Validation Studies for Models Projecting the Risk of Invasive and Total Breast Cancer Incidence

Joseph P. Costantino, Mitchell H. Gail, David Pee, Stewart Anderson, Carol K. Redmond, Jacques Benichou, H. Samuel Wieand

Background: **In 1989, Gail and colleagues developed a model for estimating the risk of breast cancer in women participating in a program of annual mammographic screening (designated herein as model 1). A modification of this model to project the absolute risk of developing only invasive breast cancer is referred to herein as model 2. We assessed the validity of both models by employing data from women enrolled in the Breast Cancer Prevention Trial.** ***Methods:*** **We used data from 5969 white women who were at least 35 years of age and without a history of breast cancer. These women were in the placebo arm of the trial and were screened annually. The average follow-up period was 48.4 months. We compared the observed number of breast cancers with the predicted numbers from the models.** ***Results:*** **In terms of absolute risk, the ratios of total expected to observed numbers of cancers (95% confidence intervals [CIs]) were 0.84 (0.73–0.97) for model 1 and 1.03 (0.88–1.21) for model 2, respectively. Within the age groups of 49 years or less, 50–59 years, and 60 years or more, the ratios of expected to observed numbers of breast cancers (95% CIs) for model 1 were 0.91 (0.73–1.14), 0.96 (0.73–1.28), and 0.66 (0.52–0.86), respectively. Thus, model 1 underestimated breast cancer risk in women more than 59 years of age. For model 2, the risk ratios (95% CIs) were 0.93 (0.72–1.22), 1.13 (0.83–1.55), and 1.05 (0.80–1.41), respectively. Both models exhibited a tendency to overestimate risk for women classified in the higher quintiles of predicted 5-year risk and to underestimate risk for those in the lower quintiles of the same.** ***Conclusion:*** **Despite some limitations, these methods provide useful information on breast cancer risk for women who plan to participate in an annual mammographic screening program. [J Natl Cancer Inst 1999;91:1541–8]**

Gail et al. *(1)* used data from the Breast Cancer Detection Demonstration Project (BCDDP) to develop a model for estimating the risk of breast cancer for women in a program of annual mammographic screening who have had no previous breast cancer and who have no evidence of breast cancer at the time of their initial screening mammogram. The model estimates the absolute risk (probability) that a woman in a program of annual screening will develop invasive or *in situ* (ductal carcinoma *in situ* [DCIS]) or lobular carcinoma *in situ* [LCIS]) breast cancer over a defined age interval. The risk factors in this model, in addition to age, include age at menarche, age at first live birth, number of previous breast biopsies, presence of atypical hyperplasia on biopsy, and number of affected first-degree relatives. Estimates of the relative risks associated with these factors are combined with estimates from the BCDDP of the baseline hazard and attributable risk to obtain estimates of the probability of developing breast cancer. This model is referred to as model 1. An interactive computer program *(2)* and graphic approaches *(3)* to make risk projections based on model 1 have been distributed to health care providers to assist in counseling. Recently, Gail and Rimer *(4)* proposed using the original model as an aid to counseling women in their forties on when to initiate regular mammographic screening.

Statisticians of the National Surgical Adjuvant Breast and Bowel Project (NSABP) modified model 1 to project the absolute risk of developing only invasive breast cancer *(5)*. This model, referred to as model 2, was used to define eligibility criteria for the Breast Cancer Prevention Trial (BCPT), a trial that demonstrated a reduction in breast cancer risk by almost 50% among women given tamoxifen *(6)*. The modification of model 1 to model 2 was accomplished by substituting age-specific invasive breast cancer rates from the Surveillance, Epidemiology, and End Results (SEER)[1] Program of the National Cancer Institute (NCI) for the breast cancer incidence rates used in the BCDDP and by use of attributable risk estimates from SEER to obtain the baseline hazard rates (*see* "Appendix" section). The NCI has distributed a computer diskette that projects the risk of invasive breast cancer based on model 2 and provides other information relevant to deciding whether a woman would benefit from tamoxifen *(7)*.

In view of the widespread use of these two models for projecting breast cancer risk, it is important to provide data on validity. Gail et al. *(1)* stressed that projections would be most reliable for women who participate in a program of annual screening because model 1 was based on women in annual screening in the BCDDP. With the use of data from the Cancer and Steroid Hormone (CASH) Study *(8)*, they showed that the model would overpredict risk in unscreened younger women. Gail et al. *(1)* and Gail and Benichou *(9,10)* argued that screening allows one to look into the future, effectively aging the woman by the "lead time" of the screening procedure. Thus, since the age-specific incidence of breast cancer increases rap-

Affiliations of authors: J. P. Costantino, S. Anderson, H. S. Wieand, National Surgical Adjuvant Breast and Bowel Project, Pittsburgh, PA, and Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh; M. E. Gail, Division of Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; D. Pee, Information Management Services, Inc., Bethesda, MD; C. K. Redmond, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh; J. Benichou, Department of Biostatistics, University of Rouen Medical School, France.

Correspondence to: Joseph P. Costantino, Dr.P.H., 230 McKee Place, Suite 403, Pittsburgh, PA 15213 (e-mail: costan+@pitt.edu).

See "Notes" following "References."

idly with age, screening increases the observed age-specific incidence, especially in the young. Several studies confirmed that the original model overpredicted risk in young women who were not in a program of regular mammographic screening *(9–12)* and seemed to perform well for women who were being screened regularly *(11)*. Only relatively small numbers of women in regular screening have been studied *(11)*.

The purpose of this study was to assess the validity of the two breast cancer models based on the application to women who were screened annually in the BCPT. Information from the literature pertaining to the validity assessments of these two models as applied to other populations is also included for comparison.

METHODS

The models were evaluated by use of data from the placebo group of the BCPT. To be eligible for the BCPT, women needed to be at least 35 years old with a life expectancy of at least 10 years, to have had no history of invasive breast cancer, to have had a negative mammogram within 180 days before randomization, and to have had a negative breast examination as part of the prerandomization clinical assessment. Women with DCIS were excluded from the BCPT but not those with LCIS. In addition, to be eligible for the BCPT, women under 60 years of age needed to have a projected 5-year risk of invasive breast cancer no less than that of an average 60-year-old woman (1.66%) based on model 2. Other inclusion criteria for the BCPT were as follows: informed consent; no current or planned pregnancy; normal endometrial biopsies if randomized after July 8, 1994, if the uterus was present; no history of pulmonary embolism or deep-vein thrombosis; and no use of estrogen or progesterone replacement therapy, oral contraceptives, or androgens since at least 3 months before randomization.

The BCPT participants included in this assessment were the subset of the placebo group included in the original publication of the BCPT results *(6)* who were white and without a history of LCIS. This population consists of 5969 women. At the time of randomization, 2332 of these women were 49 years of age or less, 1807 50–59 years old, and 1830 were 60 years or older. The average time of follow-up of this population was 48.4 months (range, 1–70 months). About 38% of the women had more than 60 months of follow-up, and about 8% had less than 1 year of follow-up. During the course of follow-up, 155 cases of invasive breast cancer and 49 cases of *in situ* breast cancer were diagnosed. In addition, 59 other women died of causes not related to breast cancer.

Statistical Methods

Two aspects of the risk models, the relative risk function and the absolute risk projection, were considered. The relative risk is the ratio of the age-specific hazard of breast cancer for a woman with given risk factors to the hazard for a woman of the same age without risk factors. The absolute risk is the probability that a woman with given risk factors will develop breast cancer over a defined age interval.

The relative risk function based on model 1 was obtained for the BCPT population from a proportional hazards model *(13)* that included an interaction between number of biopsies and an indicator that age equals or exceeds 50 years. This model had the same functional form for the log hazard as in the model of Gail et al. *(1)*. The estimates based on the BCPT were contrasted to those based on the BCDDP, the CASH Study, and the Nurses' Health Study (NHS). A comparison of the study design and other features of these four investigations is shown in Table 1. The publications of relative risks from the CASH Study and NHS *(9,12)* did not include an estimate for the effect of the diagnosis from a breast biopsy of atypical hyperplasia. Thus, in the comparison of the relative risk estimates from the four studies, data pertaining to the number of breast biopsies were not categorized by presence of atypical hyperplasia.

Projections of the absolute risk of breast cancer for the BCPT women were made by use of models 1 and 2, which incorporate all risk factors, including the diagnosis of atypical hyperplasia. Equations 5 and 6 in the study by Gail et al. *(1)* were used to calculate the absolute risk of breast cancer, p, from age at randomization, a_1, to the age at diagnosis or to last follow-up, a_2 (*see* "Appendix" section). The expected number (E) of breast cancers for a given category of women is then the sum of the values, p, for the women in that category, and E can be compared with the observed number (O) of women with breast cancer in that category. Confidence intervals (CIs) on the ratio of expected to observed numbers (E/O) were obtained by use of the exact theory under the assumption that the Os have a Poisson distribution. This was accomplished by first solving for the 95% CI for the expectation of O, namely, O_L for the lower limit and O_U for the upper limit, then dividing the E by the values of O_L and O_U to obtain the upper and lower CIs for the ratio, respectively. These analyses were performed for the categories of risk factors used in the two models and, as a composite assessment, on categories of predicted breast cancer risk by age. For the latter assessment, it was decided *a priori* to use categories of breast cancer risk based on quintiles of the distribution of the expected risks among the total population for each model. This would provide a reasonable number of categories for assessment with approximately equal numbers of women at risk. The resulting

Table 1. Selected comparative features of the four studies used to assess the validity of the breast cancer prediction model of Gail et al. (*1*)

Feature	BCDDP*	CASH† Study	NHS‡	BCPT§
Study design	Nested case–control study from a multicenter screening program of 284 780 women	Nested population-based, case–control study in 8 Surveillance, Epidemiology, and End Results Program regions (validation based on 83% of women for whom risk factor data were available)	Self-administered questionnaire follow-up study of 115 172 nurses	Multicenter, randomized control clinical trial of women at increased risk for breast cancer (validation based on 5969 women without a history of lobular carcinoma *in situ* in the placebo group)
No. of incident breast cancer cases‖	2852	4715	2396	204
Race	White	White	Predominantly white	White
Age, y¶	31 through 81	20 through 54	29 through 61	35 through 79
Calendar period of follow-up#	1973 through 1980	1980 through 1982	1976 through 1988	1992 through 1998
Prespecified frequency of mammography	Annual	Not prespecified; rare	Not prespecified; rare before 1983	Annual

See (1); BCDDP = Breast Cancer Detection Demonstration Project.

†*See (8)*; CASH (Cancer and Steroid Hormone) study data were used only to check relative risk features of the model.

‡*See (12)*; NHS = Nurses' Health Study.

§*See (6)*; BCPT = Breast Cancer Prevention Trial.

‖Includes all *in situ* and invasive cancers. For the validation of model 2 by use of BCPT data, 49 *in situ* cases were excluded.

¶Age at diagnosis for BCDDP and CASH. Age at baseline for NHS and BCPT.

#Period for the CASH case–control study was the period that the cases were diagnosed.

quintile distributions of predicted breast cancer risk yielded numbers of women in each category that were not exactly the same because of the nature of duplicate values in the distribution. Global chi-square (χ^2) goodness-of-fit tests on the basis of the squared Pearson residuals, $(O - E)^2/E$, were also calculated. All statistical tests were two-sided.

RESULTS

Relative Risks

The logistic model in equation 1 of Gail et al. *(1)* defines multivariate relative risks for the risk factors shown in the first column of Table 2. The factor-specific relative risks as originally developed from the BCDDP by the use of model 1 are provided in the second column of Table 2. To obtain an estimate of the relative risk for a woman with a particular breast cancer risk profile, one multiplies three factor-specific relative risks in Table 2 corresponding to category A (age at menarche), category B (number of biopsies and age), and category C (number of affected first-degree relatives and age at first live birth). For example, a nulliparous 55-year-old woman who began menstruating at age 12 years, who has had one biopsy, and who has one affected first-degree relative has a relative risk of 1.10 × 1.27 × 2.76 = 3.86. Note that the risks associated with number of biopsies are smaller for a woman more than 49 years of age than for a younger woman, reflecting a negative interaction between those factors in the logistic model. Similarly, the risk ratio for a woman with two affected first-degree relatives compared with a woman with no affected first-degree relatives decreases with the age at first live birth, reflecting a negative interaction.

The relative risks from the logistic model were shown to fit the original BCDDP data well, but a more rigorous test is to assess the fit of the model on different datasets. Gail and Benichou *(9)* evaluated the fit of the model to data from the CASH study, and Spiegelman et al. *(12)* reported on women who developed breast cancer in the NHS. The estimates of factor-specific relative risks from these assessments are shown in columns 3 and 4 of Table 2. It should be noted that, since detailed information was not available in the NHS evaluation, the authors *(12)* coded number of biopsies as 0 or 1 for none or one biopsy, unlike Gail et al. *(1)*, who coded 0, 1, or 2 according to whether there were none, one, or two biopsies specimens. Also, the estimates of relative risks for the NHS came from a proportional hazards model with the same functional form for the log hazard as in the logistic model of Gail et al. *(1)*. Our relative risk estimates for the BCPT, also from a proportional hazards model, are shown in the last column of Table 2. Because the analysis of BCPT data is based on only 204 incident breast cancers, the 95% CIs for the relative risks show considerable variability for the point estimates.

With a few exceptions, the data in Table 2 demonstrate good agreement among relative risk estimates obtained from these

Table 2. Comparison of factors affecting relative risk (RR) for total breast cancers (invasive and all *in situ*) estimated from data for four independent studies

	RR: BCDDP*	RR: CASH† Study	RR: NHS‡	RR (95% confidence interval): BCPT§
Age at menarche, y				
⩾14	1.00	1.00	1.00	1.00 (referent)
12–13	1.10	1.14	1.09	1.21 (0.98–1.50)
<12	1.21	1.29	1.20	1.47 (0.96–2.26)
No. of biopsies				
Age <50 y				
0	1.00	1.00	1.00	1.00 (referent)
1	1.70	1.97	1.67‖	1.11 (0.85–1.45)
⩾2	2.88	3.89	N/A‖	1.23 (0.73–2.09)
Age ⩾50 y				
0	1.00	1.00	1.00	1.00 (referent)
1	1.27	1.77	1.72‖	1.34 (1.09–1.65)
⩾2	1.62	3.13	N/A‖	1.80 (1.18–2.73)
No. of affected first-degree relatives				
Age at first live birth <20 y				
0	1.00	1.00	1.00	1.00 (referent)
1	2.61	2.04	1.78	2.13 (1.33–3.39)
⩾2	6.80	4.16	3.17	4.52 (1.78–11.49)
Age at first live birth 20–24 y				
0	1.24	1.21	1.16	1.53 (1.09–2.16)
1	2.68	2.40	2.03	2.52 (1.46–4.35)
⩾2	5.78	4.75	3.55	4.16 (1.88–9.16)
Age at first live birth 25–29 y or nulliparous				
0	1.55	1.47	1.34	2.35 (1.18–4.67)
1	2.76	2.82	2.31	3.00 (1.55–5.80)
⩾2	4.91	5.43	3.98	3.82 (1.84–7.93)
Age at first live birth ⩾30 y				
0	1.93	1.77	1.56	3.60 (1.29–10.08)
1	2.83	3.32	2.64	3.56 (1.60–7.90)
⩾2	4.17	6.20	4.47	3.51 (1.62–7.61)

**From* Table 1 in *(1)* BCDDP = Breast Cancer Detection Demonstration Project.
†*From* analysis in *(9)*, CASH = Cancer and Steroid Hormone.
‡*From* Table 4 in *(12)*, NHS = Nurses' Health Study.
§BCPT = Breast Cancer Prevention Trial.
‖The precise number of biopsies was not available in this study; the RRs displayed are for biopsies ⩾1.

four datasets. Three points can be made. First, the association with age at menarche is similar in all four datasets. Second, each of the datasets indicates a negative interaction between the number of affected first-degree relatives and age at first live birth, a feature also noted by Bondy et al. *(11)*. Last, there is some indication that the nature of the quantitative interaction between age and number of biopsies may be different among the datasets, but all studies indicate an increasing risk of disease with an increasing number of biopsies.

Absolute Risk

The expected versus observed counts for all breast cancers predicted from model 1 are shown in Table 3 according to levels of projected 5-year risk. Overall, 171.34 cancers were expected compared with 204 observed. This corresponds to an expected/observed ratio (*E/O*) of 0.84 (95% CI = 0.73–0.97). When women in the age groups of 49 years or less, 50–59 years, and 60 years or more are considered, the *E/O* ratios (95% CIs) are 0.91 (0.73–1.14), 0.96 (0.73–1.28), and 0.66 (0.52–0.86), respectively. Thus, although model 1 provided reasonable estimates of absolute risk for women under age 60 years, it underestimated risk for women 60 years of age or older. The data shown in the "all ages" category in Table 3 indicate that model 1 underestimated risk for women predicted to be in the lower quintiles of risk. The *E/O* ratios (95% CI) for the lowest to highest quintiles are 0.57 (0.40–0.84), 0.73 (0.52–1.06), 0.67 (0.50–0.93), 0.98 (0.71–1.37), and 1.07 (0.83–1.41), respectively.

Similar analyses were performed for model 2 (Table 4). Overall, 158.99 invasive cancers were predicted compared with 155 observed. This corresponds to an *E/O* ratio (95% CI) of 1.03 (0.88–1.21). The *E/O* ratios (95% CI) by age groups are 0.93 (0.72–1.22), 1.13 (0.83–1.55), and 1.05 (0.80–1.41) for the age groups of 49 years or less, 50–59 years, and 60 or more years, respectively. There is no statistically significant evidence that these *E/O* ratios differed from 1.0. Model 2 predictions by quintiles of projected 5-year breast cancer risk in the "all ages" category show a pattern similar to that found with model 1. The number of cancers is overestimated for women in the highest quintile by 21% and is underestimated for women in the lowest quintile by about 30%. The *E/O* ratios (95% CI) for the lowest to highest quintiles are 0.70 (0.47–1.11), 0.62 (0.44–0.89), 1.36 (0.88–2.22), 1.22 (0.85–1.82), and 1.21 (0.92–1.64), respectively.

To gain additional insight, we calculated *E*'s and *O*'s for categories defined by breast cancer risk factors (Table 5). For this analysis, the data for the number of breast biopsies were also stratified by history of atypical hyperplasia. The results for models 1 and 2 are similar. Agreement between the expected and observed numbers of cancers is good in most categories. The models overestimate risk in women aged less than 50 years with two or more biopsies and in women whose first birth occurred before age 20 years. The models underestimate risk somewhat in women aged less than 50 years with one biopsy and for most categories of women without affected first-degree relatives. However, none of the *E/O* ratios for model 2 exhibit a statistically significant deviation from unity and, for model 1, only the ratios for those less than 50 years of age with one biopsy, those with first live birth between 25 and 29 years of age or nulliparous without affected relatives, and those with an age at menarche less than 12 years show significant deviation from unity.

We also examined summary measures of goodness of fit based on the squared Pearson residuals. Tests were performed by summing over the 15 categories of age group by predicted risk quintiles in Tables 3 and 4 and summing individually over each of the three major categorizations of risk factors in Table 5 (three

Table 3. Comparison of the expected cases of total breast cancer (invasive and all *in situ*) predicted from model 1 to the observed cases among white women in the placebo arm of the Breast Cancer Prevention Trial

Age group, y	Predicted 5-year risk, %	No. of women	Observed (*O*) breast cancers	Expected (*E*) breast cancers	*E/O*	95% confidence intervals
⩽49	<2.32	111	1	1.93	1.93	0.35–76.25
	2.32–2.65	499	11	9.60	0.87	0.49–1.75
	2.66–3.28	521	25	12.89	0.52	0.35–0.80
	3.29–4.73	614	17	19.32	1.14	0.71–1.95
	>4.73	587	29	31.42	1.08	0.75–1.62
	Total	2332	83	75.16	0.91	0.73–1.14
50–59	<2.32	304	8	5.35	0.67	0.34–1.55
	2.32–2.65	468	14	9.80	0.70	0.42–1.28
	2.66–3.28	362	6	8.47	1.41	0.65–3.85
	3.29–4.73	326	13	10.43	0.80	0.47–1.15
	>4.73	347	13	17.69	1.36	0.80–2.56
	Total	1807	54	51.75	0.96	0.73–1.28
⩾60	<2.32	784	21	9.75	0.46	0.30–0.75
	2.32–2.65	232	8	4.77	0.60	0.30–1.38
	2.66–3.28	308	12	7.61	0.63	0.36–1.23
	3.29–4.73	244	9	8.30	0.92	0.49–2.02
	>4.73	262	17	14.01	0.82	0.51–1.41
	Total	1830	67	44.44	0.66	0.52–0.86
All ages	<2.32	1199	30	17.03	0.57	0.40–0.84
	2.32–2.65	1199	33	24.17	0.73	0.52–1.06
	2.66–3.28	1191	43	28.97	0.67	0.50–0.93
	3.29–4.73	1184	39	38.05	0.98	0.71–1.37
	>4.73	1196	59	63.13	1.07	0.83–1.41
Grand total		5969	204	171.34	0.84	0.73–0.97

Table 4. Comparison of the expected cases of invasive breast cancer predicted from model 2 to the observed cases among white women in the placebo arm of the Breast Cancer Prevention Trial

Age group, y	Predicted 5-year risk, %	No. of women	Observed (*O*) breast cancers	Expected (*E*) breast cancers	*E/O*	95% confidence intervals
⩽49	⩽1.93	592	13	8.40	0.65	0.38–1.21
	1.94–2.41	512	16	9.20	0.58	0.35–1.01
	2.42–3.10	425	8	9.35	1.17	0.59–2.71
	3.11–4.17	474	7	13.77	1.97	0.95–4.89
	⩾4.18	329	16	15.15	0.95	0.58–1.66
	Total	2332	60	55.87	0.93	0.72–1.22
50–59	⩽1.93	165	3	2.59	0.86	0.29–4.19
	1.94–2.41	523	15	9.25	0.62	0.37–1.10
	2.42–3.10	460	6	10.42	1.74	0.80–4.73
	3.11–4.17	289	10	8.51	0.85	0.46–1.77
	⩾4.18	370	9	17.63	1.96	1.03–4.28
	Total	1807	43	48.40	1.13	0.83–1.55
⩾60	⩽1.93	432	7	5.19	0.74	0.36–1.84
	1.94–2.41	152	3	2.49	0.83	0.28–4.02
	2.42–3.10	312	6	7.37	1.23	0.56–3.35
	3.11–4.17	435	12	12.99	1.08	0.62–2.10
	⩾4.18	499	24	26.68	1.11	0.75–1.74
	Total	1830	52	54.72	1.05	0.80–1.41
All ages	⩽1.93	1189	23	16.18	0.70	0.47–1.11
	1.94–2.41	1187	34	20.95	0.62	0.44–0.89
	2.42–3.10	1197	20	27.14	1.36	0.88–2.22
	3.11–4.17	1198	29	35.27	1.22	0.85–1.82
	⩾4.18	1198	49	59.45	1.21	0.92–1.64
Grand total		5969	155	158.99	1.03	0.88–1.21

Table 5. Expected (*E*) and observed (*O*) cancers for categories defined by breast cancer risk factors among white women in the placebo arm of the Breast Cancer Prevention Trial

	All breast cancers (model 1)				Invasive breast cancer (model 2)			
Variable	Observed (*O*)	Expected (*E*)	*E/O*	95% confidence intervals	Observed (*O*)	Expected (*E*)	*E/O*	95% confidence intervals
Age at menarche, y								
>13	35	29.58	0.85	0.61–1.21	29	29.32	1.01	0.70–1.51
12–13	104	95.62	0.92	0.76–1.13	81	88.68	1.09	0.88–1.38
<12	65	46.14	0.71	0.56–0.92	45	40.99	0.91	0.68–1.25
No. of biopsies and atypical hyperplasia (AH)								
Age <50 y								
0	25	21.83	0.87	0.59–1.35	23	16.43	0.71	0.48–1.13
1 (without AH)	27	17.02	0.63	0.43–0.96	19	12.79	0.67	0.43–1.12
1 (with AH)	10	4.18	0.42	0.23–0.87	6	3.07	0.51	0.24–1.39
⩾2 (without AH)	18	23.94	1.33	0.84–2.24	10	17.61	1.76	0.96–3.67
⩾2 (with AH)	3	8.18	2.73	0.93–13.23	2	5.97	2.99	0.83–24.65
Age ⩾50 y								
0	54	44.15	0.82	0.63–1.09	41	47.22	1.15	0.85–1.60
1 (without AH)	25	19.83	0.79	0.54–1.23	21	21.81	1.04	0.68–1.68
1 (with AH)	4	4.12	1.03	0.40–3.78	3	4.37	1.46	0.50–7.06
⩾2 (without AH)	30	21.81	0.73	0.51–1.08	22	23.56	1.07	0.71–1.71
⩾2 (with AH)	8	6.27	0.78	0.40–1.81	8	6.19	0.77	0.39–1.79
No. of affected first-degree relatives								
Age at first live birth <20 y								
0	0	1.64	—	—	0	1.59	—	—
1	7	11.39	1.63	0.79–4.05	6	10.13	1.69	0.78–4.60
⩾2	7	11.27	1.61	0.78–4.00	7	10.09	1.44	0.70–3.59
Age at first live birth 20–24 y								
0	11	6.04	0.55	0.31–1.10	7	6.12	0.87	0.42–2.17
1	44	34.43	0.78	0.58–1.08	38	32.09	0.84	0.62–1.19
⩾2	28	23.50	0.84	0.58–1.26	20	22.43	1.12	0.73–1.84
Age at first live birth 25–29 y or nulliparous								
0	19	10.33	0.54	0.35–0.90	15	10.52	0.70	0.43–1.25
1	46	39.60	0.86	0.65–1.18	33	35.68	1.08	0.77–1.57
⩾2	24	17.78	0.71	0.50–1.16	18	16.45	0.91	0.58–1.54
Age at first live birth ⩾30 y								
0	5	3.04	0.61	0.26–1.87	2	3.01	1.51	0.42–12.43
1	12	9.31	0.78	0.44–1.50	8	7.99	1.00	0.51–2.31
⩾2	1	3.03	3.03	0.54–119.49	1	2.89	2.89	0.52–114.15

categories of age at menarche, 10 categories of number of biopsies by hyperplasia status, and 12 categories of age at first live birth by number of affected relatives). Summing over the 15 categories in Table 3, we found a chi-square of 36.60 for model 1, indicating a lack of fit ($P = .0014$). The lack of fit for model 1 arises mainly in women more than 59 years of age and is due principally to the lower composite rates of breast cancer observed in the BCDDP population for such women (*see* "Appendix Table 1"). For model 2, the corresponding chi-square calculated from Table 4 was not statistically significant ($\chi^2 = 22.45$; $P = .097$). Likewise, for model 2, none of the goodness-of-fit tests based on the three major categorizations of risk factors in Table 5 yielded statistically significant evidences of a lack of fit ($P = .78$, .092, and .66, respectively). There was statistically significant evidence of a lack of fit for model 1 in the three categorizations in Table 5 ($P = .024$, .003, and .009, respectively). However, when women more than 59 years of age were excluded from the evaluation of model 1, the goodness-of-fit tests based on Table 3 and on the age at first live birth categories of Table 5 were no longer statistically significant.

Discussion

We have evaluated a model for projecting invasive and *in situ* breast cancer risk (model 1) and a model for projecting only invasive breast cancer risk (model 2) with the use of data from the placebo arm of the BCPT. We found good overall agreement between expected and observed counts of invasive breast cancer for model 2 (158.99 versus 155), validating the absolute risk projections over an average 4 years of follow-up. Model 2 also showed relatively good agreement between expected and observed counts in each of the age categories of 49 or less years, 50–59 years, and 60 or more years (55.87 versus 60, 48.40 versus 43, and 54.72 versus 52, respectively). Model 1 underestimated the risk of all breast cancers in women more than 59 years of age (44.44 expected versus 67 observed), but observed and predicted counts were in reasonable agreement for women younger than 60 years of age (137 versus 126.91). When predicting risk in the lower quintiles of 5-year risk, these models tended to underestimate risk; when predicting risk in the higher quintiles, they tended to overestimate risk. These deviations may partly represent random variation and partly reflect systematic biases in the multivariate regression models at the extreme levels of breast cancer risk. Considering all of the comparisons by categories of risk factors shown in Tables 3–5, relatively few *E/O* ratios for either model deviated significantly from unity. Global goodness-of-fit tests for model 2 do not demonstrate lack of agreement between observed and expected counts of invasive breast cancer. However, global goodness-of-fit tests and a comparison of the total observed and expected counts indicate that model 1 sometimes underestimated the risk of *in situ* and invasive disease, especially in women more than 59 years of age.

The main difference in the performance between models 1 and 2, which employ the same relative risk function, arises because composite age-specific rates among women more than 65 years old in the BCDDP population *(1)* were lower than in the SEER population (*see* "Appendix Table 1"). One might have expected somewhat higher rates in the BCDDP because invasive plus *in situ* cancers were counted. Perhaps the differences are partly due to random variation because the BCDDP rates were based on small numbers of cancers among older women [Table 3 in *(1)*]. Perhaps the initial BCDDP screening lowered incidence rates in years 2 and 3 of BCDDP follow-up (the years used for model 1 rates), having a greater effect in older women for whom the screening lead time is greater than in younger women. In any case, the results for model 2 indicate that use of general population SEER rates was appropriate for projecting invasive breast cancer risk. On the basis of this finding, the NCI has developed a personal computer-based software package that can be used to predict a woman's risk of invasive breast cancer from model 2. This package is available without charge and has been given to health care providers throughout the United States *(7)*.

Both models 1 and 2 predict absolute risk relatively well for women under age 60 years in the BCPT population. These findings differ from those of Spiegelman et al. *(12)*, who noted an *E/O* ratio of 1.47 with model 1 for women aged 49 years or less, which is larger than the value 0.91 for model 1 (Table 3) and 0.93 for model 2 (Table 4) seen in the BCPT population. Spiegelman et al. *(12)* analyzed NHS follow-up data for the period of 1976 through 1988. Very few women received screening mammography in the United States until the early 1980s *(14)*, and women in the NHS were not in a program of regular screening. As argued elsewhere *(9,10)*, annual screening could explain why model 1 performs so much better in the BCPT population than in women under age 50 years in the NHS. Bondy et al. *(11)* also found that model 1 overpredicted risk in women who did not adhere to American Cancer Society screening guidelines but not in those who adhered to the guidelines.

One aspect that may need further evaluation is the magnitude of the interaction between age and number of biopsies. In the past 20 years, less invasive biopsy procedures such as needle biopsy have come into use. This change may have induced more younger women with minimal evidence of disease to receive biopsies than in the 1970s. Since the 1980s, more widespread use of mammography may have also increased the use of biopsies for younger women with minimal evidence of disease. These factors might explain why the number of biopsies in women under age 50 years was less indicative of increased risk in the BCPT than in the BCDDP and in the CASH Study populations. A comparison of expected and observed frequencies of breast cancer in the BCPT for both models (Table 5) indicates that estimates of the relationship between age and number of biopsies is rather good for those 50 years of age or older but less accurate for those under 50 years of age. This may reflect changes in the use and nature of biopsies among younger women.

These validation studies on the basis of the BCPT data are subject to several limitations. First, the predictions could only be tested over a maximum follow-up period of about 6 years. It would be beneficial to test over longer follow-up periods. Second, the population in the BCPT was a high-risk population. It would be useful to have validation studies from a more representative sample of women in regular follow-up, including women with an estimated 5-year breast cancer risk less than 1.66%, the BCPT eligibility criterion. Nonetheless, the results from the BCPT are pertinent to women who are at high risk and are likely to seek counseling for breast cancer risk. Third, although the numbers of cancers observed in the BCPT are not small, larger numbers would be of value for evaluating models 1 and 2 in subgroups. Fourth, data are needed to assess the performance of these models in minority populations. Model 1 was based on the occurrence of breast cancer in white women. The NSABP statisticians, with the assistance of Gail, incorpo-

rated factors into model 2 to provide predictions for black women (*see* "Appendix" section). Among the 99 black women without a history of LCIS in the placebo arm of the BCPT, only one developed invasive breast cancer (the corresponding expected number was 0.90 cases). Thus, an in-depth assessment of the predictions from model 2 for black women was not possible, and there was even less information for other non-Caucasian women. More extensive validation for non-Caucasian women is needed before determinations can be made regarding the accuracy of predictions for this group. However, recently published data for Hispanic women *(15)* suggest that risk projections for white women may overestimate breast cancer risk among Hispanic women.

We conclude from these data that models 1 and 2 can provide useful information to assist in counseling women who are thought to be free of breast cancer following an initial screening examination with mammography and who plan to participate in a program of regular mammographic screening. The information is useful for counseling women who may be contemplating risks and benefits of preventative strategies, such as bilateral mastectomy or tamoxifen therapy. Such data may also be useful to allay unwarranted fears. Typically, women substantially overestimate their risk of getting breast cancer *(16)*. Women also overestimate their 10-year risk of death from breast cancer by as much as 20-fold *(17)*. Providing breast cancer risk estimates during counseling will help women understand the true nature of their risk and to put it into proper perspective.

As stressed elsewhere *(9,10)*, these models do not include certain risk factors that can modify risk substantially. For example, a woman who just migrated from rural China has a lower risk than implied by models 1 and 2, and a woman known to carry a disease-producing mutation of the BRCA1 or BRCA2 genes has a higher risk. The models will tend to overpredict risk in young unscreened women. Some women will have lower than predicted risk if they initiate treatment with agents such as tamoxifen *(6)*. Thus, these models are the most useful when they are employed by an experienced health care provider who is aware of the limitations of the models and the medical context.

APPENDIX

Equations to Predict Absolute Risk of Breast Cancer

The full details of the equations used to predict breast cancer risk are provided by Gail et al. *(1)*. The probability that a woman who is age a and who has age-dependent relative risk $r(t)$ will develop breast cancer by age a + is

$$\Pr\{a,\tau,r(t)\} = \int_a^{a+\tau} h_1(t)r(t)\exp\{-\int_a^t h_1(u)r(u)du\}\{S_2(t)/S_2(a)\}dt,$$

where $h_1(t)$ is the baseline age-specific hazard of developing breast cancer and where

$$S_2(t) = \exp\{-\int_0^t h_2(u)du\}$$

is the probability of surviving competing risks up to age t.

The baseline age-specific hazard rates were obtained from the average ("composite") age-specific breast cancer rates $h^*_1(t)$ using $h_1(t) = h^*_1(t)\ F(t)$, where $F(t)$ is 1 minus the attributable risk fraction for age t.

Parameters Used in Equations for Models 1 and 2

The above equations were used to make projections for both model 1 and model 2. However, the baseline hazard rates of model 2 differed from those of model 1 for three reasons. First, model 1 was designed to project the risk of all breast cancer, both invasive and *in situ*, while model 2 was designed for the BCPT to project the risk of invasive breast cancer only. Thus, the average breast cancer rates $h^*_1(t)$ used in model 1 were those for the incidence of all breast cancer, while the rates in model 2 were those for only the incidence of invasive breast cancer. Second, model 1 used BCDDP data for the average hazard rates and attributable risk fractions, whereas model 2 used data from the SEER Program. The age-specific rates used in the models are provided in Appendix Table 1. The factor $F(t)$ used in model 1 was 0.5229 for women less than 50 years of age and 0.5264 for women 50 years of age or older. This was based on the relative risks and observed exposure distributions from the cases in the BCDDP population. The factor $F(t)$ for the SEER data used in model 2 was 0.5788 for all age groups. The observed exposure distribution of cases in the CASH Study were used in this instance. For both models, the age-specific relative risk $r(t)$ was based on the logistic regression equation in Gail et al. *(1)* (*see* Table 2). Third, model 1 did not include parameters for predicting risk for black women, while model 2 included modifications to provide such predictions. This was accomplished by using race-specific SEER rates for black women and by developing estimates of the $F(t)$ for black women from the BCDDP population and converting them to estimates for the SEER data by multiplying the BCDDP estimates by the ratio of the $F(t)$ for white women in the BCDDP population to the $F(t)$ for white women in the SEER population. Although no black women were included in the assessment in this article, for completeness, we provide the rates used for black women in Appendix Table 1. The factor $F(t)$ used in model 2 for black women was 0.4146 for women under 50 years of age and 0.4228 for those age 50 years or older.

Appendix Table 1. Age-specific breast cancer incidence rates $h^*_1(t)$ used in risk calculations

	Breast cancer incidence rate per 100 000		
		Model 2†	
Age group, y	Model 1,* white women	White women	Black women
35–39	114.6	66.1	79.6
40–44	203.7	126.5	137.7
45–49	280.8	186.6	165.4
50–54	320.9	221.1	177.9
55–59	293.8	272.1	224.3
60–64	369.4	334.8	275.0
65–69	356.1	392.3	280.3
70–74	307.8	417.8	309.9
75–79	301.3	443.9	360.2

*Rates of total breast cancer (invasive and all *in situ*) from the Breast Cancer Detection Demonstration Project.

†Rates of invasive breast cancer from the Surveillance, Epidemiology, and End Results Program, 1983–1987.

REFERENCES

(1) Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81:1879–86.

(2) Benichou J. A computer program for estimating individualized probabilities of breast cancer [published erratum appears in Comput Biomed Res 1994;27:81]. Comput Biomed Res 1993;26:373–82.

(3) Benichou J, Gail MH, Mulvihill JJ. Graphs to estimate an individualized risk of breast cancer. J Clin Oncol 1996;14:103–10.

(4) Gail M, Rimer B. Risk-based recommendations for mammographic screening for women in their forties. J Clin Oncol 1998;16:3105–14.

(5) Anderson SJ, Ahnn S, Duff K. NSABP Breast Cancer Prevention Trial risk

assessment program, version 2. NSABP Biostatistical Center Technical Report, August 14, 1992.
(6) Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for the prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 study. J Natl Cancer Inst 1998;90:1371–88.
(7) Breast Cancer Risk Assessment Tool for Health Care Providers. Office of Cancer Communication. Bethesda (MD): National Cancer Institute; 1998.
(8) 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.
(9) Gail MH, Benichou J. Assessing the risk of breast cancer in individuals. In: DeVita VT Jr, Hellman S, Rosenberg SA, editors. Cancer prevention. Philadelphia (PA): Lippincott; 1992. p. 1–15.
(10) Gail MH, Benichou J. Validation studies on a model for breast cancer risk [editorial] [published erratum appears in J Natl Cancer Inst 1 994;86:803]. J Natl Cancer Inst 1994;86:573–5.
(11) Bondy ML, Lustbader ED, Halabi S, Ross E, Vogel VG. Validation of a breast cancer risk assessment model in women with a positive family history. J Natl Cancer Inst 1994;86:620–5.
(12) Spiegelman D, Colditz GA, Hunter D, Hertzmark E. Validation of the Gail et al. model for predicting individual breast cancer risk. J Natl Cancer Inst 1994;86:600–7.
(13) Cox DR. Regression models and life tables (with discussion). J R Stat Soc, Series B 1972;45:311–54.
(14) Kessler LG, Feuer EJ, Brown ML. Projections of the breast cancer burden to U.S. women: 1990–2000. Prev Med 1991;20:170–82.
(15) Miller BA, Kolonel LN, Bernstein L, Young JL Jr, Swanson GM, West D, et al., editors. Racial/ethnic patterns of cancer in the United States, 1988–1992. Bethesda (MD): National Institutes of Health, National Cancer Institute; 1996 Report No.: DHHS Publ No. (NIH)96-4104.
(16) Lerman C, Lustbader E, Rimer B, Daly M, Miller S, Sands C, et al. Effects of individualized breast cancer risk counseling: a randomized trial. J Natl Cancer Inst 1995;87:286–92.
(17) Black WC, Nease RF Jr, Tosteson AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. J Natl Cancer Inst 1995;87:720–31.

NOTES

[1]*Editor's note:* SEER is a set of geographically defined, population-based, central cancer registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Registry data are submitted electronically without personal identifiers to the NCI on a biannual basis and the NCI makes the data available to the public for scientific research.

Manuscript received November 13, 1998; revised July 8, 1999; accepted July 23, 1999.