Nurses' Health Study: Log-Incidence Mathematical Model of Breast Cancer Incidence

*Bernard Rosner, Graham A. Colditz**

Background: **In 1983, Pike et al. developed a mathematical model to quantify the effects of reproductive risk factors on the incidence of breast cancer. In 1994, we modified that model to correct some deficiencies in the original model, including a lack of terms for spacing of births and an inability to easily accommodate births after age 40 years. Our extended Pike model, while improving on the original, still had serious disadvantages, such as difficulty in translating model parameters into relative risks (RRs) and an incomplete fit to data that slightly overestimated incidence for premenopausal women with an early age at first birth and that underestimated incidence for postmenopausal women with a late age at first birth.** ***Purpose:*** **We undertook both the development of a new mathematical model to quantify the effects of reproductive risk factors on breast cancer incidence and validation of the model.** ***Methods:*** **A new log-incidence model of breast cancer incidence was developed using nonlinear regression methods, and a study population consisting of 89 132 women in the Nurses' Health Study from which a total of 2249 incident cases of breast cancer were identified. Subjects were followed from the return of the 1976 Nurses' Health Study questionnaire until June 1, 1990, or until the last questionnaire was returned, until the development of any cancer, or until death, yielding 1 148 593 person-years of follow-up. The log-incidence models were fitted using iteratively reweighted least squares analysis.** ***Results:*** **The log-incidence model provided a better fit to the data than the extended Pike model, with parameter estimates interpretable in terms of RRs. This new model can be fitted using standard commercially available statistical software. In the model, younger parous women are generally at slightly higher risk than nulliparous women, which is true for both the observed and expected RRs, and older parous women, aged 55-64 years with an early age at first birth, are at lower risk than nulliparous women, while older women with a late age at first birth are at substantially higher risk than nulliparous women.** ***Conclusion:*** **Log-incidence models, such as this one, provide an efficient framework for modeling the effect of lifestyle risk factors on breast cancer incidence that may be specifically targeted to certain time periods of a woman's reproductive life. [J Natl Cancer Inst 1996; 88:359-64]**

To better understand and summarize the relations between reproductive risk factors and breast cancer incidence, Pike et al. (*1*) proposed a mathematical model based on the observed age-incidence curve and the known relations between breast cancer risk and age at menarche, age at first birth, age at menopause, and parity. However, it did not include terms for the spacing of pregnancies, and it did not easily accommodate pregnancies after age 40 years. In the Pike et al. model, factors associated with reduced risk of breast cancer were each considered to lower the rate of "breast tissue aging," which we translate to mean on the molecular level the accumulation of genetic damage in the pathway to breast cancer. Moolgavkar et al. (*2*), conversely, used a variant of the multistage theory of carcinogenesis, based on a two-stage model. They proposed that from the time of menarche, cells mutate at a constant rate and become partially transformed. Initiated cells either proliferate, differentiate, or die; proliferation is mediated by the hormonal environment. Krailo et al. (*3*) tested the Pike et al. model based on data from a case–control study limited to premenopausal cases with age at diagnosis younger than or equal to 39 years. Kampert et al. (*4*) fitted both models to data from 1884 women with breast cancer and from 3432 control women admitted to San Francisco Bay Area hospitals from 1970 through 1977. They observed that both models gave similar results and neither fully accounted for the protective effect of early age at first full-term pregnancy among premenopausal women. Other investigations that have considered models for breast cancer incidence include De Lisi (*5*), Manton and Stallard (*6*), and Pathak and Whittemore (*7*).

We fitted an extension of the Pike et al. model to an independent prospective dataset [the prospective Nurses' Health Study from the follow-up period of 1976 to 1990 (*8*)] and added a term to summarize the spacing of births. We observed that the spacing of births was significantly related to reduced breast cancer risk; the closer subsequent births are to the first birth, the lower the risk. A transient increase in risk with the first, but not later, pregnancies is followed by a subsequent decrease in risk (*8*). Animal data also show differentiation of the breast tissue at the time of the first birth (*9*) and lower susceptibility to carcinogens after the first birth (*10*). Russo and Russo (*11*), using an animal model of breast cancer, attribute the high susceptibility of the "virginal" breast, which has a high proliferative rate before first pregnancy, to neoplastic transformation of the terminal end bud and lower rates after first pregnancy. Examination of mammary tissue, however, shows that not all terminal end buds are transformed through the first pregnancy, but those that remain may differentiate during the second or subsequent pregnancies.

While differentiation of mammary duct epithelium may be expected with the first pregnancy, epidemiologic data also suggest that not all breast cells differentiate with the first pregnancy, confirming the findings reported by Russo and Russo (*11*). Specifically, Trichopoulos et al. (*12*) reported that the spacing of births influences breast cancer risk: the closer the births are together, the lower the risk.

**Affiliations of authors:* B. Rosner, Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, and Department of Biostatistics, Harvard School of Public Health, Boston, MA; G. A. Colditz, Channing Laboratory, Brigham and Women's Hospital and Harvard Medical School, and Department of Epidemiology, Harvard School of Public Health.

Correspondence to: Bernard Rosner, Ph.D., Channing Laboratory, 180 Longwood Ave., Boston, MA 02115.

See "Notes" section following "References."

This finding suggests that cells that have not passed through differentiation with the first pregnancy may be differentiated with the subsequent pregnancy; therefore, closer pregnancies offer less time for the breast to accumulate damage to DNA.

During examination of goodness of fit for the extended Pike et al. model, we observed that the model slightly overestimated incidence for early age at first birth among women younger than 45 years and underestimated incidence for late age at first birth among women older than 54 years. In this report, we report on a new breast cancer incidence model that allows for the transient increase in risk with first pregnancy to increase with age at first pregnancy and apply the model to a cohort of 89 132 middle-aged women in the Nurses' Health Study.

Methods

The Pike et al. "breast tissue age" model (*1*) is given by incidence (I) of breast cancer at age (t)

$$I(t) = [d(t)]^k,$$

where $d(t)$ is breast tissue age at chronological age t. The constant k is determined by the rate of increase in breast cancer incidence with breast tissue age. Breast tissue age is determined by age (t), parity (s), menopausal status (m), age at menarche (t_0), age at first birth (t_1), second birth (t_2), . . ., etc., and age at menopause (t_m). In the original Pike model, breast tissue age increased at a constant rate (c) from menarche to first birth. There was then an immediate increase in breast tissue age at the time of first birth (of size k_1) and a corresponding decrease in the rate of tissue aging after first birth to a rate ($c - d_1$). Breast tissue age increased at the same rate from first birth to age 40 years after which the rate of increase diminished linearly until at menopause the rate of increase was d_3 units lower than at age 40 years.

There were several difficulties with the original Pike et al. model including *a*) ambiguities if the age at first birth was more than 40 years and *b*) the problem with modeling incidence for premenopausal women, where age at menopause is unknown. Therefore, a modified Pike "one-birth" model was constructed (*8*). Under this model, breast tissue was assumed to age at a constant rate from first birth to menopause. At menopause, we assumed that there was an immediate decrease in cumulative breast tissue age of k_3 and a decrease in the rate of tissue aging of d_3. In addition to the first birth, we found that the number and timing of subsequent births affected breast cancer incidence that was formalized in a multiple births model (Fig. 1). We assumed that the rate of tissue aging per year declined by d_2 with each birth after the first birth. In addition, we assumed that there was a one-time increase in breast tissue age of k_2 after the second birth. The multiple-births model (*8*) can be formalized using the expression

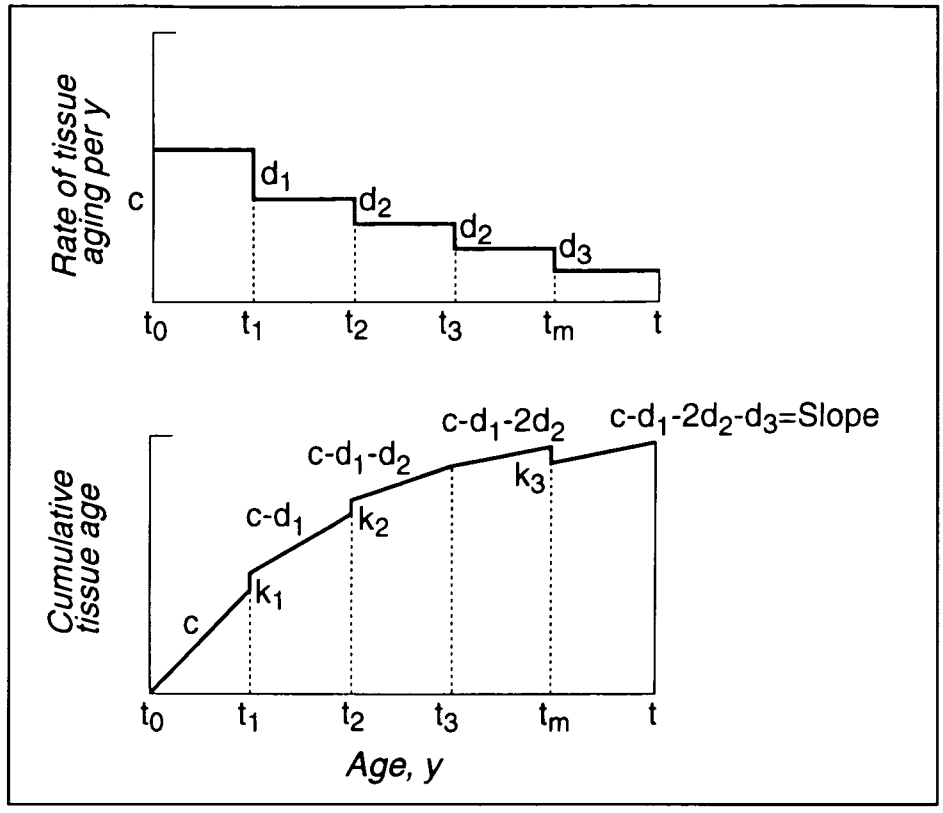

Fig. 1. Multiple-birth extended Pike et al. model of breast cancer incidence. In the top figure, the y axis is the rate of tissue aging per year, while in the bottom figure, the y axis is the cumulative tissue age. The x axis is age in years. The variable t = current age; t_0 = age at menarche; t_1 = age at first birth; t_2 = age at second birth; t_3 = age at third birth; t_m = age at menopause; c = rate of tissue aging after menarche; d_1 = decrease in the rate of tissue aging after first birth; d_2 = decrease in the rate of tissue aging after each birth after the first birth; d_3 = decrease in the rate of tissue aging at menopause; k_1 = increase in cumulative tissue age at first birth; k_2 = increase in cumulative tissue age at second birth; and k_3 = decrease in cumulative tissue age at menopause.

$$d(t) = c(t-t_0) - d_3(t - t_m)m - k_3m -$$
$$d_1(t - t_1)b_1 + k_1b_1 - d_2\sum_{i=2}^{s}(t - t_i)b_i + k_2b_2,$$

where m = 1 if postmenopausal and 0 if premenopausal; b_1 = 1 if parous and 0 if nulliparous; b_i = 1 if a woman had $\geq i$ births and 0 otherwise; s = parity; t = current age; t_i = age at ith birth; t_0 = age at menarche; and t_m = age at menopause.

Since k was unknown, generalized linear models were fit for each of 41 different values of k, ranging from 1.0 to 5.0 in increments of 0.1. The computational technique of iteratively reweighted least squares was used for parameter estimation using PROC NLIN of SAS. The log-likelihood was computed for each k and the max log-likelihood over 41 values of k was obtained.

The parameters of the breast-tissue age-type models are difficult to interpret in a relative risk (RR) context and fitting the models is computationally inconvenient, requiring separate runs for each k. In addition, the model slightly overestimated incidence for premenopausal women with an early age at first birth and underestimated incidence for postmenopausal women with a late age at first birth. Therefore, we also considered an alternative class of models that we call log-incidence models. In a broad sense, under the log-incidence models, log(incidence) is a linear function of time, whereas for the breast tissue age models log(incidence) is a linear function of log(time) or log(breast-tissue age).

In particular, the multiple-birth log-incidence model can be expressed as follows:

$$\ln(\text{incidence}) = \alpha + \beta_0 t_0 + \beta_1(t^* - t_0) + \beta_2(t - t_m)m + \beta_3(t_1 - t_0)b_1 + \beta_4 b + \beta_5 b(t - t_m)m,$$

where t^* = minimum (age, age at menopause); b = birth index = $\sum_{i=1}^{8}(t^* - t_i)b_i$ = total years from each birth to minimum (age, age at menopause) summed over all births in parous women and b = 0 for nulliparous women.

The parameters of the model can be interpreted as follows: β_0 = increase in log incidence per year from from birth to menarche; β_1 = increase in log incidence per year after menarche before menopause in nulliparous women or before first birth in parous women; β_2 = increase in log incidence per year after menopause in nulliparous women; $\beta_3(t_1 - t_0)$ = immediate increase in log incidence after the first birth that is assumed to increase linearly with $t_1 - t_0$; $\beta_1 + i\beta_4$ = increase in log incidence per year after menarche, before menopause for women with i births, i = 1, . . .8; and $\beta_2 + b\beta_5$ = increase in log incidence per year after menopause in women with birth index = b.

The derivation of the likelihood for the log-incidence model is given in the Appendix. The ration-

ale for the log-incidence models is that the number of potentially malignant cells (referred to as precancer cells) increase multiplicatively with time, but that different events in the reproductive history of a woman (e.g., menarche, first birth, and menopause) affect the rate of increase. Specifically, the number of precancer cells is assumed to increase annually at the rate of e^{β_0} prior to menarche, e^{β_1} after menarche in nulliparous premenopausal women, e^{β_2} in nulliparous postmenopausal women, $e^{\beta_1 + s\beta_4}$ after menarche in parous women with parity = s; and $e^{\beta_2 + b\beta_5}$ in parous postmenopausal women with birth index = b. Furthermore, we assume that the number of precancer cells increases immediately by a factor of $e^{\beta_3(t_1 - t_0)}$ after the first birth. Finally, the incidence rate of breast cancer is assumed to be approximately proportional to the number of precancer cells.

The log-incidence models were fit using iteratively reweighted least squares using PROC NLIN of SAS. They are much easier to fit than the breast tissue age models, since they do not require separate runs for each value of an unknown power k. Furthermore, the parameters can be easily interpreted in an RR context. Specifically:

$\exp(\beta_0 - \beta_1)$ = RR of breast cancer for a 1-year increase in age at menarche among nulliparous women

$\exp(\beta_0 - \beta_1 - \beta_3)$ = RR of breast cancer for a 1-year increase in age at menarche among parous women

$\exp(\beta_3 - \beta_4)$ = RR of breast cancer for a 1-year increase in age at first birth among parous, premenopausal women

$\exp[\beta_3 - \beta_4 - \beta_5(t - t_m)]$ = RR of breast cancer for a 1-year increase in age at first birth among parous, postmenopausal women

$\exp(-\beta_4)$ = RR of breast cancer for a 1-year increase in age at ith birth ($i \geq 2$) among multiparous, premenopausal women

$\exp[-\beta_4 - \beta_5(t - t_m)]$ = RR of breast cancer for a 1-year increase in age at ith birth ($i \geq 2$) among multiparous, postmenopausal women

$\exp(\beta_1 - \beta_2)$ = RR of breast cancer for a 1-year increase in age at menopause for nulliparous, postmenopausal women

$\exp\{\beta_1 - \beta_2 + s[\beta_4 + \beta_5(t - t_m - 1 - \bar{b})]\}$ = RR of breast cancer for a 1-year increase in age at menopause for parous postmenopausal women, where $\bar{b} = b/s$

$\exp[\beta_3(t_1 - t_0) + \beta_4 b]$ = RR of breast cancer for parous versus nulliparous premenopausal woman

$\exp\{\beta_3(t_1 - t_0) + b[\beta_4 + \beta_5(t - t_m)]\}$ = RR of breast cancer for a parous versus nulliparous postmenopausal woman

To assess the goodness of fit of the model, the total person-years were stratified by age (30-44, 45-54, and 55-64 years). Within each age group, the observed and expected (based on the multiple-births model) RR of breast cancer was computed for each of 16 subgroups of parous women defined by age at first birth (20-24, 25-29, 30-34, and 35-39 years) and parity (1, 2, 3, and 4+) versus nulliparous women. Conditional on the total observed number of events within an age group, we also computed a chi-square goodness-of-fit statistic to assess the adequacy of the model given by

$$\sum_{i=1}^{3} \sum_{j=1}^{N_i} (O_{ij} - E_{ij})^2 / E_{ij},$$

where O_{ij} = observed number of breast cancer cases for the ith age group and the jth age at first birth × parity group; E_{ij} = expected number of breast cancer cases for the ith age group and jth age at first birth × parity group, conditional on the total observed number of cases in the ith age group; $N_i = 13$ for the age group 30-44 years; and $N_i = 17$ for the age groups 45-54 and 55-64 years.

There were a total of 121 701 women in the Nurses' Health Study cohort in 1976, of whom 119 403 did not report a history of any cancer (excluding nonmelanoma skin cancer) on the 1976 questionnaire; of those women, 105 406 returned the 1978 questionnaire. We further excluded 4205 women if their number of pregnancies reported in 1976 was different by two or more children from the estimated number of pregnancies in 1976 based on reported ages of children in 1978. We excluded another 6993 women whose number of living children as derived from reported ages differed from their parity (reported in two separate questions) in 1978. We also excluded 2757 women whose number of living children in 1978 was less than their reported number of children in 1976. In addition, we excluded 412 women whose age at first birth estimated from reported children in 1978 was greater than (3+ age at first birth reported in 1976). Also, we excluded 768 women whose age at menarche was either unknown or reported to be less than or equal to 8 or greater than or equal to 22 years. Further exclusions included unknown parity (n = 199), age at any birth greater than age at menopause (n = 677), women reported to be nulliparous in 1976 whose age of the oldest child was greater than 2 years in 1978 (n = 201), and women whose menopausal status and/or age at menopause was unknown (n = 52). We also excluded 10 women whose age of death was unknown. This left 89 132 women who accrued 1 148 593 person-years of follow-up from the return of the 1976 questionnaire to June 1, 1990, yielding a total of 2249 incident cases of breast cancer, which were used in the final analyses. Subjects were followed until June 1, 1990, or until the last questionnaire was returned, until the development of any cancer, or until death.

Results

The results from fitting the multiple-births log-incidence model are shown in Table 1.

Subsequent breast cancer incidence increases 4.9% per year [$\exp(\beta_0) = \exp(0.048)$], with each year between birth and menarche. Among nulliparous women, breast cancer incidence increases 8.5% per year [$\exp(\beta_1) = \exp(0.081)$] prior to menopause and 5.1% per year [$\exp(\beta_2) = \exp(0.050)$] after menopause. The RR of breast cancer for a parous versus nulliparous, premenopausal woman is given by $\exp[\beta_3 (t_1 - t_0) + \beta_4 b] = \exp[.013 (t_1 - t_0) - .0036b]$. For a postmenopausal woman, the RR of breast cancer for a parous versus nulliparous woman is given by $\exp[\beta_3 (t_1 - t_0) + b[\beta_4 + \beta_5(t - t_m)] = \exp[.013 (t_1 - t_0) + b [-.0036 - .00020 (t - t_m)]$. Depending on the relative magnitude of (age at first birth – age at menarche) versus the birth index, parous women may be at either increased or decreased risk relative to nulliparous women. The net effect of pregnancy is a short-term increase in the incidence of breast cancer and a corresponding long-term decrease.

Table 1. Breast cancer log-incidence model based on follow-up from 1976 to 1990 fitted to Nurses' Health Study data

Parameters*	Regression coefficient	*se*†	*Z*‡	*P*§
α constant	–9.687	0.265		
β_0	0.048	0.016	3.10	.002
β_1	0.081	0.004	19.56	<.001
β_2	0.050	0.005	9.32	<.001
β_3	0.013	0.004	3.14	.002
β_4	–0.0036	0.0009	–3.87	<.001
β_5	–0.00020	0.00012	–1.73	.085

*β_0 = increase in log incidence per year from birth to menarche; β_1 = increase in log incidence per year after menarche, before menopause in nulliparous women, or before first birth in parous women; β_2 = increase in log incidence per year after menopause in nulliparous women; $\beta_3 (t_1 - t_0)$ = immediate increase in log incidence after the first birth, which is assumed to increase linearly with $t_1 - t_0$; $\beta_1 + i\beta_4$ = increase in log incidence per year after menarche, before menopause for women with i births, $i = 1, \ldots, 8$; and $\beta_2 + b\beta_5$ = increase in log incidence per year after menopause in women with birth index = b.

†*se* = standard error.

‡*Z* = regression coefficient divided by the standard error.

§Two-tailed.

The magnitude of the above increases and decreases in incidence for parous women are primarily a function of age at first birth and to a lesser extent a function of age at subsequent births. Specifically, prior to menopause the incidence of breast cancer increases 1.7% ($\exp(\beta_3 - \beta_4) = \exp(.013 + .0036)$) for a 1-year increase in age at first birth and 0.4% [$\exp(-\beta_4) = \exp(.0036)$] for a 1-year increase in age at each subsequent birth. Furthermore, the effects of age at first and subsequent births on breast cancer incidence is still greater after menopause. To illustrate the magnitude of these effects, we consider three hypothetical women with age at menarche = 13 years and age at menopause = 50 years: *a*) a nulliparous woman; *b*) a parous woman with one birth at age 35 years; and *c*) a parous woman with three births at ages 20, 23, and 26 years. The age-specific annual incidences of these three women are shown in Table 2.

A parous woman with a single birth at age 35 years [woman (*b*)] has a 34% increase in breast cancer incidence at the time of the birth relative to a nulliparous woman. Furthermore, this excess risk goes down very slowly over time. Even at age 70 years, woman (*b*) has approximately a 19% excess risk versus a nulliparous woman. Conversely, the parous woman with an early age at first birth with multiple children conceived at an early age [woman (*c*)] has a slight excess risk immediately after the first birth relative to a nulliparous woman (RR = 1.10), which slowly diminishes over time, reaching equality at age 32 years and continuing to decline until menopause (age 50 years), at which time RR = 0.82. After menopause, the incidence of breast cancer increases much more slowly for woman (*c*) (3.4%) compared with woman (*a*) (5.1%) until the RR is close to 0.60 at age 70 years.

Since the relationship between breast cancer incidence and reproductive history changes with age, a useful summary is provided by cumulative incidence rather than age-specific incidence. The cumulative incidence for woman (*a*) from age 13-70 years = 10 032 per 10^5. Woman (*b*) (one birth at age 35 years) has a 21% excess risk over the age period 13-70 years (incidence = 12 128 per 10^5), while woman (*c*) (three births at ages 20, 23, and 26 years) has a 25% decrease in risk over the similar age period (incidence = 7571 per 10^5).

Table 2. Breast cancer incidence (per 10^5) by age for three hypothetical women with age at menarche = 13 and age at menopause = 50*

	Age(s) at births, y					
	Nulliparous†		35†		20, 23, 26†	
Age, y	Incidence	RR‡	Incidence	RR‡	Incidence	RR‡
20	21	1.0	21	1.0	23	1.10
25	31	1.0	31	1.0	33	1.07
30	46	1.0	46	1.0	47	1.02
35	70	1.0	93	1.34	67	0.96
40	105	1.0	138	1.31	96	0.91
45	158	1.0	203	1.29	136	0.86
50	237	1.0	300	1.27	194	0.82
55	303	1.0	378	1.25	229	0.76
60	389	1.0	477	1.23	271	0.70
65	498	1.0	602	1.21	320	0.64
70	638	1.0	760	1.19	378	0.59
13-70	10 032	1.0	12 128	1.21	7571	0.75

*For example, consider a premenopausal 45-year-old woman who has three children at ages 20, 23, and 26 years and age at menarche = 13 years. Thus, age at first birth – age at menarche = 7, the birth index = (45 – 20) + (45 – 23) + (45 – 26) = 66 and her RR versus a nulliparous woman = exp[.013(7) – .0036(66)] = 0.86. Similarly, if the woman is a 65-year-old postmenopausal woman with age at menopause = 50 years and the same age at menarche and ages at births as above, then the birth index = (50 – 20) + (50 – 23) + (50 – 26) = 81, age – age at menopause = 15 years and her RR versus a nulliparous woman = exp{.013(7) + 81 [–.0036 – .00020(15)]} = 0.64.

†In the text, the nulliparous woman is called (*a*); the parous woman with one birth at 35 is called (*b*); and the parous woman with three births at ages 20, 23, and 26 is called (*c*).

‡Relative risk (RR) versus nulliparous women.

To assess goodness of fit of the log-incidence models, we have stratified the person-time according to age (30-44, 45-54, and 55-64 years), age at first birth (nulliparous, 20-24, 25-29, 30-34, and 35-39 years), and parity (0, 1, 2, 3, and 4+) and have computed age-specific observed and expected RRs for each combination of age at first birth and parity versus nulliparous women. These results are shown in Table 3.

The model appears to provide an adequate fit (overall chi-square = 32.22; 23 *df*; $P = .096$). Younger parous women are generally at slightly higher risk than nulliparous women, which is true for both the observed and expected RRs. Older parous women (aged 55-64 years) with an early age at first birth are at lower risk than nulliparous women, while older women with a late age at first birth are at substantially higher risk than nulliparous women.

Discussion

Through the log-incidence model, we describe a simple procedure that fits the reproductive risk factors to risk for breast cancer and allows for the interaction between age, age at first birth, and parity that has been observed in numerous epidemiologic studies (*13*). Extending the previous model that resulted in an incomplete fit to the data, we now observe a closer agreement between observed and expected RRs for breast cancer according to age and age at first birth. This 7-parameter model is easier to fit than a traditional logistic regression that would require more than 45 terms if it were to capture the interaction between age and age at first birth. Thus, the log-incidence model provides a closer fit to the data and requires far fewer terms.

Although the goodness of fit remains less than perfect, it is improved with the log-incidence model compared with the extended Pike et al. model (*8*). However, we continue to slightly underestimate the effect of late age at first birth among older women. One assumption is that the model provides a uniform estimate across all subgroups of women. The potential variation in the relation between reproductive risk factors and risk for breast cancer among subgroups of women with family history or a personal history of benign breast disease requires further exploration.

Table 3. Observed and expected relative risks (RRs) by age at first birth and parity group versus nulliparous women*

Age, y	Age at 1st birth, y	Parity 1 Observed RR (No. of cases)	Parity 1 Expected RR	Parity 2 Observed RR (No. of cases)	Parity 2 Expected RR	Parity 3 Observed RR (No. of cases)	Parity 3 Expected RR	Parity 4+ Observed RR (No. of cases)	Parity 4+ Expected RR
30-44	20-24	0.94 (7)	1.08	0.75 (55)	1.03	0.88 (71)	1.00	0.71 (44)	0.95
	25-29	1.14 (20)	1.13	1.01 (88)	1.11	0.87 (45)	1.09	0.84 (21)	1.05
	30-39	1.07 (11)	1.28	1.11 (17)	1.26	0.88 (4)	1.23	0.96 (1)	1.17
45-54	20-24	1.48 (16)	1.03	0.93 (90)	0.96	0.88 (122)	0.90	0.75 (151)	0.80
	25-29	0.92 (23)	1.12	0.87 (111)	1.05	1.02 (129)	0.99	0.67 (90)	0.90
	30-34	1.83 (30)	1.24	1.06 (38)	1.19	1.38 (30)	1.14	0.64 (6)	1.05
	35-39	1.21 (10)	1.37	1.40 (11)	1.33	2.09 (4)	1.30	2.11 (1)	1.20
55-64	20-24	0.78 (9)	0.97	0.89 (51)	0.90	0.85 (63)	0.81	0.72 (90)	0.68
	25-29	0.84 (20)	1.08	0.99 (105)	0.99	0.97 (111)	0.91	0.87 (125)	0.78
	30-34	1.44 (24)	1.21	1.43 (53)	1.13	1.21 (30)	1.05	1.08 (21)	0.94
	35-39	1.66 (16)	1.33	1.32 (13)	1.27	1.57 (7)	1.22	1.39 (2)	1.11

*No. of cases for nulliparous women was 27, 64, and 66 in age groups 30-44, 45-54, and 55-64 years.

One issue in fitting the models in this paper is the occurrence of multiple births during the follow-up period. Specifically, there was a total of 2481 multiple births, of which 2401 women had a single twin birth, 51 had two sets of twin births, 24 had a single triplet birth, four had one twin birth and one triplet birth, and one had a quadruplet birth. Multiple births were treated as a single pregnancy in model fitting. The number of multiple births is too small to consider formally in the analyses (e.g., by constructing a separate birth-index variable for the extra children associated with multiple birth pregnancies).

An important feature of this model is its ease in assessing the cumulative incidence of breast cancer to age 70 years. We note that the range in cumulative risk from ages 13-70 years is from a 21% excess risk (RR = 1.21) of breast cancer by age 70 years for women with only one birth at age 35 years (compared with nulliparous women) to a 25% decrease in risk among women with age at first birth = 20 and parity = 3 and spacing between births of 3 years (RR = 0.75). This broad range in cumulative risk of breast cancer to age 70 years emphasizes the important contribution that these reproductive risk factors (age at first birth, parity, and spacing of children) make. The cumulative incidence will obviously be more extreme, depending on the total number and the actual spacing of the children even within the range of reproductive patterns in this educated U.S. population. Were the range to vary more widely to include ages at menarche up to 20 years and parity up to 20 seen in third-world societies, the impact on breast cancer risk would be far greater. Furthermore, the data in Table 3 underestimate the true variation in cumulative incidence, since these estimates are based on the mean age at menopause (age 50 years). Age at menopause is a strong predictor of risk of breast cancer in this population and will add further variation to the distribution of cumulative risk to age 70 years when incorporated into projections of risk for individual women. The log-incidence model clearly shows that the rate of increase in breast cancer incidence slows after menopause. Furthermore, the effect of menopause has probably been underestimated in the analysis, since we did not control for the use of postmenopausal hormones, which occur frequently in this cohort. Future work will add the contribution of postmenopausal hormones, cigarette smoking, alcohol, and other lifestyle factors to determine their impact on the rate of aging among postmenopausal women.

As lifestyle or genetic factors are postulated to interact with breast tissue aging at different stages in the life cycle, these can be fitted to the model to quantify their effect. For example, if cigarette smoking or alcohol consumption before first pregnancy increases the rate of DNA damage, then we would examine this specifically through the impact of smoking or drinking on the rate of increase in precancer cells between menarche and first birth and also on the increase in risk observed immediately after first pregnancy. This approach was used in the companion paper (*14*) to fit the model separately for family history + and family history – women.

In conclusion, log-incidence models provide an efficient framework for modeling the effect of lifestyle risk factors on breast cancer incidence that can be targeted to specific time periods during the reproductive life of a woman.

Appendix: Derivation of the Likelihood

To construct the analysis dataset, for each woman we divided the total person-years from 1976 to 1990 into 1-year segments of person-time, starting at 1976 and ending either at the time of any cancer (except nonmelanoma skin cancer), the time of death, the time of the last questionnaire submitted, or June 1, 1990, whichever occurred earliest. Since questionnaires were sent out every 2 years, we also assumed that the breast cancer occurred 1 year prior to the first questionnaire in which the disease was first reported. Covariate values were computed for each 1-year segment of person-time according to responses at the most recent prior questionnaire.

Let E_i = observed number of events, T_i = number of person-years, $I_i = E_i/T_i$ = incidence rate for the ith unit of person-time, and $i = 1,\ldots N$. The computational technique of iteratively reweighted least-

squares was used for parameter estimation using the procedure PROC NLIN of SAS (SAS Institute, Cary, NC). At the jth iteration, we computed $Z_{ij} = \ln(\hat{I}_{ij}) + (I_i - \hat{I}_{ij})/\hat{I}_{ij}$, . . ., $i = 1, \ldots, N$, j = iteration number, where $\hat{I}_{ij}$ = estimated incidence for the ith unit of person-time at the jth iteration and the weight $w_{ij} = \hat{I}_{ij}T_i$ that is proportional to $1/\mathrm{var}[\ln(\hat{I}_{ij})]$ under a Poisson regression model. We then did a weighted regression of Z_{ij} on the covariates with weights w_{ij} to obtain new estimates at the $(j + 1)$th iteration. The regression was continued until the estimates at successive iterations converged.

The log-likelihood was then computed for each 1-year segment of person-time where the contribution to the log-likelihood for the ith segment was $L_i = E_i \log\hat{E}_{ij} - \hat{E}_{ij} - \log(E_i!)$ if $E_i > 0$, $= -\hat{E}_{ij}$ if $E_i = 0$ and $\hat{E}_{ij} = T_i\hat{I}_{ij}$. The overall log-likelihood was $L = \sum_{i=1}^{N} L_i$.

References

(1) 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.

(2) Moolgavkar SH, Day NE, Stevens RG. Two-stage model for carcinogenesis: epidemiology of breast cancer in females. J Natl Cancer Inst 1980;65:59-69.

(3) Krailo M, Thomas DC, Pike MC. Fitting models of carcinogenesis to a case-control study of breast cancer. J Chronic Dis 1987;40 Suppl 2:181S-9S.

(4) Kampert JB, Whittemore AS, Paffenbarger RS Jr. Combined effect of childbearing, menstrual events, and body size in age-specific breast cancer risk. Am J Epidemiol 1988;128:962-79.

(5) De Lisi C. The age incidence of female breast cancer—simple models and analysis of epidemiological patterns. Math Biosciences 1977; 37:245-66.

(6) Manton KG, Stallard E. A two-disease model of female breast cancer: mortality in 1969 among white females in the United States. J Natl Cancer Inst 1980;64:9-16.

(7) 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.

(8) 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.

(9) Russo J, Tay LK, Russo IH. Differentiation of the mammary gland and susceptibility to carcinogenesis. Breast Cancer Res Treat 1982; 2:5-73.

(10) Russo J, Russo IH. Influence of differentiation and cell kinetics on the susceptibility of the rat mammary gland to carcinogenesis. Cancer Res 1980;40:2677-87.

(11) Russo IH, Russo J. Physiological bases of breast cancer prevention. Eur J Cancer Prev 1993;2 Suppl 3:101-11.

(12) Trichopoulos D, Hsieh CC, MacMahon B, Lin TM, Lowe CR, Mirra AP, et al. Age at any birth and breast cancer risk. Int J Cancer 1983; 31:701-4.

(13) Pathak D, 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.

(14) Colditz GA, Rosner BA, Speizer FE. Risk factors for breast cancer according to family history of breast cancer. J Natl Cancer Inst 1996; 88:365-71.

Notes

Supported by Public Health Service grant CA40356 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

Manuscript received May 22, 1995; revised September 28, 1995; accepted December 13, 1995.