# Determining Carrier Probabilities for Breast Cancer–Susceptibility Genes BRCA1 and BRCA2

Giovanni Parmigiani, Donald A. Berry, and Omar Aguilar

Institute of Statistics and Decision Sciences, Duke University, Durham, NC

## Summary

**Breast cancer–susceptibility genes BRCA1 and BRCA2 have recently been identified on the human genome. Women who carry a mutation of one of these genes have a greatly increased chance of developing breast and ovarian cancer, and they usually develop the disease at a much younger age, compared with normal individuals. Women can be tested to see whether they are carriers. A woman who undergoes genetic counseling before testing can be told the probabilities that she is a carrier, given her family history. In this paper we develop a model for evaluating the probabilities that a woman is a carrier of a mutation of BRCA1 and BRCA2, on the basis of her family history of breast and ovarian cancer in first- and second-degree relatives. Of special importance are the relationships of the family members with cancer, the ages at onset of the diseases, and the ages of family members who do not have the diseases. This information can be elicited during genetic counseling and prior to genetic testing. The carrier probabilities are obtained from Bayes's rule, by use of family history as the evidence and by use of the mutation prevalences as the prior distribution. In addressing an individual's carrier probabilities, we incorporate uncertainty about some of the key inputs of the model, such as the age-specific incidence of diseases and the overall prevalence of mutations. There is some evidence that other, undiscovered genes may be important in explaining familial breast cancer. Users of the current version of the model should be aware of this limitation. The methodology that we describe can be extended to more than two genes, should data become available about other genes.**

## Introduction

Recent years have marked important progress in our understanding of inherited susceptibility to breast cancer, with the identifications of the BRCA1 and BRCA2 genes (Futreal et al. 1994; Miki et al. 1994; Wooster et al. 1995; Szabo and King 1997). Inherited mutation of one of these genes results in making its carriers at a much increased risk of developing breast cancer as well as ovarian cancer. It has become possible and increasingly common to test for the presence of these mutations. Although the availability of testing represents a major opportunity both scientifically and clinically, it also creates important challenges for women facing the testing decision and for physicians and genetic counselors interested in helping women to understand risks of breast cancer and available preventive options (Hoskins et al. 1995). The chance of carrying a genetic mutation varies markedly from woman to woman, depending on family history of breast cancer and related cancers. As a result, increasing attention needs to be given to understanding and conveying risk information in an individualized way. In this context, risk-prediction algorithms that fully explicitate our knowledge about the nature of the inheritance mechanism can contribute in an important way to a woman's decision.

Testing for BRCA1 and BRCA2 mutations is expensive (currently the cost is $2,400 for both genes), and a positive outcome can affect a person's life in important ways: in eligibility for health insurance and in potential employment discrimination, as well as in physical and psychological aspects. A positive test or simply the perception of a high risk can lead to aggressive management, ranging from more-frequent mammographies to bilateral mastectomy, again with substantive consequences on a woman's life. A crucial step in counseling a women facing these decisions is an accurate evaluation of the probability that she carries a mutation. Also, after the test(s) is performed, the relevant calculation for decision making is the posterior probability of mutation, given the outcome of the test. Accurate assessment of these probabilities requires accurate prior input.

The purpose of this article is to describe a method for finding the probability that a particular family member

(the counseland) carries a germ-line mutation at BRCA1 or BRCA2, on the basis of her family's history of breast cancer and ovarian cancer. Although the method applies to both women and men, the counseland is usually female. The family history includes the counseland and her first- and second-degree relatives. For each member, we ascertain both whether he or she has been diagnosed with breast cancer and either the age at diagnosis or, if cancer free, the current age or the age at death; we follow a similar procedure for ovarian cancer, if the member is female.

Other models have addressed the risk of breast cancer for women with a family history of the disease (Gail et al. 1989; Claus et al. 1994; Houwing-Duistermaat and Van Houwelingen 1997). Although useful, none of these models directly addresses the probability that the woman carries a mutation at BRCA1 or BRCA2, as is needed in genetic counseling and genetic-testing decisions. Shattuck-Eidens et al. (1995) have provided a summary table of estimated probabilities of carrying a BRCA1 mutation, for 11 categories of high-risk women. A logistic-regression approach has been proposed by Couch et al. (1997). These models use summary measures of family history, such as the number of affected relatives. A woman with given numbers of first- and second-degree relatives with breast cancer and/or ovarian cancer can have a wide range of probability of mutation, depending on the pedigree structure, the exact relationship of the affected members, and the family size. In practical counseling situations, it can be important to account for these fully.

We use basic Bayesian methods to incorporate all relevant family history, up to second-degree relatives. In addition, we incorporate uncertainty in the inputs. We proceed by computing a likelihood ratio for the observed family history. An estimate of the mutation frequency in the population provides the probability of a mutation in the counseland, prior to the ascertainment of family history. Bayes's rule applies to determination of the probability of a mutation, given family history. This probability is posterior to family history but is prior to genetic testing. Calculation of the likelihood is performed by use of analytic expressions, for given values of the model parameters. These are evaluated by use of software developed, in the C programming language, expressly for this project, since none of the publically available linkage-analysis software packages can handle risk predictions for mutations of two genes each of which leads to different penetrance functions for two different diseases. Uncertainty about the model parameters is addressed by use of a Monte Carlo simulation, programmed by means of the statistical package S-plus.

The relevance of Bayesian calculations for genetic counseling has been recognized for a long time (Murphy and Mutalik 1969). The use of likelihood ratios in pedigree analysis is also well established (Thompson 1986; Thompson and Guo 1991). The use of Bayesian techniques in software, for determination of the positive predictive power of family history, has a precedent in the work of Szolovits and Pauker (1992). A likelihood-approach alternative to ours, for incorporation of uncertainty in the estimates of genetic parameters, has been discussed in the work of Leal and Ott (1994). The clinical implications of our probability calculations for BRCA1 only have been developed in the work of Berry et al. (1997).

The outline of this article is as follows. In the Methods section, we introduce the notation, model assumptions, and evidence regarding the population frequency of the BRCA1 and BRCA2 mutations, as wll as the disease-incidence functions associated with these mutations. In the Results section, we present results for two family histories that arise from clinical experience. In the Appendix, we give explicit expressions for the likelihood ratio.

## Methods

### Definitions and Notation

In this section we establish the notation and review the assumptions. We assume that individuals inherit two BRCA1/BRCA2 alleles, one from each parent, and that alleles are either normal or mutated; mutations are inherited independently. We assume an autosomal dominant inheritance of mutations, empirically supported by the analysis presented Claus et al. (1991).

At each locus, an individual can have zero, one, or two mutations. Our model addresses the joint probability of a specific configuration of BRCA1 and BRCA2 genes, represented by a two-dimensional vector. We use the notation $P[\text{BRCA1} = i_1, \text{BRCA2} = i_2]$ to denote the probability of the counseland having $i_1$ mutated copies of BRCA1, $i_1 = 0, 1, 2$, and $i_2$ mutated copies of BRCA2, $i_2 = 0, 1, 2$.

The frequencies of mutations in the allele population are indicated by $f_1$ and $f_2$, for BRCA1 and BRCA2, respectively. In the absence of information about disease and family history, the probabilities that an individual inherits a given number of mutated copies of BRCA1 are $P[\text{BRCA1} = 2] = f_1^2$, $P[\text{BRCA1} = 1] = f_1(1 - f_1)$, and $P[\text{BRCA1} = 0] = (1 - f_1)^2$. The probability that an individual carries at least one BRCA1 mutation is $\pi_1 = f_1^2 + 2f_1(1 - f_1)$. Of these individuals, the fraction carrying two mutations is therefore $g_1 = f_1 / [f_1 + 2(1 - f_1)]$. Similar expressions hold for BRCA2. Joint probabilities can be obtained on the basis of independence.

*Updating*

Our goal is to compute the joint probability distribution of the BRCA1 and BRCA2 genetic-status variables of a woman, given her family history (subsequently denoted as "fam.hist." in mathematical expressions). From Bayes's rule,

$$P[\text{BRCA1}, \text{BRCA2}|\text{fam.hist.}] =$$

$$\frac{P[\text{BRCA1}]P[\text{BRCA2}]P[\text{fam.hist.}|BRCA1, BRCA2]}{P[\text{fam.hist.}]}$$

(1)

Although the BRCA1 and BRCA2 variables can be assumed to be independent a priori, they will typically not be independent conditional on family history.

On the basis of the joint probability distribution arising from expression (1), we can compute various summaries that are of interest in both decision making and genetic testing. These include the probability, given her family history, that a woman is a carrier of either a BRCA1 mutation or a BRCA2 mutation,

$$\pi^* = 1 - P[\text{BRCA1} = 0, \ \text{BRCA2} = 0|fam.hist.] \ ,$$

the marginal probabilities, given her family history, that a woman is a carrier of BRCA1,

$$\pi_1^* \equiv P[\text{BRCA1}$$

$$= 1 \text{ or } 2|\text{fam.hist.}]$$

$$= \sum_{i_1=1}^{2} \sum_{i_2=0}^{2} P[\text{BRCA1} = i_1,$$

$$\text{BRCA2} = i_2|\text{fam.hist.}] \ ,$$

and the marginal probabilities $\pi_2^*$, given her family history, that a woman is a carrier of BRCA2, obtained in a manner similar to that given above.

Determining $P[\text{fam.hist.}|BRCA1, BRCA2]$ in expression (1) can be computationally intensive. However, there are three simple building blocks that underlie this calculation: the probability of the genetic status of offspring, given those of their parents; the probability of the genetic status of parents, given those of their offspring; and the probability of disease outcome, given genetic status. The first two are discussed next; the third is discussed in the Evidence subsection. In the Appendix, we show how to compute $P[\text{fam.hist.}|BRCA1, BRCA2]$.

The probability of a configuration of the offsprings' genetic status, given the genetic status of their parents, is computed on the basis of random selection of each offspring's alleles from each parent's two alleles (see Fal-

coner and Mckay 1996). We use the notation "o" to designate the offspring's genetic status, "m" to designate the mother's genetic status, and "f" to designate the father's genetic status. Each of these is a two-dimensional vector. The two coordinates represent the number of mutated alleles at BRCA1 and BRCA2, respectively. The set of possible values is $\{0, 1, 2\} \times \{0, 1, 2\}$. Because inheritances of BRCA1 and BRCA2 mutations are assumed to be a priori independent, we can focus on marginal distributions.

The probability of any joint configuration of the parents' genetic status, given that of one of their offspring, requires a simple application of Bayes's rule. We begin by assuming independence between the genetic status of the two parents, which is tenable when random mating in the population is assumed. This implies $P[\text{m}, \text{f}] = P[\text{m}]P[\text{f}]$, and, when Bayes's rule is applied,

$$P[\text{m}, \text{f}|o] = \frac{P[\text{m}]P[f]P[o|m,f]}{\sum_{m}\sum_{f} P[\text{m}]P[f]P[o|m,f]} \ .$$

*Uncertainty about Genetic Parameters*

From expression (1) we see that inference about the genetic status of the counseland requires knowledge of the mutation frequencies and of cancer rates for both carriers and noncarriers of the mutations. These need to be evaluated for all family members whose history of breast cancer and ovarian cancer (including the lack of such history) is available. These evaluations are based on empirical studies and so are uncertain. Uncertainty can be accommodated by a Bayesian approach, as follows. We denote by $\theta$ the set of parameters indexing the cancer-rate models, and we denote by $P(\theta, f_1, f_2|\text{published data})$ the probability distribution expressing the uncertainty about $(\theta, f_1, f_2)$, on the basis of the published estimates.

We can then use $P(\theta, f_1, f_2 |\text{published data})$ to compute the distribution of the random variables

$$P[\text{BRCA1}, \text{BRCA2}|fam.hist., \theta, f_1, f_2]$$

$$= \{P[\text{BRCA1}, \text{BRCA2}|f_1, f_2]$$

$$\times \ P[\text{fam.hist.}|\text{BRCA1}, \text{BRCA2}, \theta, f_1, f_2]\}/$$

$$\times \ P[\text{fam.hist.}|\theta, f_1, f_2]$$

with BRCA1 $= 0, 1, 2$ and BRCA2 $= 0, 1, 2$. This has the feature of examining separately the uncertainty deriving from the unknown genetic status of the family members, which is integrated out, from the uncertainty about the population rates and prevalence. Alternatively, we can write the probability distribution of interest, in

CUMULATIVE RATES FOR BRCA1 CARRIERS

CUMULATIVE RATES FOR BRCA2 CARRIERS
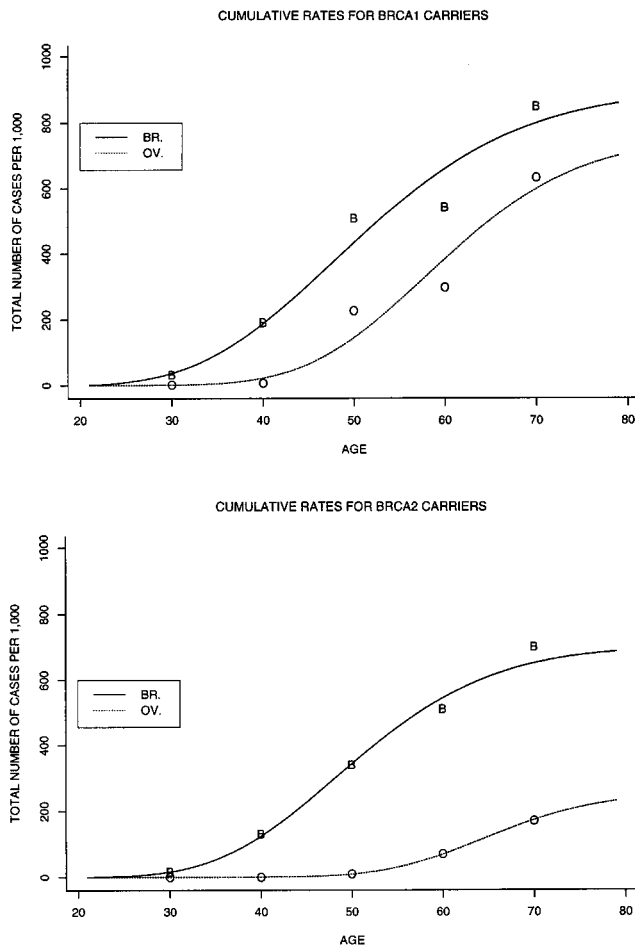
**Figure 1** Cumulative rates of breast cancer (B) and ovarian (O) cancer, for mutation carriers, for (*top*) BRCA1 (based on Easton et al. 1995) and (*bottom*) BRCA2 (based on data of D. F. Easton, personal communication). The lines indicate our interpolation, based on a three-parameter gamma c.d.f, with one of the parameters being the asymptote, which may be <1.

presence of estimation error, by integrating out the unknown parameters:

$$P[\text{BRCA1}, \text{BRCA2}|\textit{fam.hist.}] =$$

$$\int P[\text{BRCA1}, \text{BRCA2}|\textit{fam.hist.}, \theta, f_1, f_2]$$

$$p(\theta, f_1, f_2|\text{published data})df_1 df_2 d\theta \ .$$

We can evaluate the integrand ratio exactly (see the Appendix), and we perform the integration by means of Monte Carlo methods.

A calculation of interest is the probability that an individual will develop breast cancer or ovarian cancer later in life. This is straightforward and is a weighted average of the corresponding probabilities for carriers and noncarriers, when the posterior probabilities of being a carrier and a noncarrier are used as weights.

The impact that the allowance for error in the parameter estimates has, both on the carrier probabilities and on the final risk estimates, will be illustrated in the Discussion section.

*Evidence*

In this section we discuss the empirical evidence about cancer rates and mutation frequency that we will use in our calculations. Easton et al. (1995) have discussed the penetrance of breast cancer and ovarian cancer in BRCA1 carriers, as a function of age. Figure 1 summarizes the rates that they report; breast cancer rates are indicated by a "B," and ovarian cancer rates are indicated by an "O." The curves represent our interpolation. For BRCA1, ~85% of female carriers will get breast cancer, and ~65% will get ovarian cancer, by the age of 70 years. Apparently, these two cancers occur independently among carriers. By age 70 years, $\sim 1 - [(1 - .85)(1 - .65)]$, or 95%, of carriers have one disease or the other, and $\sim .85 \times .65$, or 55%, have both diseases. Similar multiplications apply for other ages.

Figure 1 also shows our interpolations of the rates reported by Easton et al. (1995). The curves are based on fitting a gamma CDF to the reported rates, by use of nonlinear least squares after an arcsine transformation of the response. The asymptote ($\alpha$) of the CDF is a free parameter, allowed to be <1 in order to accommodate incomplete penetrance: not every carrier will develop breast cancer or ovarian cancer. Formally, if $a$ is age and if $R$ is the cumulative rate under consideration, we assume that

$$R(a) = \alpha \int_0^a \frac{\beta^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-x\beta} dx \ .$$

Figure 1 also shows the results of the same type of interpolations on the penetrance functions for BRCA2, with data from D. F. Easton (personal communication; also see Easton et al. 1997). Data on the penetrance functions for male breast cancer are not available at this time. We have assumed that the shape of the penetrance function for male BRCA2 carriers is the same as that for female BRCA2 carries but that it has a smaller asymptote, .15. Penetrance functions for male BRCA1 carriers are assumed to be the same as those for normal individuals.

Although it is rare, we need to consider the possibility that an individual may carry mutations at both BRCA1 and BRCA2. No data are available concerning incidence of disease among carriers of mutations at both genes. In our model, their penetrance functions are determined as the cumulative distribution of the minimum of two in-
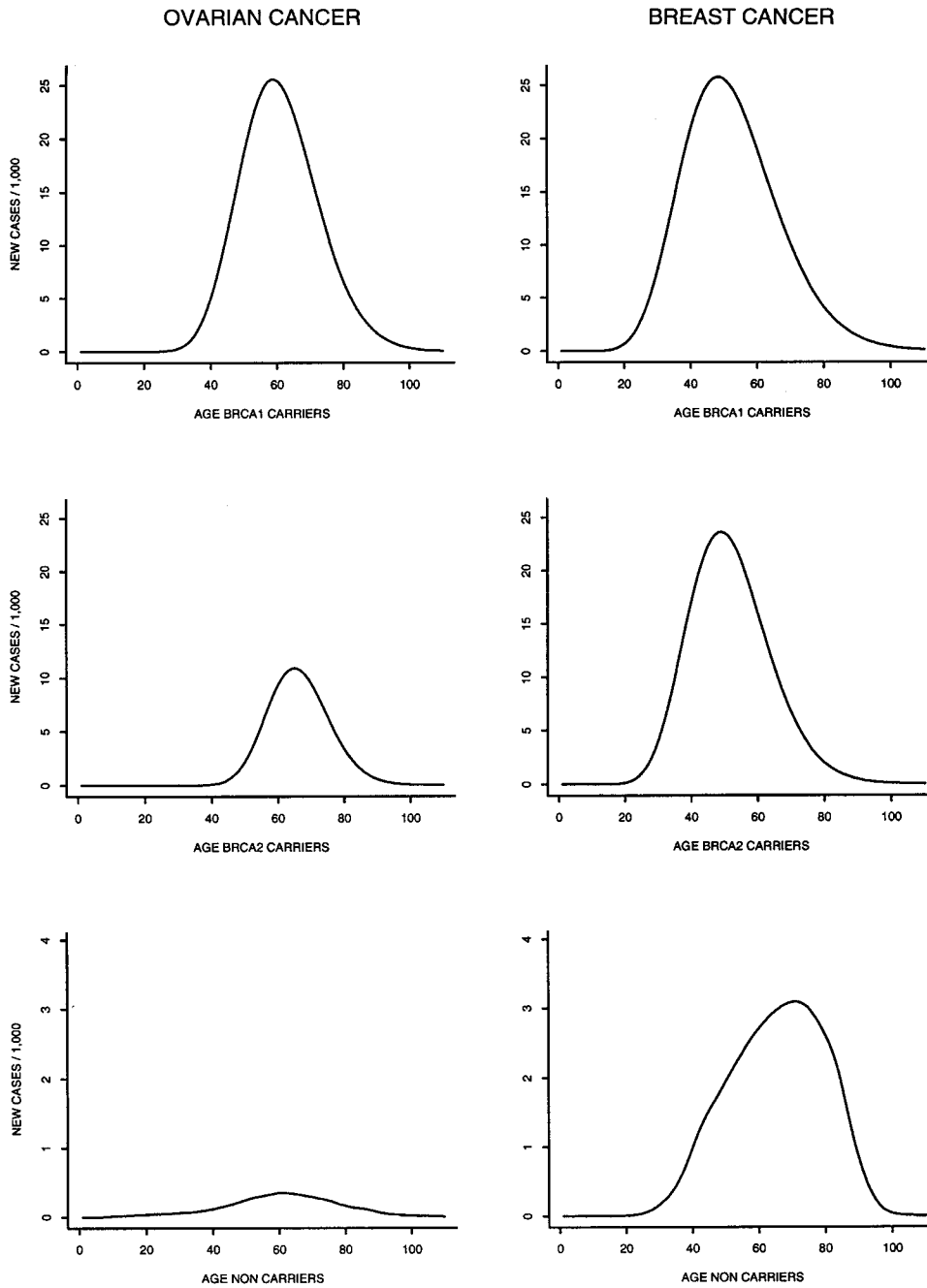
**Figure 2** Number of breast cancer cases and ovarian cancer cases, for both BRCA1 and BRCA2 carriers and noncarriers. The scales for carriers and noncarriers are different, for better resolution. All else being equal, ovarian cancer is a stronger indication of a mutation than is breast cancer, and earlier-onset breast cancer is a stronger indication than is later-onset breast cancer.

dependent events: cancer due to BRCA1 and cancer due to BRCA2. The results are not very sensitive to this assumption.

The disease history for a family member can include bilateral breast cancer. Estimation of rates of bilateral recurrence is difficult, because of the lack of good data. In the current version of our model, we approximate the
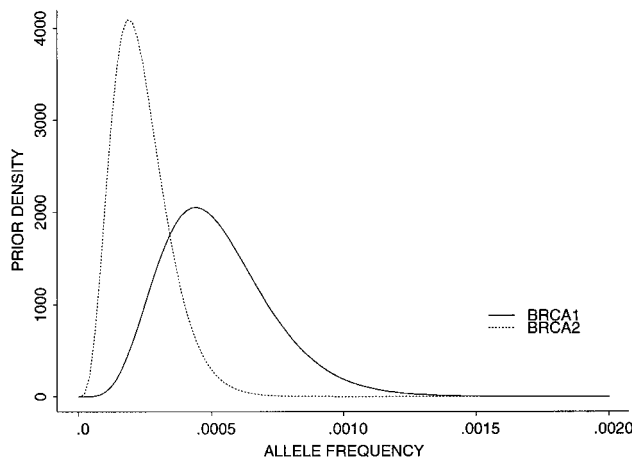
rates for bilateral recurrences by assuming that the observed breast cancer rates $R(a)$ represent the distribution of the minimum of two independent events: cancer in the left breast and cancer in the right breast. The cumulative rates of the time to cancer in a particular breast will then be $R'(a) = 1 - \sqrt{1 - R(a)}$. The likelihood for bilateral breast cancer cases that are diagnosed at ages
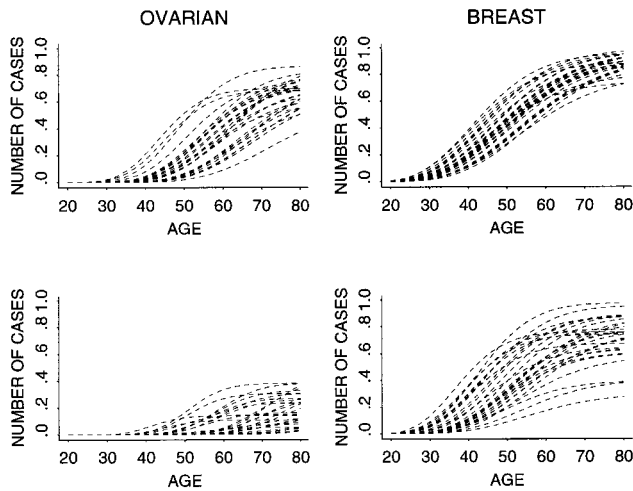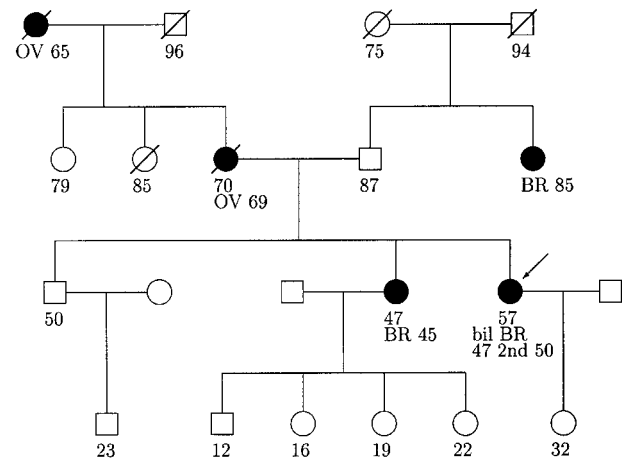
**Figure 3** Probability distribution of penetrance functions. Each graph is a sample of 30 penetrance curves. The upper two graphs refer to BRCA1 carriers, and the lower two graphs refer to BRCA2 carriers. Curves are interpreted as in figure 1. The uncertainty about the penetrance of the mutation is captured by variability in the total number of cases at age 80 years. The uncertainty about rates at young ages is high, in relative terms.

$a_1$ and $a_2$, with $a_1 \leqslant a_2$, will then be $r'(a_1)r'(a_2)$, where $r'$ is the derivative of $R'$. In addition to depending on age, the likelihood of a bilateral recurrence can depend on adjuvant treatment or preventive surgery administered after the first primary cancer. These features are not incorporated here.

Cancer rates for normal individuals can be obtained from epidemiological registries such as the SEER data base. Moolgavkar et al. (1979) used SEER data to de-
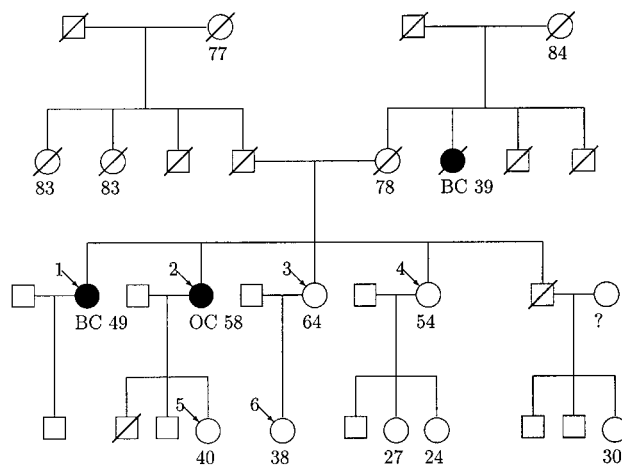


**Figure 5** Family history 1 of section 4. Age, when known, is shown for each family member; it is either age at the time of the present study or age at death.

velop a general model of incidence of breast cancer in the female population. The model accounts for a cohort effect and yields estimates of incidence and mortality as a function of age. We have used the incidence functions of Moolgavkar et al., together with life tables, to determine the breast cancer rates for the general population. Subtraction of the estimated BRCA1/BRCA2 cases leads to an age distribution for noncarriers. We have used a similar strategy for ovarian cancer, on the basis of the results reported by Yancik (1993). Figure 2 summarizes estimated carrier rates and noncarrier rates, for breast cancer and for ovarian cancer, in females. The graphs



**Figure 4** Assumed probability distributions of allele mutation frequencies. The unbroken line is the distribution of $f_1$, whereas the broken line is the distribution of $f_2$.



**Figure 6** Family history 2 of section 4. Age, when known, is shown for each female family member; it is either her age at the time of the assessment or her age at death. If there is no indication of breast cancer or ovarian cancer, then the family member is free of both.

**Table 1**

**Five Scenarios for Penetrance and Prevalence Parameters**

| | Allele Frequency | | Penetrance at Age 110 Years | | | |
| | | | Breast Cancer | | Ovarian Cancer | |
| Scenario | $f_1$ | $f_2$ | BRCA1 | BRCA2 | BRCA1 | BRCA2 |
|---|---|---|---|---|---|---|
| 1(Baseline) | .0006 | .00022 | .90 | .70 | .75 | .25 |
| 2 | .0008 | .00030 | .95 | .75 | .85 | .35 |
| 3 | .00045 | .000165 | .95 | .75 | .85 | .35 |
| 4 | .0008 | .00030 | .70 | .40 | .60 | .15 |
| 5 | .00045 | .000165 | .70 | .40 | .60 | .15 |

**Table 2**

**Probability That Counseland of Family 1 Carries BRCA1 or BRCA2, in Five Scenarios Given in Table 1**
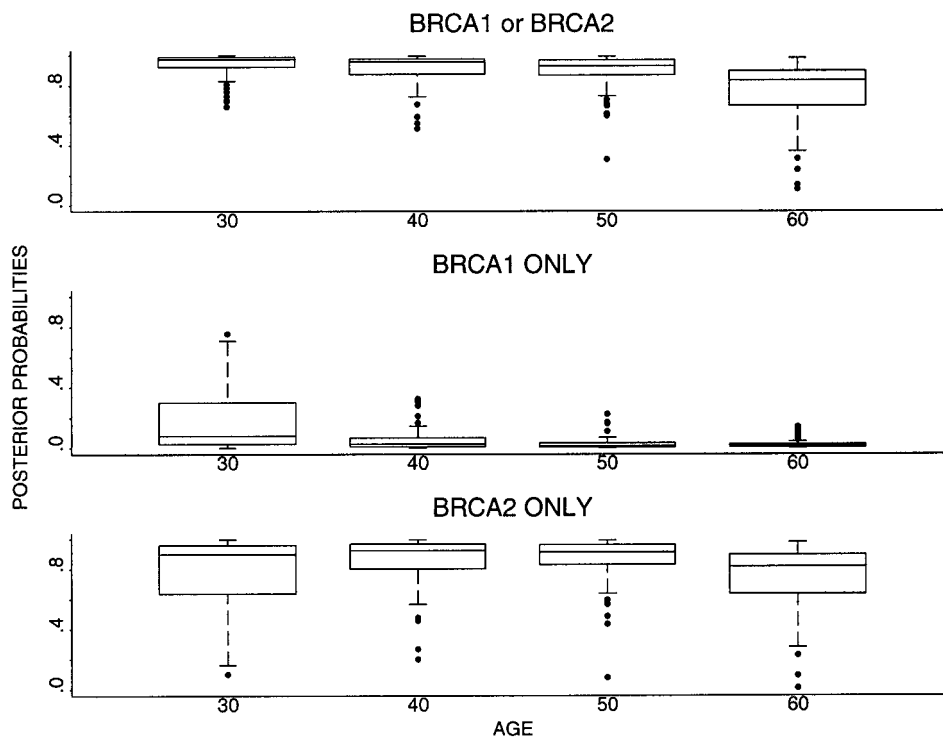
| | | Probability | |
| Scenario | $\pi_i^*$ | BRCA1 | BRCA2 |
|---|---|---|---|
| 1 | .966 | .589 | .377 |
| 2 | .978 | .429 | .549 |
| 3 | .960 | .427 | .533 |
| 4 | .966 | .843 | .123 |
| 5 | .945 | .827 | .118 |

represent the number of cases, per year, per 1,000 women. Carriers have a higher incidence and a higher propensity to develop breast cancer at a young age.

We have incorporated uncertainty about the carrier rates, as discussed in the Methods section. Easton et al. (1995) have provided confidence intervals for the incidence rates at age 50 years. On the basis of these, we have assigned probability distributions to the penetrance functions for breast cancer rates and ovarian cancer rates. We have used beta distributions for the $\alpha$ parameters and have used Gaussian distributions for the $\nu$

and $\beta$ parameters. To illustrate the implications of our specifications, we have generated a sample of the resulting curves; these are shown in figure 3.

The frequency of genetic mutations at BRCA1 has been discussed by Ford and Easton (1995), who provide a 95% confidence interval of .0002–.001. for $f_1$. On the basis of this, we have assigned a beta probability distribution on $f_1$, with parameters 6.29 and 12,000, placing ~95% of the mass on the interval (.0002–.001). The prevalence of genetic mutations at BRCA2 has been reviewed by Andersen (1996). Uncertainty assessments are



**Figure 7**    Probabilistic sensitivity analysis, family 1. Boxplots represent the distribution of the probability that family member 1 of family 1 carries a mutation of BRCA1 or BRCA2, with her age being allowed to change. We are considering the modified version of the pedigree, in which the counseland's father is assumed to have breast cancer and in which the mother is cancer free. Each boxplot is based on the same Monte Carlo sample of size 100. The variability is due to uncertainty in the cancer rates and in the allele frequency.

**Table 3**

**Probability That Counseland of Family 1 Carries BRCA1 or BRCA2, in Five Scenarios Given in Table 1**

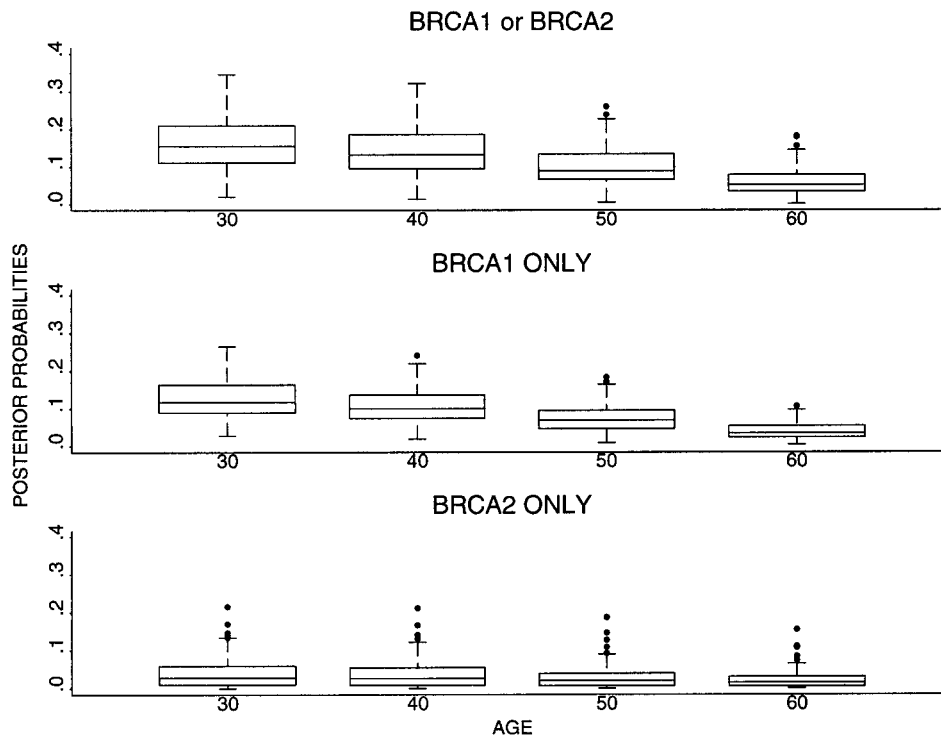| | PROBABILITY | | |
|---|---|---|---|
| SCENARIO | $\pi_i^*$ | BRCA1 | BRCA2 |
| 1 | .961 | .015 | .947 |
| 2 | .974 | .011 | .964 |
| 3 | .953 | .010 | .943 |
| 4 | .883 | .072 | .817 |
| 5 | .803 | .064 | .741 |

NOTE.—The pedigree is modified compared with that shown in figure 5; here the father of the proband has breast cancer at age 70 years, and the mother of the proband is free of cancer.

not available, and our own uncertainty assessment is based on a comparison with BRCA1. The assumed distributions of allele frequencies of BRCA1 and BRCA2 are shown in figure 4. Although the proportion of families whose cancer is explained by BRCA2 is lower, given the difference in penetrance, it is possible that the frequency of BRCA2 is similar to or even greater than that of BRCA1. Under the assumptions used in figure 4, there is a probability of ~10% that the frequency of BRCA2

**Table 4**

**Posterior Probabilities for Members of Family 2, in Five Scenarios Given in Table 1**

| SCENARIO AND TYPE OF PROBABILITY | PROBABILITY FOR FAMILY MEMBER | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1: | | | | | | |
| $\pi_i^*$ | .035 | .036 | .007 | .012 | .025 | .005 |
| BRCA1 | .017 | .018 | .002 | .005 | .019 | .003 |
| BRCA2 | .018 | .017 | .005 | .007 | .005 | .002 |
| 2: | | | | | | |
| $\pi_i^*$ | .040 | .041 | .008 | .013 | .035 | .006 |
| BRCA1 | .015 | .016 | .001 | .004 | .026 | .003 |
| BRCA2 | .025 | .025 | .006 | .009 | .009 | .003 |
| 3: | | | | | | |
| $\pi_i^*$ | .021 | .021 | .004 | .007 | .019 | .003 |
| BRCA1 | .008 | .008 | .0008 | .002 | .013 | .002 |
| BRCA2 | .013 | .013 | .003 | .005 | .005 | .001 |
| 4: | | | | | | |
| $\pi_i^*$ | .085 | .089 | .025 | .034 | .039 | .011 |
| BRCA1 | .058 | .062 | .013 | .021 | .032 | .008 |
| BRCA2 | .026 | .027 | .012 | .012 | .007 | .003 |
| 5: | | | | | | |
| $\pi_i^*$ | .047 | .049 | .014 | .019 | .021 | .006 |
| BRCA1 | .032 | .034 | .007 | .012 | .018 | .004 |
| BRCA2 | .014 | .015 | .006 | .007 | .004 | .002 |



**Figure 8** Probabilistic sensitivity analysis, family 2. Boxplots represent the distribution of the probability that family member 1 of family 2 carries a mutation of BRCA1 or BRCA2, with her age being allowed to change. Each boxplot is based on the same Monte Carlo sample of size 100. The variability is due to uncertainty in the cancer rates and in the allele frequency.

**Table 5**

**Posterior Probabilities for Member 1 of Family 2, in Five Scenarios Given in Table 1, When 30-Year-Old Niece Is Assumed to Have Had Breast Cancer at Age 30 Years**

| | | Probability | |
|---|---|---|---|
| Scenario | $\pi_i^*$ | BRCA1 | BRCA2 |
| 1 | .256 | .164 | .092 |
| 2 | .284 | .148 | .135 |
| 3 | .167 | .088 | .078 |
| 4 | .414 | .347 | .067 |
| 5 | .276 | .232 | .044 |

**Table 7**

**Posterior Probabilities for Member 1 of Family 2, in Five Scenarios Given in Table 1, When 38-Year-old Niece Is Assumed to Have Had Breast Cancer at Age 30 Years**

| | | Probability | |
|---|---|---|---|
| Scenario | $\pi^*$ | BRCA1 | BRCA2 |
| 1 | .130 | .062 | .068 |
| 2 | .140 | .047 | .093 |
| 3 | .075 | .026 | .050 |
| 4 | .289 | .218 | .071 |
| 5 | .177 | .134 | .043 |

is greater than that of BRCA1. Cancer rates for the general population are based on large datasets, and the uncertainty about their values can be ignored.

## Results

In this section we illustrate our approach by using two family pedigrees arising from actual counseling experience in the Duke Specialized Program of Research Excellence in breast cancer. We also consider small hypothetical variations of each pedigree, to show the effect on the probabilities that mutations are carried. The pedigrees are shown in figures 5 and 6.

For each family, we performed the computations under five different scenarios for the penetrance and prevalence parameters. These scenarios are specified in table 1. The baseline scenario corresponds to the best estimates of the parameters, based on the discussion in the Evidence subsection. We also performed a probabilistic analysis, using the distributions obtained in the Evidence subsection.

We begin with family 1, focusing on the woman indicated by the arrow. The probabilities that she carries mutations are shown in table 2. Both the occurrence of bilateral breast cancer and the high number of cases in the family strongly suggest the presence of a mutation.

As a result of the clear line of descent, the probability that she carries either mutation does not change significantly over the different scenarios. However, the probabilities allocated to BRCA1 mutations and to BRCA2 mutations change more markedly. In the high-penetrance scenarios, BRCA2 becomes more likely than BRCA1. This is because, with a higher penetrance for both genes, the ovarian cancer cases and the healthy cases give less support to BRCA1, compared with the baseline case.

We now consider a modified pedigree for family 1, in which the father of the proband had breast cancer at age 70 years and in which the mother is free of cancer. Table 3 shows the results. The presence of a case of male breast cancer helps discriminate between BRCA1 and BRCA2, which receives a high probability.

Figure 7 shows a probabilistic sensitivity analysis of the same calculations, based on the distributions discussed in the Results section. Boxplots represent the distribution of the probability that family member 1 carries a mutation of BRCA1 or BRCA2, with her age at onset being allowed to change. The horizontal axis indicates the age at onset of the first cancer. The second cancer is assumed to follow 3 years later. Each boxplot is based on the same Monte Carlo sample of size 100. The variability is due to uncertainty both in the cancer rates and

**Table 6**

**Posterior Probabilities for Member 1 of Family 2, in Five Scenarios Given in Table 1, When 38-Year-old Niece Is Assumed to Have Had Breast Cancer at Age 38 Years**

| | | Probability | |
|---|---|---|---|
| Scenario | $\pi^*$ | BRCA1 | BRCA2 |
| 1 | .100 | .041 | .059 |
| 2 | .112 | .032 | .080 |
| 3 | .060 | .017 | .042 |
| 4 | .217 | .151 | .066 |
| 5 | .128 | .089 | .039 |

**Table 8**

**Posterior Probabilities for Members of Family 2, in Five Scenarios Given in Table 1, When 40-Year-old Niece Is Assumed to Have Had Breast Cancer at Age 30 Years**

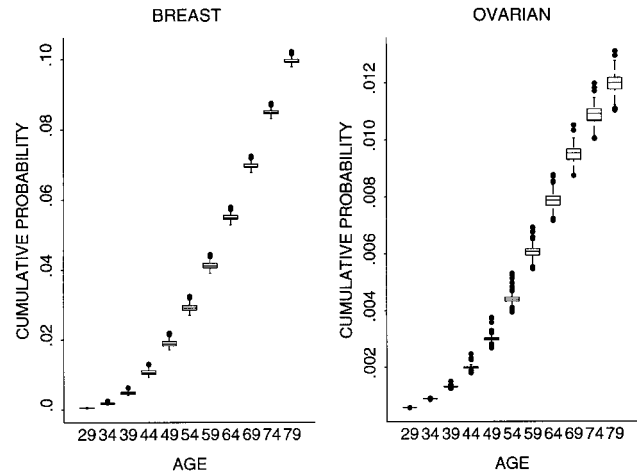| | | Probability | |
|---|---|---|---|
| Scenario | $\pi^*$ | BRCA1 | BRCA2 |
| 1 | .393 | .260 | .133 |
| 2 | .429 | .232 | .197 |
| 3 | .277 | .152 | .125 |
| 4 | .557 | .478 | .079 |
| 5 | .411 | .353 | .057 |

**Table 9**

**Posterior Probabilities for Member 1 of Family 2, in Five Scenarios Given in Table 1, When No Data Regarding Mother's Age or Cancer Status Are Assumed to Have Been Available**

| | PROBABILITY | | |
| --- | --- | --- | --- |
| SCENARIO | $\pi^*$ | BRCA1 | BRCA2 |
| 1 | .236 | .191 | .045 |
| 2 | .319 | .246 | .074 |
| 3 | .195 | .151 | .045 |
| 4 | .266 | .233 | .032 |
| 5 | .162 | .143 | .019 |

in allele frequency. Late onset decreases the overall probability that the individual carries a mutation. The presence of a male breast cancer case is a strong indication that the mutation could be at BRCA2. When the age at onset is 30 years, BRCA1 becomes comparatively more likely, and uncertainty about which of the two candidate genes is mutated becomes larger.

Consider now family 2. For the family members with cancer, we do not know either the ages at the time of the present study or the ages at death, and so, with regard to the other cancers of interest, we censored them at the time of their known cancer. For example, for member 1, we used the information that she was free of ovarian cancer at age 49 years. The age of the sister-in-law of family members 1–4 is unknown and, in any case, has little effect on the calculations, because her daughter is quite young.

When we focus on family members 1–4, there is little evidence of a mutation on their father's side of the fam-



**Figure 9** Cumulative risk of breast cancer for a woman with the family history of members 3 and 4 of family 2. The black dots represent predictions based on our model. The letters "C" represent predictions based on the model of Claus et al. (1994).



**Figure 10** Probabilistic sensitivity analysis of the cumulative risks breast and ovarian cancer for a woman with the family history of members 3 and 4 of family 2. Each boxplot is based on the same Monte Carlo sample of size 100.

ily, except for the cancer in family members 1 and 2; there is only slightly more evidence on their mother's side. The maternal aunt with breast cancer (BC39) is evidence for a mutation; but the mother was free of both breast cancer and ovarian cancer when she died at age 78 years, and this happens for <5% of carriers of BRCA1. The mother's disease-free status substantially weakens the link between the cancers of the maternal aunt and those of of family members 1 and 2. The result is that family members 1–4 have moderate probabilities of mutations. From the data shown in table 4, in the baseline scenario, it can be seen that a BRCA2 mutation is more likely than a BRCA1 mutation. This is the result of both the lack of ovarian cancer cases and the relatively late onset of the breast cancer cases. Decreasing the penetrance parameters leads to a more prominent role of the maternal aunt, resulting in an increased probability of mutation, with BRCA1 increasing more than BRCA2, because of the early onset. However, the probability that an individual carries a mutation is not monotonic in the penetrance. When the penetrance of only BRCA1 is varied, the probability of a BRCA1 mutation reaches a maximum at a penetrance of ~.5.

Figure 8 shows a probabilistic sensitivity analysis of the same calculations, again based on the distributions discussed in the Evidence subsection. Boxplots represent the distribution of the probability that family member 1 carries a mutation of BRCA1 or BRCA2, with her mother's age at death being allowed to change. Decreasing the mother's age at death also leads to a more prominent role of the maternal aunt, with BRCA1 changing more than BRCA2, because of the early onset of the aunt's cancer.
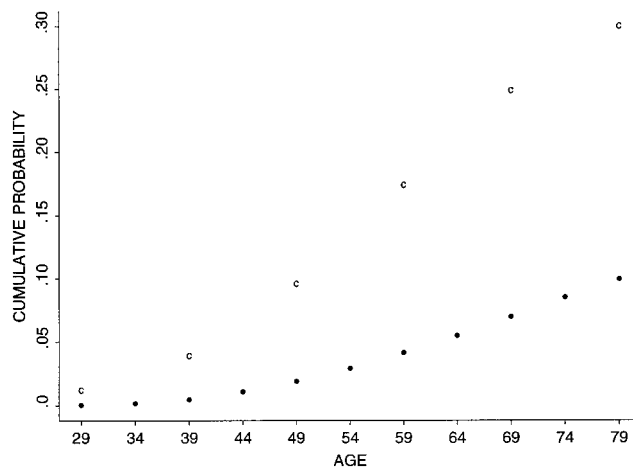
It is also interesting to focus on the probability, $\pi^*$, that family member 1 is a carrier and to consider hypothetical changes in the cancer status of other family members. We begin by assuming that the 30-year-old niece had breast cancer at age 30 years; then the probabilities are those given in table 5, showing a very substantial increase.

If the 38-year-old niece (family member 6) had breast cancer at age 38 years, then the probabilities are those given in table 6. The impact on $\pi^*$ is smaller than it is for the probabilities given in table 5, for two reasons: the older age makes it less likely that breast cancer is due to BRCA1, and the niece's mother is disease free at age 58 years and is therefore less likely to carry the mutation than are the parents of the 30-year-old niece, about whom nothing is known.

To separate the two effects, assume that the 38-year-old niece had breast cancer at age 30 years. Then the probabilities are those given in table 7.

Next, assume that the 40-year-old niece (family member 5) had breast cancer at age 30 years. Then, the probabilities are those given in table 8, which is so large because the 40-year-old niece's mother had ovarian cancer.

Finally, consider the case in which no information about either the mother's age or her cancer status was available; the results are given in table 9. This strengthens the contribution of the maternal aunt, greatly increasing $\pi^*$.

Our model can be used to make predictions about the occurrence of cancer in healthy family members. The predicted cancer rate $R$ (either breast cancer or ovarian cancer or both) at age $a$ is

$$R(a|\text{fam.hist.}) =$$

$$\sum_{i_1=0}^{2} \sum_{i_2=0}^{2} P[\text{BRCA1} = i_1, \ \text{BRCA2} = i_2|\text{fam.hist.}]$$

$$R(a|\text{BRCA1} = i_1, \ \text{BRCA2} = i_2) \ .$$

Using this expression, we can compare our model's results against the similar predictions provided by Claus et al. (1994, table 4). We focus on family 2. Because Claus et al. (1994) consider only breast cancer, we modified family 2 by assuming that family member 2 had breast cancer rather than ovarian cancer. A comparison of the cumulative incidence curves for an individual with the same family history as has been presented for family members 3 and 4 is shown in figure 9. In this case, the predictions based on our model are lower than those of Claus et al., primarily as the result of the large number of unaffected members in the pedigree. We also performed a probabilistic sensitivity analysis of the same predictions, on the basis of the distributions discussed in the Evidence subsection. The results, for both breast cancer and ovarian cancer, are shown in figure 10.

## Discussion

In this paper we have discussed a model for evaluating the probability that a woman is a carrier of a genetic mutation in the BRCA1 and BRCA2 genes, on the basis of her family's history of breast cancer and ovarian cancer in first- and second-degree relatives. We use a Bayesian approach to incorporate both the uncertainty about family members' genetic status and the uncertainty about the prevalence of the mutation in the population and about the cancer rates for carriers.

Our model is currently being used in a randomized trial comparing tailored material to standard printed material, in genetic counseling. Women in the tailored-information arm receive estimates of the probability that they carry a mutation of BRCA1 or BRCA2, computed on the basis of the model described here. The information is presented in graphical (pie chart), numerical, or verbal format, depending on patient preference. Counselees are provided with a range of probabilities reflecting uncertainty about the genetic parameters, obtained according to the procedure discussed in the Uncertainty about Genetic Parameters subsection. The software used to perform the calculations is available to interested investigators, for noncommercial purposes; thus far, it has been licensed to sites that are using it in counseling and in genetic epidemiology research.

The penetrance and prevalences used by the model are stored in an external file, which is input to the program. In this way, it is convenient to create customized computations for subpopulations with different genetic parameters. For example, recent investigations (e.g., Oddoux et al. 1996; Roa et al. 1996; Struewing et al. 1997) suggest that mutations that are specific to the Ashkenazi Jewish population may be associated with prevalence and penetrance functions different from those reported by Easton and colleagues. We have developed an analysis, parallel to that described here, that can be used in counseling women of Ashkenazi Jewish ancestry.

We have considered BRCA1 and BRCA2 and regard all other breast cancer as being sporadic. This is a limitation of our model, since additional breast cancer genes or ovarian cancer genes may exist (Håkansson et al. 1997; Vehmanen et al. 1997). Including other genes and considering all other breast and ovarian cancer as being sporadic would give a different probability—usually smaller—that a woman carries BRCA1 and BRCA2. It would also give a different probability—larger—that a counseland carries at least one mutation in one or more breast cancer genes. Our procedure can be modified to include additional genes; however, such a modification would require information, not yet available, about the

age-specific incidence of breast cancer and ovarian cancer resulting from such genes.

Spontaneous mutations of BRCA1 and BRCA2 are not currently incorporated in our model, again because of the lack of relevant empirical evidence. In the presence of spontaneous mutations, the probability that an individual carries a mutation is higher than our model predicts for cancer cases and is lower than our model predicts for noncancer cases. Also, compared with our results, the possibility of spontaneous mutations reduces the predictive power of the family history.

Environmental or other factors may modify the susceptibility to cancer, in carriers of mutations. A model explicitly incorporating such individual covariate information is likely to perform better that ours. In family 2, discussed in this section, additional covariate information may help us to infer the genetic status of the mother, a key element in the calculation. Was she a carrier exposed to factors either delaying onset or preventing cancer from developing? Unfortunately, again, knowledge about relevant factors is, at present, insufficient for quantitative analysis of this kind of question.

Although important, all these limitations are rooted in uncertainties regarding familial breast cancer. We expect important progress in this area in the near future. Models tailored to the new information could then be developed by means of the approach outlined here.

## Acknowledgments

## Appendix

### Description of the Program Used for Calculation of the Expression $P$[fam. hist. | BRCA1,BRCA2]

The structure of a pedigree, including information on first- and second-degree relatives of a given counseland individual, is specified by the following quantities: the number of siblings, $n_s$; the number of mother's siblings, $n_a$; the number of father's siblings, $n_A$; the number of offspring, $n_d$; and the number of offspring of sibling $t$, $n(t)$, where $t = 1, \ldots, n_s$. For each family member, the contribution to the likelihood function depends on the age $a$; the genetic status of BRCA1, $i$; the genetic status of BRCA2, $j$; and the disease status. We denote by

$R_B^{ij}(a)$ and $R_O^{ij}(a)$ the cumulative incidence curves for breast cancer and ovarian cancer, respectively, and we denote by $r$ the incidence curve, or the first derivative of $R$. For member $m$, since breast cancer incidence and ovarian cancer incidence are assumed to be independent, conditional on genetic status, the conditional probability of disease history, given genetic status, $\rho_m^{ij}$, factors into two disease-specific terms, as $\rho_m^{ij} = \rho_m^{B,ij}\rho_m^{O,ij}$, where

$$\rho_m^{O,ij} = \begin{cases} r_O^{ij}(a) & \text{if } m \text{ had ovarian cancer at } a \\ 1 - R_O^{ij}(a) & \text{if } m \text{ was free of ovarian cancer until } a \end{cases}$$

and $\rho_m^{B,ij}$ is defined similarly, with the exception of the possible adjustment for bilateral breast cancer, as described in the Methods section.

Incorporation of each member's contribution is based on two basic building blocks: updating of data on offspring, given data on their parents (implemented by the $d$ functions, for downward), and updating of data on parents, given data on their offspring (implemented by the $u$ functions, for upward). The downward updates will be applied, in turn, to each of the counseland's offspring, to siblings and their offspring, and to parents' siblings. The upward update will be applied to the counseland, to her parents, and to her grandparents. The computation of likelihood function requires a recursive use of these functions, as described below.

To define the basic downward function for a generic group of offspring of the same parents, let $m_1$ and $m_2$ be the BRCA1 genetic status and the BRCA2 genetic status of the offspring's mother, and let $f_1$ and $f_2$ be the BRCA1 genetic status and the BRCA2 genetic status of the offspring's father. Then, the downward step is

$$d(m_1, m_2, f_1, f_2, n) =$$

$$\prod_{m=1}^{n} \sum_{i=0}^{2} \sum_{j=0}^{2} \rho_m^{ij} P[i|m_1, f_1] P[j|m_2, f_2] \ .$$

Here, the subscript $m$ ranges over the set of siblings of the same parents, and the argument $n$ can take values $n_d$, $n_a$, $n_A$, or $n(t)$.

The contribution of the counseland's siblings, given both the genetic status of the counseland's mother and the genetic status of the counseland's father, is then

$$d_S(m_1, m_2, f_1, f_2) =$$

$$\prod_{m=1}^{n_s} \sum_{i=0}^{2} \sum_{j=0}^{2} \rho_m^{ij} P[i|m_1, f_1] P[j|m_2, f_2]$$

$$\sum_{k=0}^{2} \sum_{l=0}^{2} \rho_{q(m)}^{ij} P[k, l] d[i, j, k, l, n(t)] \ ,$$

where $m$ ranges over the set of counseland's siblings and $q(m)$ indicates the corresponding mate. If there are no

offspring of sibling $t$, then $n(t) = 0$, $d = 1$, and $\rho_{q(m)}^{ij} = 1$. The expression above assumes that no family history is available for the siblings' mates.

The information about the counseland's parents and grandparents will be incorporated via the upward structure. For both the maternal and paternal grandparents, we use

$$u_G(o_1, o_2) =$$

$$\sum_{i=0}^{2} \sum_{j=0}^{2} \sum_{k=0}^{2} \sum_{l=0}^{2} \rho_{m_1}^{ij} \rho_{m_2}^{kl}$$

$$P[i, k | o_1] P[j, l | o_2] d(i, j, k, l, n) \ ,$$

where $m_1, m_2$ is a pair of grandparents. $P[i, k | o_1]$ and $P[j, l | o_2]$ are calculated as $P[m, f | o]$ was calculated in the Updating subsection. These functions can be used to incorporate the contribution of the counseland's parents, given the counseland's own genetic status; as above, this is described by the variables BRCA1 and BRCA2. Then

$$u_p(BRCA1, BRCA2) =$$

$$\sum_{i=0}^{2} \sum_{j=0}^{2} \sum_{k=0}^{2} \sum_{l=0}^{2} \rho_{m_1}^{ij} \rho_{m_2}^{kl}$$

$$P[i, k | BRCA1] P[j, l | BRCA2] d_S(i, j, k, l) u_g(i, k) u_g(j, l) \ ,$$

where $m_1, m_2$ are the counseland's parents. Finally, we can combine all of these expressions and incorporate the information on the counseland by means of

$$P[\text{family history} | BRCA1 = i, \ BRCA2 = j]$$

$$= \rho_m^{ij} u_p(i, j) \sum_{k=0}^{2} \sum_{l=0}^{2} \rho_{q(m)}^{ij} P[k, l] d(i, j, k, l, n_d) \ ,$$

for $i, j = 0, 1, 2$. Here, $m$ is the counseland, and $q(m)$ is the counseland's mate.

## References

Andersen TI (1996) Genetic heterogeneity in breast cancer susceptibility. Acta Oncol 35(4): 407–410

Berry DA, Parmigiani G, Sanchez J, Schildkraut JM, Winer E (1997) Assessing the probability of carrying a breast cancer gene BRCA1 based on family history. J Natl Cancer Inst 89(3): 9–20

Claus EB, Risch N, Thompson WD (1991) Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. Am J Hum Genet 48:232–242

Claus EB, Risch N, Thompson WD (1994) Autosomal dominant inheritance of early-onset breast cancer: implications for risk prediction. Cancer 73:643–651

Couch FJ, DeShano ML, Blackwood MA, Calzone K, Stopfer J, Campeau L, Ganguly A, et al (1997) BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer. N Engl J Med 336:1409–1415

Easton DF, Ford D, Bishop DT, Breast Cancer Linkage Consortium (1995) Breast and ovarian cancer incidence in BRCA1-mutation carriers. Am J Hum Genet 56:265–271

Easton DF, Steele L, Fields P, Ormiston W, Averill D, Daly PA, McManus R, et al (1997) Cancer risks in two large breast cancer families linked to BRCA2 on chromosome 13q12-13. Am J Hum Genet 61:120–128

Falconer DS, Mckay TFC (1996) Quantitative genetics. Longman, Harlow, United Kingdom

Ford D, Easton DF (1995) The genetics of breast and ovarian cancer. Br J Cancer 72:805–812

Futreal PA, Liu Q, Shattuck-Eidens D, Cochran C, Harshman K, Tavtigian S, Bennett LM, et al (1994) BRCA1 mutations in primary breast and ovarian carcinomas. Science 226: 120–122

Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ (1989) Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 81:1879–1886

Håkansson S, Johannsson O, Johansson U, Sellberg G, Loman N, Gerdes A-M, Holmberg E, et al (1997) Moderate frequency of BRCA1 and BRCA2 germ-line mutations in Scandinavian familial breast cancer. Am J Hum Genet 60: 1068–1078

Hoskins KF, Stopfer JE, Calzone KA, Merajver SD, Rebbeck TR, Garber JE, Weber BL (1995) Assessment and counseling for women with a family history of breast cancer: a guide for clinicians. JAMA 273:577–585

Houwing-Duistermaat JJ, Van Houwelingen JC (1997) A family history score for breast cancer. Genet Epidemiol 14:530

Leal SM, Ott J (1994) A likelihood approach to calculating risk support interval. Am J Hum Genet 54:913–917

Miki Y, Swensen J, Shattuck-Eidens DJ, Futreal PA, Harshman K, Tavtigian S, Liu Q (1994) A strong candidate for the breast and ovarian cancer susceptibility: gene BRCA1. Science 266(7): 66–71

Moolgavkar SH, Stevens RG, Lee JAH (1979) Effect of age on incidence of breast cancer in females. J Natl Cancer Inst 62:493–501

Murphy EA, Mutalik GS (1969) The application of Bayesian methods in genetic counseling. Hum Hered 19:126–151

Oddoux C, Struewing JP, Clayton CM, Neuhausen S, Brody LC, Kaback M, Haas B, et al (1996) The carrier frequency of the BRCA2 6174delT mutation among Ashkenazi Jewish individuals is approximately 1%. Nat Genet 14:188–190

Roa BB, Boyd AA, Volcik K, Richards CS (1996) Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. Nat Genet 14:185–187

Shattuck-Eidens D, McClure M, Simard J, Labrie F, Narod S, Couch F, Hoskins K, et al (1995) A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene: implications for presymptomatic testing and screening. JAMA 273:535–541

Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, Timmerman MM, et al (1997) The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 336: 1401–1408

Szabo CI, King M-C (1997) Population genetics of BRCA1 and BRCA2. Am J Hum Genet 60:1013–1020

Szolovits P, Pauker S (1992) Pedigree analysis for genetic coun-

seling, In: Lun KC, Degoulet P, Piemme TE, Rienhoff O (eds) MEDINFO-92: proceedings of the Seventh Conference on Medical Informatics. Elsevier, New York, pp 679-683

Thompson EA (1986) Pedigree analysis in human genetics. Johns Hopkins University Press, Baltimore

Thompson EA, Guo SW (1991) Evaluation of likelihood ratios for complex genetic models. IMA J Math Appl Med Biol 8: 149–169

Vehmanen P, Friedman LS, Eerola H, Sarantaus L, Pyrhönen S, Ponder BAJ, Muhonen T, et al (1997) A low proportion of BRCA2 mutations in Finnish breast cancer families. Am J Hum Genet 60:1050–1058

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, et al (1995) Identification of the breast cancer susceptibility gene BRCA2. Nature 378:789–792

Yancik R (1993) Ovarian cancer: age contrasts in incidence, histology, disease stage at diagnosis, and mortality. Cancer 71 Suppl 2:517–523