Breast Cancer Risks and Risk Prediction Models

Christoph Engel\textsuperscript{a}  Christine Fischer\textsuperscript{b}

\textsuperscript{a}Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Germany
\textsuperscript{b}Institute of Human Genetics, University of Heidelberg, Germany

Introduction

Women with pathogenic mutations in the \textit{BRCA1} and \textit{BRCA2} genes have a considerably increased risk to develop breast and ovarian cancer during their lifetime compared to the general population. The condition is characterized by incomplete penetrance, i.e. not all mutation carriers will develop cancer during their lifetime. Moreover, for those carriers who will develop cancer, the age of onset shows large interindividual heterogeneity between families and even within families. Families with \textit{BRCA1/2} mutations are characterized by multiple affected relatives and/or cases of early cancer onset. However, not in all families with a suspect cancer history can a \textit{BRCA1/2} mutation be found. Nonetheless, individuals from such \textit{BRCA1/2}-negative families are also considered to have an elevated cancer risk.

Clinical management strategies currently comprise intensified surveillance for early breast cancer detection using multimodal imaging techniques, or prophylactic surgery such as bilateral mastectomy or salpingo-oophorectomy. Genetic counseling of carriers and other at-risk individuals and personalized clinical decision-making highly depend on knowledge, as precise as possible, on the individual age-dependent cancer risks.

Since the discovery of the \textit{BRCA1} and \textit{BRCA2} genes in the middle of the 1990s, the breast and ovarian cancer risks for mutation carriers have been investigated in numerous studies [1–23]. The results from these studies serve as a major source of information for risk communication and tailored clinical decision-making. In addition, a variety of risk prediction models were developed in the past years in order to determine personalized cancer risks based on individual features such as familial cancer history, genetic test results, and also non-genetic risk factors. These models are particularly helpful for clinical decision-making in those families without a known pathogenic mutation despite strong evidence of a hereditary background.

In this paper, we aim to give an overview of the present knowledge on empirical breast and ovarian cancer risks that were investigated in epidemiological studies. Moreover, we present and describe different risk prediction models currently used for individual risk assessment in clinical practice.

Keywords
Breast cancer · Ovarian cancer · Hereditary cancer risk · Risk prediction models

Summary
\textit{BRCA1/2} mutation carriers have a considerably increased risk to develop breast and ovarian cancer. The personalized clinical management of carriers and other at-risk individuals depends on precise knowledge of the cancer risks. In this report, we give an overview of the present literature on empirical cancer risks, and we describe risk prediction models that are currently used for individual risk assessment in clinical practice. Cancer risks show large variability between studies. Breast cancer risks are at 40–87\% for \textit{BRCA1} mutation carriers and 18–88\% for \textit{BRCA2} mutation carriers. For ovarian cancer, the risk estimates are in the range of 22–65\% for \textit{BRCA1} and 10–35\% for \textit{BRCA2}. The contralateral breast cancer risk is high (10-year risk after first cancer 27\% for \textit{BRCA1} and 19\% for \textit{BRCA2}). Risk prediction models have been proposed to provide more individualized risk prediction, using additional knowledge on family history, mode of inheritance of major genes, and other genetic and non-genetic risk factors. User-friendly software tools have been developed that serve as basis for decision-making in family counseling units. In conclusion, further assessment of cancer risks and model validation is needed, ideally based on prospective cohort studies. To obtain such data, clinical management of carriers and other at-risk individuals should always be accompanied by standardized scientific documentation.
Cancer Risks of BRCA1/2 Mutation Carriers

It has been demonstrated that the cancer risks in BRCA1/2 mutation carriers depend on the family history and also on the type of mutation [1, 13, 23–25]. Early epidemiological studies found breast cancer risks of up to 87% for BRCA1/2 mutation carriers, whereas ovarian cancer risks were reported to be 64% for BRCA1 and 27% for BRCA2 mutation carriers [5, 12, 26]. However, these cancer risk estimates were based on selected families with multiple cases and thus tended to overestimate the risks for unselected carriers or those with a less pronounced family history. Other studies that were based on less stringently selected families showed lower risks. Taken all studies together, the breast cancer penetrance is in the range of 40–87% for BRCA1 mutation carriers and 18–88% for BRCA2 mutation carriers. For ovarian cancer, the risk estimates are between 22–65% for BRCA1 and 10–35% for BRCA2. It is important to recognize that these studies are heterogeneous with regard to size, mode of ascertainment (population based vs. high-risk families), and the statistical methods of risk estimation. Consequently, there is considerable variation in the reported risks. In order to provide a comprehensive set of cancer risk estimates for BRCA mutations carriers that can be used by counselors and clinicians, Chen and Parmigiani [27] conducted a meta-analysis of 10 different studies based on strict selection criteria regarding study design and appropriate methodology. They report meta-analytic mean cumulative breast cancer risks at age 70 of 57% for BRCA1 and 49% for BRCA2 mutation carriers (table 1). The ovarian cancer risks were 40% for BRCA1 and 18% for BRCA2 mutation carriers. Lifetime risks are a convenient measure to quantify and compare disease risks in different populations. However, in daily clinical practice and at the individual level, it is more appropriate to communicate cumulative cancer risks for a comprehensible time interval, e.g. 10 years, starting from the current age of the yet unaffected counselee. Therefore, in their meta-analysis, Chen and Parmigiani [27] also provided a tabulation of the remaining cancer risks. As an example, table 2 shows the 10-year risks for breast and ovarian cancer for currently unaffected mutation carriers, depending on the current age of the counselee.

It is well known that BRCA1/2 mutation carriers are also at an elevated risk of developing contralateral breast cancer (CBC) after a previous unilateral breast cancer. To date, a variety of studies have evaluated the risks of CBC, most of which are retrospective in design [10, 22, 28–43]. A recent meta-analysis revealed CBC risks of 15% for BRCA1 mutation carriers and 9% for BRCA2 mutation carriers 5 years after the first breast cancer [44]. The 10-year risks were even higher with 27% and 19% for BRCA1 and BRCA2 mutation carriers, respectively (table 1). Moreover, a retrospective data analysis of the German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC) has shown that younger age at first breast cancer was associated with a higher risk of CBC [28, 40].

Notably, most of the aforementioned studies are retrospective in design and require appropriate methods to adjust for the non-random selection of the study populations. The inherent problem of ascertainment bias can be overcome by conducting prospective cohort studies. However, prospective studies are comparatively rare and limited in size and observation time. Recently, a larger prospective analysis of 1,887 BRCA1/2 mutation carriers from the EMBRACE (Epidemiological Study of Familial Breast Cancer) cohort study has been published [45]. The average cumulative breast cancer risks by age 70 for BRCA1 mutation carriers were estimated to be 60% for breast cancer and 59% for ovarian cancer. For BRCA2 mutation carriers, the risks were 55% for breast cancer and 17% for ovarian cancer.

As mentioned above, the study design and population ascertainment may explain some of the observed cancer risk variation. However, there is growing evidence that the increased cancer risk of BRCA1/2 mutation carriers can be modified by other genetic and non-genetic factors. In the past years, the international Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA) has identified a number of single-nucleotide polymorphisms (SNPs) that are associated with breast cancer risk [46–53]. Although the association of each single SNP is small, their combination may have a considerable effect on cancer risks, with potential impact on clinical decision-making [50]. First results from the prospective EMBRACE study indicate that individual profiles of such genetic low-risk variants may allow clinically relevant risk stratification [45]. As regards non-genetic factors, a number of exposures and lifestyle factors such as reproductive history, hormone intake, smoking, and alcohol consumption have been considered to modify the cancer risk in BRCA mutation carriers. However, the available evidence is still limited due the small sample sizes of the studies conducted so far [54].

Since the early 1990s, risk prediction models have been developed and extended to provide more personalized risk predictions. Most of the models are based on the empirical breast and ovarian cancer risk data from the aforementioned studies. Additionally,

| Table 1. Cumulative cancer risks from meta-analyses of penetrance studies [27, 44] |
|---------------------------------|------------------|------------------|
|                                  | BRCA1, %  | BRCA2, %  |
| Breast cancer (until age 70)    | 57 (47–66) | 49 (40–57) |
| Ovarian cancer (until age 70)   | 40 (35–46) | 18 (13–23) |
| CBC (within 10 years after first breast cancer) | 27 (21–33) | 19 (15–23) |

| Table 2. The 10-year cumulative breast and ovarian cancer risks for currently unaffected BRCA mutation carriers by current age (adapted from [27]) |
|-----------------|-----------------|-----------------|
| Current age, years | Breast cancer, % | Ovarian cancer, % |
|                 | BRCA1 | BRCA2 | BRCA1 | BRCA2 |
| 20               | 1.8 (1.4–2.2) | 1 (0.78–1.4) | 1 (0.68–1.8) | 0.19 (0.09–0.47) |
| 30               | 10 (8.2–13) | 6.6 (5.1–8.6) | 2.2 (1.6–3.4) | 0.52 (0.28–1) |
| 40               | 20 (16–25) | 15 (12–19) | 6.7 (5.2–8.9) | 1.9 (1.2–3.2) |
| 50               | 22 (18–27) | 18 (15–22) | 15 (12–17) | 5.2 (3.7–7.2) |
| 60               | 19 (15–24) | 17 (14–20) | 22 (20–23) | 9.8 (7.8–11) |
Breast Cancer Risk Prediction Models

Breast and ovarian cancer risk prediction models have been developed to determine individual cancer risks for arbitrary women. Typically, clinicians apply these models to women from families with multiple affected family members and to early-onset breast and ovarian cancer patients. Table 3 shows the most commonly applied risk prediction models with their main features.

The Gail model uses current age, age at menarche, age at first life birth, number of previous breast biopsies, history of atypical hyperplasia, ethnicity, and number of affected first-degree female relatives for risk prediction [55]. No assumptions about genetic traits are made; ovarian cancer cases and extended family history are not taken into account. The Gail model is also not suitable for predicting cancer risks in BRCA1/2 mutation carriers.

Other risk calculation models assume an underlying genetic model for breast cancer susceptibility and can fully utilize individual pedigree patterns. These so-called genetic risk prediction models include breast and ovarian cancer status and age of onset as well as the current age of healthy relatives of arbitrary degree. The model parameters are obtained from large epidemiological studies on age- and gene-specific empirical breast and ovarian cancer risks for mutation carriers as well as for the general population.

The earliest model, the Claus model, assumes 1 autosomal dominant gene with age-dependent penetrance [56]. It was developed before the BRCA1/2 genes had been identified and was later extended (eCLAUS model as implemented in the commercial pedigree drawing software CYRILLIC) to include ovarian cancer cases [57]. The Claus model does not take the BRCA1/2 status into account.

BRCAPRO is based on a model with the 2 genes BRCA1 and BRCA2, and it can use BRCA1/2 screening results and family histories of BRCA1/2-associated cancer types. BRCAPRO can account for multiple ethnicities, adjusts for mastectomies among relatives and, in its newest version, provides updated estimates of CBC penetrance [58].

BOADICEA (breast and ovarian analysis of disease incidence and carrier estimation algorithm) assumes that genetic susceptibility to breast cancer is due to BRCA1 and BRCA2 mutations and an additional polygenic component. It can use the mutation screening result, molecular tumor characteristics like estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and basal cytokeratin (CK) expression (CK5/6 and CK14), and family histories of BRCA1/2-associated cancers [59, 60].

IBIS assumes effects of BRCA1 and BRCA2 and models the residual genetic variability in terms of a third dominantly inherited common gene conferring moderate risks. IBIS can account for BRCA1/2 screening results and family histories of BRCA1/2-associated cancer types and family histories of BRCA1/2-associated cancers [59, 60].

As a prerequisite for clinical use, breast cancer risk prediction models need to be externally validated and compared in prospective data sets. Only few such studies exist so far and they compare selected risk models based on different cohorts. Typically, calibration and discrimination of the models are investigated. Calibration compares how many breast cancer cases would have been expected by model prediction and how many cases have actually been observed within a defined time period, e.g. 10 years, starting from the current age of the counselee (observed-to-expected (O/E) ratio). Discrimination accuracy is a measure that describes how good the model can distinguish between women who are affected or not affected by breast cancer. A common descriptive measure for the discrimina-

Table 3. Commonly used risk prediction models

<table>
<thead>
<tr>
<th>Model</th>
<th>Type of risk prediction</th>
<th>Risk factors considered in the model</th>
<th>Web link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Remaining lifetime risk</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heterozygote risk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gail</td>
<td>BC</td>
<td>none</td>
<td><a href="http://www.cancer.gov/bcrisktool">www.cancer.gov/bcrisktool</a></td>
</tr>
<tr>
<td>eCLAUS</td>
<td>BC</td>
<td>yes</td>
<td><a href="http://www.cyrilliansoftware.com">www.cyrilliansoftware.com</a></td>
</tr>
<tr>
<td>BRCAPRO</td>
<td>BC, OC</td>
<td>BRCA1/2</td>
<td><a href="http://bcb.dfci.harvard.edu/bayesmendel/brcapro.php">http://bcb.dfci.harvard.edu/bayesmendel/brcapro.php</a></td>
</tr>
<tr>
<td>IBIS</td>
<td>BC, OC</td>
<td>BRCA1/2 + single moderately penetrant gene</td>
<td><a href="http://www.ems-trials.org/riskevaluator">www.ems-trials.org/riskevaluator</a></td>
</tr>
<tr>
<td>BOADICEA</td>
<td>BC, OC</td>
<td>BRCA1/2 + polygenic component</td>
<td><a href="http://icgc.medschl.cam.ac.uk/boadicea/boadiceaweb-application">http://icgc.medschl.cam.ac.uk/boadicea/boadiceaweb-application</a></td>
</tr>
</tbody>
</table>

BC = Breast cancer, OC = ovarian cancer.
tive performance is the area under the curve (AUC) of the receiver operating characteristic (ROC). As a rule of thumb, clinical decision models should have an AUC of at least 0.7. Amir et al. [61] published a validation study based on 3,150 women comparing the models Gail, Claus, BRCAPRO, and IBIS. IBIS performed best, whereas the other models underestimated the breast cancer risk. The AUC varied between 0.72 and 0.76. Recently, Quante et al. [62] compared the Gail model with IBIS in a cohort of 1,857 women having an average breast cancer risk 3 times higher than in the general population. Regarding the 10-year breast cancer rates, IBIS showed better calibration and discrimination; the AUC was 0.7 in comparison to 0.63 for the GAIL model. In a small cohort of 358 mostly Ashkenazi high-risk women, both IBIS and BOADICEA overestimated the breast cancer risk. However, BOADICEA (O/E ratio 0.80, 95% confidence interval (CI) 0.54–0.93) was better calibrated than IBIS (O/E ratio 0.52, 95% CI 0.32–0.87) [63]. In a large Australian cohort of 4,176 women with European ancestry, BOADICEA was well calibrated (O/E ratio 0.92, 95% CI 0.76–1.10) and showed good discrimination (AUC 0.7) [64]. In summary, non-genetic risk factors are included in Gail and IBIS; only IBIS and BOADICEA include BRCA1/2 and other genetic factors into the model.

The CBC risk can be calculated with eCLAUS, BOADICEA, and BRCAPRO. eCLAUS and BOADICEA make the simplifying assumption that the first and the second breast cancer are independently conditioned on BRCA1/2 genotypes. However, evidence of a strong dependency of first and second breast cancer has recently been provided [28, 40, 65]. Thus, these risk prediction models are currently not suitable for CBC prediction and should not be used for this purpose until they have been adjusted and validated. A BRCA PRO upgrade has recently been released that explicitly accounts for the dependence between first and second breast cancer [58].

**Discussion**

Precise knowledge of individual breast and ovarian cancer risks is a major prerequisite for adequate counseling of at-risk individuals, in order to allow tailored decision-making regarding different clinical management options. Empirical data on the age-dependent risks obtained from epidemiological studies are not only important for risk communication. Such data is also important for risk calculation models as these models require valid cumulative breast cancer risks for BRCA1/2 mutation carriers and for the general population.

The current international literature reveals large heterogeneity in the reported cancer risks for BRCA1/2 mutation carriers between different studies. This might be partly due to different study designs and methods, but it may also reflect different characteristics of the study populations regarding other risk-modifying factors. Thus, the question arises as to which extent risk estimates from specific populations can be applied to other populations for risk prediction purposes. Moreover, this question of external validity must also be asked in the case of risk prediction models, since these are based on penetration functions that were derived from specific populations. It therefore seems to be advisable to obtain risk estimates in the same populations and under the same conditions in which they are later used for risk prediction. Ideally, such risk estimation studies should be prospective in design, as retrospective data analyses are prone to different sources of bias, adjustment for which is often difficult and requires sophisticated statistical methods. In order to facilitate such data collections, individuals at an elevated risk of hereditary breast and ovarian cancer should be prospectively followed up in well-designed registries. In Germany, the GC-HBOC has established a large registry providing an important source of such data for high-risk families.

In contrast to empiric cancer risk estimates, risk prediction models offer the possibility to account for family history, known genes and their mode of inheritance, and several other genetic and non-genetic factors that modify the individual cancer risk. They are important tools to identify women at high risk of developing breast cancer and can be used to offer tailored prevention and clinical management. Work is still ongoing to improve these models by including additional risk factors, such as lifestyle factors or the large number of recently identified low-risk variants. This also includes calibration of model parameters to the specific target populations. Moreover, further prospective comparative validation studies are needed to evaluate whether the individual risk estimation is sufficiently precise for clinical implementation.

Typically, risk prediction models are used by choosing a cut-off value for the definition of ‘high-risk’ persons, who are then offered participation in intensified surveillance programs. Although several of the risk prediction models show good discrimination with AUCs > 0.7, the consequences of a specific clinical decision based on a specific cut-off have to be carefully considered by evaluating the predictive values. For example, if a 3% cut-off for the BOADICEA 10-year breast cancer risk were used to offer intensified surveillance in the Australian cohort as described in [64], the corresponding predictive values would be approximately 5% for the positive predictive value and approximately 99% for the negative predictive value. The cumulative 10-year breast cancer rate in the total cohort was approximately 2.8%. As a consequence, about 95% of all women under intensified surveillance would not get breast cancer in the next 10 years, and about 5% of the women may profit from early detection.

Therefore, we recommend communicating and discussing the predictive values and the clinical consequences of risk prediction using specific cut-off values. It might be a good strategy to use risk prediction models in selected subgroups with a higher 10-year breast cancer rate and therefore better predictive values. So far, the number of observed breast cancer cases in available cohorts is still too small to provide sufficient statistical power for the required subgroup analyses. Therefore, longer follow-up is needed in future validation studies.

**Disclosure Statement**

The authors have no conflicts of interests.
References


