Risk Factors for Breast Cancer According to Estrogen and Progesterone Receptor Status

Graham A. Colditz, Bernard A. Rosner, Wendy Y. Chen, Michelle D. Holmes, Susan E. Hankinson

Background: **Evaluations of epidemiologic risk factors in relation to breast cancer classified jointly by estrogen receptor (ER) and progesterone receptor (PR) status have been inconsistent. To address this issue, we conducted a prospective evaluation of risk factors for breast cancer classified according to receptor status.** ***Methods:*** **During 1 029 414 person-years of follow-up of 66 145 women participating in the Nurses' Health Study from 1980 through 2000, we identified 2096 incident cases of breast cancer for which information on ER/PR status was available: 1281 were ER+/PR+, 318 were ER+/PR−, 80 were ER−/PR+, and 417 were ER−/PR−. We fit a log-incidence model of breast cancer and used polychotomous logistic regression to compare coefficients for breast cancer risk factors in patients with different ER/PR status. To test for differences in risk factor odds ratios based on marginal ER/PR categories, we evaluated ER status controlling for PR status and *vice versa*. The predictive ability of our log-incidence model to discriminate between women who would develop ER+/PR+ breast cancer and those who would not (and similarly for ER−/PR− breast cancer) was evaluated by using receiver operator characteristic curve analysis. All statistical tests were two-sided.** ***Results:*** **We observed statistically significant heterogeneity among the four ER/PR categories for some risk factors (age, menopausal status, body mass index [BMI] after menopause, the one-time adverse effect of first pregnancy, and past use of postmenopausal hormones) but not for others (benign breast disease, family history of breast cancer, alcohol use, and height). The one-time adverse association of first pregnancy with incidence was present for PR− but not for PR+ tumors after controlling for ER status ($P = .007$). However, the association of BMI after menopause with incidence was present for PR+ but not PR− tumors ($P = .005$). Statistically significant differences in the incidence of ER+ and ER− tumors were seen with age, both before and after menopause ($P = .003$), and with past use of postmenopausal hormones ($P = .01$). Area under the receiver operator characteristic curve, adjusted for age, was 0.64 (95% confidence interval [CI] = 0.63 to 0.66) for ER+/PR+ tumors and 0.61 (95% CI = 0.58 to 0.64) for ER−/PR− tumors.** ***Conclusions:*** **Incidence rates and risk factors for breast cancer differ according to ER and PR status. Thus, to accurately estimate breast cancer risk, breast cancer cases should be divided according to the ER and PR status of the tumor. [J Natl Cancer Inst 2004;96:218–28]**

Risk factors for the development of breast cancer, such as age, age at menarche, parity, and age at menopause, have been summarized and integrated into comprehensive models of incidence *(1–6).* Previous studies of risk factors for estrogen receptor (ER) status among breast cancer patients have typically considered age alone *(7,8)* or age and risk factors one at a time *(9–20).* Many of these studies, however, did not classify cancer cases jointly by both ER and progesterone receptor (PR) status. Potter et al. *(18)* used a joint classification system and a polychotomous logistic regression approach to analyze the prospective Iowa Women's Health Study, and they reported that parity was inversely associated with ER+/PR+ tumors but not with ER−/PR− tumors. Other studies *(14,16,20,21,22)* have used a case–case approach and have, in large part, observed no statistically significant differences among risk factors for the tumor ER/PR subtypes. However, all of these studies had fewer than 620 cancer cases in the largest subgroup studied (ER+/PR+). In prior studies, few risk factors have shown any consistent difference in association between ER+ and ER− breast cancer, although parity has been shown to be inversely associated with ER+ tumors in some studies *(14–16,18)* but not in others *(13).*

Biomathematical models that relate epidemiologic risk factors to cancer incidence can provide a context in which to study

Affiliations of authors: Cancer Epidemiology Program, Dana-Farber/Harvard Cancer Center, and Channing Laboratory, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA.

Correspondence to: Graham Colditz, MD, DrPH, Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Ave., 3rd Fl., Boston, MA 02115–5899 (e-mail: graham.colditz@channing.harvard.edu).

See "Notes" following "References."

DOI: 10.1093/jnci/djh025

the process of carcinogenesis. These types of models can also summarize the impact of multiple variables and can provide a means to identify areas of research that require more study *(23)*. In more recent clinical applications *(24,25)* of these models, patients at high risk for breast cancer have been identified so that they can be recruited to prevention trials. The classical models of carcinogenesis proposed by Armitage and Doll *(26)* and by Moolgavkar and Knudson *(27)* are the most well known. In addition, Pike et al. *(2)* reviewed the epidemiologic evidence for breast cancer in the early 1980s and proposed a model of tissue aging to account for the relationship between reproductive risk factors and breast cancer incidence. The work of Rosner et al. *(4,5)*, and similar work by Moolgavkar et al. *(1,27)* and Pathak et al. *(3,28)*, extended the analytic approach proposed by Pike et al. *(2)* by relating the timing of reproductive events, which are established risk factors for breast cancer, to the incidence of disease.

In the original Pike model of breast cancer incidence *(2)*, breast tissue age increased at a constant rate c from menarche through first birth. At the time of first birth, there was an immediate increase in breast tissue age (of magnitude k_1) and a corresponding decrease in the rate of breast tissue aging (after first birth) to a rate of $c - d_1$. Breast tissue age then increased at the same rate from first birth through age 40 years, after which time the rate of increase in breast tissue aging diminished linearly until, at menopause, the rate of increase in breast tissue age was d_3 units lower than the rate at age 40 years.

Early versions of the Pike model did not accommodate terms for the spacing of pregnancies, for premenopausal women (who, by definition, have no age at menopause), or for pregnancies after age 40 years. Such problems led Rosner and Colditz *(5)* to consider an alternative class of models (i.e., log-incidence models) in which the natural log of breast cancer incidence is a linear function of time [as compared with the Pike models *(2)*, in which log breast cancer incidence is a linear function of log time or log breast tissue age].

Results from fitting this modified Pike model to breast cancer incidence in the Nurses' Health Study cohort have been reported by Rosner and Colditz *(5)*. Briefly, among nulliparous women, breast cancer incidence was found to increase 8.5% per year before menopause and 5.1% per year after menopause. Depending on the relative magnitude of age at first birth minus age at menarche versus the birth index (defined as the sum of [minimum (age, age at menopause) minus age at *i*th birth]) over all births, parous women may be at either an increased or a decreased risk of breast cancer compared with nulliparous women. The net effect of pregnancy is a short-term increase in breast cancer incidence followed by a subsequent long-term decrease in breast cancer incidence. The magnitude of such changes in incidence for parous women is primarily a function of age at first birth and, to a lesser extent, age at each subsequent birth. Specifically, before menopause, the incidence of breast cancer increases 1.7% for each 1-year increase in age at first birth and 0.4% for each 1-year increase in age at each subsequent birth. This modified model was then further extended to include additional terms for benign breast disease, body mass index (BMI), height, alcohol use, and use of postmenopausal hormones *(6)*.

In this article, we build on our previous work in the modeling of breast cancer incidence to evaluate risk factors for ER+/− and PR+/− tumors using breast cancer modeling approaches that account for the timing of exposure to lifestyle factors. We extend follow-up in the Nurses' Health Study cohort through June 1, 2000, and evaluate established breast cancer risk factors (e.g., age at menarche, parity, age at each birth, age at menopause and type of menopause, use of postmenopausal hormones, alcohol use, history of benign breast disease, and family history of breast cancer) and their association with incident cases of breast cancer according to ER and PR status.

PARTICIPANTS and METHODS

Participant Characteristics

The Nurses' Health Study cohort was established in 1976, when 121 701 female, U.S. registered nurses between the ages of 30 and 55 years responded to a mailed questionnaire that inquired about risk factors for cancer and heart disease, with a specific focus on reproductive history, menopausal status, contraceptive practices, hormone use, cigarette smoking, and use of permanent hair dyes. The details of the establishment of this cohort have been previously reported *(29)*. Briefly, in 1976, women reported their age at first full-term pregnancy and the number of pregnancies lasting 6 months or more. In 1978 this information was updated, and the women were asked to record the ages of their living children. Every 2 years thereafter, follow-up questionnaires have been mailed to the women to bring the information about risk factors up to date and to ascertain whether major medical events have occurred. Reproductive history has been updated through 1984, and other breast cancer risk factors have been updated through 1998. Deaths in the cohort have been reported by family members or identified through the postal service or by a search of the National Death Index. It is estimated that mortality ascertainment in this cohort of women is 98% complete *(30,31)*. This study was approved by the Brigham and Women's Hospital institutional review board for the protection of human subjects.

Identification of Breast Cancer Cases

On each questionnaire, the woman was asked whether breast cancer had been diagnosed and, if so, the date of diagnosis. All women who reported having breast cancer (or the next of kin for decedents) were contacted for permission to review their relevant medical records to confirm the diagnosis. Pathology reports were also reviewed to obtain information on ER and PR status. Cases of invasive breast cancer for which we had a pathology report were included in this analysis. Receptor status was determined by either biochemical or immunoperoxidase assay, with the immunoperoxidase assay more commonly used than the biochemical assay on the more recent breast cancer cases. We excluded 750 breast cancer cases from the analysis because of missing ER and/or PR status. We also excluded cases of *in situ* tumors from the analysis. A total of 2096 incident cases of breast cancer—1281 ER+/PR+ tumors, 417 ER−/PR− tumors, 318 ER+/PR− tumors, and 80 ER−/PR+ tumors—were identified among women for whom complete information on breast cancer risk factors was available.

Population for Analysis

The endpoint for the analysis was incident invasive breast cancer with reported ER and PR status. We excluded from the analysis all women (n = 2270) who reported breast or other cancer (excluding nonmelanoma skin cancer) on the 1976 ques-

tionnaire. This left a total of 119 431 women eligible for follow-up. A total of 105 450 women returned the 1978 questionnaire in which age at each pregnancy was first ascertained. A total of 4204 women were excluded from this cohort because their number of pregnancies reported in 1976 differed by two or more children from the estimated number of pregnancies in 1976 based on reported ages of children in 1978. We also excluded 6993 women whose number of living children, as derived from the reported ages of their children, differed from their parity in 1978 and 2756 women whose number of children in 1978 was less than their reported number of children in 1976. In addition, we excluded 765 women whose age at first birth (estimated from the reported ages of children in 1978) was greater than 3 years plus the age at first birth reported in 1976. Another 763 women whose age at menarche was less than or equal to 8 years of age or greater than or equal to 22 years of age were excluded from the analysis. Reasons for further exclusions included having unknown parity (n = 83), having an age at any birth greater than the age at menopause (n = 671), having an unknown age at menopause (n = 23), being male or an invalid participant in 1976 (n = 73), and having no follow-up beyond 1978 (n = 17). These exclusions left a cohort of 89 102 women eligible for follow-up. From this follow-up cohort, we further excluded women who first became eligible in 1994 or beyond (n = 9), women with unknown duration or type of postmenopausal hormone use (n = 3776), women with unknown weight at age 18 years, women with unknown weight or height in 1976 (n = 8871), and women who had a hysterectomy with either one or no ovaries removed (n = 10 301), because these women do not have a precise age at menopause.

After all exclusions, a total of 66 145 women were followed for 1 029 414 person-years from 1980 through 2000, during which time 2846 cases of incident breast cancer occurred (as noted above, 750 women were omitted because of missing information on ER and/or PR status). Analysis began in 1980, because this is the year when weight at age 18 years and alcohol use were first reported. Compared with that of the population of women used in the model analysis, the breast cancer incidence rate ratio for the excluded subpopulation of women was 0.96 (95% confidence interval [CI] = 0.92 to 1.00), reflecting the fact that the women who were excluded from the analysis were mostly multiparous women with inconsistent pregnancy information (i.e., a low-risk population), women who had a hysterectomy (who generally have an earlier but unknown time of menopause; also considered a low-risk population), and women with unknown BMI at age 18 years.

Description of the Log-Incidence Model of Breast Cancer

We fit our log-incidence model of breast cancer to incident cases of invasive breast cancer that were identified during follow-up of the Nurses' Health Study cohort. The approach to model fitting was to assume that incidence at time $t (I_t)$ is proportional to the number of cell divisions (C_t) accumulated throughout life up to age t, that is, $I_t = kC_t$.

The cumulative number of breast cell divisions is calculated as follows:

$$C_t = C_0 \times \prod_{i=0}^{t-1} (C_{i+1}/C_i) = C_o \times \prod_{i=0}^{t-1} \lambda_i$$

Thus, $\lambda_i = C_{i+1}/C_i$ represents the rate of increase in the number of breast cell divisions from age i to age $i + 1$. Log (λ_i) is assumed to be a linear function of risk factors that are relevant at age i. The set of relevant risk factors and their magnitude may vary according to the stage of reproductive life. The details of the representation of C_i are given in *(6)*. The overall model is given by

$$\begin{aligned}\log I = {} & \alpha + \beta_0(t^* - t_o) + \beta_1 b + \beta_2(t_1 - t_0)b_{l,t-1} + \gamma_1(t - t_m)m_A \\ & + \gamma_2(t - t_m)m_B + \delta_1\, pmh_A + \delta_2\, pmh_B + \delta_3\, pmh_C + \delta_4\, pmh_{cur,t} \\ & + (\delta_4 + \delta_5)\, pmh_{past,t} + \beta_3 BMI_1 + \beta_3^* BMI_2 + \beta_4 h_1 + \beta_4^* h_2 \\ & + \alpha_1 bbd + \alpha_2\, bbd\, t_0 + \alpha_3\, bbd(t^* - t_0) + \alpha_4 bbd\, (t - t_m)m_t \\ & + \phi\, fhx + \beta_5 alc_1 + \beta_5^*\, alc_2 + \beta_5^{**}\, alc_3\end{aligned}$$

where t = age; t_o = age at menarche; t_m = age at menopause; t^* = minimum (age, age at menopause); $m_t = 1$ (if postmenopausal at age t, 0 otherwise); s_t = parity at age t; t_i = age at i^{th} birth, $i = 1,\ldots, s_t$; b = birth index $= \sum_{i=1}^{s_t}(t^* - t_i)b_{it}$; ($b_{it} = 1$ if parity $\geq i$ at age t, 0 otherwise); $m_A = 1$ (if natural menopause, 0 otherwise); $m_B = 1$ (if bilateral oophorectomy, 0 otherwise); $bbd = 1$ (if benign breast disease = yes, 0 otherwise); $fhx = 1$ (if family history of breast cancer in mother or sister = yes, 0 otherwise); pmh_A = number of years on oral estrogen; pmh_B = number of years on oral estrogen and progestin; pmh_C = number of years on other types of postmenopausal hormones; $pmh_{cur,t} = 1$ (if current user of postmenopausal hormones at age t, 0 otherwise); $pmh_{past,t} = 1$ (if past user of postmenopausal hormones at age t, 0 otherwise); BMI_j = BMI at age j (kg/m^2); alc_j = alcohol use (grams) at age j; h = height (inches).

β_o represents the rate of increase in incidence before menopause among nulliparous women with no benign breast disease and no family history. β_1 and β_2 represent modifications to the rate of increase in incidence for parous women according to the number and precise spacing of births. γ_1 and γ_2 represent rates of increase in incidence after menopause according to type of menopause among women without benign breast disease not currently using postmenopausal hormones. δ_1, δ_2, and δ_3 represent modifications to the rate of increase in incidence after menopause among women currently using postmenopausal hormones according to the duration of the specific types of postmenopausal hormones used. δ_4 and δ_5 represent the immediate effect of starting and stopping postmenopausal hormone use on rates of increase in incidence after menopause. ϕ represents the effect of family history of breast cancer on the number of breast cell divisions at birth (i.e., C_o).

The terms for BMI, height, and alcohol use in relation to menopause and postmenopausal use of hormones are summarized below:

$$BMI_1 = \sum_{j=t_0}^{t^*-1}(BMI_j - 21.8) + \sum_{j=t_m}^{t-1}(BMI_j - 24.4)pmh_{cur,j}m_j$$

$$BMI_2 = \sum_{j=t_m}^{t-1}(BMI_j - 24.4)(1 - pmh_{cur,j})m_j$$

$$h_1 = (h - 64.5)(t^* - t_0) + (h - 64.4)\sum_{j=t_m}^{t-1} pmh_{cur,j}m_j$$

$$h_2 = (h - 64.4)\sum_{j=t_m}^{t-1}(1 - pmh_{cur,j})m_j$$

$$alc_1 = \sum_{j=18}^{t^*-1} alc_j$$

$$alc_2 = \sum_{j=t_m}^{t-1} alc_j pmh_{cur,j} m_j$$

$$alc_3 = \sum_{j=t_m}^{t-1} alc_j(1 - pmh_{cur,j})m_j$$

β_3 (β_4) represents the effect of BMI (height) either before menopause or after menopause on breast cancer incidence while currently using postmenopausal hormones. β_3^* represents the effect of BMI (height) after menopause on breast cancer incidence while not using postmenopausal hormones. β_5, β_5^*, and β_5^{**} represent the effects of alcohol before menopause, after menopause while currently using postmenopausal hormones, and after menopause while not using postmenopausal hormones, respectively. The rationale for the separate terms is the finding in exploratory analyses driven by previous literature *(32)* that 1) the effects of BMI and possibly height and alcohol use on breast cancer incidence are different before and after menopause and 2) the effect of BMI on breast cancer incidence after menopause differs according to whether a woman is or is not currently using postmenopausal hormones *(33)*. α_1, α_2, α_3, and α_4 represent modifications, among women with benign breast disease, to 1) the number of breast cell divisions at birth, 2) the rates of increase in the number of cell divisions after birth but before menarche, 3) the rates of increase in the number of cell divisions after menarche but before menopause, and 4) the rates of increase in the number of cell divisions after menopause. The rationale for the extra terms involving benign breast disease ($\alpha_1, \ldots, \alpha_4$) is that the relative risk for benign breast disease varies according to age, is strongest among younger women, and diminishes over time.

The general rationale for a log-incidence model of a specific cancer is that the number of precancerous cells increases multiplicatively with time, but that the risk factor profile from birth through current age differentially affects the rate of increase in incidence. Specifically, in the breast cancer incidence model described above the number of precancerous cells is assumed to increase annually at the rate of $\exp(\beta_0)$ before menopause for nulliparous women, at the rate of $\exp(\beta_0 + \beta_1 s)$ before menopause for parous women with parity = s, and so forth. Finally, the number of precancerous cells increases immediately after the first birth by $\exp[\beta_2(t_1 - t_0)]$. The incidence rate of breast cancer is assumed to be approximately proportional to the number of precancerous cells.

Model Fit and Analyses

The log-incidence model was fit using iteratively reweighted least squares, with PROC NLIN in SAS, version 6.12 *(34)*. The parameters of the model are readily interpretable in a relative risk (RR) context. For example, $\exp(-\beta_0)$ = RR for a 1-year increase in age at menarche among nulliparous women, $\exp[-(\beta_0 + \beta_2)]$ = RR for a 1-year increase in age at menarche among parous women, and so forth.

To evaluate the consistency of risk estimates among the four tumor receptor categories (i.e., ER+/PR+, ER+/PR−, ER−/PR+, and ER−/PR− breast cancers), we ran a model that allowed estimates to vary for all exposure variables using polychotomous logistic regression *(35)*. If β_{1++} (i.e., the effect of the duration of premenopause on ER+/PR+ tumors), β_{1+-} (i.e., the effect of the duration of premenopause on ER+/PR− tumors), β_{1-+}, β_{1--} are defined similarly, then we tested the null hypothesis (H_0): $\beta_{1++} = \beta_{1+-} = \beta_{1-+} = \beta_{1--}$ versus H_1: the effect of the duration of premenopause is different on at least two tumor receptor types; the effects of all other risk factors are also assumed to be different among the four tumor subtypes under either H_0 or H_1. A similar test was performed for all other risk factors (i.e., duration of menopause, pregnancy history, benign breast disease, postmenopausal hormone use, BMI, height, and alcohol use).

On the basis of log-likelihood analyses, we calculated a heterogeneity chi-square and *P* value for each risk factor (Table 1). In performing the heterogeneity analyses, some risk factors were tested as a group (e.g., natural menopause and bilateral oophorectomy) because they are interdependent. To test for differences in risk factor odds ratios based on the marginal ER and PR categories, we evaluated ER status while controlling for PR status and *vice versa*. Specifically, to test for the effect of ER status on risk factors while controlling for PR status, we calculated d_{ER}, which is the weighted average of the effect of ER status among women who are PR+ and PR−, respectively, with weights inversely proportional to the variance: $d_{ER} = [(\beta_{1++} - \beta_{1-+})w_1 + (\beta_{1+-} - \beta_{1--})w_2]/(w_1 + w_2)$, where $w_1 = 1/[\text{var}(\beta_{1++}) + \text{var}(\beta_{1-+})]$, $w_2 = 1/[\text{var}(\beta_{1+-}) + \text{var}(\beta_{1--})]$, standard error (SE) $(d_{ER}) = [1/(w_1 + w_2)]^{1/2}$, and the variances (var) were obtained from Table 1. The test statistic (i.e., $Z_{ER} = d_{ER}/SE_{ER}$) was compared with a standard normal distribution N(0,1) to obtain a *P* value. A similar approach was used to assess the marginal effect of PR status based on d_{PR}.

Finally, interaction effects between ER and PR status were obtained from the test statistic $Z_{int} = d_{int}/\text{SE}(d_{int})$, where $d_{int} = [(\beta_{1++} - \beta_{1-+}) - (\beta_{1+-} - \beta_{1--})]$ and SE $(d_{int}) = [(\text{var}(\beta_{1++}) + \text{var}(\beta_{1-+}) + \text{var}(\beta_{1+-}) + \text{var}(\beta_{1--})]^{1/2}$, which was also compared with an N(0,1) distribution. We also estimated cumulative risk ratios for each type of breast cancer for women with typical risk profiles (i.e., nulliparous, one birth at age 35, and four births at ages 20, 23, 26, and 29 years) from age 30 to 70 years. Cumulative incidence and risk ratios are presented to compare both the absolute incidence of different tumor receptor categories of breast cancer and the contributions of the risk profile to the cumulative incidence.

Results

We fit log-incidence models separately for ER+/PR+ (n = 1281 cases) and ER−/PR− invasive breast cancers (n = 417). These same models were also fit separately to the cases of invasive breast cancers with discordant receptor status: ER+/PR− breast cancers (n = 318) and ER−PR+ breast cancers (n = 80). ER+/PR+ tumors were more common than the other three tumor receptor categories with increasing age (Fig. 1). The breast cancer incidence model was applied to the incident cases for each tumor receptor category, and the results are shown in

Table 1. Fitted breast cancer incidence model from the Nurses' Health Study (1980–2000) for predicting breast cancer incidence according to estrogen receptor (ER) and progesterone receptor (PR) status*

Variable	ER+/PR+ (n = 1281)			ER+/PR− (n = 318)			ER−/PR+ (n = 80)			ER−/PR− (n = 417)			Heterogeneity†	
	Beta	SE	P‡	Beta	SE	P‡	Beta	SE	P‡	Beta	SE	P‡	χ^2 (df)	P
Intercept	−11.016	0.340		−13.30	0.687		−12.06	1.23		−10.07	0.51			
Duration of premenopause	0.104	0.009	<.001	0.117	0.019	<.001	0.070	0.035	.047	0.048	0.015	.001	12.32 (3)	.006
Menopause (duration of menopause, y)														
Natural	0.045	0.006	<.001	0.060	0.011	<.001	−0.043	0.032	.17	0.013	0.011	.22	20.07 (6)	.003
Bilateral oophorectomy	0.041	0.008	<.001	0.046	0.016	.004	−0.076	0.044	.09	0.026	0.014	.070		
Pregnancy history														
Age at first birth − age at menarche (y)	−0.005	0.006	.37	0.028	0.011	.012	−0.0023	0.022	.92	0.016	0.011	.13	8.48 (3)	.037
Birth index§	−0.0037	0.0009	<.001	−0.0038	0.0018	.034	−0.0028	0.0036	.44	−0.00052	0.0016	.75	2.52 (3)	.47
Benign breast disease (BBD)														
BBD (yes vs. no)	0.217	0.643	.74	0.193	1.31	.88	2.855	2.288	.21	−0.315	1.06	.77	7.38 (12)	.83
BBD × age at menarche	0.079	0.029	.007	0.104	0.059	.076	−0.116	0.115	.31	0.062	0.053	.24		
BBD × duration of premenopause	−0.014	0.012	.24	−0.027	0.025	.28	−0.024	0.044	.58	0.0002	0.020	.99		
BBD × duration of menopause	−0.028	0.007	<.001	−0.015	0.014	.28	0.009	0.030	.76	−0.011	0.012	.38		
Postmenopausal hormone use														
Duration, oral estrogen alone	0.033	0.010	<.001	0.0004	0.020	.98	0.109	0.045	.016	0.028	0.018	.13	10.50 (9)	.31
Duration, oral estrogen plus progesterone	0.082	0.015	<.001	0.036	0.030	.24	0.086	0.068	.21	0.074	0.030	.012		
Duration, other types of hormones	0.031	0.014	.03	0.044	0.024	.065	−0.042	0.093	.65	0.034	0.025	.18		
Current use	0.124	0.088	.16	0.045	0.176	.80	0.178	0.349	.61	0.0026	0.159	.99	0.57 (3)	.90
Past use	−0.110	0.090	.22	−0.029	0.165	.86	−1.14	0.59	.056	−0.540	0.183	.003	8.57 (3)	.036
Body mass index (BMI), kg/m^2														
(avg. BMI during premenopause − 21.6) × duration of premenopause + (avg. BMI while on postmenopausal hormones − 24.4) × duration of postmenopausal hormone use	−0.00082	0.00029	.004	−0.00154	0.00062	.013	−0.00149	0.00109	.17	−0.00067	0.00051	.19	1.56 (3)	.67
(avg. BMI after menopause while not on postmenopausal hormones − 24.4) × duration of menopause while not on postmenopausal hormones	0.0044	0.00072	<.001	0.0013	0.0016	.39	0.0094	0.0029	<.001	0.00046	0.0016	.78	10.37 (3)	.016
Height (in)														
(Height − 64.5) × duration of premenopause + (height − 64.4) × duration of postmenopausal hormone use	0.0011	0.0004	.002	0.00009	0.00075	.90	0.0029	0.0013	.025	0.00055	0.00064	.39	5.12 (6)	.53
(Height − 64.4) × duration of menopause, while not on postmenopausal hormones	−0.00015	0.0017	.76	0.0006	0.003	.85	−0.0126	0.0080	.117	0.0008	0.0031	.79		
Alcohol consumption														
Cumulative ounces before menopause	0.00029	0.00009	.001	0.00022	0.00017	.20	0.00015	0.00037	.68	−0.00003	0.00017	.86	1.44 (3)	.70
Cumulative ounces after menopause														
With use of hormones	−0.00003	0.00034	.93	0.00087	0.00042	.04	0.00047	0.00122	.70	−0.0010	0.0009	.27	4.40 (3)	.22
Without use of hormones	0.00014	0.00025	.58	−0.00016	0.00053	.76	−0.0007	0.00194	.72	0.00067	0.00039	.08	2.01 (3)	.57
Family history of breast cancer in a first-degree relative (yes vs. no)	0.38	0.08	<.001	0.58	0.14	<.001	0.59	0.29	.040	0.54	0.13	<.001	2.40 (3)	.49

*Beta was derived from the log-incidence model. SE = standard error.

†Obtained from comparing the polychotomous logistic regression model where the specific variable(s) were allowed to be the same for all tumor subtypes and all other variables were allowed to be different for different tumor subtypes versus a polychotomous logistic regression model where all variables were allowed to be different for different tumor subtypes.

‡*P* values were obtained from a polychotomous logistic regression model comparing each tumor subtype with the control group.

§Birth index = the sum of (minimum [age, age at menopause] minus age at *i*th birth) over all births for parous women; = 0 for nulliparous women.

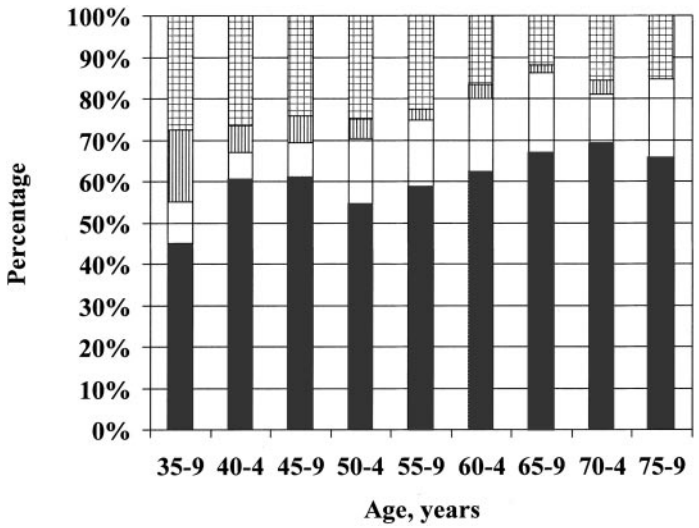

Fig. 1. The distribution of estrogen receptor (ER) and progesterone receptor (PR) tumors among incident invasive breast cancer cases in the Nurses' Health Study (1980–2000) by age at cancer diagnosis. The percentage of each of four tumor receptor categories was determined; ER−/PR− (**square-hatched bars**), ER−/PR+ (**vertical-stripe bars**), ER+/PR− (**open bars**), and ER+/PR+ (**solid bars**).

Table 1. Incidence of ER+/PR+ tumors increased by 11.0% per year during premenopausal years ($e^{0.104} = 11.0\%$), at 4.6% per year after natural menopause ($e^{0.045} = 4.6\%$), and at 4.2% per year after surgical menopause (i.e., bilateral oophorectomy). In contrast, the incidence of ER−/PR− tumors increased by 5.0% per year during premenopausal years and by only 1.3% per year after natural menopause. The log-incidence associations with age for ER+/PR− tumors were similar to those for ER+/PR+ tumors. Similarly, the rate of increase in incidence with age for ER−/PR+ tumors was similar to that for ER−/PR− tumors.

Reproductive variables appeared to show differing associations with breast cancer incidence among the four tumor receptor categories. The one-time adverse association of first pregnancy with breast cancer incidence observed in the total cohort *(4–6)* was also present for ER+/PR− (b = .028, SE = .011, $P = .012$) and ER−/PR− (b = .016, SE = .011, $P = .13$) breast cancer, but not for ER+/PR+ (b = −.005, SE = .006, $P = .37$) or ER−/PR+ (b = −.0023, SE = .022, $P = .92$) breast cancer. There was statistically significant heterogeneity for the adverse association of first pregnancy with breast cancer incidence ($\chi^2_3 = 8.48$, $P = .037$), that is, the coefficient was statistically significantly different between at least two of the receptor subtypes of breast cancer. When we evaluated the difference in breast cancer risk according to ER and PR status (Table 2), PR status was statistically significantly associated with the adverse association of first pregnancy with breast cancer incidence after controlling for ER status ($P = .007$).

Parity and age at each birth, as summarized by the birth index, showed a strong inverse association with incidence of ER+/PR+ (b = −.0037, SE = .0009, $P<.001$) and ER+/PR− tumors (b = −.0038, SE = .0018, $P = .034$) but showed no association with incidence of ER−/PR− and ER−/PR+ tumors (Table 1). Fig. 2 shows the incidence of ER+/PR+ (Fig. 2, A) and ER−/PR− (Fig. 2, B) tumors with age for typical nulliparous women, women with one birth at age 35 years, and women with four births at ages 20, 23, 26, and 29 years. There was a clear increase in the incidence of ER−/PR− tumors among women with a first birth at age 35 years, whereas this increase in incidence was absent for ER+/PR+ tumors (Fig. 2, B). The risk of diagnosis of ER−/PR− tumors (to age 70 years) was lowest for nulliparous women, whereas risk of diagnosing ER+/PR+ tumors (again, to age 70 years) was lowest among women with four births at ages 20, 23, 26, and 29 years.

In terms of cumulative incidence from age 30 to 70 years (Table 3), for a woman with an age at menarche of 13 years, an age at natural menopause of 50 years, and four births at ages 20, 23, 26, and 29 years, the cumulative relative risk to age 70 years of ER+/PR+ breast cancer was 0.71 (95% CI = 0.60 to 0.84) and of ER−/PR− breast cancer was 1.07 (95% CI = 0.77 to 1.49) compared with that for a nulliparous woman during the same period (Table 3). For a similar woman with a single birth at age 35 years, cumulative relative risk for ER+/PR+ tumors (RR = 0.86, 95% CI = 0.69 to 1.08) was lower than that of a nulliparous woman but cumulative relative risk for ER−/PR− tumors (RR = 1.39, 95% CI = 0.89 to 2.17) was higher.

The association of duration of postmenopausal hormone use (i.e., estrogen alone or estrogen plus progestin) with breast cancer incidence was similar for ER+/PR+ and ER−/PR− tumors (Table 1), but the association of current postmenopausal hormone use with incidence was stronger for ER+/PR+ tumors (b = .124, SE = .088, $P = .16$) than for ER−/PR− tumors (b = .0026, SE = .159, $P = .99$). In addition, past postmenopausal hormone use had a strong and statistically significant inverse association with the incidence of ER−/PR− tumors (b = −.540, SE = .183, $P = .003$) but no association with the incidence of ER+/PR+ tumors (b = −.110, SE = .090, $P = .22$). Similar associations between postmenopausal hormone use and breast cancer incidence were observed for the discordant tumor receptor categories; ER+/PR− tumors were similar to ER+/PR+ tumors, and ER−/PR+ tumors were similar to ER−/PR− tumors. For example, for a typical woman with natural menopause at age 50 years, with 10 years of estrogen use only from age 50 to 60 years, the relative risk (to age 70 years) of ER+/PR+ breast cancer was 1.18 (95% CI = 1.00 to 1.38) and of ER−/PR− breast cancer was 0.96 (95% CI = 0.78 to 1.17) compared with that for a woman who never used postmenopausal hormones (Table 3). For women with 10 years of estrogen plus progestin use from age 50 to 60 years, the relative risk for ER+/PR+ breast cancer (to age 70 years) was 1.67 (95% CI = 1.33 to 2.10) and for ER−/PR− breast cancer was 1.21 (95% CI = 0.87 to 1.68) compared with that for women who never used postmenopausal hormones.

BMI after menopause was statistically significantly associated with the incidence of ER+/PR+ tumors but not with the incidence of ER−/PR− tumors ($P = .016$ for test of heterogeneity for BMI among categories of receptor status; Table 1). There was no statistically significant difference in breast cancer incidence among tumor subtypes for BMI before menopause. BMI after menopause was also statistically significantly associated with the incidence of ER−/PR+ tumors, whereas it was not associated with the incidence of ER+/PR− tumors. The difference in the association of BMI after menopause with breast cancer incidence according to PR status was statistically significant after controlling for ER status ($P = .005$; Table 2). Alcohol use before menopause had a stronger association with the incidence of ER+/PR+ tumors ($P = .001$) than with that of ER−/

Table 2. Tests of difference in relative risk for tumors based on estrogen receptor (ER) and progesterone receptor (PR) status in a cohort of women from the Nurses' Health Study from 1980–2000

Variable	*P* value*		
	ER status controlling for PR	PR status controlling for ER	Interaction effect of ER/PR status†
Intercept	.008	.066	.005
Duration of premenopause	.003	.77	.42
Menopause (duration of menopause, y)			
Natural	<.001	.095	.25
Bilateral oophorectomy	.048	.29	.053
Pregnancy history			
Age at first birth − age at menarche, y	.58	.007	.42
Birth index‡	.23	.98	.61
Benign breast disease (BBD)			
BBD (yes vs. no)	.69	.52	.28
BBD × age at menarche	.17	.32	.28
BBD × duration of premenopause	.56	.90	.51
BBD × duration of menopause	.43	.62	.36
Postmenopausal hormone use			
Duration, oral estrogen alone	.09	.045	.37
Duration, oral estrogen plus progesterone	.42	.19	.67
Duration, other types of hormones	.59	.50	.54
Current use	.95	.57	.82
Past use	.01	.49	.42
Body mass index (BMI)			
(avg. BMI during premenopause − 21.6) × duration of premenopause + (avg. BMI after menopause while using postmenopausal hormones − 24.4) × duration of postmenopausal hormone use	.59	.56	.27
(avg. BMI after menopause while not using postmenopausal hormones − 24.4) × duration of menopause while not using postmenopausal hormones	.48	.005	.11
Height (in)			
(Height − 64.5) × duration of premenopause + (height − 64.4) × duration of postmenopausal hormone use	.25	.057	.43
(Height − 64.4) × duration of menopause while not using postmenopausal hormones	.52	.38	.19
Alcohol use			
Cumulative ounces before menopause	.29	.59	.81
With use of hormones	.21	.22	.14
Without use of hormones	.29	.77	.42
Family history of breast cancer in first-degree relative (yes vs. no)	.86	.28	.46

*P values were calculated from Wald tests based on d_{ER} and d_{PR}, respectively (*see* text).

†Interaction effect P values were calculated from Wald tests based on d_{INT} (*see* text).

‡Birth index = the sum of (minimum [age, age at menopause] minus age at *i*th birth) over all births for parous women; = 0 for nulliparous women.

PR− tumors (P = .86) (Table 1), but the difference in the association of alcohol use with incidence was not statistically significant (P = .70; Table 1).

Benign breast disease, family history, and height were each consistently associated with the incidence of ER+/PR+ and ER−/PR− tumors. Additional analyses were performed in which a term for calendar year was added to the model to control for possible secular trends in breast cancer incidence. There was an approximately 7% decrease in breast cancer incidence over 20 years after controlling for risk factors that were included in the model (*see* Table 1). After adjusting for secular trends, the results for breast cancer incidence according to risk factors were similar to those given in Table 1 (data not shown).

Model Fit

To evaluate the performance of the log-incidence model in predicting breast cancer incidence, we conducted goodness-of-fit and area under the receiver operator characteristic (ROC) curve analysis. We calculated the observed and expected number of ER+/PR+ and ER−/PR− breast cancer cases in 5-year age strata by using deciles of predicted age-specific breast cancer risk. The observed and expected number of cases for specific deciles of age-specific risk were then summed over all age strata and compared with a goodness-of-fit test. For prediction of the number of ER+/PR+ breast cancer cases, the overall chi-square (with 9 *df*) was 7.87 (P = .55), indicating an adequate fit. The log-incidence model provided a good spread in risk for ER+/

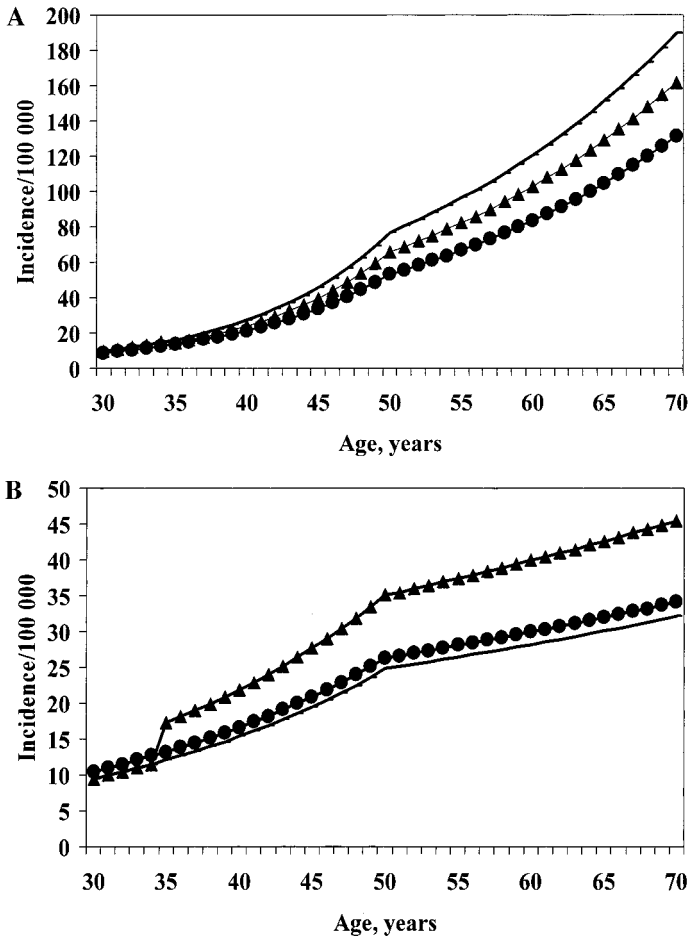

Fig. 2. Relationship between reproductive history and incidence of ER+/PR+ tumors (**A**) and ER−/PR− tumors (**B**) for a nulliparous woman (**solid line with ticks**), a woman with a birth at age 35 (**triangles**), and a woman with four births at ages 20, 23, 26, and 29 years (**circles**) for a typical woman. A typical woman had menarche at age 13 years, natural menopause at age 50 years, had no family history of breast cancer in a first-degree relative, did not have benign breast disease, did not use postmenopausal hormones, was of average height and weight, and did not drink alcohol from age 18 to 70 years.

PR+, with an observed relative risk, comparing the top decile to the bottom decile, of 7.2 (95% CI = 5.2 to 9.9) and an expected relative risk of 5.6 (95% CI = 4.1 to 7.5). For ER−/PR− breast cancer cases, the overall chi-square (with 9 *df*) was 2.99 (P = .97), with an observed relative risk, comparing the top decile to the bottom decile, of 3.9 (95% CI = 2.4 to 6.3) and an expected relative risk of 4.2 (95% CI = 2.6 to 6.9).

The predictive ability of our log-incidence model to discriminate between women who would develop ER+/PR+ breast cancer and those who would not was also evaluated using ROC curve analysis. First, we calculated the predicted absolute risk of breast cancer for each woman and stratified the data by 5-year age groups. Within each age group, we then calculated the Mann–Whitney *U* statistic, which compares the predicted risk of the case patients with the predicted risk of the control subjects (i.e., women who remained free from breast cancer). Thus, we obtained the area under the ROC curve for our predictive model for women in a specific age group; this value can be interpreted as the probability that, within a specific 5-year age group, a random case patient will have a higher predicted risk than a random control subject. We then calculated a weighted average of the age-specific Mann–Whitney *U* statistics with weights equal to the inverse variance of the age-specific statistics. Overall, the area under the ROC curve adjusted for age was 0.64 (95% CI = 0.63 to 0.66) for ER+/PR+ tumors and 0.61 (95% CI = 0.58 to 0.64) for ER−/PR− tumors, indicating adequate discriminatory accuracy.

Discussion

In these prospective data from the Nurses' Health Study, we observed that the number of ER+/PR+ breast tumors increased at a faster rate than the number of ER−/PR− tumors both before and after menopause. Parity and timing of births (i.e., early versus late) were inversely associated with ER+/PR+ tumors but not with ER−/PR− tumors, whereas the one-time adverse association of first pregnancy with incidence was evident only among ER−/PR− and ER+/PR− tumors. Women who used postmenopausal hormones had a substantially reduced risk for developing ER−/PR− tumors after stopping use of the hormones. BMI after menopause was statistically significantly more strongly associated with ER+/PR+ tumors than with ER−/PR− tumors. Overall, the four categories of tumors based on ER/PR status showed different associations with age, pregnancy history, postmenopausal hormone use, and BMI after menopause.

The distribution of the receptor status of the tumors in this study is comparable to that reported in other studies *(14,15,18,22)* using cross-classification of both ER and PR status. Furthermore, consistent with other studies *(14,15,18,22),* information on ER/PR status was available for about 70% of cancer cases. Women diagnosed with breast cancer early in follow-up were less likely than women diagnosed late in follow-up to have their tumor ER/PR status measured and their medical records obtained. Our results, similar to the results of the Surveillance, Epidemiology, and End Results (SEER) Program[1], are based on reports from individual institutions without centralized quality control *(36);* however, the prevalence of ER+/PR+ tumors in this study is consistent with the prevalence of ER+/PR+ tumors in non-Hispanic white women in the SEER database *(36).*

Our results for breast cancer incidence according to age are consistent with previous literature *(7,8).* Findings for reproductive variables are somewhat difficult to interpret in previous studies because data on ER and PR status have been presented separately. However, several studies *(10,12,18)* have shown that, among postmenopausal women, nulliparity is inversely associated with risk for developing ER−/PR− tumors, whereas other studies *(12,14–16)* have shown that parity is inversely associated with ER+ tumors but not with ER− tumors, suggesting that there are differing influences of parity on incidence, according to ER status of tumors. In general, previous studies have had low statistical power to evaluate the relationship between incidence and reproductive variables and have often not cross-classified ER and PR status.

Interestingly, Potter et al. *(18),* in analyzing the prospective Iowa Women's Health Study, noted that ER+/PR− tumors had a risk profile somewhat different than that of the other three ER/PR categories. In particular, they noted that family history was not associated with risk of ER+/PR− tumors. In our larger study, the association of family history with ER/PR status was consistent across all tumor receptor categories. Furthermore, Potter et al. *(18)* proposed that breast tumors be classified into

Table 3. Comparison of cumulative relative risk estimates of breast cancer from age 30–70 years for different estrogen receptor (ER) and progesterone receptor (PR) subtypes for hypothetical women with different risk factor profiles*

Variable	ER+/PR+		ER+/PR−		ER−/PR+		ER−/PR−	
	Incidence ($\times 10^3$)	RR (95% CI)	Incidence ($\times 10^5$)	RR (95% CI)	Incidence ($\times 10^5$)	RR (95% CI)	Incidence ($\times 10^5$)	RR (95% CI)
Age at menopause, y								
55	3145	1.50 (1.27 to 1.77)	741	1.51 (1.08 to 2.12)	212	1.79 (1.03 to 3.11)	1085	1.24 (0.99 to 1.55)
45	2096	1.00 (referent)	490	1.00 (referent)	118	1.00 (referent)	876	1.00 (referent)
Body mass index								
Avg. woman	2606	1.00 (referent)	614	1.00 (referent)	163	1.00 (referent)	989	1.00 (referent)
Stable weight woman	2319	0.89 (0.85 to 0.93)	635	1.03 (0.93 to 1.15)	149	0.92 (0.82 to 1.02)	1011	1.02 (0.94 to 1.11)
Above-avg. weight gain	3297	1.27 (1.15 to 1.39)	573	0.93 (0.76 to 1.14)	217	1.33 (1.00 to 1.76)	947	0.96 (0.83 to 1.11)
Consistently lean	2438	0.94 (0.88 to 0.99)	699	1.14 (0.99 to 1.30)	164	1.00 (0.81 to 1.24)	1053	1.07 (0.96 to 1.19)
Consistently obese	3048	1.17 (1.05 to 1.30)	501	0.82 (0.65 to 1.03)	188	1.15 (0.80 to 1.67)	894	0.90 (0.76 to 1.08)
Age at menarche, y								
11	3152	1.00 (referent)	812	1.00 (referent)	184	1.00 (referent)	1116	1.00 (referent)
15	2132	0.68 (0.62 to 0.73)	455	0.56 (0.48 to 0.66)	141	0.76 (0.56 to 1.05)	865	0.78 (0.68 to 0.89)
Age at births, y								
Nulliparous	3195	1.00 (referent)	616	1.00 (referent)	188	1.00 (referent)	901	1.00 (referent)
20	2816	0.88 (0.81 to 0.96)	670	1.09 (0.90 to 1.31)	172	0.92 (0.64 to 1.30)	995	1.10 (0.93 to 1.32)
35	2752	0.86 (0.69 to 1.08)	1064	1.73 (1.07 to 2.78)	174	0.92 (0.37 to 2.28)	1255	1.39 (0.89 to 2.17)
20, 23, 26, 29‡	2264	0.71 (0.60 to 0.84)	518	0.84 (0.57 to 1.23)	145	0.77 (0.39 to 1.52)	963	1.07 (0.77 to 1.49)
35, 38, 41, 44‡	2550	0.80 (0.63 to 1.01)	970	1.58 (0.96 to 2.59)	165	0.88 (0.35 to 2.19)	1241	1.38 (0.87 to 2.18)
Postmenopausal hormone use								
None	2606	1.00 (referent)	614	1.00 (referent)	163	1.00 (referent)	989	1.00 (referent)
Estrogen only†	3050	1.18 (1.00 to 1.38)	600	0.99 (0.72 to 1.36)	194	1.20 (0.79 to 1.84)	941	0.96 (0.78 to 1.17)
Estrogen plus progestin†	4324	1.67 (1.33 to 2.10)	775	1.28 (0.80 to 2.03)	177	1.10 (0.65 to 1.87)	1186	1.21 (0.87 to 1.68)
Benign breast disease								
No	2606	1.00 (referent)	614	1.00 (referent)	163	1.00 (referent)	989	1.00 (referent)
Yes	4257	1.64 (1.46 to 1.85)	957	1.58 (1.25 to 1.98)	291	1.80 (1.07 to 3.05)	1509	1.54 (1.24 to 1.90)
Family history of breast cancer in a first-degree relative								
No	2606	1.00 (referent)	614	1.00 (referent)	163	1.00 (referent)	989	1.00 (referent)
Yes	3757	1.45 (1.25 to 1.68)	1086	1.79 (1.36 to 2.34)	291	1.80 (1.03 to 3.17)	1673	1.70 (1.32 to 2.19)

*RR = relative risk; CI = confidence interval.

†Estimates were determined for use of hormones from age 50–60 years.

‡These two different patterns of birth portray two women with age at first birth of 20 and 35, respectively, with a spread of 3 years between each successive birth.

three receptor groups; PR+, ER−/PR−, and ER+/PR−. Our results, which had more statistical power than those from the study by Potter et al., do not support the collapsing of the PR+ tumor category across ER status because both ER and PR status were associated independently with different risk factors, as summarized in Table 2.

Similar to the findings of Potter et al. *(18)*, our findings show that epidemiologic risk factors vary by the hormone receptor expression of the breast cancer, supporting the hypothesis that these receptor expression categories represent distinct stable phenotypes in human breast cancer *(37)* rather than a single disease with a single biologic pathway. Anderson et al. *(38)* showed that among lymph node–negative women in the SEER database, each of the four ER/PR tumor subtypes was associated with separate age frequency-density plots, again suggesting that breast cancer does not represent a single disease. Treatment of breast cancer has already been divided by hormone receptor status in that hormonal agents are only used in receptor-positive cancers, and the same division of cancer cases according to receptor status should be considered for etiologic investigations.

Qualitatively comparing relative risks for breast cancer without a statistical evaluation, as reported in many of the earlier studies [e.g., *(13)*], can suggest differences in these risks but can also be problematic. Furthermore, a comparison of statistical significance between outcomes is limited by the dependence of statistical significance levels on the numbers of events for each component of the outcome. For example, small numbers of cancer cases in some of our subgroup analyses limited our ability to evaluate differences in breast cancer risk among categories defined by receptor status. In particular, the small number of ER−/PR+ cancer cases limited our ability to detect associations between this tumor receptor subtype and breast cancer risk.

The results of this study offer a rigorous comparison of risk factors for breast cancer categorized according to ER and PR status and provide flexibility in terms of allowing some risk estimates to be the same and others to be different, based on likelihood ratio methods. The Marshall and Chisholm method *(35)*, which we used to compute the polychotomous logistic regression models, has advantages over other approaches such as PROC CATMOD in SAS (SAS Institute, Cary, NC), which forces all variables to differ in the outcome categories. Moreover, if some variables have similar risks for the polychotomous outcome categories, the estimates for all exposure variables will be less accurate if estimated separately *(39)*. Because of this problem, we believe our estimates for all risk factors are more precise than estimates using PROC CATMOD.

There are also several advantages of the log-incidence model of cancer risk over conventional logistic models. First, the log-incidence model readily allows for the interaction among variables. For example, the model used in this study allowed

for the interactions among age, age at first birth, and parity, which have been observed in various epidemiologic studies *(3,28)*. Second, this model allows for a more precise timing of exposures in relation to subsequent risk of breast cancer than other models that do not account for the varying effects of risk factors with age. We note that several variables including BMI have different effects on risk before and after menopause.

We recently fit the breast cancer incidence model, derived using data from 1980 through 1994, to an independent series of cases (diagnosed from 1994 through 1998) from the Nurses' Health Study and observed a goodness of fit that was consistent across age strata and a fourfold increase in the risk of breast cancer comparing the top and bottom deciles of risk. Furthermore, an age-adjusted concordance statistic gave an area under the ROC curve of 0.62. Thus, this incidence model, when fit to an independent case series from the Nurses' Health Study, appears to be robust and to have good predictive value. When fitting the model for ER+/PR+ tumors, the age-adjusted concordance statistic value was higher (0.64) than that observed for the Gail model prediction of total breast cancer incidence (0.58) *(24)*, suggesting a modest improvement in predictive ability for the most common breast cancer tumor subtype with our log-incidence model.

In conclusion, our data indicate that some important risk factors for breast cancer differ according to ER status (e.g., age and postmenopausal hormone use) and PR status (e.g., reproductive history and BMI after menopause). These data support the hypothesis that different patterns of receptor expression correspond to different types of breast tumor. Thus, we suggest that it would be prudent to divide breast cancer cases according to both the ER and PR status of the tumor. This categorization may also be useful in understanding differences in breast cancer risk profiles among ethnic groups (e.g., Caucasian versus African American), where the mix of ER/PR types may differ.

REFERENCES

(1) Moolgavkar SH, Day NE, Stevens RG. Two-stage model for carcinogenesis: epidemiology of breast cancer in females. J Natl Cancer Inst 1980;65:559–69.

(2) Pike MC, Krailo MD, Henderson BE, Casagrande JT, Hoel DG. 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. Nature 1983;303:767–70.

(3) Pathak DR, Whittemore AS. Combined effects of body size, parity, and menstrual events on breast cancer incidence in seven countries. Am J Epidemiol 1992;135:153–68.

(4) Rosner B, Colditz GA, Willett WC. Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. Am J Epidemiol 1994;139:819–35.

(5) Rosner B, Colditz GA. Nurses' Health Study: log-incidence mathematical model of breast cancer incidence. J Natl Cancer Inst 1996;88:359–64.

(6) Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. Am J Epidemiol 2000;152:950–64.

(7) Yasui Y, Potter JD. The shape of the age-incidence curves of female breast cancer by hormone-receptor status. Cancer Causes Control 1999;10:431–7.

(8) Tarone RE, Chu KC. The greater impact of menopause on ER- than ER+ breast cancer incidence: a possible explanation (United States). Cancer Causes Control 2002;13:7–14.

(9) Hulka BS, Chambless LE, Wilkinson WE, Deubner DC, McCarty KS Sr, McCarty KS Jr. Hormonal and personal effects of estrogen receptors in breast cancer. Am J Epidemiol 1984;119:692–704.

(10) Hildreth NG, Kelsey JL, Eisenfeld AJ, LiVolsi VA, Holford TR, Fischer DB. Differences in breast cancer risk factors according to the estrogen receptor level of the tumor. J Natl Cancer Inst 1983;70:1027–31.

(11) Nomura Y, Tashiro H, Hamada Y, Shigematsu T. Relationship between estrogen receptors and risk factors of breast cancer in Japanese pre- and postmenopausal patients. Breast Cancer Res Treat 1984;4:37–43.

(12) Hislop TG, Coldman AJ, Elwood JM, Skippen DH, Kan L. Relationship between risk factors for breast cancer and hormonal status. Int J Epidemiol 1986;15:469–76.

(13) McTiernan A, Thomas DB, Johnson LK, Roseman D. Risk factors for estrogen receptor-rich and estrogen receptor-poor breast cancers. J Natl Cancer Inst 1986;77:849–54.

(14) Stanford JL, Szklo M, Boring CC, Brinton LA, Diamond EA, Greenberg RS, et al. A case-control study of breast cancer stratified by estrogen receptor status. Am J Epidemiol 1987;125:184–94.

(15) Cooper JA, Rohan TE, Cant EL, Horsfall DJ, Tilley WD. Risk factors for breast cancer by oestrogen receptor status: a population-based case-control study. Br J Cancer 1989;59:119–25.

(16) Kreiger N, King WD, Rosenberg L, Clarke EA, Palmer JR, Shapiro S. Steroid receptor status and the epidemiology of breast cancer. Ann Epidemiol 1991;1:513–23.

(17) Habel LA, Stanford JL. Hormone receptors and breast cancer. Epidemiol Rev 1993;15:209–19.

(18) Potter JD, Cerhan JR, Sellers TA, McGovern PG, Drinkard C, Kushi LR, et al. Progesterone and estrogen receptors and mammary neoplasia in the Iowa Women's Health Study: how many kinds of breast cancer are there? Cancer Epidemiol Biomarkers Prev 1995;4:319–26.

(19) Yoo KY, Tajima K, Miura S, Takeuchi T, Hirose K, Risch H, et al. Breast cancer risk factors according to combined estrogen and progesterone receptor status: a case-control analysis. Am J Epidemiol 1997;146:307–14.

(20) Huang WY, Newman B, Millikan RC, Schell MJ, Hulka BS, Moorman PG. Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. Am J Epidemiol 2000;151:703–14.

(21) Zhu K, Beiler J, Hunter S, Payne-Wilks K, Roland CL, Forbes DS, et al. The relationship between menstrual factors and breast cancer according to estrogen receptor status of tumor: a case-control study in African-American women. Ethn Dis 2002;12:S3–23–9.

(22) Britton JA, Gammon MD, Schoenberg JB, Stanford JL, Coates RJ, Swanson CA, et al. Risk of breast cancer classified by joint estrogen receptor and progesterone receptor status among women 20–44 years of age. Am J Epidemiol 2002;156:507–16.

(23) Moolgavkar SH. Cancer models. Epidemiology 1990;1:419–20.

(24) Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, 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.

(25) Costantino JP, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst 1999;91:1541–8.

(26) Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 1954;8:1–12.

(27) Moolgavkar SH, Knudson AG Jr. Mutation and cancer: a model for human carcinogenesis. J Natl Cancer Inst 1981;66:1037–52.

(28) Pathak DR, Speizer FE, Willett WC, Rosner B, Lipnick RJ. Parity and breast cancer risk: possible effect on age at diagnosis. Int J Cancer 1986;37:21–5.

(29) Colditz GA. The Nurses' Health Study: a cohort of US women followed since 1976. J Am Med Womens Assoc 1995;50:40–4.

(30) Stampfer MJ, Willett WC, Speizer FE, Dysert DC, Lipnick R, Rosner B, et al. Test of the National Death Index. Am J Epidemiol 1984;119:837–9.

(31) Rich-Edwards JW, Corsano KA, Stampfer MJ. Test of the National Death Index and Equifax Nationwide Death Search. Am J Epidemiol 1994;140:1016–9.

(32) Hunter DJ, Willett WC. Diet, body size, and breast cancer. Epidemiol Rev 1993;15:110–32.

(33) Huang Z, Hankinson SE, Colditz GA, Stampfer MJ, Hunter DJ, Manson JE, et al. Dual effects of weight and weight gain on breast cancer risk. JAMA 1997;278:1407–11.

(34) SAS version 6.12. Cary (NC): SAS Institute; 1996.

(35) Marshall RJ, Chisholm EM. Hypothesis testing in the polychotomous logistic model with an application to detecting gastrointestinal cancer. Stat Med 1985;4:337–44.

(36) Li CI, Malone KE, Daling JR. Differences in breast cancer hormone receptor status and histology by race and ethnicity among women 50 years of age and older. Cancer Epidemiol Biomarkers Prev 2002;11:601–7.
(37) Robertson JF. Oestrogen receptor: a stable phenotype in breast cancer. Br J Cancer 1996;73:5–12.
(38) Anderson WF, Chatterjee N, Ershler WB, Brawley OW. Estrogen receptor breast cancer phenotypes in the Surveillance, Epidemiology, and End Results database. Breast Cancer Res Treat 2002;76:27–36.
(39) Glynn RJ, Rosner B. Methods to evaluate risks for composite endpoints and their individual components. J Clin Epidemiol. In press 2004.

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.

Supported by grant CA87969 (to G. A. Colditz) from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services; by a Harvard Breast Cancer Specialized Projects of Oncology Research Excellence (SPORE) grant; and by the U.S. Army Center of Excellence in ER-Negative Breast Cancer. G. A. Colditz is an American Cancer Society Clinical Research Professor.

We thank Frank E. Speizer, Robert Glynn, and Walter C. Willett for critical input to the study. Marion McPhee, Barbara Egan, Gary Chase, and Karen Corsano provided technical assistance during this study.

Manuscript received April 29, 2003; revised November 21, 2003; accepted December 5, 2003.