Incorporating tumour pathology information into breast cancer risk prediction algorithms

Nasim Mavaddat*1, Timothy R Rebbeck2, Sunil R Lakhani3, Douglas F Easton1 and Antonis C Antoniou1

Abstract

Introduction: Mutations in BRCA1 and BRCA2 confer high risks of breast cancer and ovarian cancer. The risk prediction algorithm BOADICEA (Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm) may be used to compute the probabilities of carrying mutations in BRCA1 and BRCA2 and help to target mutation screening. Tumours from BRCA1 and BRCA2 mutation carriers display distinctive pathological features that could be used to better discriminate between BRCA1 mutation carriers, BRCA2 mutation carriers and noncarriers. In particular, oestrogen receptor (ER)-negative status, triple-negative (TN) status, and expression of basal markers are predictive of BRCA1 mutation carrier status.

Methods: We extended BOADICEA by treating breast cancer subtypes as distinct disease end points. Age-specific expression of phenotypic markers in a series of tumours from 182 BRCA1 mutation carriers, 62 BRCA2 mutation carriers and 109 controls from the Breast Cancer Linkage Consortium, and over 300,000 tumours from the general population obtained from the Surveillance Epidemiology and End Results database, were used to calculate age-specific and genotype-specific incidences of each disease end point. The probability that an individual carries a BRCA1 or BRCA2 mutation given their family history and tumour marker status of family members was computed in sample pedigrees.

Results: The cumulative risk of ER-negative breast cancer by age 70 for BRCA1 mutation carriers was estimated to be 55% and the risk of ER-positive disease was 18%. The corresponding risks for BRCA2 mutation carriers were 21% and 44% for ER-negative and ER-positive disease, respectively. The predicted BRCA1 carrier probabilities among ER-positive breast cancer cases were less than 1% at all ages. For women diagnosed with breast cancer below age 50 years, these probabilities rose to more than 5% in ER-negative breast cancer, 7% in TN disease and 24% in TN breast cancer expressing both CK5/6 and CK14 cytokeratins. Large differences in mutation probabilities were observed by combining ER status and other informative markers with family history.

Conclusions: This approach combines both full pedigree and tumour subtype data to predict BRCA1/2 carrier probabilities. Prediction of BRCA1/2 carrier status, and hence selection of women for mutation screening, may be substantially improved by combining tumour pathology with family history of cancer.

Introduction

Genetic testing for BRCA1 and BRCA2 has important clinical implications: individuals found to carry mutations in these genes can be carefully monitored and receive preventive therapies including oophorectomy or mastectomy [1-6]. As genetic testing is expensive and may be associated with adverse psychological effects for the individual and their family, testing is only appropriate for those at highest risk of carrying mutations.

Several models have been developed for predicting risk of carrying BRCA1 or BRCA2 mutation and subsequently developing breast cancer [7]. These models generally include information on occurrence and age of diagnosis of breast cancer and other cancers in individuals and their families [7]. Breast tumours arising in carriers of BRCA1 or BRCA2 mutations, however, also differ from one another and from sporadic tumours in terms of their pathological characteristics, including those assessed morphologically or by immunohistochemistry [8-12]. Incorporating information about the pathology of breast
tumours in the proband or family members into risk-prediction algorithms may result in improved discrimination between BRCA1 mutation carriers, BRCA2 mutation carriers and nonmutation carriers, and provide a more accurate basis for identifying those individuals that may benefit from genetic testing.

Numerous studies have linked the absence of oestrogen receptor (ER) expression in breast tumours with BRCA1 mutation carrier status using different laboratory methods, anti-ER antibodies, and cut-off points for ER staining [13–22]. In the largest study to date, the Breast Cancer Linkage Consortium (BCLC) reported that ER-negativity defined by <1% of breast cancer cells expressing detectable ER by immunohistochemistry - conferred an age-adjusted odds ratio of 13.98 (95% confidence interval, 7.1 to 27.7; P < 0.0001) for BRCA1 mutation carrier status compared with controls unsel ected for family history (FH) [10]. Furthermore, the ER status was found to be the most significant risk factor in multiple regression analyses including other factors that are individually significant predictors of BRCA1 status. These factors included progesterone receptor, human epidermal growth factor receptor 2 (HER2), mitotic count, lymphocyte infiltration, and continuous pushing margins [10].

The triple-negative (TN) tumours are negative for ER, progesterone receptor, and HER2, and they define a subset of ER-negative disease. Subsets of TN tumours also express high-molecular-weight cytokeratins (CKs) (for example, CK5/6 and CK14), also referred to as basal CKs. Expression of basal CKs in TN tumours has been shown to provide additional information predictive of BRCA1 status [11]. The distribution of ER status in breast tumours and CK expression amongst TN tumours from BRCA2 mutation carriers appears to be similar to that in tumours from the population overall [10,11]. HER2-positive tumours have been shown less common in both BRCA1 and BRCA2 mutation carriers than in noncarriers [10].

We previously developed a risk-prediction algorithm for familial breast cancer and ovarian cancer - the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) - which can be used to compute the probabilities of carrying BRCA1 and BRCA2 mutations and the probabilities of developing breast cancer or ovarian cancer in the future [23]. BOADICEA models the effects of BRCA1 and BRCA2 mutations as well as a polygenic component representing the joint multiplicative effect of a large number of genes each of small effect [24,25]. BOADICEA has recently been extended to incorporate the associations between BRCA1 and BRCA2 mutations and other cancer risks (such as prostate cancer, pancreatic cancer and male breast cancer) [23], and has been shown to discriminate well between mutation carriers and noncarriers in a large series of families identified through UK genetic clinics [7,26]. In the present report we propose a method for incorporating breast tumour pathology information into the BOADICEA risk-prediction model.

Materials and methods
Age-specific distribution of ER status and distribution of other markers
Data from the BCLC [10,11] were used to obtain the age-specific proportions of ER-negative and ER-positive tumours, and the proportions of TN tumours and CK5/6-expressing and/or CK14-expressing tumours for BRCA1 or BRCA2 carriers or controls. Breast cancer patients with BRCA1 and BRCA2 mutations were identified through families with multiple relatives diagnosed with breast cancer or ovarian cancer. Data on 182 tumours in BRCA1 mutation carriers from 119 families, 62 tumours from BRCA2 mutation carriers from 35 families, and 109 control women with breast cancer who were unsel ected for FH and not tested for mutation carrier status were available.

Details of the methods used to determine tumour marker expression are described by Lakhani and colleagues [10,11]. Briefly, assays were performed at a single, quality-assured laboratory, and both the proportion of tumour cells staining with ER, progesterone receptor, HER2, CK5/6 or CK14 antibodies and the intensity of staining as assayed by immunohistochemistry were recorded [10,11]. The present work had been carried out under local ethical approval. Owing to the sparse data on BRCA2 tumours, and since the proportion of ER-positive tumours were not significantly different from that in controls [10], we assumed in the model that the age-specific proportions of ER-positive tumours and ER-negative tumours were the same as the proportions in the general population.

Age-specific proportions of ER-negative cancers and ER-positive cancers in the general population were obtained from the Surveillance Epidemiology, and End Results (SEER) database, a large population-based US breast cancer registry [27]. Data for the years 1990 to 2006 (n = 326,839) were used. Data were coded as ER-positive, ER-negative, or borderline. All analyses were restricted to females of white ethnicity with invasive breast cancer.

Statistical methods
Extending BOADICEA to incorporate tumour pathology
BOADICEA models the genetic susceptibility to breast cancer in terms of BRCA1, BRCA2 and a polygenic component representing the combined multiplicative effect of several loci, each of small effect [23].
The breast cancer incidence for individual \( i \) is assumed to depend on the underlying genotype through a model of the form:

\[
\lambda_i(t) = \lambda_0(t) \exp(G_i(t) + P_i(t))
\]

where \( \lambda_0(t) \) is the baseline incidence at age \( t \), \( G_i(t) \) is the log relative risk corresponding to the major genotype (that is, \( BRCA1 \) carrier, \( BRCA2 \) carrier or nonmutation carrier) at age \( t \), and \( P_i(t) \) is the age-dependent polygenic effect assumed to be normally distributed with mean zero and variance \( \sigma^2_i(t) \). The polygenic component is approximated by the hypergeometric polygenic model [28].

The incidence of ovarian cancer is modelled in a similar manner, but without a polygenic component.

The probabilities of developing breast cancer and ovarian cancer are assumed to be independent, conditional on the underlying genetic effects. Cancer incidences in BOADICEA are calendar and cohort specific. The overall age-specific incidences, averaged over all major genotypes and polygenotypes, are constrained to agree with the population incidences for England and Wales [29-35]. The BOADICEA model is implemented in the pedigree analysis program MENDEL v3.3 [36]. Additional model description and the model parameter estimates can be found in our previous publication [23].

To incorporate tumour phenotypes, we considered breast cancer subtypes as different disease end points. For example, to incorporate the ER status, the total observed breast cancer incidence was assumed to be:

\[
\lambda(t) = \mu(t) + \nu(t)
\]

where \( \mu(t) \) is the incidence of ER-positive disease and \( \nu(t) \) is the incidence of ER-negative disease. We also assumed that the probabilities of developing ER-negative or ER-positive breast cancer are independent, conditional on the underlying genotype. We assumed that the polygenic component \( P_i(t) \) was identical for both ER-negative and ER-positive disease (that is, represented by the same set of polygenes conferring the same relative risks), although in principle this assumption could be relaxed. The incidence for each individual \( i \) at age \( t \) therefore follows a model of the form:

\[
\lambda_i(t) = (\mu_i(t) + \nu_i(t)) \exp(P_i(t))
\]

where \( \mu_i(t) \) and \( \nu_i(t) \) are the major genotype-specific incidences for ER-positive and ER-negative disease, respectively, for individual \( i \).

Given the existing genotype-specific (\( BRCA1 \) carriers, \( BRCA2 \) carriers, and noncarriers) incidences in BOADICEA and the age-specific distribution of ER status in breast tumours from the BCLC and SEER data, our aim was to derive separate genotype-specific incidences for ER-positive disease and ER-negative disease. These were obtained by constraining the overall \( BRCA1 \) and \( BRCA2 \) incidences over ER status and polygenic effects to equal the average breast cancer incidence for \( BRCA1 \) and \( BRCA2 \) mutation carriers, estimated previously in BOADICEA; and by constraining the overall incidence over ER status, major gene (\( BRCA1 \) carriers, \( BRCA2 \) carriers and noncarriers) and polygenic effects to agree with the population breast cancer incidences. Separate genotype-specific incidences were derived for TN ER-negative disease and non-TN ER-negative disease as well as for CK5/6- and/or CK14-expressing and non-expressing TN tumours using a similar approach. Details of the method are provided in Additional file 1.

The derived incidences were used in the penetrance calculations whenever information on the relevant tumour marker status was available. For individuals with no marker status information, penetrance calculations use the total breast cancer incidences (as in the standard BOADICEA model implementation [23]). The genetic model is therefore fully specified by the \( BRCA1 \) and \( BRCA2 \) mutation frequencies, the polygenic variance and genotype-specific incidences for each type of disease.

### \( BRCA1 \) and \( BRCA2 \) mutation carrier probabilities

The probability that an individual carries a \( BRCA1 \) or \( BRCA2 \) mutation given FH and marker status (for example, ER status), was computed as follows:

\[
P(BRCAj / FH, ER) = P(BRCAj | FH, ER) / P(BRCAj | FH, ER) + P(BRCA2, FH, ER) + P(\neg \text{carrier}, FH, ER)
\]

where \( P(BRCAj, FH, ER) \) is the probability of observing the family with the particular FH and ER status, and the proband carrying a \( BRCA j \) mutation (\( BRCA j = BRCA1 \) carrier, \( BRCA2 \) carrier, or noncarrier). These correspond to pedigree likelihoods generated using the MENDEL program [36].

### Calculations for example pedigrees

We investigated the effect of incorporating tumour phenotypic information on the mutation carrier probabilities using a variety of different scenarios. Carrier probabilities were calculated in simple pedigrees for scenarios where the marker status of tumours is unknown, or is known in one or more members of a family. For simplicity, we assumed that there was no follow-up after the age of diagnosis of cancer in each case.

### Results

**ER status of tumours from mutation carriers and unselected breast cancers in the general population**

Age-specific proportions of ER-negative tumours in \( BRCA1 \) mutation carriers and \( BRCA2 \) mutation carriers derived from the BCLC dataset are presented in Table 1. These proportions were obtained for ER status defined as
positive where >1% of cells stained with anti-ER antibody in immunohistochemistry assays. Similar results were obtained when a combination of the proportion of cells staining and a staining intensity equivalent to an Allred score of 2 - a score shown to correspond to clinical response to Tamoxifen [37-39] - was used (data not shown). Data for age categories >70 years were sparse for BRCA1 carriers, and we extrapolated data from the age category 60 to 69 years for this group. Data for BRCA2 carriers were sparse in all age groups, resulting in jumps in proportions between adjacent intervals (Table 1). We therefore used population data [27] for BRCA2 carriers in our analyses, as BRCA2 carriers have been shown to have similar ER status distribution to unselected controls [10].

The age-specific proportions of ER-negative invasive breast cancers from white females in the general population obtained from the SEER database are shown in Figure 1. As previously reported, the proportion of ER-negative tumours decreases with age [40].

### Expression of other markers in tumours from mutation carriers and unselected controls

The proportions of TN tumours amongst ER-negative tumours and of CK-expressing tumours among TN tumours derived from the BCLC dataset are presented in Table 2. These proportions were obtained for progesterone receptor, CK5/6 or CK14 - defined as positive when >1% of cells stained with the specific antibodies in immunohistochemistry - and for HER2, where the majority of cells showed a strong complete membrane staining (equivalent to score 3 on the DAKO scoring system) [10,11]. Owing to sparse data, we used a constant proportion for expression of these tumour markers across all ages at diagnosis of breast cancer. As CK expression among TN tumours in BRCA2 carriers was similar to the distribution among TN tumours in breast cases unselected for FH and not tested for mutation (P > 0.05), we used data from these controls for BRCA2 carriers in our analyses.

### Breast cancer subtype-specific risks

ER-specific incidences were calculated for each of the five birth cohorts considered in BOADICEA. The resulting incidences (averaged over all polygenotypes) of ER-positive and ER-negative breast cancers in BRCA1 carriers, BRCA2 carriers and noncarriers for women born after 1950 are shown in Figure 2. The breast cancer incidence curve in BRCA1 mutation carriers follows closely the shape of the incidence curve for ER-negative disease (Figure 2a). This increases rapidly with age until about age 50 years but decreases gradually after this age. This pattern is similar to the incidence of ER-negative disease in noncarriers (Figure 2c). In contrast, the shape of the overall breast cancer incidence curve in BRCA2 mutation carriers is similar to that for ER-positive disease, where incidence increases with age, with an inflexion at about age 50 years (Figures 2b). This is also consistent with the age-specific pattern of ER-positive disease in noncarriers (Figure 2c) and in unselected controls [40].

Table 3 shows the corresponding average cumulative risks of developing ER-positive and ER-negative breast cancer by genotype. The probabilities of developing ER-positive and ER-negative breast cancer for BRCA1 mutation carriers by age 70 were calculated to be 18% and 55%, respectively. The corresponding risks for BRCA2 mutation carriers were 44% and 21% for ER-positive disease and ER-negative disease, respectively. Subtype-specific incidences for TN tumours and basal CK-expressing or non-expressing TN tumours, and the corresponding average cumulative risks of developing these tumours, are provided in Additional file 2 (Supplementary figures S1 to S6).

### Prediction of BRCA1 and BRCA2 carrier probabilities

Figure 4 shows the predicted BRCA1 and BRCA2 carrier probabilities for an individual, ignoring FH information. Compared with the risks when the ER status of the tumour is unknown, the BRCA1 carrier probabilities are higher at all ages when the tumour is known to be ER-negative, and are lower when the tumour is known to be ER-positive (Figure 4a). For example, for a breast cancer diagnosed at age 30, the carrier probability is estimated to be 0.05 when the ER status is unknown, 0.11 when the tumour is ER-negative and 0.01 when the tumour is ER-positive.
Knowledge of the TN status and basal marker expression further influences mutation carrier probabilities (Figure 4b to 4d). For example, for breast cancer diagnosed at age 30, the \emph{BRCA1} carrier probability is estimated to be 0.19 when the tumour is TN and 0.02 when the tumour is ER-negative but not TN (Figure 4b). If both basal markers CK5/6 and CK14 are expressed on a TN tumour, the probability of carrying a \emph{BRCA1} mutation is 0.64 (Figure 4c). If neither CK is expressed on a TN tumour when both have been tested for, however, the carrier probability is 0.04. Carrier probabilities for the scenario where a test is available only for CK5/6 are also shown (Figure 4c). Carrier probabilities for CK14-positive and CK14-negative tumours when a test is available only for CK14 were almost identical to those obtained when a test was available only for CK5/6 (data not shown).

Table 2: Proportion of each tumour subtype in \emph{BRCA1} and \emph{BRCA2} mutation carriers and unselected controls

| Tumour phenotype                                    | Controls\(^a\) | \emph{BRCA1} carriers | \emph{BRCA2} carriers |
|-----------------------------------------------------|-----------------|-----------------------|----------------------|
| TN (among ER-negative cases)                        | 0.53            | 0.90                  | 0.86                 |
| CK5/6-expressing and CK14-expressing TN\(^b, c\)    | 0.14            | 0.49                  | 0.38\(^d\)           |
| Either CK5/6-expressing or CK14-expressing TN\(^b, c\) | 0.24            | 0.30                  | 0.13\(^d\)           |
| CK5/6-expressing TN\(^b\) (only CK5/6 tested)       | 0.26            | 0.64                  | 0.38\(^d\)           |
| CK14-expressing TN\(^b\) (only CK14 tested)         | 0.27            | 0.63                  | 0.50                 |

\(^a\)Controls were breast cancer cases unselected for family history and untested for mutation. \(^b\)As a proportion of all TN tumours. \(^c\)Both the CK5/6 status and the CK14 status have been tested. \(^d\)As numbers were sparse, proportions from controls were used in updating the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm.

---

\emph{Figure 1 Proportion of oestrogen receptor-negative tumours from the general population.} Oestrogen receptor (ER)-negative tumours as a proportion of all invasive breast cancers tested for ER expression in unselected females from the general population. Data from the Surveillance Epidemiology, and End Results database 1990 to 2006 [27].
shown). The TN mutation status also influences \textit{BRCA2} mutation carrier probabilities (Figure 4d).

We also estimated carrier probabilities considering both tumour markers and FH. Table 3 and Additional file 2 (Supplementary tables S1 and S2) show mutation carrier probabilities in a simple family with an affected mother and an affected (proband) daughter. Knowledge of the ER status of either tumour substantially influences the mutation carrier probabilities. Notably, the later the age at breast cancer diagnosis, the greater the relative change in the \textit{BRCA1} mutation carrier probability. For example, if the mother is 70 years old when diagnosed with breast cancer, there is a threefold increase in mutation carrier probability when the tumour is ER-negative relative to when the ER status of the tumour is unknown. If the mother is 40 years old when diagnosed, the increase is less than twofold. In some cases, carrier mutation probabilities for \textit{BRCA2} are also altered when information on

---

Figure 2 Incidence rates of oestrogen receptor-negative disease, oestrogen receptor-positive disease and all breast cancers. Incidence rates per 100,000 individuals of oestrogen receptor (ER)-negative disease, ER-positive disease and all breast cancers in (a) \textit{BRCA1} carriers, (b) \textit{BRCA2} carriers and (c) noncarriers, corresponding to women born after 1950.

Figure 3 Average cumulative risk for oestrogen receptor-negative disease, oestrogen receptor-positive disease and all breast cancers. Average cumulative risk for oestrogen receptor (ER)-negative disease, ER-positive disease and all breast cancers in (a) \textit{BRCA1} carriers, (b) \textit{BRCA2} carriers, and (c) noncarriers, corresponding to women born after 1950.
Figure 5 shows mutation carrier probabilities for a hypothetical pedigree in which sequentially more pathological information is available for a family member. Mutation carrier probabilities are influenced by additional information. For example, if the tumour is TN, the probability of carrying a BRCA1 mutation is increased. If the tumour is ER-negative but not TN, however, the probability of carrying a BRCA1 mutation is even less relative to when the tumour is ER-negative but of unknown TN status.

Mutation screening for BRCA1 and BRCA2 is not 100% sensitive and can vary depending on the mutation screening methods used. BOADICEA takes into account the reduced sensitivity of mutation testing (assumed to be 70% for BRCA1 and 80% for BRCA2 for the purposes of the example [23]). Figure 6 shows a pedigree in which the proband had been tested for mutations in BRCA1 or BRCA2. As expected, the residual probability of carrying a BRCA1 mutation is markedly higher if the tumour in either the proband or the mother is known to be ER-negative.

Discussion
We extended the BOADICEA model of genetic susceptibility to breast cancer and ovarian cancer to incorporate breast tumour pathology information - in particular, ER...
### Table 3: Carrier mutation probabilities when the ER status of the mother is unknown, negative or positive

| Proband age at diagnosis | Mother’s age at diagnosis | 30 years | 40 years | 50 years | 60 years | 70 years |
|--------------------------|---------------------------|---------|---------|---------|---------|---------|
|                          | BRCA1 | BRCA2 | BRCA1 | BRCA2 | BRCA1 | BRCA2 | BRCA1 | BRCA2 |
| 30 years Unknown         | 0.45  | 0.16  | 0.24  | 0.14  | 0.15  | 0.10  | 0.09  | 0.09  | 0.06  | 0.10 |
| ER-negative              | 0.63  | 0.11  | 0.46  | 0.10  | 0.36  | 0.08  | 0.25  | 0.08  | 0.19  | 0.09 |
| ER-positive              | 0.13  | 0.25  | 0.08  | 0.17  | 0.04  | 0.11  | 0.04  | 0.10  | 0.03  | 0.10 |
| 40 years Unknown         | 0.23  | 0.14  | 0.11  | 0.10  | 0.07  | 0.07  | 0.04  | 0.06  | 0.03  | 0.07 |
| ER-negative              | 0.38  | 0.11  | 0.24  | 0.08  | 0.18  | 0.06  | 0.12  | 0.05  | 0.09  | 0.06 |
| ER-positive              | 0.05  | 0.16  | 0.03  | 0.10  | 0.02  | 0.07  | 0.02  | 0.06  | 0.01  | 0.07 |
| 50 years Unknown         | 0.12  | 0.08  | 0.06  | 0.06  | 0.04  | 0.04  | 0.02  | 0.04  | 0.02  | 0.04 |
| ER-negative              | 0.22  | 0.08  | 0.14  | 0.05  | 0.11  | 0.04  | 0.07  | 0.03  | 0.06  | 0.04 |
| ER-positive              | 0.02  | 0.09  | 0.01  | 0.06  | 0.01  | 0.04  | 0.01  | 0.04  | 0.01  | 0.04 |
| 60 years Unknown         | 0.06  | 0.07  | 0.03  | 0.04  | 0.02  | 0.03  | 0.01  | 0.03  | 0.01  | 0.04 |
| ER-negative              | 0.11  | 0.06  | 0.07  | 0.04  | 0.06  | 0.03  | 0.04  | 0.03  | 0.03  | 0.04 |
| ER-positive              | 0.01  | 0.07  | 0.01  | 0.05  | 0.00  | 0.03  | 0.00  | 0.03  | 0.00  | 0.04 |
| 70 years Unknown         | 0.03  | 0.07  | 0.01  | 0.05  | 0.01  | 0.03  | 0.01  | 0.03  | 0.00  | 0.04 |
| ER-negative              | 0.06  | 0.07  | 0.04  | 0.05  | 0.03  | 0.03  | 0.02  | 0.03  | 0.02  | 0.04 |
| ER-positive              | 0.01  | 0.07  | 0.00  | 0.05  | 0.00  | 0.03  | 0.00  | 0.03  | 0.00  | 0.04 |

Both mother and proband are affected and the age of diagnosis of cancer in each case is the same as the age of last follow-up. Figures in italics represent genotype-specific risk when the oestrogen (ER) status is unknown. Bold figures indicate genotype-specific risk <10% when the ER status is unknown, but >10% when the ER status is known.

status. We divided breast cancer into distinct disease end points and used data on the proportion of ER-negative and ER-positive tumours in BRCA1 and BRCA2 carriers and the general population to derive age-specific incidences of ER-negative and ER-positive disease. Information on the TN status and expression of basal markers on TN tumours was also included in the model. Incorporating information on tumour pathology influences the predicted probabilities of carrying a BRCA1 or BRCA2 mutation, in particular BRCA1, and may therefore have implications for genetic testing and clinical decision-making. The methodology presented here can in principle be applied to incorporate information in risk models for other diseases where the disease can be divided into distinct phenotypes such as different tumour sites in colorectal cancer [41,42] or subphenotypes of Crohn's disease [43].

Data on the distribution of ER status for BRCA1 or BRCA2 breast cancer tumours were obtained from a study conducted by the BCLC [10]. The ER status was measured by immunohistochemistry in a single research laboratory. A potential concern is that this assay may not be representative of typical assessment of ER status conducted in routine practice. The overall proportion of ER-positive tumours in the control series of the BCLC study, albeit relatively small, was however similar to that in the SEER series, suggesting there is little relative bias in the prevalence of ER-positivity. In addition, the cases in the BCLC study were predominantly drawn from multiple breast cancer families and this might have influenced the ER status, although there is no evidence to support an association between ER status and FH [44]. The number of mutation carriers was also relatively small (182 BRCA1 mutation carriers and 64 BRCA2 mutation carriers). The age-specific proportions of ER prevalence are therefore somewhat imprecise. In the final analyses we based the BRCA2 estimates on the proportion of ER-positive and ER-negative tumours in the general population data, due to lack of precision in the age-specific estimates in the BCLC data. Data from larger studies, such the Consortium of Investigators of Modifiers of BRCA1/2 [45], may provide more precise estimates and improve the modelling of BRCA2 in the future.
To derive ER disease-specific incidences for nonmutation carriers we used the general population data from the SEER database combined with the BCLC data on mutation carriers. Data on ER status in SEER were obtained from medical records. The ER status was recorded as either positive or negative without reference to how these definitions were derived. These data may, however, be more representative of clinical situations where laboratories differ in the sensitivity with which they measure ER expression and cut-off points and scoring systems for ER positivity vary. As age-specific data were used, variation in ER expression according to age at diagnosis in control tumours is captured in our model. The observed proportion of ER-negative tumours in BCLC data among unselected controls under age 70 were 33.5%. The corresponding expected number of ER-negative tumours in the SEER database would be 35.9%. Data on TN status and cytokeratin expression in the general population were not available in these larger datasets and were derived from unselected controls from the BCLC.

We derived genotype-specific incidences for developing different pathological subtypes of breast cancer using estimates of the genotypic relative risks, mutation frequencies in the general population, estimates of the polygenic distribution and the distributions of tumour pathology markers. For this purpose, we used estimates from published studies and registry data. Uncertainty surrounds each of these parameters, however, and a degree of measurement error may also be associated with the determination of pathological markers. Although in principle it may be possible to calculate confidence intervals for the derived disease-specific incidences or carrier probabilities, this would be complex due to the unknown correlation structure between the parameters entering the model. The derived incidences are therefore based on the best currently available estimates. The cumulative risks presented in Figure 3 do not take into account competing risks of dying from nonbreast cancer causes. These risks may therefore be somewhat higher than a woman would face in reality as they assume survival to the relevant age.

Our results show that incorporating tumour marker information into BOADICEA may result in the better discrimination between BRCA1 carriers, BRCA2 carriers and nonmutation carriers. As expected, individuals with ER-negative tumours, or with ER-negative tumours diag-

![Figure 5](image)

**Figure 5** Mutation carrier probabilities for a pedigree where sequentially more tumour pathology information is available. CK, cytokeratin; ER, oestrogen receptor; TN, triple negative.

| Tumour marker status          | BRCA1 | BRCA2 |
|------------------------------|-------|-------|
| Unknown                      | 0.04  | 0.05  |
| ER-negative                  | 0.11  | 0.05  |
| ER-positive                  | 0.01  | 0.05  |
| Triple negative              | 0.18  | 0.07  |
| ER-negative non-TN           | 0.03  | 0.02  |
| TN CK5/6 and CK14-positive†  | 0.47  | 0.05  |
| TN Either CK5/6 or CK14-positive† | 0.21  | 0.06  |
| TN CK-negative†              | 0.07  | 0.07  |

*Both mother and proband are affected and the age of diagnosis of cancer in each case is the same as the age of last follow-up. Marker status of the mother’s tumour are shown.
† Both CK 5/6 and CK14 have been tested.
Box indicates that genotype-specific risk is less than 10% when tumour marker status is unknown but higher than 10% where marker status is known.

**Figure 6** Influence of oestrogen receptor status on carrier probabilities when the proband tested negative for mutations. ER, oestrogen receptor.

| ER status            | BRCAL | BRCAL | BRCAL | BRCAL |
|----------------------|-------|-------|-------|-------|
| Unknown              | 0.16  | 0.21  | 0.07  | 0.06  |
| Proband ER-negative  | 0.35  | 0.17  | 0.17  | 0.05  |
| Proband ER-positive  | 0.04  | 0.24  | 0.01  | 0.06  |
| Unknown              | 0.16  | 0.21  | 0.07  | 0.06  |
| Mother ER-negative   | 0.38  | 0.16  | 0.18  | 0.05  |
| Mother ER-positive   | 0.05  | 0.24  | 0.02  | 0.06  |

* Sensitivity 70% for BRCAL and 80% for BRCAL mutations.
Box indicates that genotype-specific risk is less than 10% when ER status is unknown but higher than 10% where ER status is known.
nosed in their family members, are predicted to have a higher BRCA1 mutation carrier probability. The current National Institute for Clinical Excellence guidelines recommend that an affected family member is screened for mutation if the predicted carrier probability for mutations in BRCA1 or BRCA2 combined is at least 20% [46]. Knowledge of the ER status may therefore affect the decision to offer testing. For example, in a pedigree with an affected daughter and mother with ages of diagnosis at 40 and 50 years, respectively, the combined mutation carrier probabilities are 0.14 when the ER status is unknown but 0.24 if the mother’s tumour is ER-negative. Similarly, the presence of an ER-positive tumour in the family may result in combined BRCA1 and BRCA2 carrier probabilities <20%.

Information on the TN status and expression of basal cytokeratins can further influence mutation carrier probabilities. In our example, the BRCA1 mutation carrier probability is increased over fourfold when the mother’s tumour is ER-CK14−CK5/6+ as compared with when it is ER-negative, but information on basal CK is unavailable. For a random 30-year-old breast cancer patient, the BRCA1 mutation carrier probability is approximately 10 times greater for a patient with ER CK14−CK5/6+ versus ER CK− breast tumour. Lakhani and colleagues argued previously that the use of CK in combination with ER status may provide a more specific test for BRCA1 carrier status than ER alone, because ER-negative tumours are more frequently observed amongst control tumours than are basal CKs [11]. Our results further highlight the potentially important role of basal CKs in addition to ER status for risk prediction.

In the present model we assumed that the polygenic component is identical for both ER-negative and ER-positive disease and for tumours of other subtypes; that is, the polygenes confer the same relative risk by disease subtype. Studies that have evaluated the familial relative risks of breast cancer by the ER status of the proband have in general found no significant differences in the risks for ER-negative disease and ER-positive disease [47-54]. In addition, the segregation analysis of Antoniou and colleagues estimated similar polygenic variances of breast cancer risk for BRCA1 and BRCA2 mutation carriers [23]. Recent studies, however, have demonstrated that many of the common breast cancer susceptibility variants are associated more strongly with ER-positive disease [55-58]. Further, Antoniou and colleagues have shown that modifying loci associated with BRCA2 cancers, but not with BRCA1 carriers, parallel those associated with ER-positive disease in the general population [55,59]. The polygenotypes for ER-positive and ER-negative thus cannot be perfectly correlated, but the extent of the correlation is not known. This may affect risk prediction in circumstances where the tumour subtype is available for more than one individual in a family. In principle, the methods we presented can be extended to allow for different polygenic components on ER-negative and ER-positive disease or other tumour subtypes once these can be estimated.

**Conclusions**

We have developed a method for incorporating tumour marker information into risk-prediction models by subdividing the overall disease into different disease end points and have implemented this method in BOADICEA to incorporate tumour ER status, TN status, and expression of basal markers. This will be implemented in the BOADICEA web interface for use in genetic counselling. The inclusion of phenotypic markers associated with BRCA1 status should improve risk prediction in breast cancer.

**Additional material**

**Additional file 1** Extending BOADICEA to incorporate tumour pathology. Details of methods for extending BOADICEA to incorporate tumour pathology [60].

**Additional file 2** Subtype-specific risks and carrier probabilities. Supplementary figure S1 shows the incidence rate of ER-negative, TN disease, and ER-negative non-TN disease: incidence rate per 100,000 individuals in (a) BRCA1 carriers, (b) BRCA2 carriers and (c) noncarriers, corresponding to women before after 1950. Supplementary figure S2 shows average cumulative risk for ER-negative disease, TN disease and ER-negative non-TN disease: average cumulative risk for (A) BRCA1 carriers, (b) BRCA2 carriers and (c) noncarriers, corresponding to women born after 1950. Supplementary figure S4 shows the incidence rate of ER-negative, TN, CK-expressing and non-expressing TN disease (both CKs tested): incidence rate per 100,000 individuals in (a) BRCA1 carriers, (b) BRCA2 carriers and (c) noncarriers, corresponding to women born after 1950. Supplementary figure S5 shows the average cumulative risk for ER-negative, TN and CK-expressing and non-expressing disease (both CKs tested): average cumulative risk for (a) BRCA1 carriers, (b) BRCA2 carriers and (c) noncarriers, corresponding to women born after 1950. Supplementary figure S6 shows the average cumulative risk for ER-negative, TN, and CK5/6-expressing and non-expressing TN disease (only CK5/6 tested): average cumulative risk for (a) BRCA1 carriers, (b) BRCA2 carriers and (c) noncarriers, corresponding to women born after 1950. Supplementary table S1 shows carrier mutation probabilities when the ER status of the proband is unknown, negative or positive. Supplementary table S2 shows carrier mutation probabilities when the ER status of the proband and the mother are unknown, negative or positive.

**Abbreviations**

BCLC: Breast Cancer Linkage Consortium; BOADICEA: Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; CK: cytokeratin; ER: oestrogen receptor; FH: family history; HER2: human epidermal growth factor receptor 2; SEER: Surveillance Epidemiology; and End Results; TN: triple negative.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

NM, DFE and ACA contributed to the conception and execution of the project. NM carried out the analysis and wrote the manuscript. SRL was lead pathologist for the BCLC study. TRR contributed to discussions. ACA and DFE supervised the analysis and participated in writing the manuscript.
Acknowledgements

The authors thank the BCLC for pathology data. NM was funded by a scholarship from the Medical Research Council. ACA is a Cancer Research-UK Senior Cancer Research Fellow. DFE is a CR-UK Principal Research Fellow.

Author Details

1Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK, 2Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics and Epidemiology, University of Pennsylvania, 217 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, USA and 3Molecular and Cellular Pathology, University of Queensland Centre for Clinical Research, Level 6 Building 71/918, University of Queensland, The Royal Brisbane & Women’s Hospital, Herston, 4029 Brisbane, Queensland, Australia

Received: 22 January 2010 Revised: 4 May 2010
Accepted: 18 May 2010 Published: 18 May 2010

References

1. Freibiel TM, Domchek SM, Neuhouser SL, Wagner T, Evans DG, Isaacs C, Garber JE, Daly MB, Eilers R, Matloff E, Tomlison G, Lynch HT, Tung N, Blum JL, Weitl J, Rubinstein WS, Gang PA, Couch F, Rebbeck TR. Bilateral prophylactic oophorectomy and bilateral prophylactic mastectomy in a prospective cohort of unaffected BRCA1 and BRCA2 mutation carriers. J Clin Breast Cancer 2007, 3:785-882.

2. Hartmann LC, Sellers TA, Schaid DJ, Frank TS, Soderberg CL, Sitta DL, Frost MH, Grant CS, Donohue JH, Woods JE, McDonnell SK, Vockley CW, Deffenbaugh A, Couch FJ, Jenkins RB. Efficacy of bilateral prophylactic mastectomy in BRCA1 and BRCA2 gene mutation carriers. J Natl Cancer Inst 2001, 93:1633-1637.

3. Meijers-Heijboer H, van Geel B, van Putten WL, Henzen-Logmans SC, Bleeker EM, Blijham GV, Verbeek AL, van der Zwan PMS, Verhoef C, van der Velden JD, Klijn JG. Breast cancer after prophylactic bilateral mastectomy in women with a BRCA1 or BRCA2 mutation. Br J Surg 2001, 88:159-164.

4. Rebbeck TR, Lynch HT, Neuhouser SL, Narod SA, Van’t Veer L, Garber JE, Evans G, Isaacs C, Daly MB, Matloff E, Olopade OI, Weber BL. Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations. N Engl J Med 2002, 346:1616-1622.

5. Rebbeck TR, Freibiel T, Lynch HT, Neuhouser SL, van’t Veer L, Garber JE, Evans GR, Narod SA, Isaacs C, Matloff E, Daly MB, Olopade OI, Weber BL. Bilateral prophylactic mastectomy reduces breast cancer risk in BRCA1 and BRCA2 mutation carriers: a prospective study. J Clin Oncol 2004, 22:1055-1062.

6. Rebbeck TR, Kauf ND, Domchek SM. Meta-analysis of risk reduction estimates associated with risk-reducing salpingo-oophorectomy in BRCA1 or BRCA2 mutation carriers. J Natl Cancer Inst 2009, 101:80-87.

7. Antoniou AC, Easton DF. Risk prediction models for familial breast cancer. Future Oncol 2006, 2:527-574.

8. Lakhan SR, Jacquemier J, Slotine JP, Gusterson BA, Anderson TJ, Van De Vijver MJ, Fand LM, Venter D, Antoniou A, Storfer-Isser A, Smyth E, Steel CM, Haines N, Scott RJ, Goldgar D, Neuhouser S, Daly PA, Orrston W, McNamara R, Scherneck S, Ponder BA, Ford D, Petro J, Stoppa-Lyonnet D, Biggs JN, Stuweing JP, Spurt NK, Bishop OT, Klijn JG, Devilee P, et al. Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. J Natl Cancer Inst 1998, 90:1138-1145.

9. Lakhan SR. The pathology of familial breast cancer: morphological aspects. Breast Cancer Res 1999, 1:31-35.

10. Lakhan SR, Van De Vijver MJ, Jacquemier J, Anderson TJ, Ossin PP, McGuffog L, Easton DF. The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progestrone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2. J Clin Oncol 2002, 20:3110-2138.

11. Lakhan SR, Reis-Filho JS, Fulford L, Renaud-Luba F, van der Vijver M, Parry S, Bishop T, Benitez J, Rivas C, Bignon YJ, Chang-Chung JI, Hamann U, Cornelisse CJ, Devilee P, Bedmag MW, Nestle-Kramling C, Daly PA, Haines N, Varley J, Laloo F, Evans G, Maugard C, Meijers-Heijboer H, Klijn JG, Olah E, Gusterson BA, Piloti S, Radice P, Scherneck S, Sobo H, et al. Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. Clin Cancer Res 2005, 11:5175-5180.

12. Lynch BJ, Holden JA, Buys SS, Neuhouser SL, Gaffney DK. Pathobiologic characteristics of hereditary breast cancer. Hum Pathol 1998, 29:1140-1144.

13. Cortesi L, Turchetti D, Bertoni C, Belloi R, Mangione L, Vinceti M, Federico M, Siliardy V, Ferranti S. Comparison between genotype and phenotype identifies a high-risk population carrying BRCA1 mutations. Genes Chromosomes Cancer 2000, 27:130-135.

14. Amos CI, Trute L, White D, Southey MC, Hamm F, Tesoriero A, Hutchins AM, Dite GS, McCredie MR, Giles GG, Hopper JL, Venter DJ. Distinct molecular pathogeneses of early-onset breast cancers in BRCA1 and BRCA2 mutation carriers: a population-based study. Cancer Res 1999, 59:2011-2017.

15. Foulkes WD, Metcalfe K, Sun P, Hanna WM, Lynch HT, Ghadranian P, Tung N, Olopade OI, Weber BL, McLennan J, Olivatto IA, Begin LR, Narod SA. Estrogen receptor status in BRCA1- and BRCA2-related breast cancer: the influence of age, grade, and histological type. Clin Cancer Res 2004, 10:2029-2034.

16. Palacios J, Honorado E, Orosco A, Cazorla A, Soriano D, Barros A, Rodríguez S, Cipodosa JC, Diez O, Alonso C, Lema E, Sanchez L, Rivas C, Benitez J. Immunohistochemical characteristics defined by tissue microarray of hereditary breast cancer not attributable to BRCA1 or BRCA2 mutations: differences from breast carcinomas arising in BRCA1 and BRCA2 mutation carriers. Clin Cancer Res 2003, 9:3606-3614.

17. Karp SE, Tomin PN, Begin LR, Martinez JJ, Zhang JC, Pollack MH, Foulkes WD. Influence of BRCA1 mutations on nuclear grade and estrogen receptor status of breast carcinoma in Ashkenazi Jewish women. Cancer 1997, 80:435-441.

18. Liderreau R, Eisinger F, Champmme MH, Nogues C, Bieche I, Birnbaum D, Pallud C, Jacquemier J, Sobo H. Major improvement in the efficacy of BRCA1 mutation screening using morphoclinical features of breast cancer. Cancer Res 2000, 60:1206-1210.

19. Eisinger F, Nogues C, Guenabretrieve JM, Peyrat JP, Bardou VJ, Noguchi T, Vennin P, Sauvan R, Liderreau R, Birnbaum D, Jacquemier J, Sobo H. Novel indications for BRCA1 screening using individual clinical and morphological features. Int J Cancer 1999, 84:263-267.

20. Loman N, Johannsson O, Bendahl PO, Berg A, Ferno M, Olsson K. Steroid receptors in hereditary breast carcinomas associated with BRCA1 or BRCA2 mutations or unknown susceptibility genes. Cancer 1998, 83:310-319.

21. Ossin PP, Lakhan SR. The pathology of familial breast cancer: immunohistochemistry and molecular analysis. Breast Cancer Res 1999, 1:36-40.

22. Johannsson OT, Idevall I, Anderson C, Borg A, Barkaddettir RB, Eglisson V, Olsson H. Tumour biological features of BRCA1-induced breast and ovarian cancers. Eur J Cancer 1999, 33:362-371.

23. Antoniou AC, Cunningham AM, Petto J, Evans DG, Laloo F, Narod SA, Risch HA, Efayde JF, Joppe H, Southey MC, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubarinski J, Groonwald J, Gorski B, Tryggvadottir L, Syjakoski K, Kalionisem OP, Eleriha H, Nevanlinna H, Pharoah PO, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. Br J Cancer 2008, 99:1457-1466.

24. Antoniou AC, Pharoah PO, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancers. Br J Cancer 2004, 91:1580-1590.

25. Antoniou AC, Easton DF. Models of genetic susceptibility to breast cancer. Onco genes 2006, 25:5898-5905.

26. Antoniou AC, Hardy R, Walker L, Evans DG, Shenton A, Eilers R, Shanley S, Pichert G, Iazz L, Rose S, Douglas F, Eccles D, Morrison PJ, Scott J, Zimmern RL, Easton DF, Pharoah PD. Predicting the likelihood of carrying a BRCA1 or BRCA2 mutation: validation of BOADICEA, BRACAPRO, IBIS, Myriad and the Manchester scoring system using data from UK genetics clinics. J Med Genet 2009, 46:425-431.

27. Surveillance, Epidemiology, and End Results (SEER) Program Limited-Use Data (1973-2006) [http://seer.cancer.gov]

28. Lange K. An approximate model of polygenic inheritance. Genetics 1997, 147:1423-1430.

29. Doll R, Payne P, Waterhouse A. J Cancer Incidence in Five Continents Volume I. Berlin: Springer-Verlag; 1966.
30. Doll R, Muir CS, Waterhouse AJ: Cancer Incidence in Five Continents Volume II. Geneva: IARC, 1970.

31. Muir C, Waterhouse AH, Mack TM, Powell J, Whelan S: Cancer Incidence in Five Continents Volume IV. Lyon: IARC Scientific Publications, 1987.

32. Parkin DM, Muir CS, Whelan S, Gao YT, Ferlay J, Powell J: Cancer Incidence in Five Continents Volume VI. Lyon: IARC Scientific Publications, 1992.

33. Parkin DM, Whelan S, Ferlay J, Teppo L, Thomas DB: Cancer Incidence in Five Continents Volume VIII. Lyon: IARC Scientific Publications, 2002.

34. Waterhouse AH, Muir C, Sharmagratnam K, Powell J: Cancer Incidence in Five Continents Volume VIII. Lyon: IARC Scientific Publications, 1982.

35. Waterhouse AH, Muir C, Correa P, Powell J: Cancer Incidence in Five Continents Volume VIII. Lyon: IARC Scientific Publications, 1976.

36. Lange K, Weeks D, Boehnike M: Programs for pedigree analysis: MENDEL, FISHER, and dGENE. Genet Epidemiol 1988, 5:471-472.

37. Allred DC, Harvey JM, Berardo M, Clark GM: Prognostic and predictive factors in breast cancer by immunohistochemical analysis. Mod Pathol 1998, 11:155–168.

38. Gown AM. Current issues in ER and HER2 testing by IHC in breast cancer. Mod Pathol 2008, 21(Suppl 2):S58–S51.

39. Harvey JM, Clark GM, Osborne CK, Alred DC: Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol 1999, 17:1474–1481.

40. Anderson WF, Chatterjee N, Ershler WB, Brawley OW: Risk factors for breast cancer by immunohistochemical analysis. Int J Cancer 2007, 120:535–542.

41. Gervaz P, Bouzourene H, Cerottini JP, Chaubert P, Benhattar J, Secic M, Chapman JM, Onnie CM, Prescott NJ, Fisher SA, Mansfield JC, Mathew CG, Waterhouse JAH, Muir C, Shanmugaratnam K, Powell J: Cancer Incidence in Five Continents Volume VIII. Lyon: IARC Scientific Publications, 2002.

42. Magnusson C: Breast cancer risk factors according to joint estrogen receptor and progesterone receptor status. Cancer Detect Prev 2000, 24:248–258.

43. Rosenberg LU, Einarsdottir K, Friman EI, Wedren S, Dickman PW, Hall P, Magnusson C: Risk factors for hormone receptor-defined breast cancer in postmenopausal women. Cancer Epidemiol Biomarkers Prev 2006, 15:2482-2488.

51. Rosenberg LU, Einarsdottir K, Friman E, Wedren S, Dickman PW, Hall P, Magnusson C: Risk factors for hormone receptor-defined breast cancer in postmenopausal women. Cancer Epidemiol Biomarkers Prev 2006, 15:2482-2488.