ORIGINAL ARTICLE

A risk prediction algorithm for ovarian cancer incorporating BRCA1, BRCA2, common alleles and other familial effects

Sarah Jervis,1 Honglin Song,2 Andrew Lee,1 Ed Dicks,2 Patricia Harrington,2 Caroline Baynes,2 Ranjit Manchanda,3,4 Douglas F Easton,1,2 Ian Jacobs,3,5 Paul P D Pharoah,1,2 Antonis C Antoniou1

ABSTRACT

Background Although BRCA1 and BRCA2 mutations account for only ~27% of the familial aggregation of ovarian cancer (OvC), no OvC risk prediction model currently exists that considers the effects of BRCA1, BRCA2 and other familial factors. Therefore, a currently unresolved problem in clinical genetics is how to counsel women with family history of OvC but no identifiable BRCA1/2 mutations.

Methods We used data from 1548 patients with OvC and their relatives from a population-based study, with known BRCA1/2 mutation status, to investigate OvC genetic susceptibility models, using segregation analysis methods.

Results The most parsimonious model included the effects of BRCA1/2 mutations, and the residual familial aggregation was accounted for by a polygenic component (SD 1.43, 95% CI 1.10 to 1.86), reflecting the multiplicative effects of a large number of genes with small contributions to the familial risk. We estimated that 1 in 630 individuals carries a BRCA1 mutation and 1 in 195 carries a BRCA2 mutation. We extended this model to incorporate the explicit effects of 17 common alleles that are associated with OvC risk. Based on our models, assuming all of the susceptibility genes could be identified we estimate that the half of the female population at highest genetic risk will account for 92% of all OvCs.

Conclusions The resulting model can be used to obtain the risk of developing OvC on the basis of BRCA1/2, explicit family history and common alleles. This is the first model that accounts for all OvC familial aggregation and would be useful in the OvC genetic counselling process.

INTRODUCTION

Ovarian cancer (OvC) is the third most common gynaecological cancer (http://www.cancerresearchuk.org/cancer-info/cancerstats/). It is well-established that OvC has a significant genetic component, with the risk to first-degree relatives of patients with OvC estimated to be approximately three times greater than the risk to women in the general population.1,2 High-penetrance mutations in BRCA1 and BRCA2 account for ~27% of these familial cancers3 and another 10% are accounted for by rare variants in the MMR genes, RAD51C, RAD51D and BRIP1 (http://www.nature.com/icogs/primer/common-variation-and-heritability-estimates-for-breast-ovarian-and-prostate-cancers/).

Risk models that incorporate both BRCA1 and BRCA2 mutations and other sources of variation are required to provide accurate estimates of mutation carrier probabilities and cancer risk for use in genetic counselling. Existing risk-prediction models for familial OvC such as Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) or BRCAPRO4 assume that all familial aggregation to OvC is due to BRCA1 and BRCA2 mutations but this does not reflect our understanding of OvC genetic susceptibility. As a consequence, these models may underestimate OvC risks in women without mutations in these genes. Therefore, how to counsel women with family history of OvC but without BRCA1 or BRCA2 mutations has remained a major unresolved question in clinical cancer genetics.

We have used data from a large, population-based series of cases diagnosed with OvC, the Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH), and segregation analysis methods to develop genetic models for OvC that incorporate the effects of BRCA1 and BRCA2 mutations and model the residual familial aggregation to OvC. The explicit effects of 17 common OvC susceptibility alleles, identified through genome-wide association studies (GWAS), were then incorporated into the algorithm. We finally considered the implications of our risk prediction model for OvC risk stratification in the general population and its use in OvC prevention.

MATERIALS AND METHODS

Study population

We used data on 1548 OvC cases (probands) recruited between 1999 and 2010, along with information on their first-degree and second-degree relatives ascertained through an epidemiological questionnaire. The probands were drawn from SEARCH, a large population-based study with cases ascertained through the Eastern Cancer Registration and Information Centre.1,3 Half-sibling status and relative type to the proband, age at cancer diagnosis, cancer site, vital status, status age (the age at death if deceased, the current age if alive) and year of birth were recorded for all probands and relatives.

**Cancer genetics**

**BRCA1 and BRCA2 mutation screening**
SEARCH OvC probands were screened for BRCA1 and BRCA2 mutations as part of a separate project to evaluate the contribution of rare, high-risk and moderate-risk variants to overall OvC risk in the general population. Briefly, this involved targeted sequence library preparation using multiplexed 48.48 Fluidigm access arrays and sequencing on an Illumina HiScan. BRCA1 and BRCA2 mutation status information was available on all 1548 probands. The following alterations were considered pathogenic: protein-truncating insertion/deletion variants, nonsense mutations, consensus splice-site variants and missense variants with reported damaging effect on protein function. For the purpose of our analysis, BRCA1 and BRCA2 mutation status were both recorded simply as mutation-positive or negative, with no distinction between different mutation types by location or functional effect.

**Statistical analysis**

**Segregation analysis of OvC**
Complex segregation analysis was used to fit genetic models to the occurrence of OvC in families, incorporating the explicit effects of BRCA1 and BRCA2 mutations on OvC risk. Female family members were followed from birth until the first of OvC diagnosis age, age at questionnaire, death age or age 80. We also considered breast cancer occurrence, but individuals were continued to be followed up for OvC after a breast cancer diagnosis in the analysis. Data on risk-reducing surgeries were not available in relatives of probands, and we were therefore unable to censor at these events. However, since this is a population-based study in which women with OvC diagnosis were recruited soon after diagnosis, and participants were not aware of their mutation status at the time of recruitment, we do not expect a high prevalence of risk-reducing surgeries at the time of pedigree collection.

To incorporate the effects of BRCA1 and BRCA2 mutations and to take account of changes in cancer incidences over time, the OvC incidence for a female i was assumed to depend on the underlying genetic effects through a model of the form

\[
\lambda_i(g, t, k) = \lambda_0(t, k) \exp(G_i(t, k) + M_i + P_i),
\]

where \(\lambda_0(t, k)\) is the baseline incidence for individuals born in birth cohort k and \(G_i(t, k)\) is the logarithm of the relative risk associated with BRCA1/BRCA2 mutation status \(g\) for age t and cohort k; \(M_i\) is the logarithm of the relative risk associated with a third hypothetical major gene and \(P_i\) is the polygenic component. \(P_i\) is assumed to have a normal distribution with variance \(\sigma^2\) and mean zero and is approximated by the hypergeometric distribution to make it amenable to ‘peeling’. Eight sets of birth cohort and calendar-period-specific incidences for OvC in the general population were derived on the basis of incidences for England and Wales as described previously for the BOADICEA model. The eight cohorts included individuals born pre-1920 then in 10-year intervals up to post 1970. As the number of BRCA1 and BRCA2 mutation carriers in the SEARCH dataset was too small to obtain reliable cancer risk estimates for mutation carriers, we also used external estimates of the OvC and breast cancer relative risks for BRCA1 and BRCA2 carriers relative to population incidences, based on some of the largest studies available. Hence, the average BRCA1 and BRCA2 OvC and breast cancer incidences over all possible genetic effects in the model were fixed. The cohort-specific baseline incidences for BRCA1 and BRCA2 carriers were obtained for each cohort separately by constraining the average incidences over all possible genetic effects to agree with the external estimates. Similarly, the baseline incidences for non-mutation carriers were obtained by constraining the incidences over the BRCA1, BRCA2, other major gene and polygenic effects to agree with the population incidences (see online supplementary material methods).

Since BRCA1 and BRCA2 mutations are also associated with increased breast cancer risks, we incorporated the effect of these mutations on breast cancer incidence. We assumed a similar model for the breast cancer incidence; however, the breast cancer incidence was assumed to depend on only the effects of BRCA1 and BRCA2 mutations.

In our analyses, we considered models with just the BRCA1 and BRCA2 effects, and models that additionally included a dominant, recessive or co-dominant hypothetical major gene and/or a polygenic component.

All the families used in the analysis consisted of women ascertained on the basis of OvC. Thus, to adjust for ascertainment bias, we employed an ascertainment assumption-free approach in which the likelihood of each family’s joint phenotype was modelled as \(P(y)/P(y_i)\), where \(y\) is the vector of all the family phenotypes including all phenotypic and genotypic information on the proband and \(y_i\) is the phenotype of the proband. A sensitivity parameter was introduced, giving the probability of detecting a mutation if one existed, to take account of the fact that mutation screening methods used cannot detect large rearrangements in BRCA1 and BRCA2. A fixed value of 0.9 for both BRCA1 and BRCA2 was used in all models, but additional sensitivity analyses were performed.

Maximum-likelihood estimates of the gene frequencies, polygenic standard deviation and the log relative risk for the hypothetical major gene were calculated using pedigree analysis software MENDEL. SEs for each parameter were obtained from the observed information matrix and were used to calculate 95% CIs. To assess goodness of fit, all of the models with a polygenic or major gene component were compared with the baseline model with just BRCA1 and BRCA2 effects using likelihood ratio tests (LRTs). Further LRTs were used to test for differences between the fit of nested models and the Akaike information criterion (AIC) equal to \(-2(\text{log-likelihood}) - \text{no. of parameters})\) was used to compare non-nested models.

**OvC risk, mutation frequency and carrier numbers prediction**

We used each of the models fitted to predict BRCA1 and BRCA2 mutation carrier frequencies and the risk of developing OvC in the future using the methods previously described in ref [11]. The predictions were used to compare the fit of the models as part of an internal validation. Although goodness-of-fit tests are not valid using the data generating dataset, we calculated \(\chi^2\) goodness-of-fit tests that compared the observed and expected number of mutations and used these as an indicator of the model fit to the data. The expected number of mutation carriers was computed as the sum of the predicted BRCA1 and BRCA2 carrier probabilities across all SEARCH families.

We used the most parsimonious model to estimate risk of developing OvC for a 50-year-old woman to demonstrate the possible clinical implications for different scenarios of BRCA1 and BRCA2 carrier status and extent of family history. The results were compared with the corresponding predictions from the current BOADICEA model.
Incorporating SNPs into the risk prediction algorithm
We extended the most parsimonious model to also incorporate
the explicit effects of the known common OvC susceptibility
alleles following the methodology already published in the
context of prostate cancer. The residual familial aggregation
of OvC was accounted for in this model by a polygenic com-
ponent reflecting the additive effects of a large number of genetic
variants. The polygenic component \( P_i \) for each individual was
divided into two parts for this purpose: a known-variant poly-
genic component \( P_{i,k} \), reflecting the polygenic risk score (PRS)
due to 17 SNPs known to be associated with \( \text{BRCA1} \) and an
unknown residual polygenic component \( P_{i,U} \). The two compo-
nents were assumed to be independent and normally distributed
with mean 0 and variance \( \sigma_k^2 \) and \( \sigma_U^2 \), respectively (see details in
online supplementary material; methods). \( \sigma_k^2 \) was calculated
using previously described methods, based on the known
allele frequencies and per-allele OR estimates. \( \sigma_U^2 \) is then
obtained as the difference between the total polygenic variance
and variance of the PRS.

Distribution of OvC risk and implications for OvC prevention
The OvC risk associated with any individual common genetic
variant is very small compared with rare variants like \( \text{BRCA1} \).
However, as there are thought to be many as yet undiscovered
common variants and their effects are assumed to be additive
on the logarithmic scale a woman with a high polygenic load is
likely to have a greatly increased risk of OvC compared with
someone with a low polygenic load. Being able to distinguish
between high-risk and low-risk individuals in the population
could be a valuable tool in clinical practice. Therefore, we con-
sidered the potential for risk prediction based both on known
common variants and the total hypothesised polygenotype. We
followed a similar approach to the methods described in
ref. [18] (see online supplementary material for more details).
We calculated the proportions of the population and of cancer
cases at different levels of SNP risk and polygenic risk and
plotted against each other for comparison purposes. This pro-
vides an informative measure of the relationship between risk
distribution in the population and among cancer cases. In the
hypothetical future when an individual’s polygenic risk can be
estimated with a high degree of accuracy, either from family
history or because most of the currently theoretical polygeno-
type is accounted for by known variants, these measures could
be used to estimate what proportions of the population would
need to be monitored/screened/followed in order to detect a
particular percentage of OvCs. This could also potentially con-
tribute to stratifying population by OvC risk to enable targeting
of effective screening/preventive intervention strategies for
appropriate risk groups.

RESULTS
Data from 1548 OvC cases recruited into the SEARCH study
were used for our analyses. Female relatives of probands
included 1340 mothers, 1404 sisters and 1144 daughters, of
whom 80 were also diagnosed with OvC and 191 with breast
cancer. The numbers of probands and their first-degree relatives,
the number of OvCs diagnosed in each group and other sample
characteristics are summarised in online supplementary table S1.
All probands were screened for \( \text{BRCA1} \) and \( \text{BRCA2} \) mutations,
identifying 44 and 62 carriers, respectively. The loci, minor
allele frequencies and ORs of the 17 SNPs used in incorporating
their effects into the final model are displayed in online supple-
mentary table S2.

Segregation analyses for OvC incorporating the effects of
\( \text{BRCA1} \) and \( \text{BRCA2} \) mutations
The results for the seven models that incorporate the explicit
effects of \( \text{BRCA1} \) and \( \text{BRCA2} \) on OvC risk and that assume
cohort-specific incidences are summarised in table 1. All the
seven models that accounted for the residual familial aggrega-
tion to OvC (in addition to \( \text{BRCA1} \) and \( \text{BRCA2} \)) provided sig-
nificantly better fit than the model that included only \( \text{BRCA1} \)
and \( \text{BRCA2} \) (\( p <2.8\times10^{-5} \)). The worst-fitting model for the
residual familial aggregation of OvC, other than \( \text{BRCA1} \) and
\( \text{BRCA2} \), was the major recessive and the most parsimonious was
the polygenic model, with AICs of 5772.244 and 5764.372,
respectively. Although the mixed models of inheritance all had
slightly larger log-likelihoods, they did not improve the fit sig-
nificantly over the model with only a polygenic component in
addition to the \( \text{BRCA1} \) and \( \text{BRCA2} \) effects (LRT \( p \) values
>0.14). In all models that included a hypothetical third major
gene, the relative risk for the susceptible women was very high
(ranging between ~54 and ~122). The estimated population
allele frequency for \( \text{BRCA1} \) and \( \text{BRCA2} \) under the polygenic
model were 0.08% (95% CI 0.06% to 0.11%) and 0.26% (95%
CI 0.002% to 0.33%), respectively, with a SD of the polygenic
component of 1.43 (95% CI 1.1 to 1.86).

Predicted number of \( \text{BRCA1} \) and \( \text{BRCA2} \) mutation carriers
and family members diagnosed with OvC
The expected numbers of \( \text{BRCA1} \) and \( \text{BRCA2} \) mutation carriers
computed under each of the models are displayed in table 2. In
line with magnitude of the log-likelihoods, all seven models
gave similar predictions that were noticeably more accurate than
the model that did not allow for additional residual familial
aggregation other than the effects of \( \text{BRCA1} \) and \( \text{BRCA2} \). The
polygenic model performed best for predicting the number of
\( \text{BRCA2} \) mutation carriers and there was a slight improvement in
accuracy of \( \text{BRCA1} \) number under the mixed models. In com-
parison, under the current implementation of BOADICEA, the
predicted \( \text{BRCA1} \) numbers were very close to the observed
values but the number of \( \text{BRCA2} \) carriers was substantially
underpredicted (\( p \) value for difference between observed and
expected number of mutations=4.6E-16).

Similarly, when computing the expected number of families
with a mother, a sister or mother and sister diagnosed with
OvC, the predicted numbers were closer to that observed for
the polygenic and mixed models of inheritance (see online sup-
plemetary table S3).

Predicting future OvC risks
We estimated the probabilities of developing OvC for a
50-year-old woman born in 1940, with the following family
histories: (i) no information on relatives; (ii) having a mother and
sister cancer free at ages 65 and 50; (iii) mother and sister diag-
nosed with OvC at ages 65 and 50; and (iv) and (v) with both
combinations of one diagnosed and one cancer free at these
same ages. We compared these estimates with the risk estimates
from the current version of BOADICEA.

Figure 1 displays the probabilities of developing OvC for a
50-year old woman without a \( \text{BRCA1} \) or \( \text{BRCA2} \) mutation.
Under the best-fitting model, the risk of OvC increases with
increasing number of relatives diagnosed with OvC. In contrast,
the corresponding predictions under BOADICEA remain the
same under all assumptions about family history. Similar pat-
terns are observed when the index female is assumed to carry a
\( \text{BRCA1} \), or a \( \text{BRCA2} \) mutation, where the risks in mutation

Table 1 Parameter estimates, goodness-of-fit measures and likelihood ratio tests (LRTs) of the seven cohort-specific models for breast and ovarian cancer

<table>
<thead>
<tr>
<th>Model</th>
<th>Minor gene frequency (95% CI)</th>
<th>Polygenic SD (95% CI)</th>
<th>Polygenic logrelative risk (95% CI)</th>
<th>Polygenic • (95% CI)</th>
<th>Log-Likelihood</th>
<th>AIC</th>
<th>LRT</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>0.00081 (0.00061 to 0.0011)</td>
<td></td>
<td>0.0026 (0.0020 to 0.0033)</td>
<td>−</td>
<td>−881.237</td>
<td>5788.474</td>
<td></td>
<td>5.11E-06</td>
</tr>
<tr>
<td>Major dominant</td>
<td>0.00079 (0.00060 to 0.0011)</td>
<td></td>
<td>0.0026 (0.0020 to 0.0032)</td>
<td>0.00025 (0.000041 to 0.0015)</td>
<td>4.8 (3.3 to 6.2)</td>
<td>2880.343</td>
<td>5768.686</td>
<td>0.047</td>
</tr>
<tr>
<td>Major recessive</td>
<td>0.00080 (0.00060 to 0.0011)</td>
<td></td>
<td>0.0026 (0.0020 to 0.0032)</td>
<td>0.085 (0.017 to 0.33)</td>
<td>4.0 (2.1 to 6.0)</td>
<td>2882.122</td>
<td>5772.244</td>
<td>0.0079</td>
</tr>
<tr>
<td>Major general</td>
<td>0.00079 (0.00060 to 0.0011)</td>
<td></td>
<td>0.0026 (0.0020 to 0.0032)</td>
<td>0.00025 (0.00020 to 0.0033)</td>
<td>4.8 (3.3 to 6.3)</td>
<td>2879.186</td>
<td>5764.372</td>
<td>0.28</td>
</tr>
<tr>
<td>Polygenic</td>
<td>0.00079 (0.00060 to 0.0011)</td>
<td></td>
<td>0.0026 (0.0020 to 0.0033)</td>
<td>1.43 (1.10 to 1.86)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>0.28</td>
</tr>
<tr>
<td>Mixed dominant</td>
<td>0.00079 (0.00059 to 0.0011)</td>
<td></td>
<td>0.0026 (0.0020 to 0.0032)</td>
<td>0.00023 (0.000023 to 0.0022)</td>
<td>4.7 (2.8 to 6.6)</td>
<td>2877.289</td>
<td>5764.576</td>
<td>0.91</td>
</tr>
<tr>
<td>Mixed recessive</td>
<td>0.00079 (0.00060 to 0.0011)</td>
<td></td>
<td>0.0026 (0.0020 to 0.0032)</td>
<td>0.076 (0.020 to 0.25)</td>
<td>3.7 (1.5 to 5.9)</td>
<td>2878.374</td>
<td>5768.806</td>
<td>0.14</td>
</tr>
<tr>
<td>Mixed general</td>
<td>0.00079 (0.00059 to 0.0011)</td>
<td></td>
<td>0.0026 (0.0020 to 0.0032)</td>
<td>0.00023 (0.000023 to 0.0023)</td>
<td>4.7 (2.8 to 6.6)</td>
<td>2877.283</td>
<td>5766.566</td>
<td></td>
</tr>
</tbody>
</table>

Implications of the polygenic model for OvC prevention

For a polygenic log-risk with the SD of 1.434, estimated under the best-fitting segregation analysis model, and assuming a baseline population OvC risk of 0.02 by age 80, the half of the population at higher risk accounts for 92% of all OvCs. Figure 3 displays the proportion of the population that have a risk greater than a given level and the proportion of the cases predicted to occur within this subgroup. From these curves, it can be seen that 50% of all cancers occur in the 7.7% of the population with a risk of 5.6% or more.

In figure 4, the population proportions are plotted against the case proportions accounted for, for the polygenic log-risk distributions and the combined SNP-effect distributions. The total known variance of the effects of 17 known SNPs is \( \sigma^2_K = 0.0915 \) (see online supplementary material and methods). Due to the low known variance, the distinction between population and case risk is very low for the 17 SNPs alone.

DISCUSSION

We used complex segregation analysis to develop a risk-prediction model for familial OvC that incorporates the effects of BRCA1 and BRCA2 mutations, family history and several newly established common OvC susceptibility alleles using data from a population-based study of OvC cases in the UK. Our model accounts for the familial aggregation of OvC and helps inform a major unresolved clinical question on how to counsel women with family history of OvC but without BRCA1 or BRCA2 mutations.

The most parsimonious model included the effects of BRCA1 and BRCA2 mutations together with a polygenic component. This suggests that most of the familial aggregation not accounted for by BRCA1 and BRCA2 consists of the effects of a large number of genetic variants, each having small contributions to the OvC familial risk. This is in line with results from recent OvC GWAS13-22 that have demonstrated that common carriers also depend on the exact family history information (see online supplementary figures S1 and S2). Under BOADICEA, the risks in mutation carriers are not modified by family history and are all very close to the corresponding risks predicted by our polygenic model algorithm for a women with no family history information.

Incorporating common alleles into the model

The loci, minor allele frequencies and ORs for the 17 SNPs considered are displayed in online supplementary table S2. Under the assumptions that the effects of the SNPs on OvC are all mutually independent and the same for BRCA1 carriers, BRCA2 carriers and non-carriers, each observed SNP profile was translated into a PRS. This PRS was assumed to have a centralised normal distribution with a variance of 0.0915, explaining about 4.5% of the total polygenic variance in our model.

The lifetime risks of OvC to a 20-year-old non-BRCA1/2 mutation carrier, conditional on known PRS and family history of OvC, are shown in figure 2. As expected, the lifetime risk of developing OvC rose exponentially with increasing PRS. For example, the lifetime risk of OvC for a woman without a BRCA1 or BRCA2 mutation but with two affected first-degree relatives is predicted to be >5% if she is at the top 50% of the PRS distribution.

Examples of age-specific risks for a 50-year-old woman at the 5th and 95th percentiles of the PRS and by different family history assumptions are shown in online supplementary figures S3–S5.
low-risk OvC susceptibility alleles exist. Parallel studies in breast cancer suggested that thousands of such genetic susceptibility alleles are likely to exist, which explain a substantial fraction of the unexplained genetic variability. A similar model is likely to apply to OvC. A model that included an additional, dominantly inherited, high-penetrance gene had the highest likelihood. Such a model could reflect the joint effects of other rare OvC susceptibility variants that confer higher risks collectively. However, our analysis may be underpowered as this model did not fit significantly better than the polygenic model.

Previous OvC segregation analyses that accounted for BRCA1 and BRCA2 mutations were based on 10-fold smaller sample set of high-risk OvC families and did not investigate polygenic models for the residual familial aggregation of OvC. In contrast to the present study, those studies found no significant evidence of a third high-penetrance gene in addition to BRCA1 and BRCA2. The difference could be explained primarily by the much lower power of those analyses due the smaller sample size but also due to the fact the ascertainment adjustment involved conditioning on all family phenotypes that imposed a much greater penalty in comparison to the present analysis that used families selected only on the OvC status of the index case.

Under the best-fitting model, the BRCA1 and BRCA2 mutation frequencies were estimated to be 0.00079 and 0.0026, respectively, corresponding to a carrier frequency of 1 in 630 population for BRCA1 and 1 in 195 population for BRCA2. These were higher than the BOADICEA estimates of 0.0006 for BRCA1 and 0.001 for BRCA2, but the difference was significant only for BRCA2 (p values 0.13 and 0.00002). This was also reflected in the significant underprediction of BRCA2 mutations under the BOADICEA model in the current dataset. This

### Table 2: Number of mutation carriers predicted by each model and comparison with observed numbers

<table>
<thead>
<tr>
<th>Model for the residual familial aggregation</th>
<th>Observed BRCA1 carriers</th>
<th>Expected BRCA1 carriers</th>
<th>Observed BRCA2 carriers</th>
<th>Expected BRCA2 carriers</th>
<th>χ² value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>44</td>
<td>56.95</td>
<td>62</td>
<td>63.59</td>
<td>2.98</td>
</tr>
<tr>
<td>Polygenic</td>
<td>44</td>
<td>49.32</td>
<td>62</td>
<td>61.98</td>
<td>0.57</td>
</tr>
<tr>
<td>Dominant major</td>
<td>44</td>
<td>55.62</td>
<td>62</td>
<td>63.08</td>
<td>2.45</td>
</tr>
<tr>
<td>Recessive major</td>
<td>44</td>
<td>55.97</td>
<td>62</td>
<td>63.11</td>
<td>2.58</td>
</tr>
<tr>
<td>General major</td>
<td>44</td>
<td>55.62</td>
<td>62</td>
<td>63.08</td>
<td>2.45</td>
</tr>
<tr>
<td>Dominant mixed</td>
<td>44</td>
<td>48.07</td>
<td>62</td>
<td>61.01</td>
<td>0.36</td>
</tr>
<tr>
<td>Recessive mixed</td>
<td>44</td>
<td>49.08</td>
<td>62</td>
<td>61.10</td>
<td>0.54</td>
</tr>
<tr>
<td>General mixed</td>
<td>44</td>
<td>48.05</td>
<td>62</td>
<td>61.02</td>
<td>0.36</td>
</tr>
<tr>
<td>BOADICEA</td>
<td>44</td>
<td>45.76</td>
<td>62</td>
<td>23.03</td>
<td>66.01</td>
</tr>
</tbody>
</table>

*χ² value, value of χ² goodness-of-fit test.

BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm.

Figure 1: Predicted risks of ovarian cancer over time to a woman born in the 1940 birth cohort without a BRCA1 or BRCA2 mutation by family history. The predicted ovarian cancer risks under the most parsimonious model vary by extent of family history of ovarian cancer. In contrast, under the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm the predicted ovarian cancer risks remain the same under all scenarios.
Figure 2  Lifetime risks of ovarian cancer to a 20-year-old born in the 1940 birth cohort without a BRCA1 or BRCA2 mutation with different polygenic risk score (PRS) and family history. Graph of the change in probabilities of developing ovarian cancer by age 80 as PRS increases from −0.8 to 0.8, to a 20 year old with five different family histories. The two dotted lines, at −0.496 and 0.496, indicate the PRS of those at the 5th and 95th percentile of risk.

Figure 3  Proportion of population above a specified absolute risk of ovarian cancer and proportion of cases occurring in that fraction of the population. Half the population have an absolute risk of ovarian cancer greater than 0.72% by age 80 and 92% of all cases occur in this half of the population. Half of all cancers occur in the 7.7% of the population with risk higher than 5.6%.
The difference between the studies is probably partly due to the data sources and differences in the mutation screening techniques. The 2785 families used to fit the BOADICEA algorithm were ascertained primarily through population-based patients with breast cancer. This source of difference would be in line with the fact that BOADICEA was found to predict BRCA1 and BRCA2 mutations and breast cancer risk well in independent datasets of families with breast cancer.10 26–28 BOADICEA has not been evaluated so far in families ascertained on the basis of OvC only. Another possible factor is the mutation screening methods. The current study is based on currently available sequencing technologies6 that are estimated to be more sensitive in detecting mutations than the techniques used in the late 1990s.29 Moreover, the knowledge of which mutations are actually pathogenic has improved substantially over time.30 Both of these factors could contribute to higher mutation frequencies, although it is unclear why the difference is only significant for BRCA2. An alternative explanation could be a differential response rate for participating in the present study between mutation carriers and non-carriers. BRCA1 and BRCA2 mutations have both been associated with improved short-term OvC survival. In particular, BRCA2 mutation carriers have been reported to have a better prognosis.31 32 If women with short prognosis are more likely to participate in the study, this could potentially lead to an overestimation of the mutation frequency. However, data on response differences by prognostic characteristics are not available to assess this.

In the long term, we expect that these differences will be resolved by fitting a single algorithm to all available data that models comprehensively both the genetic susceptibility to breast cancer and OvC. However, at this stage this is not feasible based on current technologies due to computational complexities (in particular, the number of underlying genotypes in the models). The current approach aims to develop separate algorithms for the susceptibility to breast cancer and OvC that individually incorporate the explicit effects of all observed and unobserved genetic variants such that we obtain accurate risks of each cancer. Validation studies in independent datasets will determine the most appropriate model for use in each context. As technologies evolve over time, in the long term we expect to synthesise the models into a single algorithm.

In our analyses, we took account of OvCs occurring after a breast cancer diagnosis, assuming the OvC incidence remains the same before and after the breast cancer diagnosis. Repeating the analysis but censoring at the first cancer yielded similar results (eg, under the polygenic model BRCA1 mutation frequency was estimated to be 0.083% and BRCA2 mutation frequency was 0.27% with a polygenic SD of 1.46). Therefore, our results were not sensitive to these assumptions.

In our analysis, we aimed to include only epithelial OvCs. However, subsequent to the model fitting process, additional pathology information became available, which revealed 41 of the probands' tumours to be non-epithelial OvCs. This consisted of one BRCA2 carrier and 40 non-carriers, were non-epithelial OvCs. Refitting the models using only epithelial OvCs had very little effect on results. Under the polygenic model, the estimated BRCA1 and BRCA2 mutation frequencies were 0.081% and 0.26%, polygenic SD was 1.44 and the estimated numbers of BRCA1 and BRCA2 carriers were 48.6 and 60.8, respectively.

Our models assumed that the mutation testing sensitivities were 0.9 for both BRCA1 and BRCA2. Obtaining exact estimates is difficult, but in practice mutation sensitivities could be lower. We refitted the models using a sensitivity parameter of 0.83 for

Figure 4  Proportion of cases accounted for by a given proportion of the population above a specified risk of ovarian cancer according to the total polygenic risk and the observed 17 SNP distribution. Under the total polygenic risk distribution, 50% of cancers occur in the 7.7% of the population at highest risk and 92.4% of cancers occur in the half of the population at greater-than-average risk, whereas under the 17 SNP only 62% of cancers occur in the 50% at higher risk and 50% of cases are spread among almost 40% of the population at highest risk.
BRCA1 and 0.76 for BRCA2. Under the polygenic model, the estimated BRCA1 and BRCA2 mutation frequencies were slightly higher at 0.086% and 0.3%, respectively, and the polygenic SD decreased slightly to 1.375, but none of these were significantly different than the results under a sensitivity of 0.9. These patterns are expected as the mutation frequency and mutation screening sensitivity parameters are confounded.

One possible source of bias in our analysis is the possibility of errors in the reporting of family cancer history. However, previous studies have found reported OvC history in first-degree relatives to be reasonably accurate (83.3% probability of agreement between reported OvC status in first-degree relatives and established status). Therefore, the fact that the OvC diagnoses in relatives are not confirmed is unlikely to have a great impact on our results. Another possible weakness of our study is the usage of external estimates of breast cancer and OvC relative risks to BRCA1 and BRCA2 mutation carriers. However, due to the small number of mutation carriers in the SEARCH dataset, it was not possible to estimate reliably the cancer risks for BRCA1 and BRCA2 mutation carriers. The estimates used were based on some of the largest studies available. Future studies should aim to analyse all the data jointly.

Under our models, the probabilities of developing OvC increase with the number of OvCs in relatives, while under BOADICEA the risks remain invariant, at values very close to zero. Therefore, the fact that the OvC diagnoses in relatives are not confirmed is unlikely to have a great impact on our results. Another possible weakness of our study is the usage of external estimates of breast cancer and OvC relative risks to BRCA1 and BRCA2 mutation carriers. However, due to the small number of mutation carriers in the SEARCH dataset, it was not possible to estimate reliably the cancer risks for BRCA1 and BRCA2 mutation carriers. The estimates used were based on some of the largest studies available. Future studies should aim to analyse all the data jointly.

In all models incorporating a polygenic component or known SNPs, the effects were assumed to be the same for carriers of a BRCA1 or BRCA2 mutation and non-carriers. This assumption is supported by recent studies where all but one of the OvC loci identified through GWAS were found to be associated with risk to a similar relative extent in BRCA1 and BRCA2 carriers and non-carriers. Future studies identify additional BRCA1-specific or BRCA2-specific modifiers of risk, it should be possible to extend the present model to allow for this level of complexity.

Although we have incorporated the explicit effects of the common low-risk alleles, future efforts should focus on incorporating the explicit effects of other intermediate risk OvC susceptibility variants such as RAD51C, RAD51D and BRIP1. However, prior to incorporating those into risk prediction models, it is critical to obtain precise estimates of the risks conferred by such mutations that currently are not available.

We also used our models to investigate the possible implications for OvC risk stratification in the population. Using the parameters from the polygenic model, we estimate that 50% of OvCs occur within 7.7% of the population at highest risk. Meanwhile, half of the population at lower risk is forecast to contain only 1 in 13 cancer cases. Targeting the 10% at highest polygenic risk for preventative measures or excluding the low-risk half could therefore lead to a much more efficient distribution of resources. However, to achieve this will require that we identify all the genetic factors that contribute to polygenic inheritance. The almost flat curve in figure 4 from the SNP log-risk distribution, with 50% of the population at higher risk predicted to contain around only 60% of cases, suggests very low power to discriminate between high-risk and low-risk individuals on SNP profiles alone. It is perhaps not surprising as currently only 4.5% of the OvC polygenic variance is accounted for by known low-penetrance genetic variants. However, the currently known SNP profiles in combination with family history information and other risk factors for the disease are expected to have a greater impact for individualised OvC risk prediction, as demonstrated by our model.

Our model can be used in the genetic counselling process of women with family history of OvC as well as for counselling women both with and without BRCA1 or BRCA2 mutations. This would be helpful to both BRCA1 and BRCA2 carriers and non-carriers while making decisions regarding clinical interventions following counselling. Probabilities of developing OvC based on family history, BRCA1, BRCA2 mutation status and/or polygenic risk could be used to assess the risk to an individual and to discriminate between high-risk and low-risk individuals, which may in time prove useful for targeting appropriate interventions.

FUTURE RESEARCH
Although the mutation carrier probability algorithms produced very accurate estimates of the number of carriers in the SEARCH data, an external validation is needed to establish the performance of the model in independent datasets and to assess the model performance in predicting OvC risk in prospective studies. Future plans to extend the models include the addition of lifestyle and reproductive factors such as parity, breast feeding and oral contraceptive use, mutations in genes such as RAD51C, RAD51D and BRIP1 that are known to be associated with OvC risk competing causes of mortality and differences in the associations of the various risk factors with different OvC morphological subtypes. The ultimate goal is to combine the models within the BOADICEA framework and develop a comprehensive breast and ovarian risk user-friendly prediction tool.

Correction notice The license of this article has changed since publication to CC BY 4.0.

Acknowledgements We thank all the study participants who contributed to this study and all the researchers, clinicians and technical and administrative staff who have made possible this work. In particular, we thank Craig Luccarini, the SEARCH team and the Eastern Cancer Registration and Information Centre. ACA is a Cancer Research UK Senior Cancer Research Fellow. IJ is a National Institute for Health Research Senior Investigator.

Contributors Conception and design: ACA and PPDP. Analysis: SJ, HS, AL, ED, PH and CB. Interpretation of data: SJ, ACA, PPDP, RM, DFE and U. Acquisition of data: PPDP and DFE. Drafting the article: SJ, PPDP and ACA. Critical revision: all authors. Final approval: all authors.

Funding This work has been supported by grants from Cancer Research UK (C1005/A12677, C12292/A11174, C490/A10119, C490/A10124) including the PROMISE research programme, the Eve Appeal and the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge.

Competing interests UJ is a director of Abcodia.

Patient consent Obtained.

Ethics approval Cambridgeshire 4 research ethics committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/
Supplementary Material

A Risk Prediction Algorithm for Ovarian Cancer Incorporating 
BRCA1, BRCA2, Common Alleles and Other Familial Effects

Sarah Jervis¹, Honglin Song², Andrew Lee¹, Ed Dicks², Patricia Harrington², Caroline Baynes, Ranjit Manchanda⁴,⁵, Douglas F. Easton¹,², Ian Jacobs³,⁴, Paul P.D. Pharoah¹,², Antonis C. Antoniou¹*

1 Centre for Cancer Genetic Epidemiology, Department of Public and Primary Care, University of Cambridge
2 Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge
3 Institute of Human Development, Faculty of Medical & Human Sciences, The University of Manchester and Manchester Academic Health Science Centre
4 Institute for Women’s Health, University College London
5 Department of Gynaecological Oncology, St Bartholomew’s Hospital

*Address for correspondence:

Antonis C Antoniou

Strangeways Research Laboratory

Department of Public Health and Primary Care

Worts Causeway

Cambridge CB1 8RN

UK
**Methods**

**Constraining the overall incidences.**

The population incidences of ovarian cancer were constrained to agree with national incidence rates for England and Wales. Let \( i(t) \) denote the population incidence rate at age \( t \) and denote the value of the polygenotype \( P \) by \( p_r \) to highlight its dependence on \( R \), taking value \( r \). Then

\[
i(t) = \frac{\sum_{g,p_r} \Pr(g,p_r)f_{g,p_r}(t)}{\sum_{g,p_r} \Pr(g,p_r)S_{g,p_r}(t-1)}
\]

where \((g,p_r)\) denotes the major-locus genotype \( g \) and polygenotype \( p_r \). \( f_{g,p_r}(t) \) and \( S_{g,p_r}(t) \) are the probability of developing ovarian cancer at age \( t \) and the probability of surviving ovarian cancer by age \( t \). The major-gene genotype and polygenotype are assumed to be independent so that \( \Pr(g,p_r) = \tau_g \phi_{p_r} \), where \( \tau_g \) is the probability of major genotype \( g \) and \( \phi_{p_r} \) is the polygenotype probability, given by the binomial distribution. \( f_{g,p_r}(t) \) can be rewritten as \( \lambda_{g,p_r}(t)S_{g,p_r}(t-1) \). Thus, using the relationship

\[
S_{g,p_r}(t-1) = \exp\left[\sum_{u=0}^{t-1} \lambda_{g,p_r}(u)\right] = \exp\left[\sum_{u=0}^{t-1} \lambda_g(u)\right] e^{r_r} = \exp\left[\sum_{u=0}^{t-1} \lambda_g(u)\right] = S_g(t-1)^{e^{r_r}}
\]

\[
i(t) = \frac{\sum_{g,p_r} \tau_g \phi_{p_r} \lambda_g(t) e^{r_r} S_g(t-1)^{e^{r_r}}}{\sum_{g,p_r} \tau_g \phi_{p_r} S_g(t-1)^{e^{r_r}}}
\]
The above equation is then used to estimate the baseline hazard function $\lambda_b(t)$

**Including BRCA1 and BRCA2 ovarian cancer incidence rates**

The external incidence rates of ovarian cancer in the BRCA1 and BRCA2 mutation-positive populations, denoted $b_1(t)$ and $b_2(t)$, are constrained by the following equation, similar to that for general population incidences. $\lambda_b(t)$ and $S_b(t)$ are the baseline incidence rate and survival probability of a BRCAi carrier, “free” of any polygenic effects.

$$b_i(t) = \frac{\sum p_r \Pr(p_r) f(t \mid BRCAi, p_r)}{\sum p_r \Pr(p_r) S(t-1 \mid BRCAi, p_r)} = \frac{\sum p_r \phi_{p_r} f_{b_i}(t \mid p_r)}{\sum p_r \phi_{p_r} S_{b_i} (t-1 \mid p_r)} = \frac{\sum p_r \phi_{p_r} S_{b_i} (t-1 \mid p_r) \lambda_b(t) \exp(p_r)}{\sum p_r \phi_{p_r} S_{b_i} (t-1 \mid p_r)} = \frac{\lambda_b(t) \sum p_r \phi_{p_r} S_{b_i} (t-1) e^{p_r} \exp(p_r)}{\sum p_r \phi_{p_r} S_{b_i} (t-1)^{e^{p_r}}},$$

Solving this equation for the baseline incidence gives us

$$\lambda_b(t) = \frac{b_i(t) \sum p_r \phi_{p_r} S_{b_i} (t-1) e^{p_r}}{\sum p_r \phi_{p_r} S_{b_i} (t-1)^{e^{p_r}}},$$

This can be solved iteratively, starting at $S_{b_i}(0) = 1$ and updating the baseline hazard and survival functions alternatively.

The population incidence rate equation becomes:

$$i(t) = \frac{\sum_{g,p_r} \Pr(g,p_r) f_{g,p_r}(t)}{\sum_{g,p_r} \Pr(g,p_r) S_{g,p_r} (t-1)} = \frac{\sum_{g,p_r} \Pr(g,p_r) f_{g,p_r}(t)}{\sum_{p_r} \phi_{p_r} \left( \sum_{k=0}^{K-1} \tau_k S_k (t-1)^{e^{p_r}} \right)},$$

where $k$ represents the major genotypes (0=non-carrier, 1=BRCA1 carrier, 2=BRCA2 carrier, 3...K-1= carriers of
other hypothetical ovarian cancer susceptibility genetic variants). Multiplying both
sides by the quotient gives:

\[
i(t) \sum_{p_r} \phi_{p_r} \left( \sum_{k=0}^{K-1} r_k S_k (t-1)^{e_{kn}} \right) = \sum_{g, p_r} \text{Pr}(g, p_r) f_{g, p_r}(t)
\]

\[
= \sum_{p_r} \phi_{p_r} \left( \lambda_0 \tau_0 S_0 (t-1)^{e_{n0}} + \sum_{k=3}^{K-1} \lambda_0 \tau_k S_k (t-1)^{e_{kn}} r(k) \right) e^{p_r} + \sum_{p_r} \phi_{p_r} \left( \tau_1 f_{1, p_r}(t) + \tau_2 f_{2, p_r}(t) \right),
\]

Rearrangement of this and use of parts of the \(BRCA1\) and \(BRCA2\) incidence rate
equations produce the following equation for the baseline hazard function.

\[
\lambda_0(t) = \frac{i(t) \sum_{p_r} \phi_{p_r} \left( \sum_{k=0}^{K-1} r_k S_k (t-1)^{e_{kn}} \right) - \sum_{p_r} \phi_{p_r} \left( \tau_1 f_{1, p_r}(t) + \tau_2 f_{2, p_r}(t) \right)}{\sum_{p_r} \phi_{p_r} \left( \tau_0 S_0 (t-1)^{e_{n0}} + \sum_{k=3}^{K-1} \tau_k S_k (t-1)^{e_{kn}} r(k) \right) e^{p_r}}
\]

\[
i(t) \sum_{p_r} \phi_{p_r} \left( \sum_{k=0}^{K-1} r_k S_k (t-1)^{e_{kn}} \right) - \sum_{p_r} \phi_{p_r} \left( \tau_1 \lambda_{b_1}(t) S_{b_1}(t-1)^{e_{n1}} + \tau_2 \lambda_{b_2}(t) S_{b_2}(t-1)^{e_{n2}} \right) \exp(p_r)
\]

\[
\sum_{p_r} \phi_{p_r} \left( \tau_0 S_0 (t-1)^{e_{n0}} + \sum_{k=3}^{K-1} \tau_k S_k (t-1)^{e_{kn}} r(k) \right) \exp(p_r)
\]

which can be solved as part of a multi-step iterative process along with the baseline
\(BRCA1\) and \(BRCA2\) hazard functions and the survival functions for each major
genotype.

For the major gene and mixed models of inheritance the major gene component was
based on three genes: \(BRCA1\), \(BRCA2\), and a third hypothetical gene. Because the
probability of having both a \(BRCA1\) and a \(BRCA2\) mutation is very small, we coded
\(BRCA1\) and \(BRCA2\) as a single locus with three alleles: \(BRCA1\) positive, \(BRCA2\)
positive, and a normal allele. The third gene was diallelic with a normal and a mutant
allele, and was assumed to be unlinked to \(BRCA1\) and \(BRCA2\). For simplicity, \(BRCA1\)
mutations were assumed to be dominant over all other alleles and \(BRCA2\) mutations
were assumed to be dominant over hypothetical third locus disease alleles. There were, therefore, five potential risk categories based on the major genotype: BRCA1 carriers, BRCA2 carriers, heterozygotes for the third locus risk allele, homozygotes for the third locus allele and non-carriers.

Incorporating breast cancer into the model

Like ovarian cancer, the baseline hazard functions and survival functions are obtained from constraining the national incidence rates.

Under the assumption that breast cancer incidence is independent of the polygenotype in these models, the baseline hazard functions $\lambda_1^b(t)$ and $\lambda_2^b(t)$, for the BRCA1 and BRCA2-mutation-positive populations respectively, are just the incidence rates $inc_1^b(t)$ and $inc_2^b(t)$.

The breast cancer incidence rates for the general population are constrained by the equation:

$$i^b(t) = \sum_{g,p_r} \Pr(g, p_r) f_{g}^b(t) S_{g,p_r}^o (t-1) = \sum_{g} \lambda_{g}^b(t) \sum_{p_r} \phi_{p_r} S_{g,p_r}^o (t-1)$$

$$= \sum_{g} \lambda_{g}^b(t) \left( \sum_{p_r} \phi_{p_r} S_{g,p_r}^o (t-1) \right)$$

Multiplying by the right-hand quotient and utilising information on the ovarian cancer survival functions gives the equation

$$i^b(t) \sum_{g} \tau_{g} S_{g}^b (t-1) \sum_{p_r} \phi_{p_r} S_{g,p_r}^o (t-1) e^{r_r} = \sum_{g} \lambda_{g}^b(t) \sum_{p_r} \phi_{p_r} S_{g,p_r}^o (t-1) e^{r_r}$$

$$= \lambda_1^b(t) \left( \tau_0 S_0^b (t-1) \sum_{p_r} \phi_{p_r} S_0^o (t-1) e^{r_r} + \sum_{k=3} \tau_k S_k^b (t-1) \sum_{p_r} \phi_{p_r} S_k^o (t-1) e^{r_r} \right) + \lambda_2^b(t) \tau_2 S_2^b (t-1) \sum_{p_r} \phi_{p_r} S_2^o (t-1) e^{r_r}$$
Rearranging this gives the following equation for the baseline hazard function:

\[
\lambda_0^b(t) = \frac{i^b(t) \sum_{k=0} \tau_k S_k^b(t-1) \sum_{p_r} \phi_{p_r} S_0^r(t-1)^{\phi_{p_r}} - \lambda_1^b(t) \sum_{p_r} \phi_{p_r} S_1^r(t-1)^{\phi_{p_r}} - \lambda_2^b(t) \sum_{p_r} \phi_{p_r} S_2^r(t-1)^{\phi_{p_r}}}{\tau_0 S_0^b(t-1) \sum_{p_r} \phi_{p_r} S_0^r(t-1)^{\phi_{p_r}} + \sum_{k=3} \tau_k S_k^b(t-1) \sum_{p_r} \phi_{p_r} S_k^r(t-1)^{\phi_{p_r}}}
\]
Incorporating SNPs into the risk prediction algorithm

Given that the risk $P(y^*_1 \mid y)$ of an individual developing ovarian cancer between ages $t_0$ and $t_f$ is equal to

$$
\frac{\sum_{i=t_0}^{t_f} P(y^*_i)}{P(y)}
$$

where $P(y^*_i)$ is the probability of the family phenotypes including the individual developing OvC at time point $i$, the probability $P(y^*_i \mid y, P_{K_i})$ of the same event conditional on the family genotype and the observed SNP genotypes, is equal to

$$
\frac{\sum_{i=t_0}^{t_f} P(y^*_i, P_{K_i})}{P(y, P_{K_i})} \frac{\sum_{i=t_0}^{t_f} P(y^*_i \mid P_{K_i}) P(P_{K_i})}{P(y \mid P_{K_i}) P(P_{K_i})}.
$$

This can then be rewritten in terms of the $n$th total polygenotype of the proband as

$$
\sum_{i=t_f}^{t_0} P(y^*_i, P_{K_i}) = \sum_{i=t_f}^{t_0} \sum_{n=0}^{2N+1} P(y^*_i \mid P_{K_i}) P(P_{K_i} \mid P^n_i) P(P^n_i).
$$

where $P(P_{K_i} \mid P^n_i)$ is the conditional normal density function given by

$$
P_{K_i} \mid P^n_i \sim N\left(P^n_i, \frac{\sigma^2_k}{\sigma^2_k + \sigma^2_U}, \frac{\sigma^2_k}{\sigma^2_k + \sigma^2_U}\right)
$$

$P(y^*_i \mid y)$ can also be written in terms of $P_i$ as

$$
\frac{\sum_{i=t_0}^{t_f} P(y^*_i)}{P(y)} = \frac{\sum_{i=t_0}^{t_f} \sum_{n=0}^{2N+1} P(y^*_i, P^n_i)}{\sum_{n=0}^{2N+1} P(y, P^n_i)}.
$$

Thus the probability of an individual developing OvC conditional on their observed SNP genotype is the ratio of two likelihood function sums, each term of which can be
computed as the corresponding term in the family history-conditional risk probability multiplied by a conditional normal density.

**Distribution of ovarian cancer risk and implications for ovarian cancer prevention.**

Based on our model, given a log-normal polygenic risk of $e^{\gamma_p}$ in the general population, where $y_p$ is predicted to follow a normal distribution with standard deviation $\sigma$ and mean $-\sigma^2 / 2$, rescaled so that the average risk $E(e^{\gamma_p})$ is equal to 1, it is easily established that the distribution of initial risk among individuals diagnosed with cancer is also log-normal with the log-risk $y_c \sim N\left(\frac{\sigma^2}{2}, \frac{\sigma^2}{2}\right)$ (see [1], methods). Computing the area under the two normal curves to the right of any given risk point gives us an estimate of the proportion of the population with risk greater than a given level and of the proportion of all cancer cases which will occur within this subgroup. Comparing these values gives a potentially informative measure of the relationship between risk distribution in the population and among cancer cases.

A measure of the predictive power of a 17-SNP genotype risk score could be informative as an indicator of how useful these SNPS are in combination for predicting OvC risk-distribution in the general population – a risk score that can be used to identify a high proportion of all cancers in a relatively low proportion of the population is very useful while one which would need almost half the population to be closely monitored to detect little more than 50% of cancers is almost as costly and ineffective as following 50% of the population at random. A comparison of the predictive power of the SNP risk with that of a total polygenic risk based on explicit family history could also give an indication of how much familial OvC still remains unaccounted for.
The combined log-effects of the seventeen SNPs were assumed to have a normal distribution with variance \( V = \sum V_i \) where \( V_i = \log \left( \frac{1 - p_i + p_i \exp(2r_i)}{(1 - p_i + p_i \exp(r_i))^2} \right) \) is the variance of the log-risk distribution from the \( i^{th} \) SNP, with frequency \( p_i \) and log-risk \( r_i \) [2] [3]. The proportions of the population and of cancer cases at different levels of SNP risk and polygenic risk were plotted against each other for comparison purposes.
Results: Sample statistics, pathology and genetic information.

Supplementary Table 1. Sample size, age and case distribution for the probands and their relatives

<table>
<thead>
<tr>
<th></th>
<th>Probands</th>
<th>Mothers</th>
<th>Sisters</th>
<th>Daughters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;=50</td>
<td>&gt;50</td>
<td>Proband&lt;=50</td>
<td>Proband&gt;50</td>
</tr>
<tr>
<td>Individuals</td>
<td>415</td>
<td>1133</td>
<td>356</td>
<td>984</td>
</tr>
<tr>
<td></td>
<td>1548</td>
<td>1340</td>
<td>1404</td>
<td></td>
</tr>
<tr>
<td>Families (no. with 1 or more)</td>
<td>415</td>
<td>1133</td>
<td>356</td>
<td>984</td>
</tr>
<tr>
<td></td>
<td>1548</td>
<td>1340</td>
<td>824</td>
<td></td>
</tr>
<tr>
<td>Mean age (SD)</td>
<td>43.9 (6.4)</td>
<td>60.4 (5.7)</td>
<td>69.9 (10.8)</td>
<td>74.5 (11.7)</td>
</tr>
<tr>
<td></td>
<td>55.9 (9.4)</td>
<td>73.3 (11.7)</td>
<td>57.2 (13.0)</td>
<td></td>
</tr>
<tr>
<td>No. ovarian cancers</td>
<td>415</td>
<td>1133</td>
<td>20</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>1548</td>
<td>51</td>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>
Supplementary table 2: SNPs and associated Odds Ratio estimates used in the construction of the Polygenic Risk Score.

<table>
<thead>
<tr>
<th>Locus</th>
<th>SNP</th>
<th>Minor allele frequency</th>
<th>Per-allele odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>3p157</td>
<td>rs15789171</td>
<td>0.049</td>
<td>1.45</td>
</tr>
<tr>
<td>9p16</td>
<td>rs3814113</td>
<td>0.32</td>
<td>0.83</td>
</tr>
<tr>
<td>8p129</td>
<td>rs1400482</td>
<td>0.13</td>
<td>0.85</td>
</tr>
<tr>
<td>19p17</td>
<td>rs4808075</td>
<td>0.30</td>
<td>1.12</td>
</tr>
<tr>
<td>17p40</td>
<td>rs62065444</td>
<td>0.18</td>
<td>1.15</td>
</tr>
<tr>
<td>17p43</td>
<td>rs10069690</td>
<td>0.26</td>
<td>1.09</td>
</tr>
<tr>
<td>5p1</td>
<td>rs2252894</td>
<td>0.66</td>
<td>0.89</td>
</tr>
<tr>
<td>2p176</td>
<td>rs12450786</td>
<td>0.26</td>
<td>1.13</td>
</tr>
<tr>
<td>8p82</td>
<td>rs74544416</td>
<td>0.067</td>
<td>1.20</td>
</tr>
<tr>
<td>17p33</td>
<td>rs3744763</td>
<td>0.59</td>
<td>0.94</td>
</tr>
<tr>
<td>10p22</td>
<td>rs12779865</td>
<td>0.34</td>
<td>1.09</td>
</tr>
<tr>
<td>1p36</td>
<td>rs56318008</td>
<td>0.15</td>
<td>1.11</td>
</tr>
<tr>
<td>1p34.3</td>
<td>rs58722170</td>
<td>0.23</td>
<td>1.08</td>
</tr>
<tr>
<td>4q26</td>
<td>rs17329882</td>
<td>0.24</td>
<td>1.09</td>
</tr>
<tr>
<td>6p22.1</td>
<td>rs1161331104</td>
<td>0.31</td>
<td>0.93</td>
</tr>
<tr>
<td>9q34.2</td>
<td>rs635634</td>
<td>0.19</td>
<td>1.11</td>
</tr>
<tr>
<td>17q11.2</td>
<td>chr17:29181220:i5</td>
<td>0.28</td>
<td>0.91</td>
</tr>
</tbody>
</table>
Supplementary Table 3: Predicted number of families with ovarian cancer under each model fitted

<table>
<thead>
<tr>
<th>Model</th>
<th>Only mother diagnosed with Ovarian cancer</th>
<th>Only 1 sister diagnosed with ovarian cancer</th>
<th>Mother and 1 sister diagnosed with ovarian cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed number of families</td>
<td>38</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>Base</td>
<td>19.9</td>
<td>7.98</td>
<td>0.39</td>
</tr>
<tr>
<td>Major Dominant</td>
<td>29.83</td>
<td>10.73</td>
<td>2.20</td>
</tr>
<tr>
<td>Major Recessive</td>
<td>26.76</td>
<td>18.90</td>
<td>1.24</td>
</tr>
<tr>
<td>Major General</td>
<td>29.87</td>
<td>10.74</td>
<td>2.20</td>
</tr>
<tr>
<td>Polygenic</td>
<td>36.27</td>
<td>14.54</td>
<td>1.35</td>
</tr>
<tr>
<td>Mixed Dominant</td>
<td>35.11</td>
<td>13.50</td>
<td>1.95</td>
</tr>
<tr>
<td>Mixed Recessive</td>
<td>33.94</td>
<td>17.86</td>
<td>1.55</td>
</tr>
<tr>
<td>Mixed General</td>
<td>35.61</td>
<td>13.64</td>
<td>2.04</td>
</tr>
</tbody>
</table>

*These assume no other cancers, breast or ovarian cancer in other family members (i.e., the three scenarios are mutually exclusive)
Supplementary figures

Supplementary Figure 1: Predicted risks of ovarian cancer over time to a *BRCA1*-carrier born in the 1940 birth cohort by family history.

Supplementary Figure 2: Predicted risks of ovarian cancer over time to a *BRCA2*-carrier born in the 1940 birth cohort by family history.

Supplementary Figure 3: Estimated ovarian cancer cumulative risk to a 50-year old female born in the 1940 birth cohort in the general population (family history information not considered), by PRS percentile.

Supplementary Figure 4: Estimated ovarian cancer cumulative risk to a 50-year old female born in the 1940 birth cohort with a mother diagnosed with ovarian cancer at 65, by PRS percentile.

Supplementary Figure 5: Estimated ovarian cancer cumulative risk to a 50-year old female born in the 1940 birth cohort with mother and sister diagnosed with ovarian cancer at 65 and 50, by PRS percentile.
Supplementary Figure 1

Risk of ovarian cancer since age 50

- No family history information
- Mother and sister cancer-free at 65 and 50
- Mother with ovarian cancer at 65
- Sister with ovarian cancer at 50
- Mother and sister with ovarian cancer at 65 and 50
- BOADICEA

Age
Supplementary Figure 5

- Pink line: 5th percentile of risk
- Blue line: Population level (50th percentile) risk
- Green line: 95th percentile of risk

Risk vs Age
References

