases, cancer registries, and large health maintenance organizations databases.

First degree relatives with history of breast cancer (0,1,2+)

NHS – Whether a mother or sister(s) were diagnosed with breast cancer was collected every 4 years from 1976 to 2004. Data on whether daughter(s) were diagnosed with breast cancer was collected in 2000 and 2004.

WHI – Whether a mother, sister(s), daughter(s) were diagnosed with breast cancer was collected at baseline screening. No further updates on family history were collected during the WHI.

Age at first live birth (<20, 20-24, 25-29/nulliparous, >30]

NHS –Assessed in all questionnaires from 1976 to 1984.

WHI – Reproductive history was assessed at Main Study baseline. Age at first live birth was categorized as (<20, 20-29, ≥30); however, the Gail Model categorizes age at first live birth as (<20, 20-24, 25-29/nulliparous, ≥30). The WHI also collected data on age at first term pregnancy, categorized as (<20, 20-24, 25-29, 30-34, 35-39, 40-44, ≥45). Therefore, if age at first live birth was <20, then we defined age at first birth <20 years. If first live birth at age 20-29 and first term pregnancy at 20-24, then we defined age at first birth as 20-24 years. If first live birth was between ages 20-29 and first term pregnancy was between ages 25-29, then we defined age at first birth as 25-29 years. If first live birth was at age ≥30, then we defined age at first birth as ≥30 years. If age at first live birth was missing then we used age at first term pregnancy.

Number of Breast Biopsies (0,1,2+)

NHS – We captured biopsies reported from the 1982 through 2004 questionnaires. Each questionnaire asks if participants were diagnosed with benign breast disease since the last questionnaire and if so when and whether the benign breast disease was confirmed by breast biopsy. If a participant reported never having benign breast disease then they were categorized as never having a breast biopsy. If a participant reported having benign breast disease but it was never confirmed by biopsy then we categorized this participant as not having a breast biopsy.

The 1982 and 1984 questionnaires asked the month and year benign breast disease was diagnosed. From 1986 through 2004, questionnaires categorized date of biopsy confirmed benign breast disease as: before June of the first questionnaire year (e.g. before June 1, 2002), during the two questionnaire years (e.g., June 2002 to May 2004), and after June of the second questionnaire year (e.g., after June 2004). We considered biopsies reported during the current questionnaire period as a biopsy performed during that questionnaire period and we tallied the number of biopsies reported.

Biopsies labeled as being done before or after the questionnaire period were assigned as either had a biopsy "before current questionnaire" or "after current questionnaire." If a participant had biopsy confirmed benign breast disease "before the current questionnaire" followed by biopsy confirmed benign breast disease "after current questionnaire" on a later questionnaire, we categorized these women as having undergone two separate biopsies. If a participant had a biopsy confirmed benign breast disease "before current questionnaire" followed by biopsy confirmed benign breast disease "before current questionnaire" at a later questionnaire, we categorized this biopsy as one biopsy since each of these before responses could be referring to one biopsy performed many years before. If a participant had an "after current questionnaire" followed by a "before current questionnaire" at a later questionnaire, we categorized this biopsy as one biopsy. If a participant had a biopsy after the current time period and then subsequently gave a date during a following questionnaire time period and then gave a "before current questionnaire" in a follow-up questionnaire, we categorized this biopsy as one biopsy. If a participant had a biopsy "after current questionnaire period" followed by a biopsy "after current questionnaire period" we categorized these biopsies as two separate biopsies since a participant must return a questionnaire between these biopsies where she would be able to report that she had a biopsy before the current questionnaire or during the current questionnaire. If a participant reported having a biopsy but never reported the date of the biopsy, then she was categorized as having one biopsy.

Biopsies reported on the 2004 survey were not included if a woman was diagnosed with breast cancer within the first year of follow-up since that biopsy may have been performed as part of diagnosis.

WHI –At baseline participants were asked about their history and number of breast biopsies. Participants were then asked semiannually (CT) or annually (OS) if they had undergone a breast biopsy since their last medical history update questionnaire and the date. Number of breast biopsies were tallied. We excluded biopsies performed within one year of breast cancer diagnosis in case the biopsy was performed as part of diagnosis.

Age at menarche (</=11, 12-13, 14+)

NHS – Assessed in the 1976 questionnaire. Range in ages 6-49. We recoded outliers >21 years as missing observations.

WHI – Reproductive history was collected at baseline screening. Age at first menstrual period was collected as (<10, 11, 12, 13, 14, 15, 16, ≥17).

Atypical Hyperplasia

WHI – Histological sections were reviewed for pathology for CT participants who had at least 1 biopsy during the main study. Participants whose pathology review indicated presence of atypical ductal/lobular hyperplasia are identified as having hyperplasia; CT participants whose biopsy did not indicate presence of atypical hyperplasia are identified as not having hyperplasia; and all other CT participants who did not have a biopsy during the main study are identified as unknown for hyperplasia (consistent with the hyperplasia variable in the Gail model macro). Information on the type of benign lesions was not available for OS participants.

Mammogram in the past 2 years

NHS – In 2004, the NHS asked participants if they had a mammogram in the past 2 years. If so, a participant reported whether the mammogram was performed for screening or due to symptoms. To be consistent with data collected in the WHI, we considered a woman to have undergone a mammogram regardless if it was for screening or diagnostic purposes. Participants that completed the short version of the 2004 questionnaire were not asked to report on mammography use.

WHI –If a participant reported at least one mammogram in the past 2 years, then she was categorized as having undergone mammography. These data are collected semiannually for CT participants and annually for OS participants.

Post-menopausal hormone therapy (HT)

We categorized HT use as estrogen plus progesterone (E+P) or estrogen alone (E-alone) and we reported whether a participant was a current, past, or never user.

NHS – HT use was assessed beginning in 1982. Between 1982-1986, E+P was not differentiated from E-alone. Between 1988-1992, a woman could report if she used E+P and/or E-alone; however, there was only one variable for duration of HT. Between 1994-2004, E+P and E-alone use and duration for each was captured separately. We did not consider vaginal estrogen use alone as having used E-alone. We categorized participants first by whether or not they were currently using HT and then by type.

WHI – HT use and typewas assessed at baseline screening and at start of the extension study for both the CT and OS. For the CT group, we also considered type and duration of active use and we considered non-WHI assigned hormone therapy use reported on participants' current medication list which was assessed annually. For example, a participant may have been randomized to the E-alone arm of the study but also reported using

E+P on her medication list during the main study. For the OS group, we also considered HT use reported on participant's current medication list at three year follow-up.

Significant Illness

NHS-We considered self-reported history of diabetes (assessed since the 1976 questionnaire), myocardial infarction (assessed since the 1976 questionnaire), stroke (assessed since the 1982 questionnaire), peripheral artery disease (assessed since the 1988 questionnaire), emphysema (assessed since the 1990 questionnaire), and congestive heart failure (assessed since the 1998 questionnaire). We did not consider a history of cancer since the NHS only confirms first cancer diagnoses. NHS has confirmed cases of diabetes, myocardial infarction, stroke, and peripheral artery disease through medical record review (when participants gave permission to access their medical records) or through supplementary questionnaires completed by participants. In secondary analyses, we used confirmed diagnoses when available.

WHI – The eligibility screening for both CT and OS assessed a history of diabetes, myocardial infarction, stroke, and congestive heart failure. The baseline medical history questionnaire assessed history of peripheral artery disease, emphysema, and congestive heart failure. During the study, participants in the CT (semiannually) and OS (annually) reported if they developed cancer, myocardial infarction, stroke, peripheral artery disease, and congestive heart failure and these outcomes were adjudicated. Participants also reported if they were diagnosed with diabetes; however, diagnosis of diabetes was not adjudicated. OS participants reported emphysema in annual follow-up. CT participants did not report on emphysema after start of the trial and before the start of the extension study. Report of emphysema was not adjudicated.

Body Mass Index (BMI) (kg/m^2)

NHS – We used the most recent nurse self-reported BMI within 10 years before start of follow up. BMI has been assessed on every questionnaire since 1976 .

WHI – We used the most recent BMI measured and reported on the physical measurement form from the main study.

Oophorectomy

NHS – On every questionnaire participants report whether their ovaries are intact, one ovary was removed, both ovaries were removed, surgery and unknown number of ovaries removed, or missing. We categorized participants as having at least 1 ovary intact bilateral oophorectomy or surgery and unknown number of ovaries removed.

WHI - Participants were asked at baseline screening if they ever had an operation to have one or both ovaries taken out (no, one was taken out, both were taken out, yes and unknown number taken out, yes and part of an ovary was taken out, don't know or missing). If participants did not report that they had surgery (don't know or missing) or reported only part of an ovary was taken out, we considered them to have at least one ovary intact (see our oophorectomy categories above).

Breast cancer tumor size

NHS – Tumor size was abstracted from medical record review (categorized as 0.1-2.0 cm, 2.1-4.0 cm, or 4.1+).

WHI - Tumor size was abstracted through medical record review. To be consistent with NHS we categorized tumor size as (0.1-2.0 cm, 2.1-4.0 cm, or 4.1+ cm).

GAIL MODEL EQUATION

$$P\{a,\ \tau,\ r(t)\} = \int_a^{a+\tau} h_1\ (t) r(t) exp\left(-\int_a^t h_1(u) r(u) du\right) \left\{\frac{exp\left(-\int_0^t h_2(u)\ du\right)}{exp\left(-\int_0^a h_2(u) du\right)}\right\} dt$$

a = age at start of follow up
τ = years of follow up
t = age before end of follow up
r(t) = age specific relative risk

P{a, τ, r(t)} = baseline hazard × relative risk × P(of no breast cancer between a and t) × P(alive at t | alive at a)
Baseline hazard = $h_1(t) = h_1^*(t)\ (1 - AR(t))$
$h_1^*(t)$ = population composite incidence, from Surveillance, Epidemiology and End Results Program
$AR(t)$ = attributable risk, from Gail model
Relative risk = r(t), from Gail model Relative Risks
$h_1(u)$ = Hazards of breast cancer
$h_2(u)$ = Hazards of death due to non-breast cancer related causes, from National Center for Health Statistics

DESCRIPTION OF THE DATA SOURCES

Nurses' Health Study (NHS): The NHS is a longitudinal study of 121,700 US female nurses (97% non-Hispanic white), ages 30-55 in 1976, living in 11 of the most populous US states (California, Connecticut, Florida, Maryland, Massachusetts, Michigan, New Jersey, New York, Ohio, Pennsylvania, or Texas) at baseline. Nurses were chosen to participate because of the higher accuracy of health information that they could provide and potentially greater willingness to participate in a longitudinal health study than a general cohort (71.5% of the 170,000 female nurses invited to participate returned baseline questionnaires). At baseline and in biennial follow-ups, participants provide detailed lifestyle, demographic, and medical history information through mailed questionnaires. Follow-up questionnaires are mailed in June of even-numbered years and non-responders are sent a second mailing in September. A third and fourth questionnaire are sent to those who still have not responded. A fifth mailing of a shorter questionnaire with key exposures and the list of major illnesses is finally sent. Since 1982, women who have not responded to any of the five mailings are telephoned for follow-up. Using these procedures, ~90% of those living complete a questionnaire every 2 years.[1] Data are currently available through 2010. All dates in NHS are recorded as months from the year 1900.

Women's Health Initiative (WHI): The WHI is one of the largest studies on the health of postmenopausal women done to date. Between 1993 and 1998, more than 161,000 women between 50 and 79 years of age were recruited into multiple clinical trial (CT) components or an observational study (OS). Women could enroll in the Dietary Modification Trial (DM, 48,835 participants) and/or the Hormone Replacement Trials (HT) which included The Estrogen plus Progestin (E+P) arm (16,608 participants) and the Estrogen-alone (E-alone) arm (10,739 participants). One year after randomization into the DM, HT, or both, 36,282 women were also randomized for the Calcium and Vitamin D Supplemental Trial (CaD).

Women were eligible for the WHI if they were postmenopausal, aged 50 to 79 years, able and willing to provide written informed consent, and had no plans to move from their residential area for at least 3 years. Women with any medical condition that predicted <3 years survival were excluded. Each trial component had additional exclusion criteria specific to that trial. Women were ineligible for the CTs if they did not want to discontinue taking hormone therapy upon study entry, or had a history of breast cancer; women were ineligible for the dietary component if they already followed a low-fat diet or too frequently ate away from home and they were ineligible for the Calcium/Vitamin D Supplemental trial if they had a history of kidney stones or were unwilling to limit vitamin D intake. Postmenopausal women who were screened for the clinical trials but were ineligible or unwilling to participate in randomization were asked to enroll in the OS.

Participants were recruited from areas surrounding 40 clinical centers established primarily at major academic health centers in 24 states and the District of Columbia. Recruitment areas included urban, suburban, and rural populations. Though not a probability sample, enrollment of racial/ethnic minority groups proportionate to the total minority population of women between 50 and 79 years of age (18.2% according to the 1990 US Census) was a high priority. For most clinics, initial contact with potential participants was through a mass mailing of the recruitment brochure, which provided basic information on the WHI and contained a postage-paid return postcard to indicate interest in study participation. Trained telephone interviewers conducted additional eligibility screening with age-eligible women who returned cards or called the clinical center. A total of three clinical visits were conducted to enroll women in a clinical trial. Women could be found to be ineligible or unwilling for clinical trial enrollment at any point in the screening process and were then offered the opportunity to participate in the observational study. Written informed consent was obtained for all participants.

A complete explanation of WHI data collection is described at https://cleo.whi.org/data/Pages/home.aspx. In brief, self-administered questionnaires were completed by each participant at baseline. All women underwent physical examinations at the time of study enrollment. Questionnaires were administered to collect information regarding race/ethnicity; ages at enrollment, menarche and first birth; number of breast biopsies; number of first-degree relatives with breast cancer; years since menopause, menopausal hormone therapy use; dietary intakes; energy expenditures; lifestyle factors (e.g., smoking); prevalent medical conditions (e.g., diabetes) and medications. During follow-up, women enrolled in the OS completed annual questionnaires and women in the clinical trials completed questionnaires every 6 months to update medical and other lifestyle information.

The original protocol indicated that study close-out was to occur between October 2004 and March 2005. Participants in the Dietary Modification and Calcium/Vitamin D trials remained on intervention until their last clinic visits during this interval. The estrogen plus progesterone intervention was terminated in July 2002 at the recommendation of the WHI Data and Safety Monitoring Board, following the findings that the risks outweighed the benefits for combined hormone use.[2] The estrogen alone intervention was stopped in March 2004 at the direction of the NHLBI based on an increased risk of stroke and the unlikelihood of being able to establish either coronary heart disease benefit or an adverse effect on breast cancer.[3] Participants in both hormone trials continued to be followed through mailed questionnaires every 6 months and annual mammography, through the scheduled close-out visits. Participants in the OS were followed annually by mail until the final cycle of mailings that began in spring of 2004. Additional details on the design of the WHI, WHI participants, and major study findings for the DM and CaD are described elsewhere.[3-11]

At the final transition visit or contact for the CTs which generally occurred between October 2004 and March 2005, WHI participants were invited to join the extension study (ES). Enrollment of OS participants was initiated following the centralized mailing of transition ("close-out") packets in April 2004 to spring 2005. ES participants were required to provide written informed consent. Once enrolled in the ES, participants completed annual data collection forms primarily by mail with follow-up for additional details typically completed by phone. The annual mailing obtained data on self-reported outcomes, hormone use among Hormone Therapy (HT) Trial participants, and quality of life. Women that participated in the HT trials were encouraged to undergo annual mammography for the first two years of the ES. Outcomes data from the WHI includes centrally verified, locally verified and self-reported outcomes occurring on or before September 30, 2010 and adjudicated through March 31, 2011 for all ES participants. Dates in the WHI are presented in days from main study randomization (CT) or enrollment (OS)

References for data description.

1. Colditz GA, Manson JE, Hankinson SE. The Nurses' Health Study: 20-year contribution to the understanding of health among women. J Womens Health 1997;6:49-62.

2. 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 2002;288:321-33.
3. Anderson GL, Limacher M, Assaf AR, et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. JAMA 2004;291:1701-12.
4. Anderson GL, Manson J, Wallace R, et al. Implementation of the Women's Health Initiative study design. Ann Epidemiol 2003;13:S5-17.
5. Hays J, Hunt JR, Hubbell FA, et al. The Women's Health Initiative recruitment methods and results. Ann Epidemiol 2003;13:S18-77.
6. Howard BV, Manson JE, Stefanick ML, et al. Low-fat dietary pattern and weight change over 7 years: the Women's Health Initiative Dietary Modification Trial. JAMA 2006;295:39-49.
7. Jackson RD, Wactawski-Wende J, LaCroix AZ, et al. Effects of conjugated equine estrogen on risk of fractures and BMD in postmenopausal women with hysterectomy: results from the women's health initiative randomized trial. J Bone Miner Res 2006;21:817-28.
8. Prentice RL, Caan B, Chlebowski RT, et al. Low-fat dietary pattern and risk of invasive breast cancer: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. JAMA 2006;295:629-42.
9. Prentice RL, Langer RD, Stefanick ML, et al. Combined analysis of Women's Health Initiative observational and clinical trial data on postmenopausal hormone treatment and cardiovascular disease. Am J Epidemiol 2006;163:589-99.
10. Wactawski-Wende J, Kotchen JM, Anderson GL, et al. Calcium plus vitamin D supplementation and the risk of colorectal cancer. N Engl J Med 2006;354:684-96.
11. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. Int J Epidemiol 1994;23:991-9.
12. 1970 census: Characteristics of Populations, Volume I, Chapter B, Table 48.
13. Stampfer MJ, Willett WC, Speizer FE, et al. Test of the National Death Index. Am J Epidemiol 1984;119:837-9.