

# A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes

AC Antoniou<sup>1</sup>, PDP Pharoah<sup>2</sup>, G McMullan<sup>3</sup>, NE Day<sup>4</sup>, MR Stratton<sup>5</sup>, J Peto<sup>5</sup>, BJ Ponder<sup>2</sup> and DF Easton\*<sup>1</sup>

<sup>1</sup>CRC Genetic Epidemiology Unit, Institute of Public Health, Strangeways Research Laboratory, Worts Causeway, University of Cambridge, Cambridge CB1 8RN, UK; <sup>2</sup>CRC Human Cancer Genetics Research Group, University of Cambridge, Cambridge CB1 8RN, UK; <sup>3</sup>High Performance Computing Facility, Cavendish Laboratory, University of Cambridge, Cambridge CB1 8RN, UK; <sup>4</sup>Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; <sup>5</sup>Institute of Cancer Research, Sutton, Surrey, UK

In computing the probability that a woman is a BRCA1 or BRCA2 carrier for genetic counselling purposes, it is important to allow for the fact that other breast cancer susceptibility genes may exist. We used data from both a population based series of breast cancer cases and high risk families in the UK, with information on BRCA1 and BRCA2 mutation status, to investigate the genetic models that can best explain familial breast cancer outside BRCA1 and BRCA2 families. We also evaluated the evidence for risk modifiers in BRCA1 and BRCA2 carriers. We estimated the simultaneous effects of BRCA1, BRCA2, a third hypothetical gene 'BRCA3', and a polygenic effect using segregation analysis. The hypergeometric polygenic model was used to approximate polygenic inheritance and the effect of risk modifiers. BRCA1 and BRCA2 could not explain all the observed familial clustering. The best fitting model for the residual familial breast cancer was the polygenic, although a model with a single recessive allele produced a similar fit. There was also significant evidence for a modifying effect of other genes on the risks of breast cancer in BRCA1 and BRCA2 mutation carriers. Under this model, the frequency of BRCA1 was estimated to be 0.051% (95% CI: 0.021–0.125%) and of BRCA2 0.068% (95% CI: 0.033–0.141%). The breast cancer risk by age 70 years, based on the average incidence over all modifiers was estimated to be 35.3% for BRCA1 and 50.3% for BRCA2. The corresponding ovarian cancer risks were 25.9% for BRCA1 and 9.1% for BRCA2. The findings suggest that several common, low penetrance genes with multiplicative effects on risk may account for the residual non-BRCA1/2 familial aggregation of breast cancer. The modifying effect may explain the previously reported differences between population based estimates for BRCA1/2 penetrance and estimates based on high-risk families.

*British Journal of Cancer* (2002) **86**, 76–83. DOI: 10.1038/sj/bjc/6600008 www.bjcancer.com

© 2002 The Cancer Research Campaign

**Keywords:** segregation analysis; BRCA3; Polygenes; high-risk families; population studies

Of the breast cancer susceptibility genes which have been identified to date, BRCA1 and BRCA2 are the most important 'high risk' genes accounting for the majority of families with multiple cases of breast and ovarian cancer. Other breast cancer susceptibility alleles include mutations in the TP53 gene (Malkin *et al*, 1990), mutations in the PTEN gene (Nelen *et al*, 1996) and rare HRAS1 mutations (Krontiris *et al*, 1993). There is also evidence that heterozygous carriers of the ATM gene are at increased risks of breast cancer (Easton, 1994; Swift *et al*, 1991; Olsen *et al*, 2001). But even if the effects of all the identified genes are taken together, it is estimated that they can explain only 20–25% of the familial relative risk for breast cancer (Easton, 1999; Rebbeck, 1999). Evidently other breast cancer susceptibility genes may exist. However, it is unclear whether the residual familial aggregation is due to rare mutations conferring a high risk of the disease, similar to BRCA1/2, or to common but lower risk variants.

In genetic counselling it is important to have an accurate genetic model on which to base estimates of mutation carrier probabilities and cancer risks. Such a model needs to incorporate both the effects of both the known susceptibility genes (BRCA1 and BRCA2) and of other possible breast cancer susceptibility genes. The existing genetic models are either based on a single susceptibility gene (Claus *et al*, 1991) or take into account only the effects of BRCA1 and BRCA2 (Parmigiani *et al*, 1998) and hence they do not describe the familial breast cancer adequately.

In this study we have combined data from a population-based series of breast cancer cases with data from high risk families, to investigate the genetic models that can explain familial breast cancer outside BRCA1 and BRCA2 families, and evaluate the evidence for risk modifiers in BRCA1 and BRCA2 mutation carriers. In a previous analysis of the population based Anglian Breast Cancer study (Antoniou *et al*, 2001), we have demonstrated that the familial aggregation of breast cancer can be adequately explained by the effects of BRCA1, BRCA2 and a polygenic component (i.e. a large number of low risk polymorphisms acting multiplicatively). Here we investigate how these models fit data on high risk families, and

\*Correspondence: DF Easton; E-mail: doug.easton@srl.cam.ac.uk  
Received 19 July 2001; revised 25 October 2001; accepted 25 October 2001

evaluate the modifying effects of other genes in BRCA1/2 carriers.

## MATERIALS AND METHODS

### Anglian Breast Cancer (ABC) study

The data collection for this study is described in more detail elsewhere (The Anglian Breast Cancer Study Group 2000; Antoniou *et al*, 2001). One thousand four hundred and eighty four cases diagnosed with breast cancer under the age 55 years were identified in the region served by the Anglian Cancer Registry between 1991 and 1996. The cases were invited to provide blood sample and to complete an epidemiological questionnaire including information on family history of cancer in all first degree relatives. Genomic DNA, extracted from the blood, was screened for BRCA1 and BRCA2 mutations by Conformation Sensitive Gel Electrophoresis (CSGE), and mutations were confirmed by sequencing.

### Multiple case families

One hundred and fifty six families were ascertained in response to national publicity in the UK and by referral by oncologists or general practitioners. Each family contained two or more breast cancer cases, at least one of which was diagnosed under age 50 years. Cancer occurrence and follow up was recorded on all family members. At least one individual from each family had DNA samples analyzed for BRCA1 and BRCA2 mutations by CSGE. These families are referred to as the 'B' families.

### Statistical methods

The analysis was based on breast and ovarian cancer occurrence in the combined dataset of the 'B' families and the ABC study. All individuals were censored at age 80 years, the age of their first cancer or their age of death, whichever occurred first. Thus, only information on the first cancer was included in the analysis. Individuals known to have died but with unknown age at death were censored at age last at follow up or at age 70 years, whichever was smaller. If it was known that an individual had developed breast or ovarian cancer, but the age at diagnosis was unknown, we treated the age at interview or age at death, whichever was smaller, as the age of cancer. Individuals with no age information were treated as lost at follow up and were censored at age zero years. Females who had had prophylactic oophorectomy were censored for both breast and ovarian cancer at the age of the surgery since oophorectomy is known to reduce the risk of breast cancer as well as ovarian cancer. Females who had had mastectomy were censored for breast cancer at the age of the surgery. The analyses were implemented using the computer program MENDEL (Lange and Weeks, 1988) which we parallelized to run on a Hitachi SR2201 parallel computer of the University of Cambridge High Performance Computing Facility.

### Models

We modelled the simultaneous effects of BRCA1, BRCA2, a third hypothetical gene BRCA3 which was assumed to have increased risks of breast cancer, and a polygenic effect. The models for this analysis are described in detail elsewhere (Antoniou *et al*, 2001). Briefly, the breast cancer incidence rate  $\lambda(t)$  was assumed to be based on the Cox model (Cox and Oakes, 1984),  $\lambda(t) = \lambda_0(t) \exp(G+P)$ .  $\lambda_0(t)$  is the baseline incidence rate,  $G$  depends on the major genotype and  $P$  is the polygenic component, which is assumed to be normally distributed with zero mean and variance  $\sigma_p^2$ . The polygenic component  $P$  was assumed to act in BRCA1 and BRCA2 carriers as well as non-carriers. However, we also fitted models in which  $P$  had no effect in BRCA1/2 carriers, and models in which the variance of the polygenic 'modifying' component was

allowed to take a different value  $\sigma_m^2$ . The Hypergeometric Polygenic Model (HPM) (Lange, 1997a,b) was used to approximate the polygenic and modifying components, where  $P$  was approximated by

$$P = \frac{R - N}{\sqrt{\frac{N}{2} \sigma_p}}$$

$R$  has a binomial distribution  $(2N, \frac{1}{2})$  according to the HPM and  $N$  is the number of loci used in the HPM, which was set equal to 3 in our analyses.

The ovarian cancer incidence rates were based on a similar model, but we assumed that the polygenotype and the modifying genes did not have any effect on the ovarian cancer risks. BRCA3 carriers were assumed to develop ovarian cancer according to the general population incidence rates. Breast and ovarian cancers were assumed to occur independently conditional on genotype.

The overall breast and ovarian cancer incidence rates were constrained to agree with the national incidence rates for England and Wales (1983–1987). The implementation of these constraints is described elsewhere (Antoniou *et al*, 2001). In order to estimate the BRCA1 and BRCA2 cancer risks, we assumed that the observed incidence rates for BRCA1 and BRCA2 mutation carriers were a constant multiple ( $RR$ ) of the Breast Cancer Linkage Consortium (BCLC) rates (Easton *et al*, 1993; Ford *et al*, 1998). We then investigated the effect of allowing the relative risks ( $RR$ ) to vary with age. For the models allowing for the modifying effect of other genes on the breast cancer risks of BRCA1 and BRCA2, we assumed that the observed incidence rate was the average over all possible modifying effects.

In the ABC families, only the index case was tested for BRCA1 and BRCA2 mutations. In the 'B' families, mutation testing was first performed on an index affected individual. If a mutation was found, other relatives were then screened for the same mutation. For the index cases we assumed that the sensitivity of the mutation testing was 64%, previously estimated by the Breast Cancer Linkage Consortium for similar methodologies (Ford *et al*, 1998). For any other relatives the mutation screening was assumed to be 100% sensitive and 100% specific. In the ABC study only the index patient was tested, and the screening sensitivity was assumed to be 64%.

### Estimation

The genetic models were specified in terms of the breast and ovarian cancer incidence rates in BRCA1/2 carriers, the allele frequencies, the number of loci ( $N$ ) in the HPM and the standard deviation for the polygenic and modifying components,  $\sigma_p$  and  $\sigma_m$ . Maximum likelihood estimation was used to estimate the parameters. For the ABC families we maximized the conditional likelihood of observing the disease phenotypes and mutation status given the disease phenotype of the index case. For the 'B' families, which were ascertained on the basis of multiple affected individuals, we conditioned on the disease phenotypes of all family members.

The variances of the parameters were estimated by inverting the observed information matrix. To compute confidence intervals we used transformed values of the parameters which provide estimates which are normally distributed. Gene frequencies were transformed using the logit function  $\log(p/(1-p))$ , while for relative risks and  $\sigma_p^2$  and  $\sigma_m^2$  we used the log transformation.

### Goodness of fit

The goodness of fit of the models was tested using likelihood based tests. Starting from a 'sporadic' model (i.e. only the BRCA1 and BRCA2 effects taken into account), we tested if the addition of parameters improved the fit significantly. The models were

constructed by adding parameters that were significant in a step-wise fashion. Non-nested models were compared using the Akaike Information Criterion (Akaike, 1974):

$$AIC = -2\ln \text{lik} + 2[\text{no. of independent parameters}].$$

The model that minimizes AIC was chosen (Elston, 1990). For the best fitting models we also compared the predicted number of mutations in index cases with the number observed.

### Carrier probabilities

To calculate the number of mutations predicted by each model we first computed the probability that the index case was a mutation carrier, on the basis of family history (FH) of breast or ovarian cancer. Thus, the probability  $\pi_{j1}$  that the index case in family  $j$  was a BRCA1 carrier was given by:

$$\pi_{j1} = P(\text{BRCA1}|\text{FH}_j) = \frac{P(\text{BRCA1}, \text{FH}_j)}{P(\text{FH}_j)} \quad (1)$$

$$= \frac{P(\text{BRCA1}, \text{FH}_j)}{P(\text{BRCA1}, \text{FH}_j) + P(\text{BRCA2}, \text{FH}_j) + P(\text{Non} - \text{BRCA1}/2, \text{FH}_j)} \quad (2)$$

These probabilities were computed in MENDEL as the ratio of the likelihood of observing the family with the BRCA1 mutation divided by the likelihood of observing all possibilities. The total number of BRCA1 mutations was then obtained by summing  $\pi_{j1}$  over all families. BRCA2 predictions were computed similarly.

## RESULTS

### ABC families

Table 1 summarizes the pedigrees by the number of breast and ovarian cancer cases diagnosed under age 80 years. One thousand two hundred and eighty probands did not have any family history of breast or ovarian cancer in the first degree relatives. One hundred and sixty-nine cases had one first degree relative diagnosed with breast cancer and 11 had two first degree relatives diagnosed with breast cancer. Twenty-three cases had one relative that developed ovarian cancer under age 80 years, and one case had one breast cancer case and one ovarian cancer case in first degree relatives. There were some differences between the present analysis and the previous analysis of the ABC dataset (Antoniou *et al*, 2001), with regard to the number of mutations. Three mutations previously thought to be disease causing have been reclassified as non-disease causing, while five new mutations have been found by reevaluating the CSGE gels. In total there were eight BRCA1 mutations and 15 BRCA2 mutations. Six of the BRCA1 mutation carriers did not have any family history of breast or ovar-

**Table 1** ABC pedigrees by number of female first cancers under age 80 years and number of mutations

Number of cases	Number of pedigrees	Number of mutations	
		BRCA1	BRCA2
1 BrCa	1280	6	9
2 BrCa	169	2	5
3 BrCa	11	–	1
1 BrCa, 1 OvCa	23	–	–
2 BrCa, 1 OvCa	1	–	–
Total	1484	8	15

ian cancer in first degree relatives and two had a relative diagnosed with breast cancer under age 80 years. Nine BRCA2 mutation carriers had no family history of breast or ovarian cancer, five had one relative that developed breast cancer and one had two first degree relatives diagnosed with breast cancer. Mutations were not identified in any of the 24 index cases with a family history of ovarian cancer.

### 'B' families

Table 2 shows the families grouped by the number of breast cancer cases diagnosed under age 50 years and the number of ovarian cancer cases diagnosed under age 80 years. Eighteen families had at least three breast cancer cases diagnosed under age 50 years and at least one ovarian cancer case. Twenty families had at least four breast cancer cases diagnosed under age 50 years and no ovarian cancer cases. Twenty five families had only three breast cancer cases diagnosed under age 50 years, three families had two breast cancer cases and at least one ovarian cancer case, 66 families had two breast cancer cases under age 50 years and 24 families had one breast cancer case diagnosed under age 50 years. In total there were 618 cases of breast cancer diagnosed at any age and 32 cases of ovarian cancer. 21 (13%) of the index cases were tested positive for BRCA1 mutations and 18 (11.5%) tested positive for BRCA2 mutations. The majority of the mutations were in families with at least three breast cancer cases. In total, 318 individuals were tested for BRCA1 and BRCA2 mutations (including the index cases). Fifty seven BRCA1 mutation carriers and 74 BRCA2 mutation carriers were identified.

### Segregation analysis

**Major gene models** Table 3 summarizes the results for the major gene models for BRCA3 incorporating BRCA1 and BRCA2 mutations into the models but no allowance for polygenic effects. In these, we assumed that BRCA3 mutations conferred a constant relative risk over all age groups relative to the background incidence rates. The major gene models included a dominant, a recessive and a general model where the homozygous and heterozygous carriers were assumed to have different relative risks. All three models fitted significantly better than the sporadic model ( $P < 0.0001$ ). The general model converged to the recessive model which had the lowest AIC among these models. Under the recessive model, BRCA3 was estimated to have an allele frequency of 33% with a cumulative breast cancer risk of 27.6% by age 70 years in the homozygous carriers. The mutation frequencies were estimated to be 0.042% for BRCA1 and 0.054% for BRCA2. The breast cancer risk by age 70 years was estimated to be 49.6% for BRCA1 and 82.2% for BRCA2. The corresponding ovarian cancer risks were 32.8% for BRCA1 and 17.8% for BRCA2.

**Polygenic models** The results for the polygenic models are shown in Table 4. The inclusion of the polygenic component ( $\sigma_p^2$ ) improved the fit significantly over the sporadic model ( $P < 0.0001$ ). There was also significant evidence for a modifying effect on the breast cancer risks of BRCA1 and BRCA2 mutation carriers ( $\sigma_p^2, \sigma_m^2, P=0.0238$ ). But there was no evidence for a difference between the 'modifying variance'  $\sigma_m^2$  and the polygenic variance in non carriers. Thus the model with a single polygenic variance term was the most parsimonious. We also tested a model with different modifying variances in BRCA1 and BRCA2 carriers. Under this model, the BRCA1 modifying variance reached an artificial upper boundary of 2.0, but the BRCA2 modifying variance converged to a lower value (1.423). This was due to the absence of family history of breast cancer in BRCA1 mutation carriers in the ABC families. However, the model was not a significantly better fit over the model with a common variance in BRCA1 and BRCA2 carriers.

**Table 2** Observed (OBS) and expected number of mutations for the 'B' and ABC Families. The 'B' families are categorized according to the number of breast cancer cases diagnosed under age 50 years, and the number of ovarian cancer cases. The ABC families are categorized according to family history (FH) of breast or ovarian cancer. All the predictions are under the recessive (REC) and polygenic (POL) models ( $\sigma_p^2 = \sigma_m^2$ ) of Tables 3 and 4

Mutation	'B' families			ABC families				
	Fam type	OBS	REC	POL	Fam type	OBS	REC	POL
BRCA1	≥ 3 brca	9	6.05	5.86		6	6.57	6.43
BRCA2	≥ 1 ovca	3	3.13	3.05	-ve FH	9	8.48	8.24
No mutation		6	8.82	9.09		1265	1264.95	1265.34
BRCA1	≥ 4 brca	5	3.45	3.64		2	5.01	4.83
BRCA2	0 ovca	6	4.58	4.75	+ve FH	6	7.37	7.27
No mutation		9	11.98	11.61		196	191.62	191.90
BRCA1	3 brca	3	2.42	2.70				
BRCA2	0 ovca	3	3.15	3.46				
No mutation		19	19.42	18.85				
BRCA1	2 brca	1	0.86	0.82				
BRCA2	≥ 1 ovca	0	0.09	0.12				
No mutation		2	2.05	2.06				
BRCA1	2 brca	1	1.69	1.87				
BRCA2	0 ovca	4	4.31	4.52				
No mutation		61	60.01	59.61				
BRCA1	1 brca	2	0.37	0.41				
BRCA2	Any ovca	2	1.16	1.21				
No mutation		20	22.47	22.38				
BRCA1	Totals	21	14.83	15.30	Totals	8	11.57	11.26
BRCA2		18	16.42	17.10		15	15.85	15.51
No mutation		117	124.75	123.60		1461	1456.57	1457.23

**Table 3** Segregation analysis with MENDEL; major gene models

Parameters	Sporadic	Dominant	Recessive	General
$\hat{p}_1$	0.0005829 (0.026–0.132%)	0.0004451 ( $1.5 \times 10^{-7}$ –0.57)	0.0004192 (0.015–0.11%)	0.0004192 (0.015–0.11%)
$\hat{p}_2$	0.0006760 (0.033–0.14%)	0.0005558 ( $8.4 \times 10^{-8}$ –0.786)	0.0005426 (0.021–0.14%)	0.0005426 (0.021–0.14%)
$\hat{p}_3$	NA	0.1462 ( $1.27 \times 10^{-9}$ –1.0)	0.3316 (11.2–66.2%)	0.3316 (11.2–66.2%)
$\hat{R}$ BRCA3:				
Heterozygotes	NA	15.77 (0.00–inf)	NA	1.00
Homozygotes	NA	15.77	15.61 (14.99–16.25)	15.61
$\hat{R}$ BRCA1:				
Breast	0.3200 (0.19–0.56)	0.4894 (0.00–inf)	0.5532 (0.23–1.35)	0.5532 (0.23–1.35)
Ovarian	0.7562 (0.24–2.40)	0.7386 (0.007–74.17)	0.7439 (0.23–2.36)	0.7439 (0.23–2.36)
$\hat{R}$ BRCA2:				
Breast	0.5578 (0.34–0.91)	0.8430 (0.0–inf)	0.9404 (0.39–2.25)	0.9404 (0.39–2.25)
Ovarian	0.6482 (0.24–2.40) (0.14–2.42)	0.6339 (0.007–74.17) (0.003–134.54)	0.6252 (0.23–2.36) (0.18–2.19)	0.6252 (0.23–2.36) (0.18–2.19)
Log-likelihood	-1885.017	-1856.748	-1854.833	-1854.833
Data contributions:				
ABC families	-1712.436	-1683.309	-1681.502	-1681.502
'B' families	-172.581	-173.439	-173.331	-173.331
Degrees of freedom	6	8	8	9
AIC	3782.034	3729.496	3725.666	3727.666

$\hat{p}_1, \hat{p}_2, \hat{p}_3$ : BRCA1, BRCA2 and BRCA3 allele frequencies.  $\hat{R}$ : The relative risks as compared to non-carriers for BRCA3 and relative to the BCLC rates for BRCA1 and BRCA2 carriers (Easton et al., 1993; Ford et al., 1998). 95% Confidence intervals in parentheses.

**BRCA1 and BRCA2 risks** Under the most parsimonious model, ( $\sigma_p^2 = \sigma_m^2$ ) the allele frequency for BRCA1 mutations in the general population was estimated to be 0.056% (95% CI: 0.023–0.132%)

and the allele frequency for BRCA2 mutations was 0.072% (95% CI: 0.037–0.143%). The standard deviation for the polygenic and modifying effects was estimated to be 1.291 (95% CI: 1.096–

**Table 4** Segregation analysis with MENDEL. Polygenic models

Parameters	Sporadic	Parameters included in the model			
		$\sigma_p^2$	$\sigma_p^2, \sigma_m^2, a$	$\sigma_p^2 = \sigma_m^2$	$\sigma_p^2, \sigma_{m1}^2, \sigma_{m2}^2$
$\hat{p}_1$	0.0005829 (0.026–0.132%)	0.0004208 (0.020–0.090%)	0.0006387 (0.027–0.15%)	0.0005567 (0.02–0.13%)	0.0006639 (0.03–0.16%)
$\hat{p}_2$	0.0006760 (0.033–0.14%)	0.0005401 (0.030–0.097%)	0.0007950 (0.04–0.16%)	0.0007240 (0.04–0.14%)	0.0007491 (0.04–0.16%)
RR BRCA1:					
Breast	0.3200 (0.19–0.56)	0.5357 (0.335–0.882)	0.3104 (0.163–0.591)	0.3664 (0.197–0.681)	0.2930 (0.147–0.586)
Ovarian	0.7562 (0.24–2.40)	0.7457 (0.246–2.262)	0.6065 (0.212–1.734)	0.6274 (0.216–1.823)	0.5843 (0.202–1.693)
RR BRCA2:					
Breast	0.5578 (0.34–0.91)	0.913 (0.616–1.354)	0.5532 (0.324–0.945)	0.6335 (0.386–1.039)	0.6076 (0.316–1.168)
Ovarian	0.6482 (0.14–2.42)	0.6305 (0.196–2.024)	0.5746 (0.187–1.764)	0.5801 (0.186–1.806)	0.5537 (0.176–1.745)
$\hat{\sigma}_p$	NA	1.269 (1.076–1.497)	1.286 (1.091–1.516)	1.291 (1.096–1.521)	1.280 (1.085–1.511)
$\hat{\sigma}_m$	NA	NA	2.00 NA	1.291 (1.096–1.521)	NA
$\hat{\sigma}_{m1}$	NA	NA	NA	NA	2.00
$\hat{\sigma}_{m2}$	NA	NA	NA	NA	1.423 (0.481–4.212)
Log-likelihood	–1885.017	–1856.371	–1853.818	–1854.340	–1853.675
Data contributions:					
ABC families	–1712.436	–1683.059	–1681.398	–1682.035	–1681.508
'B' families	–172.581	–173.312	–172.420	–172.305	–172.167
Degrees of freedom	6	7	8	7	9
AIC	3782.034	3726.742	3723.636	3722.680	3725.350

<sup>a</sup>0 ≤  $\sigma_m$  ≤ 2.0.  $p_1, p_2$ : BRCA1 and BRCA2 allele frequencies. RR: The relative risks as compared to the BCLC rates (Easton et al., 1995; Ford et al., 1998). 95% Confidence intervals in parentheses.

1.521). The relative risks of breast and ovarian cancer in BRCA1 and BRCA2 carriers, relative to the BCLC risks (Easton et al, 1993; Ford et al, 1998) were all estimated to be less than one. The cumulative breast cancer risk for a BRCA1 mutation carrier, based on the observed incidence, was estimated to be 36.5% by age 70 years and the ovarian cancer risk 28.5%. The cumulative breast cancer risk for a BRCA2 mutation carrier was estimated to be 68.7% and the corresponding ovarian cancer risk was 16.6%.

To estimate more precisely the shape of the incidence curves for BRCA1 and BRCA2 mutation carriers, we fitted a model where separate breast cancer relative risks were assumed for ages 20–39, 40–49, 50–59 and 60–79 years, for both BRCA1 and BRCA2 mutation carriers. We also allowed for separate ovarian cancer relative risks in the age groups 30–49, 50–59, and 60–69 years for BRCA1 and in the age groups 30–49 and 50–69 years for BRCA2. The number of ovarian cancer cases were not sufficient to allow the age specific ovarian cancer risks to be estimated in any more detail. Where the relative risks applied to two decades of age, we assumed that the ratio of incidence rates in those decades was the same as the ratio from the analysis of the Breast Cancer Linkage Consortium (Easton et al, 1995; Ford et al, 1998). The log-likelihood in this case was maximized over 16 parameters and the estimates are shown in Table 5 together with the age specific incidence rates and cumulative risks. The model fits significantly better than the model with constant relative risk ( $P=0.03$ ). The BRCA1 and BRCA2 mutation frequencies were estimated to be 0.051% (95% CI: 0.021–0.125%) and the BRCA2 mutation frequency was 0.068% (95% CI: 0.033–0.141%) respectively. The standard deviation for the polygenic and modifying effects was estimated to be 1.291 (95% CI: 1.097–1.519), identical to the model with constant relative risks. The estimated breast cancer incidence rates for BRCA1 increase up to the age group 40–49 years but then

decrease, while the incidence rates for BRCA2 mutation carriers increase with age. The relative risk of breast cancer as compared with the population incidence rates decrease sharply with age for BRCA1 mutation carriers, from 23.88 in the 30–39 year age group to 2.31 in the age group 60–69 years ( $P_{trend}=0.005$ ). In contrast, the estimated breast cancer relative risk for BRCA2 mutation carriers as compared to population incidence was 17.52 for the age group 30–39 years but virtually constant over age 40 years at around 11–12 ( $P_{trend}=0.697$ ). The ovarian cancer incidence rates in BRCA1 carriers rise to a peak in the 40–49 years age group, followed by a slight decline, while the BRCA2 ovarian cancer incidence rates were highest in the 50–59 year age group.

The average risk of breast cancer for non-carriers, over all polygenic effects, is estimated to be 4.89% by age 70 years. Based on the 10 and 90% percentiles of the distribution of the polygenic component the cumulative breast cancer risks are 0.3 and 17.9% respectively. The cumulative breast cancer risk in BRCA1 mutation carriers over all modifiers is estimated to be 35.26% by age 70 years. This is lower than the estimate based on the Breast Cancer Linkage Consortium data of 71% (Easton et al, 1993). The estimated breast cancer risks at the 10 and 90% percentiles of the distribution of the modifying effect are 3.7 and 92.1% respectively. The average breast cancer cumulative risk by age 70 years in BRCA2 mutation carriers is 50.26% (7.2 and 99.3% at the 10 and 90% percentiles respectively). This is also lower than the BCLC estimate of 84% (Ford et al, 1998).

### Predicted number of mutations

The predicted number of mutations under the best fitting major gene and polygenic model (with a single relative risk parameter over all age groups) are shown in Table 2. The predicted numbers

**Table 5** Parameters for the polygenic model with  $\sigma_p^2 = \sigma_m^2$  and different breast and ovarian cancer relative risks over age groups

	BRCA1			BRCA2			Non-carriers risk
	Incidence	Risk	RR <sup>a</sup>	Incidence	Risk	RR <sup>a</sup>	
Breast cancer							
30–39	0.011222	0.1286	23.88	0.008236	0.0831	17.52	0.0036
40–49	0.016621	0.2620	12.40	0.014471	0.2067	10.80	0.0146
50–59	0.008255	0.3205	4.91	0.020352	0.3527	12.11	0.0298
60–69	0.004843	0.3526	2.31	0.026326	0.5026	12.53	0.0489
Ovarian cancer							
30–39	0.000206	0.0041	3.43	0.000220	0.0032	3.67	0.0007
40–49	0.011730	0.1143	53.32	0.000440	0.0076	2.00	0.0023
50–59	0.008134	0.1835	20.86	0.004622	0.0524	11.85	0.0058
60–69	0.009756	0.2594	19.51	0.004161	0.0910	8.32	0.0104
Frequency		0.00051		0.0068			
$\sigma_p^2, \sigma_m^2$	1.291						
Log-likelihood	–1844.975						
Contributions; ABC:	–1679.324, 'B': –165.561						
AIC:	3721.950						

<sup>a</sup>Relative risk as compared to population incidence rates.

were very similar under the two models. The predicted number of mutations for the ABC families were similar to the observed number for index cases without family history of breast or ovarian cancer but they were slightly higher than observed for the index cases with positive family history of the diseases. For the 'B' families, the predicted number of BRCA2 mutations is very similar to the observed number, but the number of BRCA1 mutations is somewhat higher than predicted (21 vs 15.30 respectively for the polygenic model). The predicted number of mutations did not change much when separate relative risk parameters were assumed over all age groups for the polygenic model. In the ABC families, 10.44 BRCA1 mutations and 15.67 BRCA2 mutations were predicted (compared to eight and 15 observed respectively), while in the 'B' families 15.96 BRCA1 mutations and 16.72 BRCA2 mutations were predicted (compared to 21 and 18 observed respectively).

## DISCUSSION

We investigated the genetic models for familial breast cancer in a combined dataset consisting of both high-risk families and a population based series of breast cancer cases. We found the most parsimonious model to be one with a polygenic component with an equal variance in BRCA1/2 carriers and in non-carriers. This corresponds to a model in which several polymorphisms each act multiplicatively on risk (by for example, increasing the rate at which key mutational events will occur) to the same extent on carriers and non-carriers. A common recessive allele with moderate penetrance had the best fit among the major gene models fitted for 'BRCA3'. However, the evidence for a major gene is weakened by the fact that the contribution of the 'B' families to the overall log-likelihood is always smaller under the major gene models for BRCA3 than their contribution under the sporadic model. On the other hand, the contribution of the 'B' families is larger under the polygenic models (with modifiers) than under the sporadic model.

Under the best fitting model ( $\sigma_p^2 = \sigma_m^2$ ), the frequency of BRCA1 mutations in the general population was estimated to be 0.051% (95% CI: 0.021–0.125%) which corresponds to about one in 974 individuals being a BRCA1 carrier. This is very similar to the estimate of Ford *et al.* (1995) of one in 800. Higher estimates were obtained by Whittemore *et al.* (1997) (0.14%) and Antoniou *et al.* (2000) (0.13%). These studies were, however, based on segregation analysis of ovarian cancer and

may be inflated by not allowing for other familial effects. The estimate from the ABC dataset alone was 0.023% (Antoniou *et al.*, 2001), lower than the estimate from our combined analysis but included in the confidence interval. The mutation frequency for BRCA2 mutations was estimated to be 0.068% (95% CI: 0.033–0.141%). This corresponds to one in 734 individuals being a BRCA2 carrier in the general population. Again this is somewhat lower than 0.17% estimated in the ovarian cancer segregation analysis of Antoniou *et al.* (2000) (0.17%). The frequency from the ABC analysis alone was again somewhat lower, 0.041% (Antoniou *et al.*, 2001).

The fact that the observed number of BRCA1 mutations in the 'B' families was higher than predicted might suggest that we have slightly underestimated the BRCA1 allele frequency as a consequence of finding (by chance) too few BRCA1 mutations in the ABC study. Alternatively it may be that the sensitivity of mutation testing was lower in the much larger ABC study, although it is not clear why this should be true only for BRCA1.

Under our best fitting model the breast cancer risk for BRCA1 carriers, based on the average incidence over all modifying effects, was estimated to be 35.3% by age 70 years. This is moderately lower than the estimates derived by the BCLC studies of high risk families, which are in the range 71–85% (Easton *et al.*, 1993, 1995) and also slightly lower than the estimates derived from risks to relatives of mutation carriers in population based series of breast cancer patients (Struewing *et al.*, 1997; Hopper *et al.*, 1999). These differences are broadly consistent with our best model, since cases in high risk families will be expected to have a much larger polygenic component than average, while relatives of early onset breast cancer cases will be expected to have only a slightly higher polygenic component than average. The ovarian cancer risk for BRCA1 mutation carriers is estimated to be 26% by age 70 years. This is somewhat higher than the estimate of Struewing *et al.* (1997) but lower than the BCLC estimates (50.6–63.3%) (Easton *et al.*, 1993, 1995). It is also lower than our previous estimate based on ovarian cancer families (66%) (Antoniou *et al.*, 2000). The higher estimates from the last two studies are based on families that were selected on the basis of either ovarian cancer cases or number of breast and ovarian cancer cases, and could be the result of modifying genes or other familial factors increasing ovarian cancer risks. In principle, our model could be extended to account for such modifying factors.

The cumulative risk of breast cancer for BRCA2 mutation carriers was estimated to be 50.3% by age 70 years. Again this is

lower than the BCLC estimates (Ford *et al*, 1998) but more similar to the estimates of the population based study of Hopper *et al*. (1999). The corresponding ovarian cancer risk by age 70 years was estimated to be 9.1%. This is lower than the first BCLC estimate (27%) (Ford *et al*, 1998) and the estimate based on ovarian cancer families (31%) (Antoniou *et al*, 2000). It is also somewhat lower than the estimate from a recent BCLC study of genotype-phenotype correlations (14%) (Thompson and Easton, 2001) and the estimate of Struewing *et al*. (1997) based on carriers of Ashkenazi Jewish founder mutations.

We found that the breast cancer incidence rates in BRCA1 mutation carriers relative to the population incidence rates decrease sharply with age. This contrasts with the pattern in BRCA2 carriers, in whom the relative risk is approximately constant over age 40 years. Thus, BRCA2 mutations are estimated to confer a breast cancer risk which is similar to the population risk but 10–12 times greater. This difference in the pattern of incidence rates is mirrored in the very different histopathology of tumours in BRCA1 and BRCA2 carriers (Lakhani *et al*, 1998), and reflects important mechanistic differences. A practical implication of these different age-incidence patterns, is that BRCA1 mutation carriers are more likely to develop breast cancer at a younger age than BRCA2 mutation carriers.

The polygenic model suggests that several genes, each having small but multiplicative effects on risk can account for the residual non-BRCA1/2 familial clustering of breast cancer. Genes other than BRCA1 and BRCA2 implicated in the aetiology of hereditary breast cancer include the TP53 gene (Malkin *et al*, 1990), and the PTEN gene (Nelen *et al*, 1996). Mutations in these genes, however, predispose to rare autosomal dominant syndromes and their contribution to familial aggregation is very small. Heterozygous carriers of the ataxia-telangiectasia gene ATM have been reported to be at increased risks of breast cancer (Swift *et al*, 1991; Easton, 1994; Olsen *et al*, 2001) and there is also evidence that rare HRAS1 alleles may be associated with moderate risks for breast cancer (Krontiris *et al*, 1993). It has been suggested that genes involved in steroid hormone metabolism and transport could act together defining a high-risk profile for breast cancer (Henderson and Feigelson, 2000). Genes included in this pathway include the HSD17B1 gene, the CYP17 and CYP19 genes and the Estrogen Receptor (ER) gene (Henderson and Feigelson, 2000). At present, however, there is no clear evidence that polymorphisms in these genes are associated with a significant risk (Dunning *et al*, 1998). In a meta-analysis of all the studies of low risk polymorphisms, significant evidence was found for carriers of the (TTTA)<sub>10</sub> Cyp19 polymorphism, the GSTP1 Ile105Val polymorphism and the TP53 Arg72Pro polymorphism (Dunning *et al*, 1998) but the evidence was not conclusive.

Rebbeck *et al*. (1999b) found that BRCA1 carriers with long repeat lengths of the (AG)<sub>n</sub> polymorphism at the Androgen Receptor gene may have earlier age at diagnosis of breast cancer. Similarly, the steroid hormone metabolism gene AIB1 has been

reported to modify the breast cancer risk in BRCA1 mutation carriers (Rebbeck *et al*, 1999a). Carriers of a variant AIB1 allele were found to have an age of diagnosis of breast cancer which is on average 3 years younger than non-carriers. However, neither of these results have been replicated.

The Hypergeometric Polygenic Model is equivalent to a fully additive polygenic trait, with no dominance or epistatic variance (Lange, 1997b). In practice some degree of dominance may exist, given the slightly higher familial relative risks among sibs (Pharoah *et al*, 1997). Among the major gene models fitted, the recessive model was the most parsimonious, and a recessive effect was also found by Cui *et al*. (2001) in BRCA1/2 negative families. However, some of the evidence for a recessive effect may be due to secular increase in incidence, which would give rise to a higher risk to siblings of probands, and to the protective effect of parity which slightly reduces the risk to mothers of probands (Antoniou *et al*, 2001). It is also, of course, possible that some of the familial aggregation modelled here in terms of genetic susceptibility may be due to clustering of lifestyle risk factors (for example diet or reproductive factors) within families.

In conclusion, the present results suggest that a number of low penetrant genes may account for the familial clustering of breast cancer outside BRCA1 and BRCA2 families. The modifying effect on the breast cancer risks of BRCA1 and BRCA2 carriers may explain some of the differences between the risk estimates from population based studies and high risk families. The resulting model provides a framework for risk estimation to counsel women with a family history of breast cancer, allowing one to estimate carrier probabilities (separately for BRCA1 and BRCA2) and incidence rates in the same analysis. The model can also be extended straightforwardly to include risks of other cancers, second breast/ovarian cancers and tumour pathology. In principle it should also be possible to include lifestyle factors, although this would depend on assumptions about their effects in susceptible individuals.

## ACKNOWLEDGEMENTS

Our calculations were performed on the Hitachi SR2201 located at the University of Cambridge High Performance Computing Facility. This work was supported by grants from The Cancer Research Campaign, the NCI (1R01 CA81203–01A1), the Cambridge Commonwealth Trust, the Overseas Research Studentships, the A.G. Leventis Foundation, and the Raymond and Beverly Sackler Medical Research Centre. The Anglian Breast Cancer Study was funded by the NHS Research and Development. We would like to thank Julian Liscombe (ABC study Coordinator), Karen Redman (Research Nurse), Vickie Basham and Jane Gregory (mutation analysis). PDP Pharoah is a senior Clinical research Fellow of The Cancer Research Campaign (CRC). BAJ Ponder is a Gibbs Fellow of the CRC. DF Easton is a principal research fellow of the CRC.

## REFERENCES

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automatic Control* **19**: 716–726
- Antoniou AC, Gayther SA, Stratton JF, Ponder BAJ, Easton DF (2000) Risk models for familial ovarian and breast cancer. *Genet Epidemiol* **18**: 173–190
- Antoniou AC, Pharoah PDP, McMullan G, Day NE, Ponder BAJ, Easton DF (2001) Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population based study. *Genet Epidemiol* **21**: 1–18
- Claus EB, Risch N, Thompson D (1991) Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. *Am J Hum Genet* **48**: 232–242
- Cox DR, Oakes D (1984) Analysis of survival data. London: Chapman and Hall
- Cui J, Antoniou AC, Dite GS, Southey MC, Venter D, Easton DF, Giles GG, McCredie MRE, Hopper JL (2001) 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
- Dunning AM, Healy CS, Pharoah PDP, Foster NA, Lipscombe JM, Redman KL, Easton DF, Day DF, Ponder BAJ (1998) No association between a polymorphism in the steroid metabolism gene CYP17 and the risk of breast cancer. *Br J Cancer* **77**: 2045–2047

- Easton DF, Bishop DT, Ford D, Crockford GP and the Breast Cancer Linkage Consortium (1993) Genetic linkage analysis in familial breast and ovarian cancer. *Am J Hum Genet* **52**: 678–701
- Easton DF (1994) Cancer risks in A-T heterozygotes. *Int J Radiat Biol* **66**: S177–S182
- Easton DF, Ford D, Bishop T and the Breast Cancer Linkage Consortium (1995) Breast and Ovarian Cancer incidence in BRCA1 mutation carriers. *Am J Hum Genet* **56**: 265–271
- Easton DF (1999) How many more breast cancer predisposition genes are there. *Breast Cancer Res* **1**: 14–17
- Elston RC (1990) Models for discrimination between alternative models of inheritance. In *Advances in statistical methods for generic livestock* Gianola D and Hammond F (eds) pp 41–45 New York: Springer Berlin, Heidelberg
- Ford D, Easton DF, Peto J (1995) Estimates of the gene frequency of BRCA1 and its contributions to breast and ovarian cancer incidence. *Am J Hum Genet* **57**: 1457–1462
- Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, Sobol H, Teare MD, Struewing J, Arason A, Scherneck S, Peto J, Rebbeck TR, Tonin P, Neuhausen S, Barkardottir R, Eyfjord J, Lynch H, Ponder BA, Cayther SA, Zelada-Hedman M (1998) Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *Am J Hum Genet* **62**: 676–689
- Henderson BE, Feigelson HS (2000) Hormonal carcinogenesis. *Carcinogenesis* **21**: 427–433
- Hopper JL, Southey MC, Dite GS, Jolley DJ, Giles GG, McCredie MRE, Easton DF, Venter DJ and the Australian Breast Cancer Family Study (1999) Population-based estimate of the average age specific cumulative risk of breast cancer for a defined set of protein-truncating mutations in BRCA1 and BRCA2. *Cancer Epidemiol Biomarkers Prev* **8**: 741–747
- Krontiris TG, Devlin B, Karp DD, Robert NJ, Risch N (1993) An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. *N Engl J Med* **329**: 517–523
- Lakhani SR, Jacquemier J, Sloane JP, Gusterson BA, Anderson TJ, van de Vijver MJ, Farid LM, Venter D, Antoniou A, Storer-Isser A, Smyth E, Steel CM, Haites N, Scott RJ, Goldgar D, Neuhausen S, Daly PA, Ormiston W, McManus R, Scherneck S, Ponder BA, Ford D, Peto J, Stoppa-Lyonnet D, Easton DF (1998) Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. *J Natl Cancer Inst* **90**: 1138–1145
- Lange K, Weeks D (1988) Programs for pedigree analysis: MENDEL, FISHER and dGENE. *Genet Epidemiol* **5**: 471–472
- Lange K (1997a) An approximate model of polygenic inheritance. *Genetics* **147**: 1423–1430
- Lange K (1997b) *Mathematical and Statistical methods for genetic analysis*. New York, Berlin, Heidelberg: Springer-verlang
- Malkin D, Li FP, Strong LC, Fraumeni JF, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA, Friend SH (1990) Germ line p53 mutations in a familial syndrome of breast cancer sarcomas and other neoplasms. *Science* **250**: 1233–1238
- Nelen MR, Padberg GW, Peeters EA, Lin AY, van den Helm B, Frants RR, Coulon V, Goldstein AM, van Reen MMM, Easton DF, Eeles RA, Hodgson S, Mulvihill JJ, Murday VA, Tucker MA, Mariman ECM, Starink TM, Ponder BAJ, Ropers HH, Kremer H, Longy M, Eng C (1996) Localisation of the gene for Cowden disease to chromosome 10q22-23. *Nat Genet* **13**: 114–114
- Olsen JH, Hahneemann JM, Borresen-Dale A, Brondum-Nielsen K, Hammarstrom L, Kleinerman R, Kaariainen H, Lonnqvist T, Sankila R, Seersholm N, Tretli S, Yuen J, Boice JD, Tucker M (2001) Cancer in Patients with Ataxia-Talangiectasia and in their relatives in the Nordic countries. *J Natl Cancer Inst* **93**: 121–127
- Parmigiani G, Berry DA, Aguilar O (1998) Determining carrier probabilities for breast cancer susceptibility genes BRCA1 and BRCA2. *Am J Hum Genet* **62**: 145–158
- Pharoah PP, Day NE, Duffy S, Easton DF, Ponder BAJ (1997) Family history and the risk of breast cancer: A systematic review and meta analysis. *Int J Cancer* **71**: 800–809
- Rebbeck TR (1999) Inherited genetic predisposition in breast cancer. *Cancer* **86**: 2493–2501
- Rebbeck T, Kantof P, Krithivas K, Neuhausen A, Godwin, Daly MB, Narod SA, Garber JE, Lynch HT, Weber BL, Brown MA (1999a) Modification of breast cancer risk in BRCA1 mutation carriers by the AIB1 gene. *Carcinogenesis* **40**: 194
- Rebbeck TR, Kantof PW, Krithivas K, Neuhausen S, Blackwood MA, Godwin AK, Daly MAB, Narod SA, Garber JE, Lynch HT, Weber BL, Brown M (1999b) Modification of BRCA1-associated breast cancer risk by the polymorphic Androgen-Receptor CAG repeat. *Am J Hum Genet* **64**: 1371–1377
- Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, Timmerman MM, Brody LC, Tucker MA (1997) The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N Engl J Med* **336**: 1401–1408
- Swift M, Morell D, Massey R, Chase CL (1991) Incidence of cancer in 161 families affected by Ataxia-Talangiectasia. *N Engl J Med* **325**: 1831–1836
- The Anglian Breast Cancer Study Group (2000) Prevalence and Penetrance of BRCA1 and BRCA2 mutations in a large population based series of breast cancer cases. *Br J Cancer* **83**: 1301–1308
- Thompson D, Easton DF on behalf of the Breast Cancer Linkage Consortium (2001) Variation in cancer risks by mutation position in BRCA2 mutation carriers. *Am J Hum Genet* **68**: 410–419
- Whittemore AS, Gong G, Intyre J (1997) Prevalence and contribution of BRCA1 mutations in breast cancer and ovarian cancer: results from three U.S. population-based case-control studies of ovarian cancer. *Am J Hum Genet* **60**: 496–504