Genetic Analysis of Breast Cancer in the Cancer and Steroid Hormone Study

Elizabeth B. Claus,* Neil Risch,*'† and W. Douglas Thompson‡

*Department of Epidemiology and Public Health and †Department of Human Genetics, Yale University School of Medicine, New Haven, CT; and ‡University of Southern Maine, Portland

Summary

The familial risk of breast cancer is investigated in a large population-based, case-control study conducted by the Centers for Disease Control. The data set is based on 4,730 histologically confirmed breast cancer cases aged 20 to 54 years and on 4,688 controls who were frequency matched to cases on the basis of both geographic region and 5-year categories of age, and it includes family histories, obtained through interviews of cases and controls, of breast cancer in mothers and sisters. Segregation analysis and goodness-of-fit tests of genetic models provide evidence for the existence of a rare autosomal dominant allele (q = .0033) leading to increased susceptibility to breast cancer. The effect of genotype on the risk of breast cancer is shown to be a function of a woman's age. Although, compared with noncarriers, carriers of the allele appear to be at greater risk at all ages, the ratio of age-specific risks is greatest at young ages and declines steadily thereafter. The proportion of cases predicted to carry the allele is highest (36%) among cases aged 20–29 years. This proportion gradually decreases to 1% among cases aged 80 years or older. The cumulative lifetime risk of breast cancer for women who carry the susceptibility allele is predicted to be high, approximately 92%, while the cumulative lifetime risk for noncarriers is estimated to be approximately 10%.

Introduction

Numerous studies have investigated the genetic transmission of breast cancer. In most instances, the transmission is reported to be explained by an autosomal dominant gene with sporadic cases (Williams and Anderson 1984; Bishop et al. 1988; Newman et al. 1988), although Goldstein et al. (1988) found evidence for a recessive subgroup defined by synchronous bilateral probands. Go et al. (1983), using high-risk families and no prevalence constraints in their analysis, found that a number of different transmission models fit the data equally well as did the Goldstein et al. (1987) model using bilateral breast cancer cases. In some studies, affected individuals have been defined solely by the presence of breast cancer, while in other studies breast cancer cases have been divided into subgroups based on menopausal status (Go et al. 1983; Bishop et al. 1988), age at onset (Sattin et al. 1985; Schwartz et al. 1985; Lynch et al. 1988*a*, 1988*b;* Claus et al. 1990*a*, 1990*b*), bilaterality (Goldstein et al. 1987, 1988), time interval between first and second primary tumors for bilateral cases (Goldstein et al. 1988), and occurrence of other cancers (Go et al. 1983; Schildkraut et al. 1989).

The present study investigates the ability of a number of possible genetic models to fit the observed patterns of transmission of breast cancer in a population-based, case-control study conducted by the Centers for Disease Control (CDC). This data set is the largest of its kind ever collected. In this data set the pattern of breast cancer among first-degree relatives, as well as the ages at which those relatives with breast cancer are affected, appear to be the most informative risk factors for the prediction of a woman's risk of breast cancer (Claus et al. 1990*a*). Risk of breast cancer to a mother or sister of a breast cancer case increases with decreasing age at onset of the case. The increase in risk to relatives that is associated with early age at onset appears not to be due entirely to correlation among

Received June 1, 1990; revision received October 4, 1990.

Address for correspondence and reprints: Dr. Elizabeth B. Claus, Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, New Haven, CT 06510.

0002-9297/91/4802-0008$02.00

ages at onset within families, indicating that age at onset and disease status may both share the same genetic determinant (Claus et al. 1990*b*). Furthermore, the risk to an individual increases with the number of family members affected with breast cancer, including a sharp increase in risk to women with at least two affected first-degree relatives. This finding lends support to a hypothesis that the distribution of breast cancer cases in the general population includes a small number of genetic cases combined with a larger number of nongenetic cases.

Based on these findings, genetic models are fit to the age-specific familial recurrence data incorporating information on family history of breast cancer as well as the ages at onset of relatives affected with breast cancer. Goodness-of-fit tests are used to compare the observed age-specific risk patterns with those predicted under the best-fitting genetic models. In addition, the age-specific risks of breast cancer for postulated carriers and noncarriers of the disease are compared in a test of proportionality of hazards by genotype. An analysis of this sort is important for its possible insight into the etiology of this disease, as well as for its potential to provide information for genetic linkage studies.

Subjects and Methods

Study Population

Data were obtained from the Cancer and Steroid Hormone Study, a multicenter, population-based, case-control study conducted by the CDC. The data set consists of 4,730 histologically confirmed breast cancer cases aged 20–54 years and of 4,688 controls. The cases were registered between December 1, 1980, and December 31, 1982, at eight Surveillance, Epidemiology, and End Results (SEER) Centers of the National Cancer Institute. The eight centers include the cities and metropolitan areas of Atlanta, Detroit, San Francisco, and Seattle, the four urban counties of Utah, and the states of Connecticut, Iowa, and New Mexico. Controls were frequency matched to cases according to geographic region and 5-year categories of age. Cases and controls with a previous history of breast cancer or with a breast biopsy of unknown outcome were excluded from the study. Cases and controls were interviewed about the occurrence of breast cancer in specific female relatives. For all of the analyses in this study, only the mothers and sisters of white cases and controls were included. Second-degree relatives and nonwhites were not included because of underreporting of the disease for these groups (Claus et al. 1990*a*) and because of the relatively small number of nonwhite cases and controls. Daughters were also excluded, as only two daughters of cases and two daughters of controls were reported to have had breast cancer. During the interview, in addition to family history an extensive amount of additional information was collected concerning pregnancy and menstrual history, history of benign breast disease, age at onset of disease, alcohol and cigarette usage, breast surgery, sociodemographic variables, and use of oral contraceptives. A detailed description of the study may be found elsewhere (Wingo et al. 1988).

Analysis

Segregation analysis is performed using POINTER (Lalouel and Yee 1980). All individuals in the analysis are assigned to liability classes according to both age at last observation and sex. The liability classes, based on the observed cumulative risk of breast cancer at 10-year age intervals among first-degree relatives of the controls, are .0002 by age 29 years, .0045 by age 39 years, .0157 by age 49 years, .0293 by age 59 years, .0458 by age 69 years, .0725 by age 79 years and .0997 at age 80 years or older. Males were given a liability value of zero, as no information regarding breast cancer exists for these individuals within this data set. Because of the 2-year time span of the study, the probability that any of the families were included in the study because of having more than one proband was very small, and, in fact, no such families were observed. Therefore, the ascertainment probability used is .001, corresponding to single ascertainment. Given a particular mode of inheritance, POINTER estimates the probability or likelihood of observing the sample data (in the form of nuclear families) and allows for the estimation of genetic parameters as well as for statistical comparison of various transmission models by using either a joint or conditional likelihood. When conditional likelihood is used, the phenotypes of offspring are calculated conditional on parental phenotypes,while the joint likelihood considers a nuclear family as a whole, with the phenotypes of parents and offspring considered jointly (Lalouel and Morton 1981). Both joint and conditional models are examined in the present study. Note that, although POINTER allows for testing among the various transmission models, an overall goodness-of-fit test for a set of data is not available.

In addition to the POINTER analyses, several genetic models are fit to the age-specific recurrence patterns of breast cancer among first-degree relatives by using maximum likelihood. The likelihoods are computed as a joint analysis of mothers and sisters of cases and controls (also defined as the probands). The likelihood for the mother of a case is calculated conditional on the age at which the case was affected. The likelihood for sisters of cases is calculated conditional on both the age at which the case was affected and the breast cancer status of the mother. The age at which a mother with breast cancer was affected or, in the case of an unaffected mother, the current age or the age at time of death is also incorporated into the model. For relatives of controls, the likelihood for mothers and sisters is calculated conditional on the current age of the control. As was done for sisters of cases, the likelihood for sisters of controls is calculated conditional on both the mother's breast cancer status and age. In these analyses, the age-at-onset distribution as well as the risk of breast cancer are assumed to be dependent on the same diallelic major locus, with alleles A (abnormal) and a (normal).

Let P, M, and S represent the breast cancer status of a proband, mother, and sister, respectively, and define each as 1 if the individual is affected at a given age and as 0 if the individual is not affected by their current age or by age at death. Let a_p, a_m, *and* a_s represent age at onset for a case, an affected mother, and an affected sister, respectively, and let ca_p, ca_m, and ca_s represent current age or age at death for a control, an unaffected mother, and an unaffected sister, respectively. Let G_k represent the probability that an individual has genotype *k;* let G_{jk} represent the probability that a mother has genotype *j,* given that the proband has genotype *k;* and let G_{ijk} represent the probability that a sister has genotype *i,* given that the mother has genotype *j* and that the proband has genotype *k.* Finally, define $R(x,k,\theta)$ as the cumulative risk of breast cancer up to age *x* for an individual with genotype *k,* where θ represents the parameter vector; and let $r(x,k,\theta)$ represent the corresponding density function. Set

$$H(\mathrm{rel},k,\theta) = \begin{bmatrix} r(x,k,\theta) & \text{if rel} = 1 \\ 1 - R(x,k,\theta) & \text{if rel} = 0 \end{bmatrix},$$

and then

$$\ln L = \ln \sum_i \mathrm{pr}(\mathrm{M}_i = m / \mathrm{P}_i = p) + \ln \sum_i \mathrm{pr}(\mathrm{S}_i = s / \mathrm{M}_i = m, \mathrm{P}_i = p) ,$$

where *i* is summed over the cases and controls and where *p, m,* and *s* ε 0,1. Let *i* be implicit; then it can be shown that the first term, $\mathrm{pr}(\mathrm{M} = m / P = p)$, is equal to

$$\sum_j \sum_k H(m,j,\theta) \times \mathrm{G}_{jk} \times \left[\frac{\mathrm{G}_k \times H(p,k,\theta)}{\sum_k \mathrm{G}_k \times H(p,k,\theta)} \right]$$

and that $\mathrm{pr}(\mathrm{S} = s / \mathrm{M} = m, \mathrm{P} = p)$ is equal to

$$\sum_i \sum_j \sum_k H(s,i,\theta) \times \mathrm{G}_{ijk} \times \left[\frac{\mathrm{G}_{jk} \times H(p,k,\theta)}{\sum_k \mathrm{G}_{jk} \times H(p,k,\theta)} \right] \times$$

$$\left(\left\{ H(m,j,\theta) \times \mathrm{G}_j \times \left[\frac{\sum_k \mathrm{G}_{jk} \times H(p,k,\theta)}{\sum_k \mathrm{G}_k \times H(p,k,\theta)} \right] \right\} \div \right.$$

$$\left. \left\{ \sum_j H(m,j,\theta) \times \mathrm{G}j \times \left[\frac{\sum_k \mathrm{G}_{jk} \times H(p,k,\theta)}{\sum_k \mathrm{G}_k \times H(p,k,\theta)} \right] \right\} \right) ,$$

where *i, j, k* ε AA, Aa, aa.

In the first model the age-at-onset distribution is assumed to be a genotype-dependent step function. If FRAC is defined as the fractional portion of a given number, the cumulative risk function takes on the form of a step function as follows:

$$R(x,k,\theta) = \sum_{i=1}^{i_x - 1} \lambda_{ik} + \{[\mathrm{FRAC}(x/10) + 1/10] \times \lambda_{i_x k}\}$$

.

$$r(x,k,\theta) = [\mathrm{FRAC}(x/10) + 1/10] \times \lambda_{i_x k}$$

The parameters of interest, which are estimated using the maximum-likelihood computer program MAXLIK (Kaplan and Elston 1972), include the gene frequency of the high-risk allele A, denoted as *q,* and each genotype's age-specific interval penetrances—i.e., $\lambda_{i,\mathrm{AA}}$, $\lambda_{i,\mathrm{Aa}}$, and $\lambda_{i,\mathrm{aa}}$—where *i* denotes the seven age categories 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, and 80–89 years, and where i_x represents the interval which contains the age *x.* Because of the extremely low occurrence of breast cancer in women before the age of 20 years, the probability of becoming affected with breast cancer before 20 years of age is assumed to be zero.

The second model mimics a Cox proportional hazards model in an attempt to test the effect that genotype has on risk. If time *t* is divided into *i* intervals each with lower endpoint t_{i-1} and upper endpoint t_i and if γ_i represents the hazard in interval i for genotype aa, then the cumulative hazard function for genotype aa can be written as

$$H(t) = \sum_{j=1}^{i-1} \gamma_j(t_j - t_{j-1}) + \gamma_i(t - t_{i-1}) \ .$$

The penetrance or the probability that a woman is affected with breast cancer within the ith interval is therefore

$$\lambda_i = \mathrm{F}(t_i) - \mathrm{F}(t_{i-1}) = \exp[-\sum_{j=1}^{i} \gamma_j(t_j - t_{j-1})] - \exp[-\sum_{j=1}^{i-1} \gamma_j(t_j - t_{j-1})] \ .$$

Let $(t_j - t_{j-1})$ be set equal to a constant of 10 years; then λ_i is given by

$$\lambda_i = \exp(-10 \sum_{j=1}^{i} \gamma_j) - \exp(-10 \sum_{j=1}^{i-1} \gamma_j) \ ,$$

where i is defined as above. For individuals with genotype Aa and AA, let α_{Aa} and α_{AA}, respectively, represent the proportion by which the hazard increases over that for individuals with genotype aa. The hazard constant for interval i is therefore represented by γ_i for an individual with genotype aa, by $\alpha_{\mathrm{Aa}} \times \gamma_i$ for an individual with genotype Aa, and by $\alpha_{\mathrm{AA}} \times \gamma_i$ for an individual with genotype AA, where A is the susceptibility allele. Under an autosomal dominant major-locus model $\alpha_{\mathrm{Aa}} = \alpha_{\mathrm{AA}}$. For an intermediate model, α_{Aa} and α_{AA} are estimated separately, while, for an autosomal recessive model, α_{Aa} is set equal to 1 and α_{AA} is estimated. Maximum-likelihood estimates using MAXLIK are calculated for a maximum of 10 parameters: γ_i, $i = 1,7$, the hazard constant for each age interval; α_{Aa} and α_{AA}, the hazard proportionality constants for genotype Aa and aa, respectively; and q, the gene frequency. Hence, a general equation for the age- and genotype-specific penetrance is written as

$$\lambda_{ik} = \exp(-10\alpha_k \sum_{j=1}^{i} \gamma_j) - \exp(-10\alpha_k \sum_{j=1}^{i-1} \gamma_j) \ ,$$

where it is assumed that α_{aa} is set equal to 1. The cumulative risk function and corresponding density function, $R(x,k,\theta)$ and $r(x,k,\theta)$, respectively, are defined as before.

Note that the proportional hazards model is nested within the first model. To test whether the assumption of proportional hazards holds, it is therefore assumed that two times the difference between the log likelihood from the proportional hazards model and the log likelihood for the first model is distributed as a χ^2 random variable with df equal to the number of age intervals.

On the basis of the shape of the derived age-at-onset step functions from the first model, a model assuming a normal age-at-onset distribution for each genotype is examined. For this model, the parameters include q, the gene frequency of allele A; μ_j and σ_j, the mean and SD of the age-at-onset distribution for genotype j; and λ_j, the cumulative lifetime penetrance for an individual with genotype j, where j is defined to be AA, Aa, or aa. Define $F(x,k,\theta)$ as the cumulative normal age-at-onset distribution for individuals with genotype k, and define $f(x,k,\theta)$ as the corresponding normal density function, and let λ_k be the lifetime penetrance for an individual with genotype k. Then

$$R(x,k,\theta) = \lambda_k \times F(x,k,\theta)$$
$$r(x,k,\theta) = \lambda_k \times f(x,k,\theta) \ .$$

Goodness-of-fit tests are used to compare the observed age-specific risk patterns with those predicted under the best-fitting genetic models. Both observed age-specific breast cancer risk among relatives and standard errors are calculated using the Kaplan-Meier method with BMDP1L (Dixon 1983). Separate plots are computed for mothers and sisters of probands whose age at onset is 20–29, 30–39, 40–49, or 50–54 years. The plots for sisters are also stratified by the breast cancer disease status (i.e., affected or not) of the mother. Plots for mothers and sisters of the controls are also derived. In an effort to determine the extent to which the models with highest likelihood obtained from POINTER and from the likelihood analyses actually fit the observed survival curves, the expected survival curves under the most likely models are compared graphically with the various observed survival curves described above. The expected curves under POINTER are not examined for the relatives of the controls, as the controls were used to define the liability classes.

Relatives with unknown current age or unknown age at death are eliminated from all analyses. For affected relatives with known current age but unknown age at onset, their age at onset is estimated by using the average age at onset for affected individuals whose current age matches that of the relative.

Table 1

Segregation Analysis with POINTER, under Joint Likelihood

Model	d^a	t^b	q	h^2	Z^c	$-2\ln L + c$
Sporadic, $q = H = 0$						−46,511.36
Multifactorial, $q = 0$:						
No generation difference				.254 (.013)	[1]	−46,716.87
Generation difference				.236 (.022)	1.236 (.267)	−46,717.38
Major locus, $H = 0$:						
Recessive	[0]	1.627 (.074)	.1612 (.0143)			−46,741.72
Intermediate	[0.5]	3.817 (.170)	.0023 (.0008)			−46,807.21
Dominant	[1]	1.916 (.090)	.0023 (.0005)			−46,807.76
Mixed	1	1.916	.0023	0		−46,807.76

[a] Degree of dominance; square brackets indicate that parameter was fixed at given value.

[b] Distance between the means of the two homozygous genotypic distributions.

[c] Ratio of adult h^2 to childhood h^2; square brackets indicate that parameter was fixed at given value.

Results

Table 1 presents results from the analyses using POINTER under a joint likelihood. Under a major-locus model, an autosomal dominant model provides the best fit to the data. The intermediate model converges to the dominant model. A recessive model is rejected ($\chi^2 = 66.04$, $P < .01$). A major-locus model provides a better fit to the data than does either a sporadic model or a model with only a polygenic component ($\chi^2 = 296.40$, $P < .01$ and $\chi^2 = 90.89$, $P < .01$, respectively). The mixed model converges to the dominant model. Under the dominant model, q is estimated to be .0023. The lifetime risk of breast cancer is estimated to be 69% for carriers of the abnormal susceptibility allele, versus approximately 10% for noncarriers. By age 29 years the proportion of total cases who are carriers is estimated to be 47%, while by age 80 years and older the proportion of cases who are carriers has gradually decreased to 2.5%. The results for the conditional model (not presented here) are similar.

The results from the likelihood model for which age at onset is assumed to be distributed as a step function are presented in table 2. The data are best fit by an autosomal dominant model. Under this model q is estimated to be .0033. The lifetime risk of breast cancer for carriers of the abnormal allele is estimated to be nearly 100%, versus 12% for noncarriers. The results from the model for which age at onset is normally distributed are presented in table 3. As for the step-function model, an autosomal dominant model provides the best fit. The value of q is once again estimated to be .0033. The lifetime risk of breast cancer to carri-

Table 2

Cumulative Probability of Individual Being Affected with Breast Cancer by a Given Age, Under Step-Function Model

Age (Years)	Cumulative Probability for Genotype		
	AA	Aa	aa
20–29	.0167	.0167	.0002
30–39	.1444	.1444	.0027
40–49	.3758	.3758	.0138
50–59	.5477	.5477	.0275
60–69	.6743	.6743	.0497
70–79	.9452	.9452	.0798
80+	1.0000	1.0000	.1254

Note.—The estimated value of $q = .0033$.

Table 3

Parameter Estimates and Standard Errors When Age at Onset Is Assumed to Be Normally Distributed

Parameter	Estimate	Standard error
μ_{AA} ($=\mu_{Aa}$)	55.435	1.742
μ_{aa}	68.990	1.532
σ_{AA} ($=\sigma_{Aa}=\sigma_{aa}$)	15.387	.669
λ_{AA} ($=\lambda_{Aa}$).	.928	.163
λ_{aa}	.100	.009
q	.0033	.0012

ers is 92%, versus a lifetime risk of 10% for noncarriers. The median age at onset for carriers of the abnormal gene is estimated to be approximately 55 years, versus 69 years for noncarriers.

The null hypothesis of proportional hazards across genotypes is strongly rejected in this study (χ^2 = 344.02, df = 6, $P << .01$), indicating that the ratio of the hazards functions for individuals with different genotypes is not constant but very much time dependent. Table 4 presents the age-specific hazard ratios computed by comparing the estimated risk for the heterozygotes with the estimated risk for the normal homozygotes, under both the step-function model and the model for which age at onset is normally distributed. The increase in risk starts off high; for the 20–29-year age interval the ratio of the hazards is estimated at approximately 98 under the step-function model and at approximately 76 under the normal distribution model, and it gradually decreases over time, to 2 and 3, respectively, for the 80–89-year age interval. The pattern of hazards ratios gives strength to the argument that age at onset and disease status are indeed strongly related for breast cancer and that familial cases are increasingly represented among cases with early onset.

The predicted and observed age-specific breast cancer risk curves for mothers and sisters are presented in figures 1–10. At all times, for both mothers and sisters, the predicted risk curves for the step-function model remain within 2 SD of the observed Kaplan-Meier product limit estimates and provide a closer fit to the observed risks than do any of the predicted curves. For mothers of cases, the fit is excellent. The predicted risk curves for mothers of cases whose age at onset is 30–54 years are essentially identical to the observed risk curves. The predicted curves for mothers of very young cases (age at onset 20–29 years) slightly underestimate the observed curves, although they lie within the confidence limits. The model in which age at onset is assumed to be normally distributed also fits the risk curves for mothers quite well, particularly when the case's age at onset was 40–54 years. For cases age 20–39 years, the normal model tends to underestimate the risk to mothers, and at one point the predicted normal curve slips outside the confidence band. POINTER

Table 4

Ratio of Estimated Risk of Breast Cancer for Heterozygotes versus Estimated Risk of Breast Cancer for Normal Homozygotes

Age	Risk Ratio in	
(Years)	Step-Function Model	Normal Model
20–29	98	76
30–39	50	43
40–49	21	25
50–59	13	14
60–69	6	8
70–79	9	5
80+	2	3

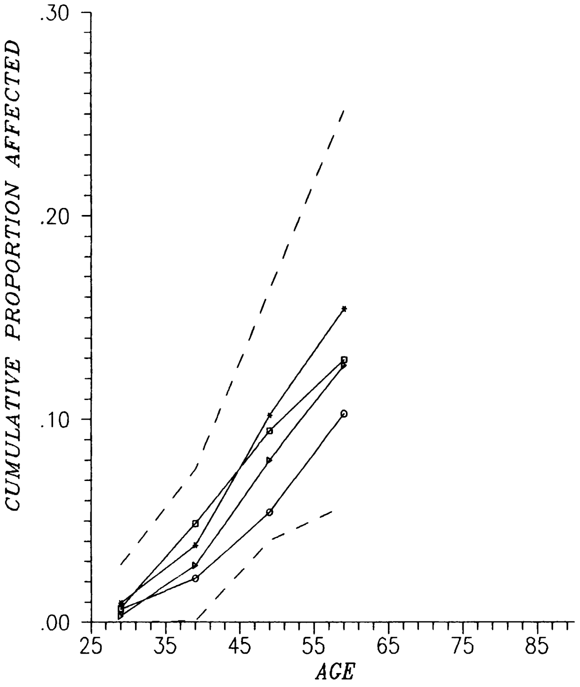

Figure 1 Cumulative risk of breast cancer for mothers of cases aged 20–29 years. Dotted line (– –) denotes 95% confidence interval; asterisked line (*—*) denotes observed values; triangled line (△–△) denotes step function values; squared line (□–□) denotes POINTER values; and circled line (○–○) denotes normal values.

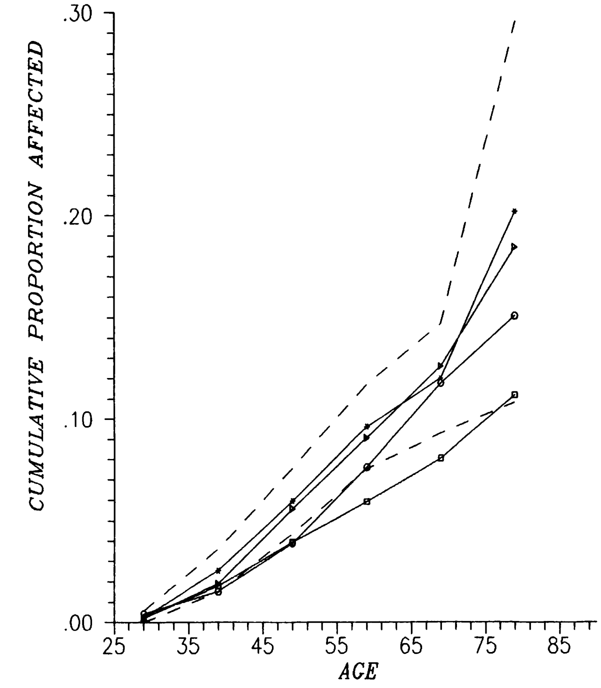

Figure 2 Cumulative risk of breast cancer for mothers of cases age 30–39 years. Scheme for plotted values is as in fig. 1.

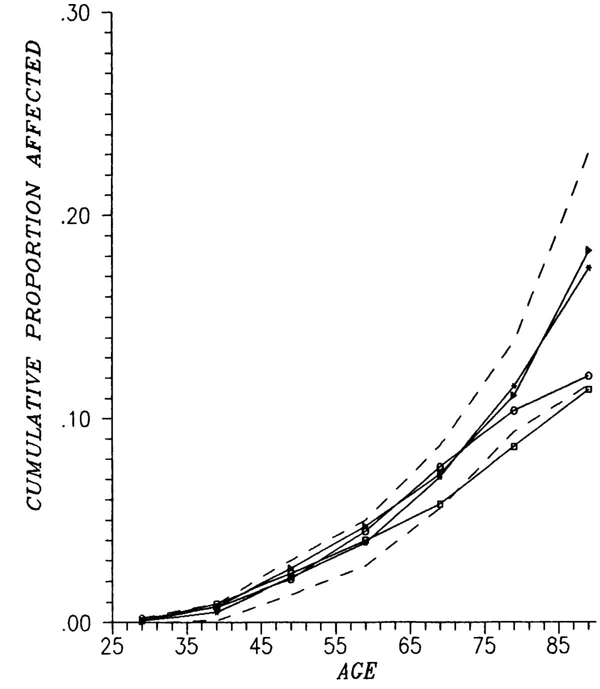

Figure 4 Cumulative risk of breast cancer for mothers of cases age 50–54 years. Scheme for plotted values is as in fig. 1.

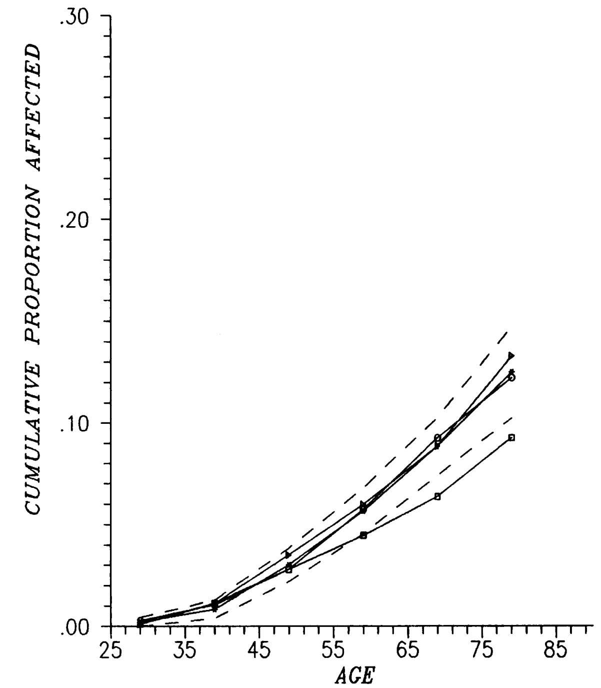

Figure 3 Cumulative risk of breast cancer for mothers of cases age 40–49 years. Scheme for plotted values is as in fig. 1.

fares poorly here, tending to underestimate the risk to mothers of cases whose age at onset is 30–54 years. For mothers of cases age 20–29 years, POINTER provides as good a fit to the observed data for mothers as does the step function. However, as the age of the case increases from 30 years onward, so does the amount by which POINTER underestimates the true risk to mothers, causing the POINTER curve to fall outside the confidence band for the observed values.

The fit for sisters of white cases, presented by both the case's age at onset and breast cancer status of the mother, is generally quite good under the step-function model but varies somewhat by the case's age at onset. Although at no time do the predicted risk curves under the step-function model fall outside the confidence bands, the ability of the predicted curve to match the observed curve increases greatly with the case's age at onset. This holds true regardless of the disease status of the mother. Sisters of young probands are so few that the Kaplan-Meier estimates for these women are not stable, as evidenced by the large standard errors—and hence wide confidence bands—around these estimates. The model tends in this instance to underestimate the risk to sisters of young

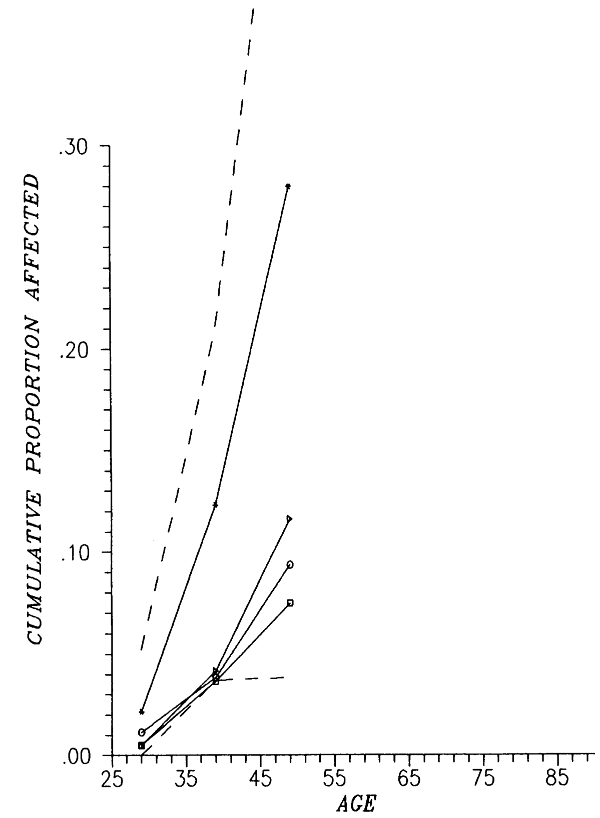

Figure 5 Cumulative risk of breast cancer for sisters of cases age 20–39 years and with an affected mother. Scheme for plotted values is as in fig. 1.

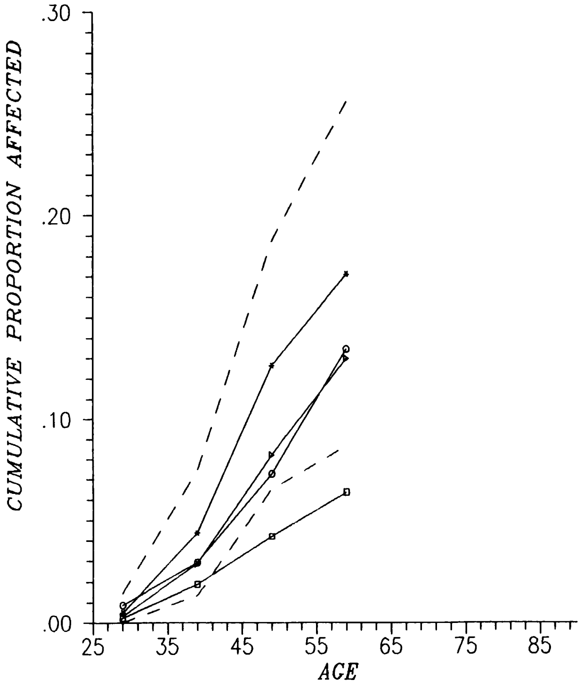

Figure 6 Cumulative risk of breast cancer for sisters of cases age 40–49 years and with an affected mother. Scheme for plotted values is as in fig. 1.

(<40 years old) cases. When the case's age at onset is >40 years, the step-function model does well. The normal plots once again slightly underestimate the true risk to sisters, and POINTER greatly underestimates the true risk, falling outside the confidence bands for all categories of sisters. The curves predicted under POINTER tend to underestimate the risk to both mothers and sisters and fall outside the confidence band for the observed values in almost every instance. For mothers and sisters of the controls, both the step-function model and the normal distribution model provide good fits to the observed values. The fact that these models fit the control series indicates that the models are successful in predicting age-specific breast cancer risk for the general population.

Discussion

The results of the present study provide evidence for the existence of a rare autosomal dominant allele segregating for increased susceptibility to breast cancer. The allele appears to confer particularly high risk at young ages. The rarity of the susceptibility allele implies that the number of women in the general population who are carriers for the allele is small and that the majority of women diagnosed with breast cancer can probably be defined as nongenetic cases. The best-fitting models in this study predict that, among those women who do carry the allele, nearly all will become affected with breast cancer if they live long enough. Women who are noncarriers are predicted to have essentially the reported lifetime risk seen for the general U.S. population, i.e., approximately 10%. These results concur with those of Newman et al. (1988), who performed segregation analyses on a subset of 1,579 breast cancer cases taken from the San Francisco Bay area and metropolitan Detroit portions of the CDC data set and who also found evidence for the existence of a rare autosomal dominant allele. Although the autosomal dominant single-major-locus model fit well here, evidence has also been presented elsewhere to suggest that there may be more than one locus underlying breast cancer and that shared gene(s)

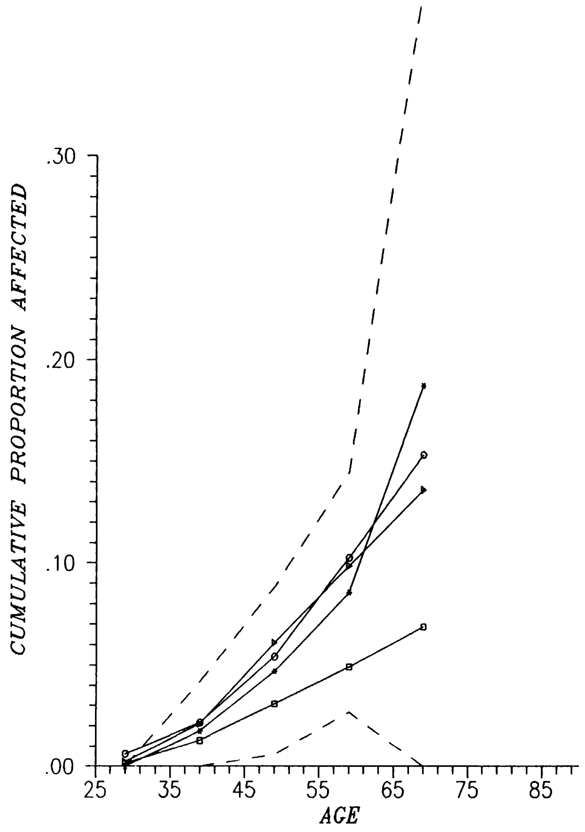

Figure 7 Cumulative risk of breast cancer for sisters of cases age 50–54 years and with an affected mother. Scheme for plotted values is as in fig. 1.

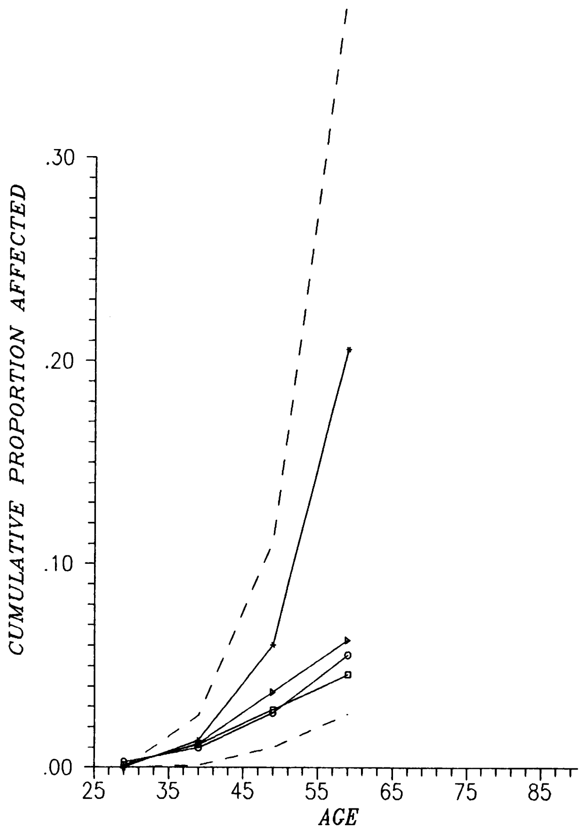

Figure 8 Cumulative risk of breast cancer for sisters of cases age 20–39 years and with an unaffected mother. Scheme for plotted values is as in fig. 1.

and unique gene(s) may exist for breast and ovarian cancer (Schildkraut et al. 1989).

In this analysis, POINTER yields consistently lower cumulative risks. This may be explained by the fact that the other models provide greater flexibility in fitting the data than does the POINTER model. POINTER analyses are constrained by liability classes so that only *d, t, q,* and h^2 (heritability) may vary. The age-specific liability classes are determined by thresholds on a mixture of normal liability components. The other models in the present study allow for estimation of parameters representing both genotype- and age-dependent risks and hence afford greater flexibility in fitting the observed age-dependent risk data.

Furthermore, it should be noted that POINTER imposes a constraint on the model specified; namely, the use of liability thresholds forces a relationship between age at onset and genotype. When more than one genotype is at risk, individuals in low-frequency liability classes will more frequently have a high-risk genotype than will those in high-frequency liability classes, and hence the risk to relatives of the former will be predicted to be greater than for relatives of the latter. To demonstrate this effect, the POINTER analysis was performed again by using the CDC pedigrees with one change. Age at onset was randomly reassigned across families within each class of relative (mothers or sisters). As was true with the nonrandomized data set, an autosomal dominant model provides the best fit to the data (t = 2.104, q = .0014). The model once again predicts that early-onset cases are more likely to be genetic cases (Aa) than are later-onset cases and that the age-specific as well as the cumulative risk of breast cancer for mothers and sisters is dependent on the proband's age at onset; however, this relationship was no longer present in the observed data set after randomization. These findings indicate that investigators working with a disease for which the relationship between age at onset and disease status is unknown—

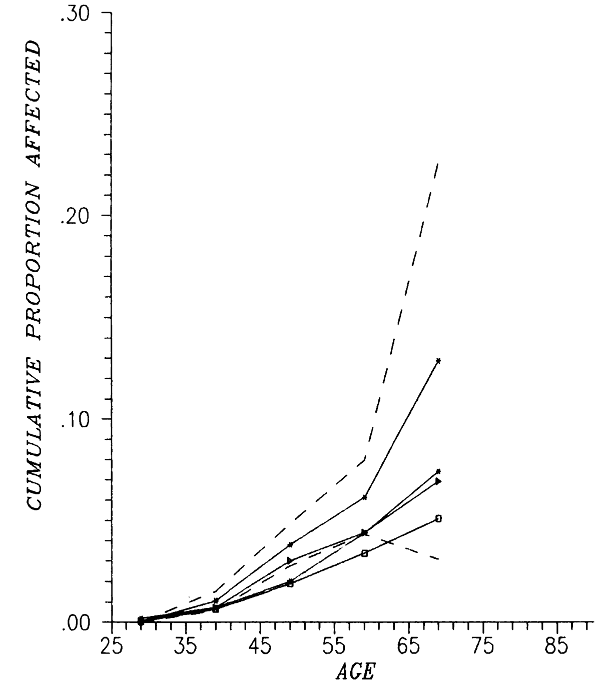

Figure 9 Cumulative risk of breast cancer for sisters of cases age 40–49 years and with an unaffected mother. Scheme for plotted values is as in fig. 1.

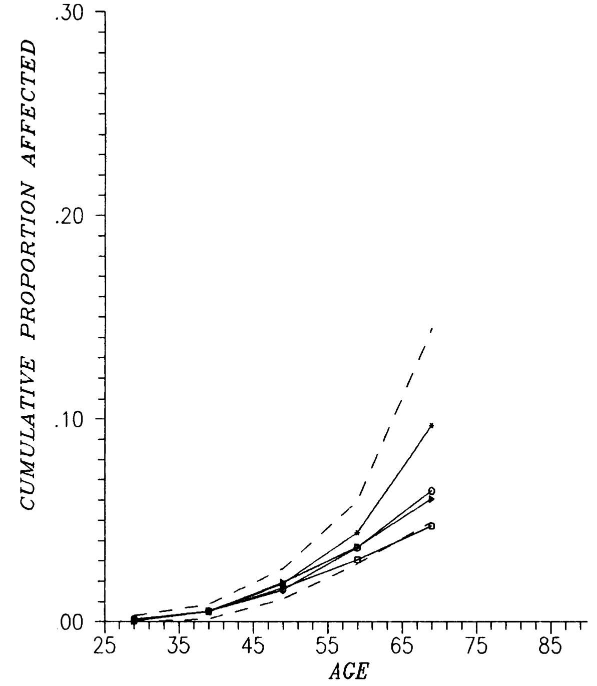

Figure 10 Cumulative risk of breast cancer for sisters of cases age 50–54 years and with an unaffected mother. Scheme for plotted values is as in fig. 1.

or known, but not as described by POINTER—should be aware of these implicit assumptions regarding age. Furthermore, the same effect occurs when liability classes are defined by other criteria, such as sex.

Breast cancer as a disease does not have quantifiable laboratory findings; one must rely on clinical observations for diagnosis. Since early detection of breast cancer is one of the most important factors in determining survival, it becomes extremely important to identify, before visible onset of the disease, those women who are at greater risk. Given the very high probability of a carrier becoming affected with breast cancer, the question arises as to how these women can be identified. It has been shown in this data set that the number of first-degree relatives affected with breast cancer, along with the ages at which these relatives became affected, are the two most important factors determining a woman's risk (Claus et al. 1990*a*). The present study predicts that the great majority of breast cancers are nongenetic. The implication of this finding is that most women in the general population are at some risk for breast cancer, and for these women it will be important to identify other risk factors.

The present study is unique in its attempt to measure how good a fit various genetic models provide to an observed data set, for it calculates the expected age-specific risk of breast cancer under a given genetic model and then compares them with the observed age-specific risk. In general, most investigators test a series of nested likelihood models and then choose the model with the best likelihood, without providing any means of testing goodness of fit.

In the past, family studies involving breast cancer have focused on small numbers of large pedigrees. Such studies have suffered from lack of power due to small size and, in general, from problems including nonrandom ascertainment of families with seemingly unusually high rates of cancers; many of these samples were obtained from clinic populations, and control series were not employed, making generalization difficult. An important strength of the present study is that the CDC data set consists of incident cases who were selected without regard to family history. An analysis of this type has not, to our knowledge, been attempted on so large a population-based series of respondents as that provided by the CDC data set. It is hoped that the results presented in the present paper will give physicians and their patients a reference point from which to begin to assess the possible risks associated with various patterns of family history of breast cancer.

Acknowledgments

The authors wish to acknowledge the contributors to the Cancer and Steroid Hormone Study. Study design and coordination were done at the Division of Reproductive Health, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control: principal investigator, George L. Rubin; project director, Phyllis A. Wingo; project associates, Nancy C. Lee, Michele G. Mandel, and Herbert B. Peterson. Data collection centers and principal investigators were as follows: Atlanta, Raymond Greenberg; Connecticut, J. Wister Meigs and W. Douglas Thompson; Detroit, G. Marie Swanson; Iowa, Elaine Smith; New Mexico, Charles Key and Dorothy Pathak; San Francisco, Donald Austin; Seattle, David Thomas; Utah, Joseph Lyon and Dee West. Pathology review principal investigators were Fred Gorstein, Robert McDivett, and Stanley J. Robboy. Project consultants were: Lonnie Burnett, Robert Hoover, Peter M. Layde, Howard W. Ory, James J. Schlesselman, David Schottenfeld, Bruce Stadel, Linda Webster, and Colin White. Pathology consultants were Walter Bauer, William Christopherson, Deborah Gersell, Robert Kurman, Allen Paris, and Frank Vellios. The Cancer and Steroid Hormone Study was supported by interagency agreement 3-Y01-HD-8-1037 between the Centers for Disease Control and the National Institute of Child Health and Human Development, with additional support from the National Cancer Institute. This work was supported by NIH grant GM39812 from NIGMS (to N.R.). E.B.C. was supported by USPHS training grant 2-T32-ES07085. W.D.T. and data collection in Connecticut were supported by contract 200-80-0561 from the Centers for Disease Control.

References

Bishop DT, Cannon-Albright L, McLellan T, Gardner EJ, Skolnick MH (1988) Segregation and linkage analysis of nine Utah breast cancer pedigrees. Genet Epidemiol 5: 151–169

Claus EB, Risch N, Thompson WD (1990*a*) Age of onset as an indicator of familial risk of breast cancer. Am J Epidemiol 131:961–972

——— (1990*b*) Using age of onset to distinguish between subforms of breast cancer. Ann Hum Genet 54:169–177

Dixon WJ (ed) (1983) BMDP statistical software. Regents of the University of California, Los Angeles

Go RCP, King MC, Bailey-Wilson J, Elston RC, Lynch HT (1983) Genetic epidemiology of breast cancer and associated cancers in high-risk families. I. Segregation analysis. J Natl Cancer Inst 71:455–461

Goldstein AM, Haile RWC, Hodge SE, Paganini-Hill A, Spence MA (1988) Possible heterogeneity in the segregation pattern of breast cancer in families with bilateral breast cancer. Genet Epidemiol 5:121–133

Goldstein AM, Haile RWC, Marazita ML, Paganini-Hill A (1987) A genetic epidemiologic investigation of breast cancer in families with bilateral breast cancer. I. Segregation analysis. J Natl Cancer Inst 78:911–918

Kaplan EB, Elston RC (1972) A subroutine package for maximum likelihood estimation (MAXLIK). Inst Stat Mimeo Ser 823. University of North Carolina, Chapel Hill

Lalouel JM, Morton NE (1981) Complex segregation analysis with pointers. Hum Hered 31:312–321

Lalouel JM, Yee S (1980) POINTER: computer programs for complex segregation analysis of nuclear families with pointers. Popul Genet Lab Tech Rep 3

Lynch HT, Conway T, Fitzgibbons RJ Jr, Schreiman J, Watson P, Marcus J, Fitzsimmons SL, et al (1988*a*) Age of onset heterogeneity in hereditary breast cancer: minimal clues for diagnosis. Br Cancer Res Treat 12:275–285

Lynch HT, Watson P, Conway T, Fitzsimmons SL, Lynch J (1988*b*) Breast cancer family history as a risk factor for early onset breast cancer. Br Cancer Res Treat 11:263–267

Newman B, Austin MA, Lee M, King MC (1988) Inheritance of human breast cancer: evidence for autosomal dominant transmission in high risk families. Proc Natl Acad Sci USA 85:1–5

Sattin RW, Rubin GL, Webster LA, Huezo CM, Wingo PA, Ory HW, Layde PM, Cancer and Steroid Hormone Study (1985) Family history and risk of breast cancer. JAMA 253:1908–1913

Schildkraut J, Risch N, Thompson WD (1989) Evaluating genetic association among ovarian, breast, and endometrial cancer: evidence for a breast-ovarian cancer relationship. Am J Hum Genet 45:521–529

Schwartz AG, King MC, Belle SH, Satariano WA, Swanson GM (1985) Risk of breast cancer to relatives of young breast cancer patients. J Natl Cancer Inst 75:665–668

Williams WR, Anderson DE (1984) Genetic epidemiology of breast cancer: segregation analysis of 200 Danish pedigrees. Genet Epidemiol 1:7–20

Wingo PA, Ory H, Layde PM, Lee NC, Cancer and Steroid Hormone Group (1988) The evaluation of the data collection process for a multicenter, population-based, case-control design. Am J Epidemiol 128:206–217