Modeling Familial Clustered Breast Cancer Using Published Data

M. A. Jonker,[1] C. E. Jacobi,[2] W. E. Hoogendoorn,[2] N. J. D. Nagelkerke,[1] Geertruida H. de Bock,[2] and Johannes C. van Houwelingen

[1]Departments of Medical Statistics, and [2]Medical Decision Making, Leiden University Medical Center, Leiden, the Netherlands

Abstract

The purpose of this research was to model the familial clustering of breast cancer and to provide an accurate risk estimate for individuals from the general population, based on their family history of breast and ovarian cancer.

We constructed a genetic model as an extension of a model by Claus *et al.* (E. B. Claus *et al.*, Am. J. Hum. Genet., *48:* 232–242, 1991), with three breast cancer genes, *BRCA1, BRCA2,* and a hypothetical *BRCAu,* in two variants, one in which *BRCAu* was dominant and one in which *BRCAu* was recessive. The model parameters were estimated using published estimates of population incidence and relative risks. Risk estimation was performed for a set of 196 counselees and for a set of simulated counselees with both the dominant *BRCAu* and the recessive *BRCAu* model, and compared relating to medical management.

Estimates of the model parameters were found. Relative risks among family members were comparable between the model of Claus *et al.* (E. B. Claus *et al.*, Am. J. Hum. Genet., *48:* 232–242, 1991) and our model. The dominant and the recessive model provided approximately similar lifetime risks for breast cancer.

Our model is suitable for breast cancer risk estimation in a health care setting.

Introduction

It has been demonstrated that a family history of breast cancer is a risk factor for breast cancer (1–3). This familial risk for breast cancer has first been modeled by Claus *et al.* (4) as a single autosomal dominant susceptibility gene.

After publication of this model, two highly penetrant breast cancer susceptibility genes, *BRCA1* and *BRCA2,* were identified (5–10). These genes were modeled by Parmigiani *et al.* (11) and Berry *et al.* (12), with the primary aim to estimate *BRCA1* and *BRCA2* carrier probabilities in counselees based on their family history of breast and ovarian cancer.

However, these two breast cancer susceptibility genes only explain a small proportion of the familial clustering of breast cancer (13). Because shared environmental risk factors presumably account for <10% of familial clustering (14), additional as yet unidentified genes are probably involved in breast cancer susceptibility. Antoniou *et al.* (15, 16) and Cui *et al.* (17) aimed to estimate the prevalence and breast cancer risk of *BRCA1* and *BRCA2,* and to model the unknown breast cancer susceptibility genes by using data from a population-based series of breast cancer families (15, 17) and from high-risk families (16).

Here, we present a model of breast cancer risk among women from the general population with a family history of breast cancer for the use of risk estimation in health care settings. It includes both the known *BRCA1* and *BRCA2* genes and an additional hypothetical gene *BRCAu,* and so extending the familial clustering of breast cancer as modeled by Claus *et al.* (4). For the quantification of model parameters we used published data, both from a meta-analysis on breast cancer risks and from population-based studies on the penetrance and mutated-allele frequency of *BRCA1* and *BRCA2*.

Materials and Methods

Design. First, a genetic model with three genes, *BRCA1, BRCA2,* and a hypothetical third gene, was constructed. To estimate model parameters, published data on breast cancer risks and incidence were used (1, 7–10, 18). Second, we explored the use of this model for individual risk estimation as provided in familial cancer clinics using both simulated pedigrees and real pedigrees.

The Hypothetical Gene *BRCAu*. In addition to the known breast cancer susceptibility genes, *BRCA1* and *BRCA2,* a third gene was hypothesized: *BRCAu,* where "u" denotes unknown. Previous research has shown that from a biological view the familial clustering not explained by the *BRCA1* and *BRCA2* genes is most likely polygenic (15–17). However, as our objective was risk estimation in health care settings, the biological correctness was not of primary importance.

The third gene was modeled to explain all of the familial clustering of breast cancer unaccounted for by the *BRCA1* and *BRCA2* genes. The total familial clustering as modeled by Claus *et al.* (4) was taken as the reference, but had to be extended with the information available on *BRCA1* and *BRCA2,* on bilateral breast cancer and on ovarian cancer, and with the population breast cancer incidence.

We constructed two variants of the genetic model, as two extremes. In the first variant *BRCAu* was assumed to be dominant (model I) and in the second *BRCAu* was assumed to be recessive (model II).

Parameters in the Genetic Model. The following model parameters were used: the mutated-allele frequencies of *BRCA1, BRCA2,* and *BRCAu,* the genetic susceptibility for breast cancer based on these genes, and breast cancer penetrance. These

Received 5/21/03; revised 8/14/03; accepted 8/20/03.
Grant Support: National Health Insurance Board of the Netherlands (College voor Zorgverzekeringen).
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Requests for reprints: Dr. Geertruida H. de Bock, Department of Medical, Decision Making, Leiden University Medical Center, P. O. Box 9600, 2300 RC, Leiden, the Netherlands. Phone: 31-71-526-4904; Fax: 31-71-526-6838; E-mail: G.H.de_Bock@lumc.nl.

parameters were estimated using relative risks for breast cancer among women with family relatives with breast cancer.

Mutated-Allele Frequencies. The mutated-allele frequencies in the population are denoted by q_1, q_2, and q_u for *BRCA1*, *BRCA2*, and *BRCAu*, respectively. Individuals inherit two *BRCA1*, *BRCA2*, and *BRCAu* alleles, one from each parent. For any randomly chosen individual, the probability that an individual has inherited zero, one, or two mutated copies on, for example, the *BRCAu* gene equal $(1-q_u)^2$, $2q_u(1-q_u)$, and q_u^2, respectively. Similar expressions hold for *BRCA1* and *BRCA2*.

The mutated allele frequencies q_1 and q_2 were both estimated at 0.0006 [by Ford *et al.* (7) for *BRCA1* and by Peto *et al.* (8) for *BRCA2*] The parameter q_u is unknown and has to be estimated, denoted by $\hat{q}_{u,I}$ for model I and by $\hat{q}_{u,II}$ for model II.

Genetic Susceptibility. The probability that a woman is genetically susceptible due to the *BRCA* genes is denoted by r. As mutations on the *BRCA1* and *BRCA2* genes were found to be dominant, a woman has to carry only one of these to be genetically susceptible. In model I (mutated allele on *BRCAu* is dominant), r_u also equals the probability that a woman carries at least one mutated allele, but in model II, in which the mutated allele on *BRCAu* is recessive, r_u equals the probability of carrying two copies.

Let the joint probabilities that a woman and her mother, her sister, her daughter, or any first degree relative both are genetically susceptible by their allelic combination on a *BRCA* gene be denoted by $r_{i,M}$, $r_{i,S}$, $r_{i,D}$, and $r_{i,FD}$, respectively, for $i = 1,2,u$. Note that $r_{i,M} = r_{i,D}$. The probability that a woman in the general population is not genetically susceptible to breast cancer by any of these genes is indicated by p.

As the mutations of *BRCA1* and *BRCA2* are highly penetrant, we assume that carrying a mutation on *BRCAu* in addition to a mutation on *BRCA1* or *BRCA2* has no additional effect on breast cancer susceptibility. Thus, $p \approx 1-r_1-r_2-r_u$.

In addition, for dominant mutations $r_i \approx 2q_i$ and $r_{i,D} = r_{i,M} \approx r_{i,S} \approx q_i$. Thus, $r_{i,FD} \approx q_i$. This is true for *BRCA1* and *BRCA2*, and also for *BRCAu* in model I. In model II the mutated allele on *BRCAu* is assumed to be recessive, and then $r_u = q_u^2$, $r_{u,D} = r_{u,M} = q_u^3$ and $r_{u,S} \approx 0.25q_u^2 + 0.5q_u^3$.

On the basis of the values for q_1 and q_2 of 0.0006, r_1 and r_2 equal 0.0012 and $r_{i,D} = r_{i,M} \approx r_{i,S} \approx 0.0006$ for $i = 1,2$. The parameters r_u and p have to be estimated, denoted by $\hat{r}_{u,I}$ and $\hat{p}_I$ for model I, and $\hat{r}_{u,II}$ and $\hat{p}_{II}$ for model II.

Breast Cancer Penetrance. For women who are genetically susceptible by their *BRCA*1, *BRCA2* or *BRCAu* gene, the penetrance for breast cancer is denoted by F_1, F_2, and F_u, respectively, and the penetrance for breast cancer for women who are not genetically susceptible on any of these genes by F_0. The population penetrance is denoted by F, *i.e.*, the risk for women from the general population at age x is $F(x)$.

The conditional risk for women at age x with a mother who is affected with breast cancer before or at age y is denoted by $F_M(x \mid y)$. For the risk of developing breast cancer between ages a and b, we use the notation $F([a,b])$, with $F([a,b]) = F(b) - F(a)$.

Using the parameters as stated above, the penetrance of breast cancer in the general population at age x equals:

$$F(x) \approx pF_0(x) + r_1F_1(x) + r_2F_2(x) + r_uF_u(x), \qquad \text{(A)}$$

and between ages a and b:

$$F([a,b]) \approx pF_0([a,b]) + r_1F_1([a,b]) + r_2F_2([a,b]) + r_uF_u([a,b]). \qquad \text{(B)}$$

An estimate of F was computed based on the age-dependent breast cancer incidence as given by the Netherlands Cancer Registry (18). Estimates of the penetrance for women carrying *BRCA1* and *BRCA2* mutations, F_1 and F_2 at the ages 40, 50, 60, and 70 were taken from Easton *et al.* (10) and Ford *et al.* (9). The parameters F_u and F_0 have to be estimated, denoted by $\hat{F}_{u,I}$ and $\hat{F}_{0,I}$ for model I, and $\hat{F}_{u,II}$ and $\hat{F}_{0,II}$ for model II.

Relative Risks for Breast Cancer. Relative risks for breast cancer are used to measure genetic effects, and these are denoted generically by λ. The relative risk for breast cancer for a woman at age x with a mother with breast cancer before or at age y is the ratio of the penetrance for such women and the penetrance for women from the general population: $\lambda_M(x \mid y) = F_M(x \mid y)/F(x)$. Analogously, λ_S, λ_D, and λ_{FD} are defined as the relative risks for women with a sister, a daughter, or any first degree relative with breast cancer, respectively.

If no breast cancer susceptibility genes existed, the relative risk $\lambda_M(x \mid y)$ would equal 1 for all values of x and y. The difference $\lambda_M - 1$ thus represents the genetic association between a woman and her mother at certain ages of breast cancer onset. This association is shown in Equation C in which the relative risk λ_M is written in terms of the genetic model (see Appendix for the derivation)

$$\lambda_M(x|y) - 1 \approx \sum_{i=1,2,u} r_{i,M}\left(\frac{F_i(x)}{F(x)} - 1\right)\left(\frac{F_i(y)}{F(y)} - 1\right). \qquad \text{(C)}$$

Similarly, the relative risk of a woman with a sister with breast cancer, λ_S, is:

$$\lambda_S(x|y) - 1 \approx \sum_{i=1,2,u} r_{i,S}\left(\frac{F_i(x)}{F(x)} - 1\right)\left(\frac{F_i(y)}{F(y)} - 1\right). \qquad \text{(D)}$$

A similar relationship holds for λ_D. In model I, $r_{i,D} = r_{i,M} \approx r_{i,S} \approx q_i$ for $i = 1,2,u$ and $\lambda_D(x \mid y) \approx \lambda_M(x \mid y) \approx \lambda_S(x \mid y)$ for all values of x and y. Consequently, in this model the first degree relative does not have to be specified and the following equation, with $r_{i,FD} \approx q_i$ for $i = 1,2,u$, holds

$$\lambda_{FD}(x|y) - 1 \approx \sum_{i=1,2,u} r_{i,FD}\left(\frac{F_i(x)}{F(x)} - 1\right)\left(\frac{F_i(y)}{F(y)} - 1\right). \qquad \text{(E)}$$

The risk of being affected with breast cancer between two ages is

$$\lambda_{FD}([a_0,b_0]|[a_1,b_1]) - 1 \approx \sum_{i=1,2,u} r_{i,FD}\left(\frac{F_i([a_0,b_0])}{F([a_0,b_0])} - 1\right) \cdot \left(\frac{F_i([a_1,b_1])}{F([a_1,b_1])} - 1\right). \qquad \text{(F)}$$

In model II the mutation on the *BRCAu* gene was assumed to be recessive, and different equations apply. For $q_u \leq 0.5$, from Equations C) and D, $\lambda_S(x \mid y) \geq \lambda_M(x \mid y)$ for all values of x and y. Consequently, $\lambda_S(x \mid y) \gtrsim \lambda_{FD}(x \mid y) \gtrsim \lambda_M(x \mid y)$ for all values of x and y. From Equations C and D and the inequalities above, it follows that

$$\sum_{i=1,2,u} r_{i,M}\left(\frac{F_i(x)}{F(x)} - 1\right)^2 \lesssim \lambda_{FD}(x|x) - 1 \lesssim \sum_{i=1,2,u} r_{i,S}\left(\frac{F_i(x)}{F(x)} - 1\right)^2, \qquad \text{(G)}$$

and similarly for $\lambda_{FD}([a,b] \mid [a,b])$.

Estimates of the relative risks were taken from a reanalysis

of combined data from 52 epidemiological studies (1), which provided a cross-classified table with estimates of $\lambda_{FD}([x_1,x_2]\,|\,[y_1,y_2])$, for $[x_1,x_2]$ and $[y_1,y_2]$ varying from (0, 40, 40, 50, 50, 60) and [60,∞). In our analysis the interval [60,∞) was interpreted as [60,70]. It also presented estimates of $\lambda_M(x\,|\,y)$ and $\lambda_S(x\,|\,y)$ (1) (for $x = 50$ and $y = 50$, separately; for other values of x and y estimates of λ_M and λ_S hardly differed, and published estimates of λ_{FD}).

The parameters of model I have to be estimated using Equations A and E. Equation E, however, can be solved for many different values of $r_{u,I}$ and $F_{u,I}$. To find a unique solution, the number of unknown parameters in this equation has to be reduced. Therefore, we chose a value of 0.03 for $\hat{q}_{u,I}$, based on Antoniou *et al.* (15).

Risk Estimation. To compare the outcomes of model I and II, an actual risk estimation was performed for a set of 196 true counselees and for a set of simulated counselees.

The risk for breast cancer of a woman was estimated using the history of all family members of unilateral, bilateral, and male breast cancer and ovarian cancer, using the BRCAPRO calculations according to Parmigiani *et al.* (11), incorporating both affected and nonaffected relatives.

Comparison of the Models Using Real Pedigrees. For 196 women who attended the department of clinical genetics at Leiden University Medical Center for genetic counseling, the family history was elicited by questionnaire and interview. All of the first, second, and third degree relatives on the paternal and maternal side were taken into consideration. An extensive description of the data are given by van Asperen *et al.*[3]. For all 196 of the women the lifetime risk of breast cancer (up to the age of 75) was computed with both model I and model II.

Comparison of the Models Using Simulated Pedigrees. Because all of the pedigrees in the population are different, it was impossible to perform simulations for all of the possible pedigrees separately. Therefore, we used one large pedigree and applied different patterns of familial clustering. The counselee in this pedigree is a woman of 35 who has one sister aged 38. She has two paternal aunts, one paternal uncle, two maternal aunts, and one maternal uncle. Her parents, and aunts and uncles on both sides are between 58 and 69 years of age. The paternal grandmother died when she was 82, and the maternal grandmother when she was 70 years of age. To check the sensitivity of the results found with respect to the pedigree that was used, we also simulated other pedigrees. Outcomes were similar.

Genotypes in the pedigree were generated 5 million times with model I and 5 million times with model II. All of the members of the pedigree inherited genotypes according to Mendelian rules. Given the genotype of an individual it was simulated whether the individual developed (bilateral) breast and/or ovarian cancer, and if so at what age, resulting in a phenotypic family history.

In general, only women with a family history of breast or ovarian cancer attend familial cancer clinics for risk estimation. Therefore, comparisons were restricted to simulated pedigrees in which the counselee was without breast symptoms and met one of the following criteria: (*a*) the counselee had at least one first-degree relative with breast cancer under the age of 50 or at least two second-degree relatives with breast cancer under the age of 50, either in the paternal or in the maternal family or on both sides; or (*b*) both ovarian and breast cancer occurred in one side or in both sides of the counselee's family.

For all of the simulated pedigrees that met one of these criteria, lifetime risk for breast cancer up to age 75 was estimated with both model I and model II, using the phenotypic family history of breast and/or ovarian cancer. The estimated risks were categorized to evaluate whether the additional diagnostics that would be offered to the counselee would depend on the model (I or II) that was used in the risk estimation. The category thresholds were based on common guidelines for additional management of counselees: 0.2 and 0.3 (19). A lifetime risk of breast cancer of 0.3 is often taken as a threshold for additional genetic assessment, such as DNA testing. A lifetime risk of 0.2 is often taken as a threshold for yearly surveillance, and a lifetime risk <0.2 is used as an indication that special interventions will not be cost-effective.

Table 1 Estimates of the penetrance for BRCA1 ($\hat{F}_1$; 10), BRCA2 ($\hat{F}_2$; 9), and for BRCAu, and noncarriers in Model I ($\hat{F}_{u,I}$ and $\hat{F}_{0,I}$) and in Model II ($\hat{F}_{u,II}$ and $\hat{F}_{0,II}$)

Age	BRCA1	BRCA2	BRCAu, Model I		BRCAu, Model II	
	F_1	F_2	$\hat{F}_{u,I}$	$\hat{F}_{0,I}$	$\hat{F}_{u,II}$	$\hat{F}_{0,II}$
40	0.19	0.12	0.082	0.0016	0.14	0.0011
50	0.51	0.28	0.17	0.012	0.25	0.0106
60	0.54	0.48	0.28	0.027	0.42	0.025
70	0.85	0.84	0.38	0.048	0.57	0.048
μ	53.9	58.5	56.3	66.3	67.7	72.0
σ	16.5	13.8	17.2	14.9	25.7	16.5
c	0.96	1.00	0.48	0.08	1.00	0.10

The formula for the penetrance function for each of these genes is given by the product of c and Φ (age $-$ μ)/σ, where Φ denotes the normal cumulative distribution.

Results

Estimation of Remaining Parameters in Model I. For model I the parameters $\hat{p}_{u,I}$, $\hat{q}_{u,I}$, $\hat{r}_{u,I}$, $\hat{F}_{u,I}$, and $\hat{F}_{0,I}$ had to be estimated. Equations A and E were available for the estimation of the parameters in model I, because $r_S \approx r_M \approx r_{FD}$. As thus three parameters were unknown, a unique solution of the two equations does not exist. However, for any given $\hat{q}_{u,I}$, $\hat{p}_{u,I}$ and $\hat{r}_{u,I}$ are fixed, and we, therefore, fixed $\hat{q}_{u,I}$ at 0.03.

Then estimates for $\hat{F}_{u,I}(40)$ and $\hat{F}_{0,I}(40)$ were found using Equations A and E (Table 1). Subsequently, we estimated $\hat{F}_{u,I}([40, 50])$ and $\hat{F}_{0,I}([40, 50])$ using Equations B and F. Because $F_u([a,b])+F_u(a) = F_u(b)-F_u(a)+F_u(a) = F_u(b)$ for all values of a and b with $a \leq b$, $F_u(50)$ and $F_0(50)$ were estimated by, respectively, $\hat{F}_{u,I}([40, 50])+\hat{F}_{u,I}(40)$ and $\hat{F}_{0,I}([40, 50])+\hat{F}_{0,I}(40)$, (Table 1). Analogous to these, estimates $\hat{F}_{u,I}(60)$,$\hat{F}_{u,I}(70)$, $\hat{F}_{0,I}(60)$, and $\hat{F}_{0,I}(70)$ were found.

Estimation of Remaining Parameters in Model II. As $\hat{p}_{u,II}$, $\hat{q}_{u,II}$, $\hat{r}_{u,II}$, $\hat{F}_{u,II}$, and $\hat{F}_{0,II}$ were unknown, the parameters $\hat{F}_{u,II}$, $\hat{F}_{0,II}$, and $\hat{q}_{u,II}$ had to be estimated. For this model: $p \approx 1-2q_1-2q_2-q_u^2$. Using the Equations A, C, and D with $x = 50$ and $y = 50$, we found $\hat{q}_{u,II} = 0.193$, and $\hat{F}_{u,II}(50) = 0.247$, $\hat{F}_{0,II}(50) = 0.0106$. The probability that a woman is genetically susceptible to breast cancer by the genotype at BRCAu in model II is $\hat{r}_{u,II} = \hat{q}_{u,II}^2 = 0.037$.

The estimates of $\lambda_{FD}(40|40)_{FD}([50,60]|[50,60])$ and $\lambda_{FD}([60,70]|[60,70])$ from the literature were used to estimate intervals (lower bound, upper bound) for $\hat{F}_{u,II}(x)$ and $\hat{F}_{0,II}(x)$ for

[3] van Asperen, C. J., Jonker, M. A., Jacobi, C. E., van Diemen-Homan, J. E. M., Bakker, E., Breuning, M. H., van Houwelingen, J. C., and de Bock, G. H. Risk estimation for healthy women from breast cancer families: new insights and new strategies. Cancer Epidemiol. Biomark. Prev., in press, 2004.

Table 2 The estimated genetic association due to BRCA1, BRCA2, and BRCAu, in Model I, related to the estimates of the relative risk $\lambda_{FD}(x|x)$ and the Claus-model

Age (x)	Genetic association due to			Relative risks	
	BRCA1	BRCA2	BRCAu	$\lambda_{FD}(x\|x)$	$\lambda_{Claus}(x\|x)$
40	0.48	0.18	4.04	5.70	5.36
50	0.30	0.084	1.41	2.79	3.04
60	0.084	0.064	0.95	2.10	1.97
70	0.079	0.077	0.62	1.78	1.51

$x = 40$, $x = 60$, and $x = 70$, using Equation G, namely [0.117;0.155],[0.396;0.440], and [0.526;0.606]. The midpoints of the intervals were taken as estimates for $\hat{F}_{u,II}$ (Table 1). Using Equation A, $\hat{F}_{0,II}$ was found (Table 1).

Genetic Association for Developing Breast Cancer. The *BRCA1*, *BRCA2*, and *BRCAu* genes together explained all familial clustering of breast cancer as observed in the general population and, thus, represent the genetic associations for developing breast cancer (Table 2, presented for model I). The total contribution of *BRCAu* in the genetic association is independent of the choice of $\hat{q}_{u,I}$. At each age, the genetic association of *BRCA1*, *BRCA2*, and *BRCAu* adds up to λ_{FD}-1. For comparison, the relative risks according to the model of Claus *et al.* (4) are presented as well (Table 2). The results show that the estimated genetic association due to *BRCA1* and *BRCA2* is much lower than of *BRCAu*, indicating that *BRCA1* and *BRCA2* only explain a small proportion of the total genetic association for developing breast cancer. The findings of Claus *et al.* (4) differ only slightly.

Risk Estimation. Table 1 shows penetrance functions for carriers of one of the *BRCA* genes and for noncarriers. These functions summarize the information needed for risk estimation.

Comparison of the Models for Real Pedigrees. The relative risks for breast cancer up to age 75, as computed for 196 women, are presented in Fig. 1. Estimation of risks was based on the family history of breast and/or ovarian cancer among the first, second, and third degree relatives, and were computed with both model I and model II.

For breast cancer risks <0.25 the two models largely agreed. For risks of ≥ 0.25 the differences between the outcomes of the two models increased slightly with lifetime risk. This finding is likely caused by the fact that in model I the penetrance for *BRCAu* at age 75 is lower than the penetrance for *BRCAu* at age 75 in model II. If both ovarian cancer and breast cancer had been diagnosed in the family, the two models provided similar lifetime risks.

Comparison of the Models for Simulated Data. In the first simulation, in which model I was used to generate genotypes of pedigrees, 4.76% of the 5 million simulated pedigrees satisfied the inclusion criteria. On the basis of selected pedigrees, the lifetime risks for breast cancer up to age 75 were computed based on the phenotypic family history for each counselee with both model I and model II. The categorized risks are presented in Table 3. Each cell provides the percentages of counselees who were selected for the corresponding category.

For 89.1% of the selected pedigrees, the lifetime risks as computed by model I and model II agreed. Thus, the additional diagnostics were independent of the genetic model that was used for risk estimation. Risk estimates from the two models differed by at most one risk category. Almost 9% of the counselees were located below the diagonal in Table 3. For these counselees the lifetime risk estimated by model II was lower than by model I. In 2% of the counselees, as categorized above the diagonal, model II estimated higher lifetime risks than model I.

Similar results were obtained in the simulation, in which model II generated genotypes of pedigrees. Of the 5 million simulated pedigrees, 4.62% satisfied the inclusion criteria. For 89.3% of the pedigrees model I and model II agreed regarding lifetime risks, for 8.8% model I estimated higher lifetime risks than model II, and for 1.9% of the counselees model II estimated higher lifetime risks than model I.

We found that where model I estimated higher lifetime risks than model II, women in either the paternal or the maternal family, but not in both, were affected with breast cancer. If the estimated lifetime risks from model II were higher than from model I, both the paternal and maternal family of the counselee were affected with breast cancer. When both ovarian cancer and breast cancer were present in the family, the estimated risks in both models agreed. Such a family history indicates that probably *BRCA1* or *BRCA2* mutated alleles are involved, and the parameters for these genes are equal in the two models.

The results for other pedigrees and for a different choice of the mutated allele-frequency $\hat{q}_{u,I}$ were similar (data not shown). We have assessed $\hat{q}_{u,I}$ in a range of 0.0006–0.03. The lowest frequency (*i.e.,* 0.0006) was based on the mutated-allele frequencies as found for *BRCA1* and *BRCA2*. The highest mutated-allele frequency (*i.e.,* 0.03) was based on the results of the analysis of Antoniou *et al.* (15). Any choice of $\hat{q}_{u,I}$ results in similar associations between first-degree relatives. For a small minority of the families, a different choice of $\hat{q}_{u,I}$ resulted in a slightly different risk estimate, caused by multiple breast cancer cases among first and second degree relatives, *i.e.,* high risk families. These differences in risk estimates depending on the choice of $\hat{q}_{u,I}$ did not result in different clinical management.

Discussion

In this article we presented two genetic models for individual breast cancer risk estimation. These two models contain either a dominant or a recessive third hypothetical gene, *BRCAu,* in addition to *BRCA1* and *BRCA2*. To estimate model parameters for this third gene we used published estimates of the population penetrance of breast cancer and of relative breast cancer risks for women with an affected first degree relative. The parameters for the hypothetical third gene were estimated such that the three genes jointly explain the relative risks and population penetrance.

The method described in this article has several advantages over likelihood methods applied to data from an epidemiological study (as *e.g.,* Ref. 15). First, as estimates of the relative risks were taken from external sources, accurate estimates based on a meta-analysis could be used. Second, in view of the simplicity of the estimation method, the method can easily be adjusted to different population penetrances. This may be important to adapt the genetic models to country-specific population risks. And third, the estimation procedure can easily be repeated when new, more accurate estimates of the mutated-allele frequency, and the penetrance for *BRCA1* and *BRCA2* become available. Moreover, if new breast cancer predisposition genes are identified, but all of the genetic association cannot yet be explained, the method described here can be use again to extend the genetic model, which then explains all of the genetic association using the new information.

To quantify our model, we used the Dutch overall breast cancer incidence and penetrance, combined with gene-specific

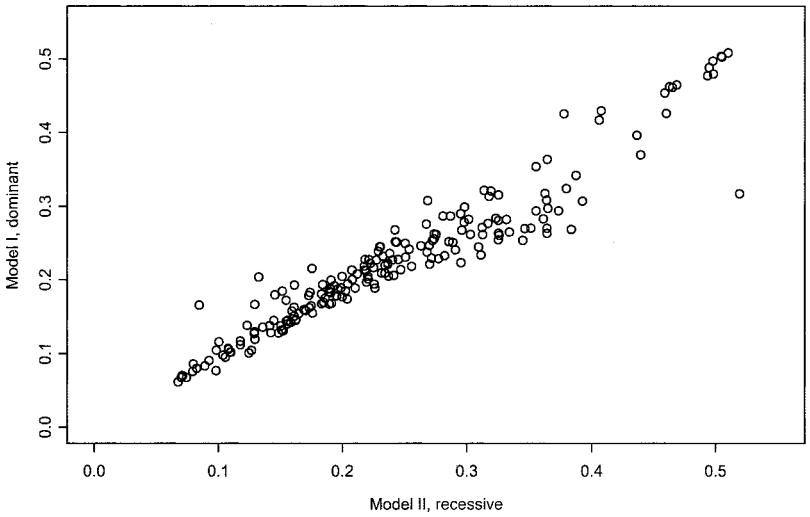

Fig. 1.

penetrances of *BRCA1* and *BRCA2* derived from other European countries, as no Dutch data are available. We think that these penetrances do not differ much between these countries, and, therefore, we expect no problems in the general use of our model. Moreover, the model can be easily fit on the overall breast cancer penetrance of other countries.

For the penetrance functions of *BRCA1* and *BRCA2*, the estimates of Easton *et al.* (10) and Ford *et al.* (9) were used. Because these estimates were derived from studies of multiple case families, they may overestimate the risks of *BRCA1* and *BRCA2* carriers in other families, as suggested by Antoniou *et al.* (20). A lower penetrance function for *BRCA1* or *BRCA2* carriers in our model would result in either a higher penetrance function for BRCAu carriers or a higher BRCAu mutated allele frequency. Biased estimates of the model parameters for *BRCA1* and *BRCA2* would, thus, be corrected by the estimates of the parameters for BRCAu, thus ensuring that the model still provides adequate risk estimates.

Available data strongly indicate that *BRCA1* and *BRCA2* mutations tend to account for most families with large numbers of cases, whereas the proportion of overall familial aggregation due to *BRCA1* and *BRCA2* is small. This indicates that if the additional gene is assumed to be dominant this gene must be quite common with a lower penetrance function than the penetrance functions for *BRCA1* and *BRCA2*. Accordingly, we chose in this study a value of 0.03 for $q_{u,I}$ to enable us to estimate the parameters of model I. This value was based on a previous study by Antoniou *et al.* (15) who considered, similar to our model I, a model with, in addition to *BRCA1* and *BRCA2*, a dominant gene named *BRCA3*. They assumed that the penetrance functions for woman *BRCA1* and *BRCA2* carriers were equal to the estimates in Easton *et al.* (10) and Ford *et al.* (9), as we did, and estimated the model parameters with a maximum likelihood method. Although different values for $q_{u,I}$ result in different estimated genetic models, all of the estimated genetic models correctly predict the genetic association of breast cancer between two first-degree relatives and also correctly estimate the penetrance for breast cancer in the total female population.

Antoniou *et al.* (15) also considered a model with a recessive gene *BRCA3* in addition to *BRCA1* and *BRCA2*, comparable with model II in this article. They estimated the remaining model parameters with a maximum likelihood method, and found a mutated-allele frequency for *BRCAu* of 0.24 and a corresponding penetrance at the age of 70 of 0.42. Our estimates are 0.193 for the mutated-allele frequency and 0.57 for the penetrance at the age of 70.

To compare the outcomes of our two models in individual risk assessment, two comparisons were performed. In the first comparison breast cancer risks using both models were computed for a set of 196 women and their families who attended our hospital for genetic counseling. In the second comparison the pedigrees were simulated. In both comparisons both model I and model II yielded similar breast cancer risks for most of the counselees, and consequently similar medical advice. The differences in breast cancer risks between the two models (see Fig. 1 and Table 3) are mainly caused by the presence of breast cancer in second degree relatives or by multiple cases among first degree relatives. It is unknown which of the two models better explains familial clustering of genetic breast cancer in larger families (first, second, and third degree relatives). The results of Cui *et al.* (17) suggest that there is a substantial recessively inherited risk of early onset breast cancer. In Antoniou *et al.* (15) the best fitting single gene model for the third hypothetical gene was a recessive model (like model II). A dominant model for the third gene (like model I) gave a somewhat poorer fit, but the difference was not statistically significant. Fortunately, both our models give similar risk estimates for women from the general population, and, thus, to similar medical decisions, as was shown in Table 3. For people with a high breast cancer risk, model II provided slightly higher risks than model I, but this would seldom affect the medical decisions. For people with lower risks, model I estimated in some cases slightly higher risks than model II. If model I were used

Table 3 Correspondence of lifetime risks of counselees according to Model I and Model II, whereby the genotypes of the family members in the pedigrees were generated by Model I

Model I ↓ |Model II→	LTR[a] ≤ 0.2	0.2 < LTR ≤ 0.3	LTR > 0.3	Total
LTR ≤ 0.2	77.9	0.7	0.0	78.6
0.2 < LTR ≤ 0.3	8.7	8.6	1.3	18.6
LTR > 0.3	0.0	0.1	2.6	2.7
Total	86.6	9.4	3.9	100.0

[a] LTR, lifetime risk.

for risk estimation, slightly more women would be offered annual surveillance.

Our model requires a computer program for risk estimation, which may not be practical in clinical practice, as pedigrees have to be entered into a computer. The so-called Claus Tables (21) are still widely used in clinical practice, as they can be used for hand-written pedigrees. On the basis of our model, a new method for practical risk estimation has been developed.[3] This method still uses the Claus Tables, but also incorporates ovarian cancer, bilateral breast cancer, and other situations that cannot be handled appropriately by the Claus Tables. Therefore, the new method should provide more accurate risks than those based on the Claus Tables alone.

Acknowledgments

We thank Lodewijk Sandkuijl for valuable discussions and helpful comments. We regret his sudden and early death (22). Furthermore, we thank Thea Vliet Vlieland for her advice.

Appendix: Derivation of Equation C.

The probability of a woman and her mother both being genetically susceptible by their allelic combinations on their *BRCA1* gene is denoted by $r_{1,M}$. Consequently, the probability that a woman is genetically susceptible on *BRCA1*, whereas her mother is not, equals $r_1 - r_{1,M}$, with r_1 as the probability that a woman is genetically susceptible on her *BRCA1* gene. *Vice versa* the same probability is found. Combining these results yields that the probability that neither a woman nor her mother are genetically susceptible on their *BRCA1* gene is equal to $1 - 2r_1 + r_{1,M}$. Similar results hold for *BRCA2* and *BRCAu*. If a woman and her mother are both genetically susceptible to breast cancer, this is because of their allelic combination on the *BRCA1*, the *BRCA2*, or on the *BRCAu* gene. So the probability that a woman and her mother are both genetically susceptible to breast cancer and are affected before the ages x and y respectively, is approximately equal to $r_{1,M}F_1(x)F_1(y) + r_{2,M}F_2(x)F_2(y) + r_{u,M}F_u(x)F_u(y)$. A similar equation can be derived for the situation that either the woman or her mother is genetically susceptible, and both women are affected before the ages x and y. In the case that both women are noncarriers the expression is approximately equal to $(1 - \Sigma_{i=1,2,u}(2r_i - r_{i,M}))F_0(x)F_0(y)$. The simultaneous population penetrance for a woman and her mother, denoted by F_M, can be written as the sum of the probabilities for the four situations just mentioned.

$$F_M(x, y) \approx \sum_{i=1,2,u} r_{i,M}F_i(x)F_i(y) + \sum_{i=1,2,u} (r_i - r_{i,M})F_i(x)F_0(y) + \sum_{i=1,2,u} (r_i - r_{i,M})F_0(x)F_i(y) + \left(1 - \sum_{i=1,2,u} (2r_i - r_{i,M})\right)F_0(x)F_0(y). \quad \text{(A1)}$$

As we subtract the term $F(x)F(y)$ from both sides of the equation and we rewrite F on the right side by the linear combination as given in Equation A we will find

$$F_M(x, y) - F(x)F(y) \approx \sum_{i=1,2,u} r_{i,M}(F_i(x) - F_0(x))(F_i(y) - F_0(y)). \quad \text{(A2)}$$

A negligible small error is made if F_0, the penetrance for noncarriers, in the preceding equation is replaced by the population incidence curve of F. Equation C is obtained by dividing the left and right hand side of the above equation by $F(x)F(y)$.

References

1. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without disease. Lancet, *358:* 1389–1399, 2001.

2. Pharoah, P. D., Lipscombe, J. M., Redman, K. L., Day, N. E., Easton, D. F., and Ponder, B. A. Familial predisposition to breast cancer in a British population: implications for prevention. Eur. J. Cancer, *36:* 773–779, 2000.

3. Pharoah, P. D., Day, N. E., Duffy, S., Easton, D. F., and Ponder, B. A. Family history and the risk of breast cancer: a systematic review and meta-analysis. Int. J. Cancer, *71:* 800–809, 1997.

4. Claus, E. B., Risch, N., and Thompson, W. D. Genetic Analysis of breast cancer in the cancer and steroid hormone study. Am. J. Hum. Genet., *48:* 232–242, 1991.

5. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., Mc Clure, M., Frye, C., Hattier, T., Phelps, R., Haugen-strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P. K., Norris, F. H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J. C., Lewis, C., Neuhausen, S., Cannonalbright, L., Goldgar, D., Wiseman, R., Kamb, A., and Skolnick, M. H. A. strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. Science (Wash. DC), *266:* 66–71, 1994.

6. Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., and Micklem, G. Identification of the breast cancer susceptibility gene *BRCA2*. Nature (Lond.), *378:* 789–791, 1995.

7. Ford, D., Easton, D. F., and Peto, J. Estimation of the frequency of *BRCA1* and its contribution to breast and ovarian cancer incidence. Am. J. Hum. Genet., *57:* 1457–1462, 1995.

8. Peto, J., Collins, N., Barfoot, R., Seal, S., Warren, W., Rahman, N., Easton, D. F., Evans, C., Deacon, J., and Stratton, M. R. Prevalence of *BRCA1* and *BRCA2* gene mutations in patients with early onset breast cancer. J. Natl. Cancer Inst., *91:* 943–949, 1999.

9. Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D. T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M. D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebeck, T. R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B. A., Cayther, S. A., and Zelada-Hedman, M. Genetic heterogeneity and penetrancy analysis of the *BRCA1* and the *BRCA2* genes in breast cancer families. Am. J. Hum. Genet., *62:* 676–689, 1998.

10. Easton, D. F., Ford, D., Bishop, D. T., and the Breast Cancer Linkage Consortium. Breast and ovarian cancer incidence in *BRCA1*-mutation carriers. Am. J. Hum. Genet., *56:* 265–271, 1995.

11. Parmigiani, G., Berry, D. A., and Aguilar, O. Determining carrier probabilities for breast cancer susceptibility genes *BRCA1* and *BRCA2*. Am. J. Hum. Genet., *62:* 145–158, 1998.

12. Berry, D. A., Iversen, E. S., Jr., Gudbjartsson, D. F., Hiller, E. H., Garber, J. E., Peshkin, B. N., Lerman, C., Watson, P., Lynch, H. T., Hilsenbeck, S. G., Rubinstein, W. S., Hughes, K. S., and Parmigiani, G. BRCAPRO validation, sensitivity of genetic testing of *BRCA1*/*BRCA2*, and prevalence of other breast cancer susceptibility genes. J. Clin. Oncol., *20:* 2701–2712, 2002.

13. Nathanson, K. L., Wooster, R., and Weber, B. L. Breast cancer genetics: what we know and what we need. Nat. Med., *7:* 552–556, 2001.

14. Hopper, J. L., and Carlin, J. B. Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. Am. J. Epidemiol., *136:* 1138–1147, 1992.

15. Antoniou, A. C., Pharoah, P. D. P., McMullan, G., Day, N. E., Ponder, B. A. J., and Easton, D. F. Evidence for further breast cancer susceptibility genes in addition to *BRCA1* and *BRCA2* in a population-based study. Genet. Epidemiol., *21:* 1–18, 2001.

16. Antoniou, A. C., Pharoah, P. D. P., McMullan, G., Day, N. E., Straton, M. R., Peto, J., Ponder, B. J., and Easton, D. F. A. comprehensive model for familial breast cancer incorporating *BRCA1*, *BRCA2* and other genes. Br. J. Cancer, *86:* 76–83, 2002.

17. Cui, J., Antoniou, A. C., Dite, G. S., Southey, M. C., Venter, D. J., Easton, D. F., Giles, G. G., McCredie, M. R. E., and Hopper, J. L. After *BRCA1* and *BRCA2* - what next? multifactorial segregation analyses of three-generation,

population-based australian families affected by female breast cancer. Am. J. Hum. Genet., *68:* 420–431, 2001.

18. Visser, O., Coebergh, J. M. W., and Schouten, L. J. Incidence of Cancer in the Netherlands 1989. Fourth Report of the Netherlands Cancer Registry. Utrecht, 1995.

19. Hoskins, K. F., Stopfer, J. E., Calzone, K. A., Merajver, S. D., Rebbeck, T. R., Garber, J. E., and Weber, B. L. Assessment and counselling for women with a family history of breast cancer. J. Am. Med. Assoc., *273:* 577–585, 1995.

20. Antoniou, A., Pharoah, P. D. P., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., Loman, N., Olsson, H., Johannsson, O., Borg, A., Pasini, B., Radice, P., Manoukian, S., Eccles, D. M., Tang, N., Olah, E., Anton-Culver, H., Warner, E., Lubinski, J., Gronwald, J., Gorski, B., Tulinius, H., Thorlacius, S., Eerola, H., Nevanlinna, H., SyrjΣkoski, K., Kallioniemi, O. P., Thompson, D., Evans, C., Peto, J., Lalloo, F., Evans, D. G., and Easton, D. F. Average risks of breast and ovarian cancer associated with *BRCA1* and *BRCA2* mutations detected in case series unselected for family history: a combined analysis of 22 studies. Am. J. Hum. Genet., *72:* 1117–1130, 2003.

21. Claus, E. B., Risch, N., and Thompson, W. D. Autosomal dominant inheritance of early-onset breast cancer: implications for risk prediction. Cancer (Phila.), *81:* 1879–1886, 1994.

22. Freimer, N. B., Heutink, P., and Wijmenga, C. Obituary: Lodewijk A. Sandkuijl (July 31, 1953 - December 4, 2002). Am. J. Human Genet., *72:* 781–784, 2003.

Modeling Familial Clustered Breast Cancer Using Published Data

M. A. Jonker, C. E. Jacobi, W. E. Hoogendoorn, et al.

Cancer Epidemiol Biomarkers Prev 2003;12:1479-1485.