Effect of Changing Breast Cancer Incidence Rates on the Calibration of the Gail Model

Sara J. Schonfeld, David Pee, Robert T. Greenlee, Patricia Hartge, James V. Lacey Jr, Yikyung Park, Arthur Schatzkin, Kala Visvanathan, and Ruth M. Pfeiffer

From the Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda; Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health; and Department of Oncology, Johns Hopkins School of Medicine, Baltimore; Information Management Systems, Rockville, MD; and Epidemiology Research Center, Marshfield Clinic Research Foundation, Marshfield, WI.

Submitted July 22, 2009; accepted February 16, 2010; published online ahead of print at www.jco.org on April 5, 2010.

Supported by the Intramural Research Program of the National Institutes of Health and the National Cancer Institute.

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

Corresponding author: Ruth M. Pfeiffer, PhD, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, 6120 Executive Blvd, Bethesda, MD 20892; e-mail: pfeiffer@mail.nih.gov.

© 2010 by American Society of Clinical Oncology

0732-183X/10/2814-2411/\$20.00

DOI: 10.1200/JCO.2009.25.2767

ABSTRACT

Purpose

The Gail model combines relative risks (RRs) for five breast cancer risk factors with age-specific breast cancer incidence rates and competing mortality rates from the Surveillance, Epidemiology, and End Results (SEER) program from 1983 to 1987 to predict risk of invasive breast cancer over a given time period. Motivated by changes in breast cancer incidence during the 1990s, we evaluated the model's calibration in two recent cohorts.

Methods

We included white, postmenopausal women from the National Institutes of Health (NIH) –AARP Diet and Health Study (NIH-AARP, 1995 to 2003), and the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO, 1993 to 2006). Calibration was assessed by comparing the number of breast cancers expected from the Gail model with that observed. We then evaluated calibration by using an updated model that combined Gail model RRs with 1995 to 2003 SEER invasive breast cancer incidence rates.

Results

Overall, the Gail model significantly underpredicted the number of invasive breast cancers in NIH-AARP, with an expected-to-observed ratio of 0.87 (95% CI, 0.85 to 0.89), and in PLCO, with an expected-to-observed ratio of 0.86 (95% CI, 0.82 to 0.90). The updated model was well-calibrated overall, with an expected-to-observed ratio of 1.03 (95% CI, 1.00 to 1.05) in NIH-AARP and an expected-to-observed ratio of 1.01 (95% CI: 0.97 to 1.06) in PLCO. Of women age 50 to 55 years at baseline, 13% to 14% had a projected Gail model 5-year risk lower than the recommended threshold of 1.66% for use of tamoxifen or raloxifene but $\geq 1.66\%$ when using the updated model. The Gail model was well calibrated in PLCO when the prediction period was restricted to 2003 to 2006.

Conclusion

This study highlights that model calibration is important to ensure the usefulness of risk prediction models for clinical decision making.

J Clin Oncol 28:2411-2417. © 2010 by American Society of Clinical Oncology

INTRODUCTION

The Gail model,¹ publicly available in a slightly modified version² as the Breast Cancer Risk Assessment Tool (<http://www.cancer.gov/bcrisktool>), predicts a woman's risk of developing invasive breast cancer over a specified period of time (eg, 5 or 10 years), given her age and risk factor profile. The probability of developing breast cancer is computed by combining relative risks (RRs) estimated from the Breast Cancer Detection Demonstration Project (BCDDP)¹ with attributable risks, age-specific breast cancer incidence rates, and competing mortality rates from all other causes from the Surveillance, Epidemiology, and End Results (SEER) program for the years 1983 to 1987.

Several large chemoprevention studies of tamoxifen (Nolvadex; AstraZeneca, Wilmington, DE) and raloxifene (Evista; Lilly, Indianapolis, IN) used a 5-year Gail model-projected breast cancer risk $\geq 1.66\%$ to determine eligibility.^{3,4} Recent American Society of Clinical Oncology (ASCO) guidelines also refer to this risk threshold in their recommendations for use of these agents and note the common use of this threshold in risk assessment.⁵ The guidelines indicate that premenopausal women and postmenopausal women with low risk of adverse effects and a 5-year projected invasive breast cancer risk $\geq 1.66\%$ may benefit from tamoxifen and/or raloxifene for prevention of estrogen receptor-positive breast cancer.⁵ Inherent in this recommendation is the assumption that the threshold comes from a

well-calibrated model, that is, there is good agreement between the expected numbers of invasive breast cancers predicted by the model and the number of breast cancers that will develop in a population.⁶⁻⁸ If the model is not well calibrated, then the threshold of 1.66% may not be appropriate.

Validation studies of the Gail model have been published previously.^{2,9-12} The model was well calibrated among white women in the placebo arm of the Breast Cancer Prevention Trial who received annual screening² and among regularly screened women with a family history of breast cancer.⁹ The model slightly underpredicted the number of breast cancers in the Nurses' Health Study¹⁰ (6%). Substantial underprediction was reported in two other, more recent, studies: 40% among women with atypical hyperplasia,¹¹ defined somewhat more stringently than in BCDDP, and 20% in the Women's Health Initiative.¹²

U.S. invasive breast cancer incidence rates increased during the 1990s, declined moderately starting in 2000, and substantially dropped in 2003.¹³⁻¹⁵ Recognizing that rate changes may affect model calibration (and therefore also explain some of the discrepancies between previous validation studies), we validated the Gail model among white, postmenopausal women from two large, population-based studies with follow-up through recent calendar years, the National Institutes of Health (NIH) –AARP Diet and Health Study (NIH-AARP), and the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO).

METHODS

Study Population

The present validation analysis was conducted among women from NIH-AARP and PLCO. As described previously in greater detail,¹⁶ the NIH-AARP study is a large, ongoing, prospective study that recruited AARP members from the states of California, Florida, Pennsylvania, New Jersey, North Carolina, and Louisiana and the metropolitan areas of Atlanta, Georgia and Detroit, Michigan from 1995 to 1996. At baseline, it included more than 200,000 women age 50 to 71 years. The PLCO study has also been described previously in detail.¹⁷ In brief, PLCO is a multicenter (n = 10) screening trial that enrolled approximately 77,500 women, age 55 to 74 years at baseline, between 1993 and 2001. Women were eligible if they had no history of lung, colorectal, or ovarian cancer and were neither undergoing cancer treatment nor participating in other screening or prevention trials. Women who had undergone bilateral oophorectomy or were taking tamoxifen were initially ineligible but were later included. From NIH-AARP and PLCO, we included white, postmenopausal women with known parity, no history of in situ or invasive breast cancer, and age younger than 90 years at the start of follow-up. Women with questionnaires completed by proxy were excluded from NIH-AARP. The NIH-AARP study was approved by the National Cancer Institute (NCI) Special Studies institutional review board. PLCO was approved by institutional review boards at the NCI and individual screening centers.

Cancer Ascertainment

In NIH-AARP, invasive breast cancers diagnosed between entry into the cohort and December 31, 2003 were ascertained through linkage with state cancer registries. In PLCO, cancers diagnosed between entry into the study and May 31, 2006 were ascertained through annual mailed questionnaires to participants and subsequently were confirmed by medical records. Reports from physicians and relatives were also collected when available. The National Death Index and cancer registries as available were periodically searched to enhance completeness of end point ascertainment.

Collection of Gail Model Variables

Risk factors included in the Gail model were obtained from baseline questionnaires. The model includes age at menarche (≥ 14 , 12 to 13, or < 12

years), number of previous breast biopsies (0, 1, or ≥ 2), presence of atypical hyperplasia in a biopsy (yes or no), age at first live birth (< 20 , 20 to 24, 25 to 29/nulliparous, or ≥ 30 years), and number of first-degree relatives with a history of breast cancer (0, 1, ≥ 2). NIH-AARP ascertained age at menarche in slightly different categories from those in the Gail model. We therefore randomly reallocated women in the 11-to-12 and 13-to-14 categories of NIH-AARP to the less-than-12 and 12-to-13 and to the 12-to-13 and ≥ 14 categories, respectively, by using the age at menarche distribution from two other cohorts.^{18,19} Number of breast biopsies (PLCO) and atypical hyperplasia (both studies) were not ascertained and were therefore set to unknown in the risk calculations. For each variable, the Gail model assigns the lowest RR to the unknown category. At baseline, NIH-AARP asked whether a woman had affected sisters and/or daughters but did not ask the number of each. If a woman answered yes, we assumed that she had one affected sister and/or one affected daughter.

Statistical Methods

To assess model calibration, we first compared RR estimates from NIH-AARP and PLCO to those used in the Gail risk prediction model. We estimated RRs and corresponding 95% CIs from Cox proportional hazards models. Women were observed from age at baseline questionnaire until age at invasive breast cancer diagnosis or censoring. Censoring events were death, loss to follow-up, or end of follow-up (NIH-AARP, December 31, 2003; PLCO, May 31, 2006). In PLCO, we also censored women at in situ breast tumor diagnosis because of incomplete follow-up of subsequent invasive breast cancer in those women. However, fewer than 1% of women in PLCO had an in situ diagnosis during follow-up, so few subsequent invasive tumors were missed. Of note, we obtained virtually identical RR estimates in NIH-AARP, whether or not we censored at in situ breast tumor diagnosis.

Next, we obtained the absolute risk estimate of invasive breast cancer for each woman from age at baseline until the end of the study, loss to follow-up, age of 90 years, or diagnosis of an in situ breast tumor (in PLCO only) by using her baseline risk factor profile. Model calibration was assessed by comparing the observed number (O) of invasive breast cancers for each cohort with the expected number (E) predicted by the model, obtained by summing individual absolute risk estimates across all women, or for women in a particular risk factor category. CIs for expected-to-observed ratios were calculated by assuming a Poisson distribution for the observed numbers of cases ($95\% \text{ CI} = \frac{E}{O} \exp \pm 1.96 \sqrt{\frac{1}{O}}$). Model discrimination was assessed by using the area under the receiver-operating characteristic curve (AUC).

Finally, we assessed the calibration of an updated model. This model combined the Gail model RRs with SEER invasive breast cancer incidence rates and competing mortality rates from the period corresponding to our cohorts, 1995 to 2003. Computation of the expected number of breast cancers was identical to that of the original Gail model.

RESULTS

In NIH-AARP, 5,665 invasive breast cancers were diagnosed among 181,979 women. In PLCO, 2,223 cancers were reported among 64,868 women. The median age at entry was similar in NIH-AARP (62.8 years) and PLCO (62.3 years). The mean prediction period was slightly shorter in NIH-AARP (7.5 years) than in PLCO (8.6 years). The distribution of Gail model factors was similar between the two cohorts (Table 1). The RRs for age at menarche and number of breast biopsies (NIH-AARP only) were similar to those in the Gail model (Table 2). The Gail model includes an interaction term between number of biopsies and age category (< 50 years and ≥ 50 years) that we could not evaluate, because all women were age 50 years or older. RRs in strata defined by family history and age at first live birth were substantially lower than those used in the Gail model (Table 2). A

Table 1. Distribution of Patients for the Gail Model Risk Factors Among White Women in NIH-AARP and PLCO

Variable	NIH-AARP (n = 181,979)				PLCO (n = 64,868)			
	Patients		Total		Patients		Total	
	No.	%	No.	%	No.	%	No.	%
Age at menarche, years								
7-11	1,253	22.1	39,026	21.5	487	21.9	13,036	20.1
12-13	3,180	56.1	101,178	55.6	1,169	52.6	34,094	54.1
≥ 14*	1,232	21.8	41,775	23.0	567	25.5	16,738	25.8
No. of first-degree relatives with breast cancer and age at first birth < 20 years†								
0‡	672	11.9	27,522	15.1	233	10.5	9,288	14.3
1	130	2.3	3,800	2.1	47	2.1	1,342	2.1
≥ 2	4	0.1	201	0.1	6	0.3	145	0.2
No. of first-degree relatives with breast cancer and age at first birth 20-24 years								
0‡	1,948	34.4	71,316	39.2	810	36.4	26,606	41.0
1	427	7.5	9,788	5.4	178	8.0	3,928	6.1
≥ 2	35	0.6	537	0.3	31	1.4	404	0.6
No. of first-degree relatives with breast cancer and age at first birth 25-29 years or nulliparous								
0‡	1,673	29.5	51,084	28.1	603	27.1	16,122	24.9
1	347	6.1	7,045	3.9	105	4.7	2,458	3.8
≥ 2	33	0.6	369	0.2	16	0.7	212	0.3
No. of first-degree relatives with breast cancer and age at first birth ≥ 30 years								
0‡	335	5.9	9,018	5.0	167	7.5	3,777	5.8
1	59	1.0	1,231	0.7	24	1.1	531	0.8
≥ 2	2	0.04	68	0.0	3	0.1	55	0.1
No. of breast biopsies in age ≥ 50 years								
0§	3,870	68.3	137,836	75.7				
1	1,134	20.0	28,845	15.9				
≥ 2	661	11.7	15,298	8.4				

Abbreviations: NIH-AARP, National Institutes of Health–AARP Diet and Health Study; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

*Includes unknown age at menarche.

†Includes unknown age at first birth.

‡Includes unknown number of first-degree relatives with breast cancer.

§Includes unknown number of breast biopsies.

second interaction term between age at first birth and number of affected first-degree relatives was not significant in either cohort ($P_{\text{interaction}} = .89$ and $.06$ for NIH-AARP and PLCO, respectively).

Overall, the Gail model significantly underestimated the number of breast cancer occurrences in both cohorts (Table 3), by 13% for NIH-AARP (expected-to-observed ratio, 0.87; 95% CI: 0.85 to 0.89) and by 14% for PLCO (expected-to-observed ratio, 0.86; 95% CI, 0.82 to 0.90). Underprediction was also observed in all strata defined by age at menarche (both studies) and number of breast biopsies (NIH-AARP). The Gail model significantly overpredicted the number of breast cancers among women with a family history of breast cancer, particularly among women with a family history of breast cancer who first gave birth before age 25 years.

The overpredictions in subgroups defined by age at first live birth and family history of breast cancer can be explained by lower RR estimates for these factors in the validation cohorts. Overall, however, the RRs of the risk factors used in the model were similar to those observed in our cohorts. We therefore suspected that the substantial, overall underprediction was due to lower age-specific rates used in the Gail model (SEER 1983 to 1987) compared with the rates in our cohorts, particularly for ages younger than 70 years (Table 4).

SEER invasive breast cancer incidence rates from 1995 to 2003 were substantially higher than those of 1983 to 1987 and generally were closer to the age-specific incidence rates of our cohorts (Table 4). The overall, age-adjusted invasive breast cancer incidence rate for SEER 1983 to 1987 was 343.0 per 100,000 person-years.²⁰ The overall, age-adjusted rate for SEER 1995 to 2003 was 389.99 per 100,000 person-years,²¹ which was similar to NIH-AARP (391.9) and somewhat lower than PLCO (422.9).

After updating the absolute risk prediction model by replacing 1983 to 1987 SEER breast cancer incidence rates with those from 1995 to 2003, the overall expected-to-observed ratio was 1.03 (95% CI, 1.00 to 1.05) for NIH-AARP and was 1.01 (95% CI, 0.97 to 1.06) for PLCO, which indicated good calibration (Table 5). The updated model slightly overpredicted cancer occurrence among women in NIH-AARP with an early age at menarche and among women who reported at least two breast biopsies. Similar to the Gail model, the number of cancers among women with a family history of breast cancer was overestimated. Recognizing that invasive breast cancer rates dropped substantially in 2003,¹³⁻¹⁵ we examined calibration of the Gail model restricted to the 2003-to-2006 period among women in PLCO who were alive and had not been diagnosed with a breast cancer as of

Table 2. RR Estimates of Gail Model Risk Factors for White Women in NIH-AARP and PLCO Compared With the RRs Used in the Gail Model

Variable	Gail Model (BCDDP) RR*	NIH-AARP (n = 181,979)		PLCO (n = 64,868)	
		RR	95% CI	RR	95% CI
Age at menarche, years					
7-11	1.21	1.11	1.03 to 1.21	1.15	1.02 to 1.30
12-13	1.10	1.06	1.02 to 1.10	1.07	1.01 to 1.14
≥ 14†	1.00‡	1.00‡		1.00‡	
No. of first-degree relatives with breast cancer and age at first birth < 20 years§					
0	1.00‡	1.00‡		1.00‡	
1	2.61	1.46	1.30 to 1.64	1.59	1.34 to 1.89
≥ 2	6.80	2.13	1.68 to 2.70	2.53	1.80 to 3.57
No. of first-degree relatives with breast cancer and age at first birth 20-24 years					
0	1.24	1.16	1.12 to 1.20	1.20	1.14 to 1.27
1	2.68	1.70	1.57 to 1.84	1.72	1.52 to 1.93
≥ 2	5.78	2.49	2.15 to 2.87	2.45	1.99 to 3.03
No. of first-degree relatives with breast cancer and age at first birth 25-29 years or nulliparous					
0	1.55	1.34	1.25 to 1.44	1.44	1.29 to 1.61
1	2.76	1.97	1.8 to 2.16	1.85	1.60 to 2.14
≥ 2	4.91	2.91	2.49 to 3.40	2.37	1.87 to 3.02
No. of first-degree relatives with breast cancer and age at first birth ≥ 30 years					
0	1.93	1.55	1.39 to 1.72	1.73	1.46 to 2.05
1	2.83	2.30	2.00 to 2.64	1.99	1.60 to 2.48
≥ 2	4.17	3.40	2.64 to 4.39	2.30	1.54 to 3.42
No. of breast biopsies at age ≥ 50 years					
0¶	1.00‡	1.00‡			
1	1.27	1.26	1.21 to 1.31		
≥ 2	1.62	1.59	1.47 to 1.71		

Abbreviations: RR, relative risk; NIH-AARP, National Institutes of Health–AARP Diet and Health Study; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; BCDDP, Breast Cancer Detection Demonstration Project.

*RRs from Breast Cancer Detection Demonstration Project, Table 1¹; 95% CIs not available.

†Includes unknown age at menarche.

‡Reference value.

§Includes unknown age at first birth.

||Includes unknown number of first-degree relatives with breast cancer.

¶Includes unknown number of breast biopsies.

January 1, 2003. The Gail model was well calibrated for that period in PLCO (overall expected-to-observed ratio, 1.00; 95% CI, 0.94 to 1.08). Discriminatory accuracy was unchanged by the recalibration; the overall AUCs for the Gail and updated models were 0.58 in NIH-AARP and 0.59 in PLCO.

We then examined the effect of the recalibration on the identification of women who could potentially benefit from chemoprevention on the basis of current recommendations (Table 6). Of women age 50 to 55 years at baseline, 14.09% (n = 2,978) in NIH-AARP and 13.0% (n = 13) in PLCO had a projected 5-year risk less than 1.66% by using the Gail model but ≥ 1.66% by using the well-calibrated model. On the basis of the Gail model estimates, these women would not be considered for tamoxifen or raloxifene; conversely, on the basis of the well-calibrated model, they might have been eligible and, in fact, benefitted from chemoprevention.⁵

DISCUSSION

The Gail model underestimated the number of invasive breast cancers by 13% to 14% among women age 50 years or older from two large,

contemporary cohorts. As the RRs of the risk factors used in the model were generally similar to those observed in our cohorts, we attributed the overall underprediction to differences between the age-specific breast cancer rates used in the Gail model (SEER 1983 to 1987) and those of our study populations. The overall calibration of the model improved greatly by replacing SEER breast cancer rates from 1983 to 1987 with those from the main follow-up period of our studies, 1995 to 2003, which captured the higher breast cancer incidence rates during this period.¹³ The model was also well calibrated in PLCO when risk was projected for the period of 2003 to 2006. This improvement in model calibration was due, after a substantial drop in invasive breast cancer rates in 2003,¹³⁻¹⁵ to rates that were closer to the SEER rates used for the Gail model than rates during the 1995-to-2003 period.

Our results highlight the importance of validating models for risk prediction in contemporary cohorts and updating model components, including baseline rates and/or RR estimates periodically if necessary. This principle needs to be applied before the model can be used in national and international populations different from those used for model development. To ensure that the Breast Cancer Risk

Table 3. Comparison of Expected to Observed Numbers of Invasive Breast Cancers Among White Women in the NIH-AARP and PLCO Studies Using the Gail Model

Variable	NIH-AARP				PLCO			
	Observed*	Expected†	Expected:Observed		Observed*	Expected†	Expected:Observed	
			Ratio	95% CI			Ratio	95% CI
Overall	5,665	4,934.2	0.87	0.85 to 0.89	2,223	1,910.4	0.86	0.82 to 0.90
Age at menarche, years								
7-11	1,253	1,145.7	0.91	0.87 to 0.97	487	410.2	0.84	0.77 to 0.92
12-13	3,180	2,746.2	0.86	0.83 to 0.89	1,169	1,038.4	0.89	0.84 to 0.94
≥ 14	1,223	1,031.3	0.84	0.80 to 0.89	565	458.1	0.81	0.75 to 0.88
No. of first-degree relatives with breast cancer and age at first birth < 20 years								
0	619	451.2	0.73	0.67 to 0.79	225	163.1	0.72	0.64 to 0.83
1	126	172.5	1.37	1.15 to 1.63	47	61.5	1.31	0.98 to 1.74
≥ 2	3	23.5	7.83	2.53 to 24.29	6	16.8	2.79	1.26 to 6.22
No. of first-degree relatives with breast cancer and age at first birth 20-24 years								
0	1,859	1,504.4	0.81	0.77 to 0.85	803	622.5	0.78	0.72 to 0.83
1	427	473.4	1.11	1.01 to 1.22	178	197.0	1.11	0.96 to 1.28
≥ 2	35	56.8	1.62	1.16 to 2.26	31	43.9	1.42	1.00 to 2.02
No. of first-degree relatives with breast cancer and age at first birth 25-29 years or nulliparous								
0	1,586	1,331.6	0.84	0.80 to 0.88	594	470.8	0.79	0.73 to 0.86
1	347	348.0	1.00	0.90 to 1.11	105	127.9	1.22	1.01 to 1.48
≥ 2	33	33.0	1.00	0.71 to 1.41	16	19.6	1.22	0.75 to 1.99
No. of first-degree relatives with breast cancer and age at first birth ≥ 30 years								
0	320	292.0	0.91	0.82 to 1.02	165	137.9	0.84	0.72 to 0.97
1	59	62.6	1.06	0.82 to 1.37	24	28.1	1.17	0.78 to 1.74
≥ 2	2	5.0	2.51	0.63 to 10.04	3	4.3	1.45	0.47 to 4.48
No. of breast biopsies								
0	3,855	3,355.0	0.87	0.84 to 0.90				
1	1,134	925.5	0.82	0.77 to 0.87				
≥ 2	661	637.0	0.96	0.89 to 1.04				

Abbreviations: NIH-AARP, National Institutes of Health–AARP Diet and Health Study; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

*Observed number of invasive breast cancers. Separate category for unknown for each factor not shown because of small numbers.

†Expected number of invasive breast cancers from the Gail model on the basis of 1983-1987 Surveillance, Epidemiology, and End Results invasive breast cancer incidence rates.

Assessment Tool remains useful to clinicians and researchers to assess breast cancer risk in white women in the United States, an updated version will be available online and will be based on the most recent breast cancer incidence rates from SEER.

Changes in breast cancer incidence rates throughout the 1990s most likely explain differences between the results of this study compared with two previous validation studies^{2,10} that concluded the model was well calibrated. The projection period for both of these

Table 4. Invasive Breast Cancer Incidence Rates per 100,000 Person-Years Among White Women

Age Group, Years	Gail Model (SEER 1983-1987)*	SEER 1995-2003	NIH-AARP (1995-2003)		PLCO (1993-2006)	
			Rate	95% CI	Rate	95% CI
50-54	221.1	262.0	280.2	225.6 to 334.9	—	—
55-59	272.1	334.02 ²¹	358.3	331.5 to 385.2	370.8	322.8 to 418.7
60-64	334.8	397.4	404.3	382.3 to 426.3	419.0	385.5 to 452.6
65-69	392.3	448.8	478.6	457.0 to 500.2	426.5	392.7 to 460.3
70-74	417.8	489.5	475.6	451.0 to 500.2	406.2	369.5 to 442.9
75-79	443.9	516.1	494.0	433.1 to 554.9	421.6	368.9 to 474.3
80-84	—	482.7	—	—	511.3	385.1 to 637.6
85-89	—	404.1	—	—	741.0	0 to 1,768.0

NOTE. Cells with — represent age groups with no cases of breast cancer in that column.

Abbreviations: SEER, Surveillance, Epidemiology, and End Results; NIH-AARP, National Institutes of Health–AARP Diet and Health Study; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

*From Constantino et al² Appendix Table 1. Rates for age groups of 80-84 years and 85-89 years were not presented.

Table 5. Comparison of Expected With Observed Numbers of Invasive Breast Cancers Among White Women in the NIH-AARP and PLCO Studies Using an Updated Model on the Basis of 1995-2003 SEER Breast Cancer Incidence Rates

Variable	NIH-AARP				PLCO			
	Observed*	Expected†	Expected:Observed		Observed*	Expected†	Expected:Observed	
			Ratio	95% CI			Ratio	95% CI
Overall	5,665	5,810.4	1.03	1.00 to 1.05	2,223	2,250.2	1.01	0.97 to 1.06
Age at menarche, years								
7-11	1,253	1,349.9	1.08	1.02 to 1.14	487	483.5	0.99	0.91 to 1.08
12-13	3,180	3,233.6	1.02	0.98 to 1.05	1,169	1,223.0	1.05	0.99 to 1.11
≥ 14	1,223	1,213.9	0.99	0.94 to 1.05	565	539.3	0.95	0.88 to 1.04
No. of first-degree relatives with breast cancer and age at first birth < 20 years								
0	619	532.9	0.86	0.80 to 0.93	225	192.6	0.86	0.75 to 0.98
1	126	202.9	1.61	1.35 to 1.92	47	72.3	1.54	1.16 to 2.05
≥ 2	3	27.4	9.12	2.94 to 28.28	6	19.6	3.26	1.47 to 7.26
No. of first-degree relatives with breast cancer and age at first birth 20-24 years								
0	1,859	1,772.1	0.95	0.91 to 1.00	803	733.8	0.91	0.85 to 0.98
1	427	555.9	1.30	1.18 to 1.43	178	231.5	1.30	1.12 to 1.51
≥ 2	35	66.2	1.89	1.36 to 2.63	31	51.3	1.65	1.16 to 2.35
No. of first-degree relatives with breast cancer and age at first birth 25-29 years or nulliparous								
0	1,586	1,569.7	0.99	0.94 to 1.04	594	555.2	0.93	0.86 to 1.01
1	347	409.3	1.18	1.06 to 1.31	105	150.4	1.43	1.18 to 1.73
≥ 2	33	38.6	1.17	0.83 to 1.65	16	22.9	1.43	0.88 to 2.33
No. of first-degree relatives with breast cancer and age at first birth ≥ 30 years								
0	320	343.7	1.07	0.96 to 1.2	165	162.5	0.98	0.85 to 1.15
1	59	73.6	1.25	0.97 to 1.61	24	33.0	1.37	0.92 to 2.05
≥ 2	2	5.9	2.94	0.74 to 11.77	3	5.1	1.69	0.55 to 5.24
No. of breast biopsies								
0	3,855	3,953.3	1.03	0.99 to 1.06				
1	1,134	1,089.2	0.96	0.91 to 1.02				
≥ 2	661	748.3	1.13	1.05 to 1.22				

Abbreviations: NIH-AARP, National Institutes of Health–AARP Diet and Health Study; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; SEER, Surveillance, Epidemiology, and End Results.

*Observed number of invasive breast cancers. Separate category for unknown for each factor not shown because of small numbers.

†Expected number of invasive breast cancers from the recalibrated Gail model on the basis of 1995-2003 SEER breast cancer incidence rates.

studies began in 1992 and ended in 1997¹⁰ or 1998.² However, breast cancer incidence increased by approximately 2% per year throughout the 1990s among women age 45 to 59 years and 0.6% per year among women age 60 years or older, and incidence began to decline in 2000 and 2001 for the age groups of 45 to 59 years and 60 years or older, respectively.¹³ This moderate decline was followed by a sharp drop in invasive breast cancer incidence rates in 2003 among women age 50 years or older.¹³⁻¹⁵ In our validation cohorts, there was little follow-up

before 1995. All women in NIH-AARP entered the cohort in 1995 or 1996. In PLCO, less than 10% of the women entered the cohort before 1995. Underprediction of the Gail model by approximately 20% was also seen in the Women's Health Initiative on the basis of an analysis of breast cancers diagnosed within 5 years of entry (September 1993 to September 2000) onto the study.¹² However, the authors did not attribute the discrepancy between their results and those of previous validation studies to differences in underlying rates but to more complete case ascertainment that resulted from higher rates of mammographic screening and biopsy.¹²

The overprediction observed among women with a family history of breast cancer may be attributed to lower RR estimates for family history of breast cancer in the validation cohorts than the Gail model. However, family history of breast cancer was reported by less than 15% of women in each cohort; therefore, differences in the RRs likely had a small impact on the overall calibration.

Our validation study has some limitations. We restricted our analysis to postmenopausal women because of limited numbers of premenopausal women in the cohorts. By design, all women in these studies were at least 50 years old at the start of follow-up. We also restricted the analysis to white women because of limited numbers of non-white women in these cohorts and because other models may

Table 6. Comparison of Projected 5-Year Risk Classification Using the Gail and Updated Models Among Women Age 50-55 Years at Entry

Gail Model Risk Classification	Updated Model Risk Classification (%)			
	< 1.66%		≥ 1.66%	
	NIH-AARP	PLCO	NIH-AARP	PLCO
< 1.66%	66.88	70.0	14.09	13.0
≥ 1.66%	0	0	19.03	17.0

NOTE. Comparison of NIH-AARP (n = 21,141 women) and PLCO (n = 100 women).

Abbreviations: NIH-AARP, National Institutes of Health–AARP Diet and Health Study; PLCO, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

better assess breast cancer risk in non-white populations.²² We did not have information about atypical hyperplasia in either cohort or the number of breast biopsies in PLCO; thus, we had to set those variables to unknown in the computation of the absolute risk estimates. This may have contributed to underprediction of risk in a small subset of women, as the Gail model program assigns the lowest RR for the specific variable to the unknown category. However, in a sensitivity analysis that used imputed values for the number of biopsies in PLCO, the Gail model still significantly underpredicted breast cancer risk.

This lack of information also highlights that an important practical limitation to model validation is the availability of model risk factors in large, contemporary cohorts. Although many of the risk factors in the Gail model, such as age at menarche or age at first live birth, are routinely obtained on questionnaires, number of biopsies or absence or presence of benign hyperplasia are not. Ideally, those questions would be added to standard epidemiologic questionnaires.

Strengths of our study include the use of large, contemporary cohorts with good ascertainment of breast cancer occurrences. The age-adjusted invasive breast cancer incidence rates were comparable to that of SEER 1995 to 2003, which suggests that women in the cohorts are representative of U.S. white women in this age group.

In conclusion, the calibration of risk prediction models is sensitive to trends in underlying population rates. In our study, higher baseline rates between 1995 and 2003 in the validation cohorts compared with those used in the model resulted in substantial underprediction of invasive breast cancers. As a consequence of the underprediction during this period, approximately 13% to 14% of women were lower than the recommended thresh-

old of 1.66% for use of tamoxifen or raloxifene; conversely, on the basis of the well-calibrated model, they might have been eligible and, in fact, benefitted from chemoprevention.⁵ This study highlights the importance of using appropriate baseline rates in absolute risk models and the importance of calibration of risk prediction models for clinical decision making.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Sara J. Schonfeld, Patricia Hartge, James V.

Lacey Jr, Kala Visvanathan, Ruth M. Pfeiffer

Financial support: Arthur Schatzkin

Administrative support: Arthur Schatzkin

Provision of study materials or patients: Robert T. Greenlee, James V. Lacey Jr, Arthur Schatzkin

Collection and assembly of data: Robert T. Greenlee, James V. Lacey Jr, Yikyung Park, Arthur Schatzkin

Data analysis and interpretation: Sara J. Schonfeld, David Pee, Patricia Hartge, James V. Lacey Jr, Kala Visvanathan, Ruth M. Pfeiffer

Manuscript writing: Sara J. Schonfeld, Robert T. Greenlee, James V.

Lacey Jr, Yikyung Park, Kala Visvanathan, Ruth M. Pfeiffer

Final approval of manuscript: Sara J. Schonfeld, David Pee, Robert T. Greenlee, Patricia Hartge, James V. Lacey Jr, Yikyung Park, Arthur Schatzkin, Kala Visvanathan, Ruth M. Pfeiffer

REFERENCES

- 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* 81:1879-1886, 1989
- Costantino JP, Gail MH, Pee D, et al: Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst* 91:1541-1548, 1999
- Fisher B, Costantino JP, Wickerham DL, et al: Tamoxifen for prevention of breast cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J Natl Cancer Inst* 90:1371-1388, 1998
- Vogel VG, Costantino JP, Wickerham DL, et al: Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: The NSABP study of tamoxifen and raloxifene (STAR) P-2 trial. *JAMA* 295:2727-2741, 2006
- Visvanathan K, Chlebowski RT, Hurley P, et al: American Society of Clinical Oncology clinical practice guideline update on the use of pharmacologic interventions including tamoxifen, raloxifene, and aromatase inhibition for breast cancer risk reduction. *J Clin Oncol* 27:3235-3258, 2009
- Gail MH, Pfeiffer RM: On criteria for evaluating models of absolute risk. *Biostatistics* 6:227-239, 2005
- Gail MH, Costantino JP, Bryant J, et al: Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. *J Natl Cancer Inst* 91:1829-1846, 1999
- Freedman AN, Graubard BI, Rao SR, et al: Estimates of the number of US women who could

benefit from tamoxifen for breast cancer chemoprevention. *J Natl Cancer Inst* 95:526-532, 2003

9. Bondy ML, Lustbader ED, Halabi S, et al: Validation of a breast cancer risk assessment model in women with a positive family history. *J Natl Cancer Inst* 86:620-625, 1994

10. Rockhill B, Spiegelman D, Byrne C, et al: Validation of the Gail, et al: Model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst* 93:358-366, 2001

11. Pankratz VS, Hartmann LC, Degnim AC, et al: Assessment of the accuracy of the Gail model in women with atypical hyperplasia. *J Clin Oncol* 26:5374-5379, 2008

12. Chlebowski RT, Anderson GL, Lane DS, et al: Predicting risk of breast cancer in postmenopausal women by hormone receptor status. *J Natl Cancer Inst* 99:1695-1705, 2007

13. Glass AG, Lacey JV Jr, Carreon JD, et al: Breast cancer incidence, 1980-2006: Combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. *J Natl Cancer Inst* 99:1152-1161, 2007

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

15. Clarke CA, Glaser SL, Uratsu CS, et al: Recent declines in hormone therapy utilization and breast cancer incidence: Clinical and population-based evidence. *J Clin Oncol* 24:e49-e50, 2006

16. Schatzkin A, Subar AF, Thompson FE, et al: Design and serendipity in establishing a large cohort with wide dietary intake distributions: The National Institutes of Health-American Association of Retired

Persons Diet and Health study. *Am J Epidemiol* 154:1119-1125, 2001

17. Prorok PC, Andriole GL, Bresalier RS, et al: Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* 21:273S-309S, 2000

18. Doody MM, Freedman DM, Alexander BH, et al: Breast cancer incidence in U.S. radiologic technologists. *Cancer* 106:2707-2715, 2006

19. Schairer C, Byrne C, Keyl PM, et al: Menopausal estrogen and estrogen-progestin replacement therapy and risk of breast cancer (United States). *Cancer Causes Control* 5:491-500, 1994

20. Surveillance, Epidemiology, and End Results (SEER): SEER Stat Database Incidence, SEER 9 Regs Limited-Use, 1973-2005: Katrina/Rita Population Adjustment Linked to County Attributes in Total United States, 1969-2005. Bethesda, MD, National Cancer Institute, DCCPS, Cancer Statistics Branch, released April 2008, based on November 2007 submission

21. Surveillance, Epidemiology, and End Results (SEER): SEER Stat Database Incidence, SEER 13 Regs Limited-Use, 1992-2005: Katrina/Rita Population Adjustment Linked to County Attributes in Total United States, 1969-2005. Bethesda, MD, National Cancer Institute, DCCPS, Cancer Statistics Branch, released April 2008, based on November 2007 submission

22. Gail MH, Costantino JP, Pee D, et al: Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst* 99:1782-1792, 2007