Dempster-Shafer Theory Based Uncertainty Models for Assessing Hereditary, BRCA1/2-Related Cancer Risk

E. Auer*1 and W. Luther2

1University of Applied Sciences Wismar, Germany
2University of Duisburg-Essen, Germany

Abstract

We investigate possibilities for modeling uncertainty in the context of assessing hereditary, BRCA1/2-related cancer risk. For that purpose, we overview and evaluate studies with large groups of participants of various ethnicities with standardized data selection policy. Risk classes for deleterious gene mutations are described there by mean values with confidence intervals or percent intervals of lower and upper bounds of risk. It is difficult to extrapolate the results obtained for the selected cohort if, for example, the families in the sample already have an increased basic risk or if risks of individual persons and their family are mixed. Often, intervals obtained from the studies which metastudies rely on are summarized without calibration of the average mutation risk, leading to dramatic overestimations and inconsistencies. In this paper, we demonstrate ways in which study results can be interpreted consistently, used for assessing personal/familial risk, and compared with the help of the Dempster-Shafer theory and ground truth (empirical) data.

Keywords: interval analysis, Dempster-Shafer theory, BRCA1/2-related risk assessment and genetic counseling

1. Introduction

Cancer is one of the deadliest diseases of our times and occurs due to mutations in cells. Mutations (now known as pathogenic variants) of two kinds can be discerned: somatic and germline. The somatic kind accounts for the most of the cancer cases. It cannot be inherited since only tissues originating directly from the gene damaged cell are affected. The germline kind is responsible for 5% to 10% of all cancer cases and is hereditary because it influences reproductive cells (e.g., a sperm cell). Despite their relative infrequency, germline mutations are important since knowledge about them allows doctors to advise whole affected families on suitable cancer detection, prevention and treatment strategies.

Breast cancer (BC) is the most common type of cancer among women worldwide (second in both sexes combined as of 2018 [1]). One of the important achievements in cancer control was the discovery of BRCA1/2 tumor suppressor genes in the 1990s. Not all mutations in BRCA1/2 are harmful. Although mutations in BRCA1/2 genes account for only 5% to 10% of all BC cases1, the cancer risk for women already carrying a harmful gene mutation increases dramatically. Pathogenic variants in BRCA1 are considered to lead to over 50% to 65% risk of hereditary BC and to 40% to 55% risk for BRCA2 [2] as compared to about 12% risk of developing BC during a lifetime in the general female population. Additionally, the likelihood for ovarian cancer (OC) increases. Mutations in BRCA1/2 genes must not necessarily be of germline kind, but the germline mutations account for nearly 50% of all hereditary BC [3].

Since there is a strong correlation between the hereditary BC and OC syndrome (HBOC) and BRCA1/2 mutations, a possible strategy for identifying persons at high risk is to take a look at whether they carry a pathogenic variant in these genes. Modern genetic tests reliably identify BRCA1/2 mutations but are not necessarily helpful for everyone. To identify persons at risk of harmful mutations either in order to provide more specific suggestions concerning individual prevention/risk mitigation or to assess the possible success of certain treatment therapies, the use of familial risk assessment tools2 is an essential preliminary step that helps to decide whether a genetic test is appropriate for a patient. Such tools implement models for computing the likelihood of BRCA1/2 mutation (IBRCAm). Although they could be employed at a non-specialized doctor’s office or in the framework of a general website, it is difficult to gauge from the outcome alone if the patient should be referred for the genetic testing or if they need a certain prevention therapy. Already from the empirical numbers about the connection between cancer and mutation risks, one can guess that the uncertainty in predicting a mutation is large.

To simplify the process of identifying persons at risk of a harmful BRCA1/2 mutation (and, therefore, a higher risk of HBOC), a number of questionnaire-type approaches were developed3 that make inquiries about certain HBOC indicators and assign scores to the answers. Researchers agree (cf. voice.ons.org) that key indicators of hereditary risk are, among others,

- cancer occurring at a young age, for example, BC at age 50 or younger (e.g., less than 40), denoted by BC50− and BC40−, respectively, in the following;
- rare cancers, for example, ovarian cancer (e.g., OC50−);
- more than one primary cancer in one person (e.g., BC50−, OC50− together, BCOC50) or bilateral cancer4, for example, BC in both breasts (nBC);
- individuals having certain family histories (FH) of cancer (e.g., cancer in a near relative CNR);

---

1©2022 ISRERM Organizers. Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5184-1_GS-03-079-cd

2For example, bayesmendel.dfci.harvard.edu/risk/
3For example, FHAT [4], MSS [5], RST [6]
4bilateral and multifocal cancers are often combined
individuals from families with a known mutation associated with increased cancer risk.

Additionally, the likelihood of carrying a $BRCA1/2$ mutation is higher for certain ethnicities. In Section 2.2, we describe further indicators (risk factors) which are important for gauging the mutation probability, based on the overview of relevant publications in Section 2.1.

Data on the indicators play a crucial role in assigning a person to a risk class (without a gene test). A simple definition of a risk class can be given as a group of people or a cohort of families having a common $BRCA1/2$ mutation probability depending on a certain combination of indicators. In Section 3, we use the data on indicators from relevant publications to compute a guess for the IBRCAm, which can serve as a basis for assigning a person or a family to their risk category. In Section 3.2, we develop Dempster-Shafer theory (DST) based models for assessing the IBRCAm for various disease patterns. In Section 3.3, we offer a scheme for comparing the results on the IBRCAm from different studies in a meaningful manner with the help of modeled cumulative IBRCAm curves. The basic concepts and references on the DST are in Section 3.1.

In a recent paper [7], we presented interval strategies for quantifying uncertainty for questionnaire-type approaches (e.g., FHAT). However, the question of how good the data on which these methods are based on remains. Now, we aim to provide an (at least partial) answer to this question and other remaining questions from [7] by considering current medical studies and metastudies with large groups of participants from different regions of the world with well-defined data selection policies. Out of the available publications, we choose those that provide criteria for selecting a test cohort using a predefined (a priori) HBOC risk class. The common approach in these studies is that the prevalence of pathogenic variants is predicted depending on disease patterns of probands and their relatives, their ethnicity, age of cancer onset and other factors. The data on patients and their relatives are usually obtained from medical, non-public databases or from publicly available, internationally accessible databases of papers (cf. Section 2.3) and are analyzed statistically. Depending on the cohort selected for the study, this approach can result in considerably different IBRCAm and, consequently, different assessments for the posterior risk classes to which the patients or their family members are assigned. It is difficult to interpret the results if the included families already have an increased basic risk (e.g., as in [8], [9]) or if the risks of individual persons and their families are mixed [9]. That is, the average numbers of $BRCA1/2$ mutations for the individual clinical pictures in a family, as well as the average value for the percentage in the cohort, actually depend on the cohort composition.

In the context of $BRCA1/2$-related cancer research, it is widely spread to combine the results of individual studies and evaluate them. Many such metastudies suffer from the lack of comprehensible requirements, quality criteria, and metrics for the personal data used in the underlying studies, for the proband selection, collection of FH data, for the used model-based familial risk assessment algorithms, choice of persons for genetic testing, and interpretation of results. In particular, the epistemic uncertainty in the data and risk models is seldom accounted for in the proper way. In the studies, risk classes for deleterious gene mutations are described by confidence intervals or percent intervals of lower and upper bounds of risk. In metastudies, such intervals are often summarized without calibration of the average mutation risk, which leads to dramatic overestimations and inconsistencies, even if individual and familial risks are distinguished properly. One of our goals is to improve this situation, which we suggest to approach as described in Section 3.

2. State of the Art: Overview

In the following, we analyze several relevant studies and metastudies that compute the IBRCAm (excluding our previous paper [7]). In Subsection 2.1.1, the chosen papers are described in short, including their strengths and shortcomings. In Subsection 2.2, we discuss what factors are relevant for estimating the posterior individual or familial mutation risk. In Subsection 2.3, the available types of databases for HBOC information are briefly commented on. Finally, the most important points are summarized in Subsection 2.4.

2.1. Relevant IBRCAm Studies

In this subsection, we provide a short overview of the chronologically ordered studies that are relevant for the models developed in Section 3 and point out their possible advantages or shortcomings from the point of view of the task in this paper. In all studies, all (or some of the) probands underwent genetic testing for $BRCA1/2$ mutations. In all studies, there is the explicit or implicit possibility of a bias due to the fact that the chosen probands initially belonged to a high risk class.

A relatively early paper [10] provides predictions for mutations in $BRCA1/2$ correlated with such risk factors as age of diagnosis, personal and FH, and ethnicity (compiled in tables, denoted Frank tables in the following). The cohort of overall 10000 participants is divided into two groups according to Ashkenazi Jewish (AJ) ethnicity and all other ethnicities (non-AJ). The participants underwent sequence analysis of $BRCA1$ followed by analysis of $BRCA2$ for those of them who agreed to it. For the non-AJ group, 4716 persons out of the tested 6724 provided enough information about their personal/FH to be able to establish a meaningful connection to IBRCAm (2233 out of 3022 in the AJ case). The median first age of onset (FAO) for 4663 BC patients is recorded as 44; 53 years of age for 779 patients with OC. Based on the results of this survey, the authors identified risk factors for $BRCA1/2$ mutations and correlated them with rates for developing BC, BC with subsequent OC, and contralateral BC. This allowed them to model the IBRCAm with the help of logistic regression analysis. However, popular questionnaire-type risk assessment tools are difficult to assess with data exclusively from the Frank tables because they have very small numbers of probands for certain disease constellations. Additionally, the postmenopausal BC ($BC_{51+}$, e.g., in the family history) is often not considered explicitly although used somehow inside the tables.

In [11], the main focus is on determining the mutation prevalence in dependence on ethnicity. From 1996-2006,
the authors studied BRCA1/2 mutations in 46276 women of varying ancestry with a high risk of cancer. The data were taken from the repository of BRCA1/2 testing maintained by Myriad Genetic Laboratories (USA). The reported IBRCAm is 12.5% overall. More information about the result of this study is in Section 2.2, where we discuss the factor of ethnicity in more detail. The reported limitations of the study, apart from the selection bias resulting from choosing women with the high risk, are a bias due to inconsistencies in the ethnicity specification and in cancer history recall.

The study [9] seems to be the most comprehensive one to date on the topic of deriving (posterior) risk classes for gene variants that have a strong connection with the development of HBOC. The authors consider families with a history of BC or OC (i.e., with clustering or early onset of BC or OC) and aim to characterize the corresponding prevalence of BRCA1/2 mutations in them. As the basis, extensive data from over 21 thousand families were collected over almost 20 years for individual family members. If available, the family member with the most severe phenotype underwent genetic testing (called there the index patient). The cohort included 617578 participants overall, 1896 with bBC and OC, 62 females with both BC and OC, 14 males with BC, and 147 individuals without BC or OC who underwent genetic testing. The authors admit that a major limitation of their study, apart from the relatively small proband group, is the lack of information about the degree of kinship among the tested groups or families, it is obvious that the majority of the studies considers participants from a certain preliminary risk group defined by disease patterns which vary from study to study. It is not surprising that the risk of a germline mutation in a group of randomly selected individuals is lower and different from that of individuals with an elevated risk. Although we also use the results from other studies, we do not rely on the same approach as in [12] and combine them without at least trying to calibrate the data first.

A common feature in [8], [9], [11] is that only families having certain disease patterns are included in the studies, which makes it possible that people from only one group of cancer are tested. The results of the studies are difficult to compare because index patients and family members are not always strictly separated and the number of tested family members is often higher than the number of patients (cf. [8]). A further problem is that information on the distribution in dependence on FAO is often given only as the mean (the usual case) or the median as in [10] over
all BC/OC patients. However, it is necessary to differentiate finer within the age groups (20y-30y, 30y-40y, etc.), especially if the mean is considered; sometimes such risk factors as BC\textsubscript{40-} are not precise enough for predicting the IBRCA\textsubscript{m}. To compare the mutation probabilities, IBRCA\textsubscript{m} frequency distributions in dependence on FAO for the patients and their family members are necessary.

2.2. Relevant Risk Factors for the IBRCA\textsubscript{m}

The references mentioned above identify additional important HBOC risk factors as founder effects/ethnicity (F/E) and presence of a triple-negative BC subtype tumor (TNBC). In the following, we consider several risk factors in more detail. Additionally, an overview of the cohort composition for the studies [8]–[11] wrt. the considered risk indicators is given in Table 1. Note that the pool of risk factors shown in the Table and the set of criteria for selecting probands in the cohort are actually not the same concepts. Therefore, in the last column of the Table, we show our assessment of the (a priori) risk class(es) to which the cohort in the respective study belongs based on its selection criteria (disease patterns). It is difficult to use the metastudy [12]: Although the aspects from Table 1 were considered in the study, the actual data are difficult to obtain without repeating the whole research process and, therefore, cannot be calibrated.

2.2.1. Age of Onset

Early age of cancer onset (mostly, BC\textsubscript{40-}) is reported as one of the main indicators for a \textit{BRCA1/2} mutation throughout all of the studies. An interesting finding in [8] is that the early BC onset in Japanese people is associated with a \textit{BRCA1/2} pathogenic variant independently of any family history of BC or OC. In [9], the authors report that \textit{BRCA1/2} mutations are found in 13.7\% of cases in families with a single case of very early BC (<36 years). Cases of unilateral or bilateral BC\textsubscript{51+} alone do not increase the probability of mutation detection. However, it is clear from [13] that it is necessary to differentiate more finely according to the age groups for the FAO.

2.2.2. Influence of Ethnicity; Founder Effects

A considerably higher \textit{BRCA1/2} mutation risk is observed within the AJ population, at least in two different countries: Israel [14] and the USA [11]. In [14], it is found that 11\% of BC and 40\% of OC are caused by certain harmful mutations in \textit{BRCA1/2} genes. Overall, 12\% of the AJ population of Israel carry one of these mutations. In the USA, 2.5\% of AJ population carry a harmful \textit{BRCA1/2} mutation, the tenfold risk as compared to the general population [11]. This fact is widely accepted and has been taken into account in such early studies as [10]. In Japanese population, a recent study [8] found that 11.2\% of cases had pathogenic \textit{BRCA 1} variants and 9.0\% \textit{BRCA 2} pathogenic variants. A conclusion from [11] is, however, that the risk is similar across diverse ethnicities in the USA if one excludes the Ashkenazi ethnicity.

According to the dictionary of the National Cancer Institute, ‘founder effect is a genetic alteration observed with high frequency in a group that is/was geographically or culturally isolated, in which one or more of the ancestors was a carrier of the altered gene’. Most cancer predisposition genes are considered heterogeneous, but certain populations display a tendency to founder mutations. A literature overview is given, for example, in [15] for the Latin American group. The \textit{BRCA1/2} variation characteristics are distinctive also in the AJ population and can be limited to three predominant specific founder mutations.

2.2.3. Family History

Generally, FH is considered to be important throughout the studies aside from a small number of exceptions (cf. early age of onset). In [9], it is reported that highest mutation frequencies are observed in families with at least two OCs (41.9\%). Further disease patterns in a family corresponding to a high frequency are at least one BC and one OC (41.6\%) and mBC with at least one female BC or OC (35.8\%). Occurrence of nBC or BC\textsubscript{50-} and OC\textsubscript{50-} in the same woman (BCOC\textsubscript{SP}) leads to higher mutation frequencies compared with the occurrence of two primary cancers in different individuals (49.0\% vs. 31.5\%).

2.2.4. TNBC

It is reported in [8] that TNBC is usually associated with a \textit{BRCA1} mutation (whereas persons carrying a \textit{BRCA2} mutation develop estrogen-positive BC). From [13] (Table 2), it can be seen that the IBRCA\textsubscript{m} for BC in the age group 25–50y is in the interval [0.08, 0.18]. However, if only TNBC is considered, the likelihood becomes higher and is in the interval [0.12, 0.28].

Harmful gene variants need to be considered separately from the variants of unknown significance (VUS). In the following, if we provide numbers or rates of positively tested persons, we mean those with a \textit{BRCA 1} and/or a \textit{BRCA 2} variant including typical AJ founder mutations; we do not differentiate further wrt. specific sequencing and VUS.

2.3. Standardized Databases

Access to modern cancer databases is essential to ensure that larger numbers of cases are available for research in order to validate mathematical models [16]. Databases might have various purposes and are maintained by national and international research institutions (e.g., universities and hospitals), companies and foundations. They help to understand and optimize the development, detection, spread, treatment and prevention of diseases as well as to support people in dealing with the disease. The main types of database purposes are:

- To register cancer in the population depending on different parameters, such as gender, age, place of residence, professions and disease occurrence (German Consortium for HBOC database\textsuperscript{3} [9], [17], Japanese HBOC Consortium database [8], not public);
- To typify cancer and treatment methods (e.g., a clinical database supported by Myriad Genetic Laboratories, Inc. used in [11], not public);
- To classify pathogenic gene variants of various rel-

\textsuperscript{3}www.konsortium-familiaerer-brustkrebs.de
Table 1. Cohort composition in relation to risk factors. A checkmark in the column denotes that the factor was considered in the study. For the BC40−, any specific age earlier than 40 is included. If multiple ethnicities are considered, we provide numbers for one group only, non-Ashkenazi for [10] (from Table 1.2 therein), Western European for [11]. ‘Overall’ means the total number of families tested for BRCA1/2 mutations (equal to the number of index patients for some studies) and can be smaller than the total number of persons. The abbreviation ‘av’ means average.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>BC50−</th>
<th>BC51+</th>
<th>nBC</th>
<th>mBC</th>
<th>OC</th>
<th>BCOC/SP</th>
<th>no C</th>
<th>TNBC</th>
<th>F/E</th>
<th>FH</th>
<th>BC40−</th>
<th>Overall</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>[10]</td>
<td>5303</td>
<td>661</td>
<td>−</td>
<td>76</td>
<td>1260</td>
<td>803 (135)</td>
<td>229</td>
<td>−</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>4716</td>
<td>av</td>
</tr>
<tr>
<td>[9]</td>
<td>26236</td>
<td>24452</td>
<td>6028</td>
<td>−</td>
<td>671</td>
<td>7250 (−1917)</td>
<td>−</td>
<td>−</td>
<td>✓</td>
<td>✓</td>
<td>21401</td>
<td>av</td>
<td></td>
</tr>
<tr>
<td>[8]</td>
<td>2054</td>
<td>−</td>
<td>14</td>
<td>89</td>
<td>−</td>
<td>(62)</td>
<td>147</td>
<td>✓</td>
<td>−</td>
<td>−</td>
<td>✓</td>
<td>2366</td>
<td>high</td>
</tr>
</tbody>
</table>

evance and VUS (e.g., BRCAexchange⁶, ClinVar⁷, GSEA database⁸);
− To provide medical evidence in form of publications on the subject (e.g., PubMed database⁹ used in [12]).

Note that the information we need for the models in this paper is not contained in publicly available databases directly and must be retrieved from the relevant publications.

2.4. Summary

The studies discussed in this paper that determine the prevalence of germline mutations differ substantially in how the probands for the cohorts were selected. The cohorts used in the studies typically include individuals from the high risk category (cf. Table 1). The selection criteria are based on disease patterns (combinations of risk factors) in the patients or their 1st/2nd degree maternal/paternal relatives. That is, the average risk for a deleterious gene variation depends on the selection criteria and disease patterns of the cohort and can range from 2.6% to 27.9% [8]. If we were to merge all the cited results, we would be dealing with cohorts consisting of people from six continents with BC and OC at any age, but also with such subtypes as TNBC, with cumulative and early onset, with bilateral or male BC, and OC at any age, but also with such subtypes as TNBC, with cumulative and early onset, with bilateral or male BC, as well as with prostate, colon, and pancreatic cancers.

That is, it is important to choose studies with clearly described cohorts of large sizes. The included (index) patients and their family members should be classified wrt.
− such risk factors as BC51+, BC50−, BC40−, nBC, mBC, OC50−;
− at least median/mean first age of onset in age groups of 5-10 years;
− standardized selection criteria (disease patterns), the a priori risk class;
− origin/ethnicity of the patient (e.g., the youngest family member with BC), FH including the degree of relationship and the first occurrence of the disease;
− detected tumor subtypes (e.g., TNBC); and, finally,
− the quality and the trustworthiness of the data (e.g., use of public databases with a disclosed search strategy).

(We limit our considerations to these factors and do not take into account, for example, hormone-related kinds of BC.) In this way, it would be possible to justify the use of the findings for certain patient groups to obtain a reliable assignment to a risk category based on the IBRCAm and to validate the results of scoring tools, for example, those newly presented or described in [7].

3. Dempster-Shafer Models

Our focus is on computing a consistent IBRCAm as a compromise of personal and family risk. For this purpose, we propose to use the DST. The main problem with using this approach is to find the ground truth since average mutation probability varies strongly depending on the cohort composition (9.9% to 27.9% [8], [11]). We estimate the IBRCAm for different disease patterns based on tables from [10] for AJ and non-AJ population in Subsection 3.2. In Subsection 3.3, we rely on the IBRCAm values modeled using Penn IIⁱ⁰ in dependence on FAO and the results for age groups given in [13] for comparing the final IBRCAm results as provided by [9]–[11].

3.1. Basic Concepts

With the help of the Dempster-Shafer theory [18], it is possible to combine evidence from different sources (e.g., experts) and to provide a measure of confidence that a certain event occurs. A classical probability measure defines the probability that a random variable X is equal to its certain realization xᵢ (real number, point value). The DST allows us to assign a probability to the event that a realization of X belongs to a given set (e.g., [xᵢ, xᵢ])ⁱ¹. The result is given in terms of the lower and upper bounds (belief and plausibility) on the probability of a subset of a sample space Ω. A random DST variable can be characterized by its basic probability assignment (BPA) m. If A₁, ..., Aₙ are the sets of interest where each Aᵢ ∈ 2Ω, then

\[ m : 2^\Omega \rightarrow [0,1], \quad m(Aᵢ) = pᵢ, \quad i = 1 \ldots n, \]

\[ m(\emptyset) = 0, \quad \sum_{i=1}^{n} m(Aᵢ) = 1. \quad (1) \]

The mass of the impossible event Θ is equal to zero, with every Aᵢ with m(Aᵢ) ≠ 0 called a focal element. The sum of masses of focal elements should be equal to one. This condition might lead to the necessity to normalize real life evidence because experts tend to provide BPAs for which

⁶brcaxchange.org/variants
⁸www.gsea-msigdb.org/gsea/index.jsp
⁹pubmed.ncbi.nlm.nih.gov/
¹⁰https://pennmodel2.pmacs.upenn.edu/penn2/
¹¹A similar interpretation is possible for continuous random variables
it does not hold. The plausibility ("worst case") and belief ("best case") functions can be defined with the help of the BPAs for all \( i = 1 \ldots n \) and \( Y \subseteq \Omega \) as

\[
P(Y) := \sum_{A_i : Y \neq \emptyset} m(A_i), \quad Bel(Y) := \sum_{A_i \subseteq Y} m(A_i).\tag{2}
\]

If two or more experts provide different estimations in the same area, the BPAs have to be aggregated. There exist several methods for this purpose [18], of which Dempster’s rule is used in this paper:

\[
m_{12}(A_k) = \frac{\sum_{\forall A_i, i \neq k} m_1(A_i) m_2(A_k)}{1 - \sum_{\forall A_i, i \neq k} m_1(A_i) m_2(A_k)}\tag{3}
\]

with \( A_k \neq \emptyset, m_1(\emptyset) = 0 \).

3.2. The DST Model for the IBRCAm

3.2.1. Non-Ashkenazi Ethnicity

In [7], we used Dempster’s rule to combine data on mutation probabilities in BRCA1/2 correlated with personal and family history of cancer based on tables from [10] for non-Ashkenazi population (cf. Table 2). Let \( \Omega \) contain \( n = 9 \) distinct elements corresponding to relevant risk factors as described in Section 2.2: \( f_1 = b_1 (\text{BC} 51), f_2 = b_2 (\text{BC} 50), f_3 = a_1 (\text{OC at any age}), f_4 = a_2 (\text{OC} 40), f_5 = b a (\text{an additive factor for } b_2 \text{ to account for } \text{BC} 40), f_6 = o a (\text{an addition to } a_1 \text{ to account for premature } \text{OC} 40), f_7 = n r (\text{non-Ashkenazi population (cf. Table 2)}), f_8 = b i l \text{ (nBC)}, f_9 = b m \text{ (mBC)}. \)

Further important risk factors can be discerned: two BC cases \( b_1, b_2 \), two OC cases \( o_1, o_2 \), and other combinations mentioned in Column 1 of Table 2. The mass of the subsets of \( \Omega \) not mentioned in Table 2 is supposed to be zero. The BPA \( m_1 \) in Column 2 is inspired by proband’s mutation probabilities (personal risk from [10]) and BPA \( m_2 \) in Column 3 by those of her family members (family risk [10]). With Dempster’s rule from Eq. (3), we can combine the BPAs for \( m_1 \) and \( m_2 \) as shown in Column 4, whereas the belief function values obtained by applying the definition in Eq. (2) for the combined BPA \( m_{12} \) are given in Column 5. For reference, the corresponding combined probabilities from the same Frank tables are given in Column 6 if available. The model allows us to compute belief values for further disease patterns, for example, those shown below.

Pattern 1: Assume the FH of a father with BC at 42 years of age and a second degree relative with BC51. Then a lower (best case) estimation of the IBRCAm corresponding to this FH can be computed as \( Bel_{m_{12}}(bm, nr, b_1, b_2) = m_{12}(b_1, b_2) + m_{12}(b_1) + m_{12}(b_2) + m_{12}(nr) = 0.257 \). The family risk (for a patient with this FH) as provided by the Penn II model is the IBRCAm 27%.

Pattern 2: Suppose that the patient has premature BC (BC40–), her mother BC30–, then a lower estimation of the IBRCAm is \( Bel_{m_{12}}(b a, nr, b_2) = m_{12}(b_2) + m_{12}(ba) + m_{12}(nr) = 0.198 \). The corresponding value from [10] is 29.7%, from Penn II 15%–35%.

Pattern 3: Suppose the patient has OC40– and her aunt BC50–, then a lower estimate of the IBRCAm is \( Bel_{m_{12}}(o_1, b_2, o a, nr) = m_{12}(b_2, o_1) + m_{12}(o_1) + m_{12}(b_2) + m_{12}(oa) + m_{12}(nr) = 0.358 \) (Penn II: 12%–40% family risk).

Pattern 4: Suppose the patient is aged 22 and is diagnosed OC and BC; her mother has bilateral BC at over 50 years of age. The lower estimation for the IBRCAm is \( Bel_{m_{12}}(b_2, o_1, b_1, nr, o a, ba, bil) = m_{12}(o_1) + m_{12}(b_1) + m_{12}(b_2) + m_{12}(nr) + m_{12}(bil) + m_{12}(b_1, o_1) + m_{12}(b_2, o_1) + m_{12}(oa) + m_{12}(ba) = 0.096 + 0.116 + 0.039 + 0.026 + 0.053 + 0.098 + 0.023 + 0.020 + 0.056 = 0.527 \) compared to 54% from Penn II (family risk).

Pattern 4, variant: We can capture different IBRCAm for age groups within BC50–, also by using intervals for ba. Based on data from [10], we can choose the interval [0.036,0.076] for the IBRCAm associated with this risk factor. Then the result for the same pattern as in the previous example would be the IBRCAm in [0.507,0.547].

3.2.2. Ashkenazi Ethnicity

The second half of Table 2 (Column 8–12) shows the data on the IBRCAm similarly to those in Section 3.2.1 but for the AJ ethnic group. Again, we might want to use intervals to better capture the IBRCAm according to the age group within BC40–: \( m_{12}(ba) = [0.069, 0.207] \) according to [10]. Examples are:

Pattern 1: Proband with BC40– (between 30 and 40 years), aunt with BC51+: \( Bel_{m_{12}}(ba, nr, b_1, b_2) = m_{12}(b_1, b_2) + m_{12}(b_1) + m_{12}(b_2) + m_{12}(oa) + m_{12}(nr) = 0.024 + 0.057 + 0.147 + 0.069 + 0.022 = 0.319 \) as compared to 31.8% from [10] and 24%–35% from Penn II.

Pattern 2: Proband with BC50– (at age 45y), sister with OC50–: \( Bel_{m_{12}}(oa, nr, b_2) = m_{12}(b_2) + m_{12}(oa) + m_{12}(nr) = 0.147 + 0.144 + 0.022 = 0.313 \) as compared to 41.5% from [10] and 31% from Penn II.

Pattern 3: Proband with OC50– and OC and BC51+, in the FH: \( Bel_{m_{12}}(o_1, o_2, b_1, nr) = m_{12}(o_1, o_2) + m_{12}(b_1, o_1) + m_{12}(o_2, b_1) + m_{12}(o_2, o_1) + m_{12}(nr) = 0.417 \) as compared to 41.2% from [10] and 41% from Penn II.

Pattern 4: Proband with OC and BC at age 50: \( Bel_{m_{12}}(b_2, o_1, b_1, nr, o a, ba, bil) = m_{12}(o_1) + m_{12}(b_2) + m_{12}(b_1) + m_{12}((nr) + m_{12}(bil) + m_{12}(oa) + m_{12}(b_2, o_1) + m_{12}(b_1, b_2) + m_{12}(oa) + m_{12}(ba) = 0.134 + 0.147 + 0.057 + 0.022 + 0.022 + 0.021 + 0.024 + 0.069 + 0.518 \) (with intervals: [0.518,0.656]) as compared to 53% given by Penn II.

This model seems to underestimate the IBRCAm somewhat if two cancers of the same kind are present in the personal/FH (cf. Pattern 2, non-AJ).
The goal is to estimate the mean \( l_{BRCAm} \) of a cohort model for the main risk factors and for TNBC from [13]. Information about the dependency on age given by the Penn II model for the index patient, who is often the youngest\(^{12} \), using the FAO at a younger age or for certain constellations (TNBC) provides the interval \([11, 16]\)% since the exact age is not specified. The overall \( l_{BRCAm} \) is in the interval \([50, 55]\)%.

In the following, we compute the average mutation risk, which has the meaning of a family risk (identical to that of the index patient, who is often the youngest\(^{12} \)), using the model above and compare it with the numbers that are given in the corresponding studies. We calculate this risk as the scalar product of numbers of BC, BCOCsp, BCOC in the patient/FH members and the first age of onset as given by the study, divided by the total number of patients/families, respectively. Note that we always use the same model for each study. All computations were performed in Maple. An example for the non-AJ case from [10] with numbers from Table 1 and average FAO from [10] is below:

\[
l_{BRCAm} = \frac{(5303 \cdot BC_{20-70}(44) + 661 \cdot BC_{20-70}(55) + 1260 \cdot OC_{20-70}(53) + 862 \cdot (BC_{20-70}(35) - BC_{20-70}(44)) + 803 \cdot (OC_{20-70}(53) + BC_{20-70}(44)) + 135 \cdot (OC_{20-70}(53) + BC_{20-70}(49) + BCOC_{sp}(49)))}{4716} = (4491 \cdot 9.4 + 661 \cdot 5.5 + 862 \cdot 13.5 + 1260 \cdot 5.8 + 803 \cdot 15.2 + 135 \cdot 15.3) / 4716 = 16.67\%.
\]

In [10] itself, the corresponding \( l_{BRCAm} \) is given as 16.77% (for 4716 participants of the non-AJ cohort). For the AJ cohort, we estimate 21.34% (compared to 20.46%).

3.3. A Possible Comparison Scheme

From the publications [8], [9], [11], [13], it is clear that there are further risk factors that play an important role for the \( l_{BRCAm} \) as compared to those considered in [10], for example, TNBC, BCOCsp, BCOC in the FH separately for the patient and her 1st/2nd degree relatives combined with age at the first occurrence of the disease FAO. Information on FAO is crucial since the \( l_{BRCAm} \) varies with age (usually, the period from 20 to 70 years is considered). With FAO at a younger age or for certain constellations (TNBC) and ethnicities (AJ) for patients, the \( l_{BRCAm} \) increases, sometimes non-linearly. Surveys conducted separately for each age group of 5 to 10 years within the time period lead to more accurate results and better comparability.

We can differentiate more precisely within the age risk factor by providing cumulative percentage curves depending on patient’s FAO. For this purpose, we combine information about the dependency on age given by the Penn II model for the main risk factors and for TNBC from [13]. The goal is to estimate the mean \( l_{BRCAm} \) of a cohort across studies and data resources, using piecewise linear risk percent curves for factors \( BC_{20-70} \) (replacing \( BC_{50-} \), \( BC_{51-} \), \( BC_{40-} \)), BC, mBC, TNBC, \( OC_{20-70} \) (replacing OC, \( OC_{50-} \), \( OC_{40-} \)) both for the patient and FH, and the FAO using the five age groups shown in Table 3 (the \( l_{BRCAm} \) itself) and in Figure 1 (the cumulative percentage curves). Note that mBC is \( BC_{20-70} \) plus 2% and nBC is \( BC_{20-70} \) plus 9% for the cumulative curves. For BCOC in FH, the corresponding BC and OC numbers are added. For BCOCsp, the growth factors 1, 2, 3, 4, 5 are added to curve sections 1–5 corresponding to five age groups from Table 3. For AJ ethnicity, it is necessary to take into account a higher risk in the model, so we add 5% to \( BC_{20-70} \) risk (as a constant), 3% to \( OC_{20-70} \), the additional growth of 2,3,4,5,6% to AJ BCOC in each age segment, and 5,6,7,8,9% for AJ mBC. (These additions roughly correspond to the additional scores for these factors in FHAT.)

With the help of these cumulative curves, the \( l_{BRCAm} \) can be computed more precisely for those examples from Section 3.2, for which the first age of onset is known exactly (e.g., in Pattern 4, AJ). In case of the AJ proband, we have BCOCsp at 35, that is, the \( l_{BRCAm} \) is 39%. For her aunt having mBC at age over 50, the corresponding curve provides the interval \([11, 16]\)% since the exact age is not specified. The overall \( l_{BRCAm} \) is in the interval \([50, 55]\)%.

Table 2. DST based data fusion for nine risk factors: Non-AJ case in Cols. 2-6, AJ case in Cols. 7-11.

<table>
<thead>
<tr>
<th>Factor</th>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_{12} )</th>
<th>( Bel_{m_{12}} ) ([10])</th>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_{12} )</th>
<th>( Bel_{m_{12}} ) ([10])</th>
</tr>
</thead>
<tbody>
<tr>
<td>( {b_1} )</td>
<td>0.02</td>
<td>0.04</td>
<td>0.039</td>
<td>0.039</td>
<td>0.06</td>
<td>0.02</td>
<td>0.057</td>
<td>0.057</td>
</tr>
<tr>
<td>( {b_2} )</td>
<td>0.1</td>
<td>0.04</td>
<td>0.116</td>
<td>0.116</td>
<td>0.12</td>
<td>0.07</td>
<td>0.147</td>
<td>0.147</td>
</tr>
<tr>
<td>( {o_1} )</td>
<td>0.07</td>
<td>0.04</td>
<td>0.098</td>
<td>0.098</td>
<td>0.12</td>
<td>0.06</td>
<td>0.134</td>
<td>0.134</td>
</tr>
<tr>
<td>( {o_2} )</td>
<td>0.07</td>
<td>0.04</td>
<td>0.096</td>
<td>0.096</td>
<td>0.12</td>
<td>0.06</td>
<td>0.144</td>
<td>0.144</td>
</tr>
<tr>
<td>( {ba} )</td>
<td>0.06</td>
<td>0.04</td>
<td>0.056</td>
<td>0.056</td>
<td>0.06</td>
<td>0.06</td>
<td>0.069</td>
<td>0.069</td>
</tr>
<tr>
<td>( {oa} )</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.022</td>
<td>0.022</td>
</tr>
<tr>
<td>( {ar} )</td>
<td>0.02</td>
<td>0.04</td>
<td>0.026</td>
<td>0.026</td>
<td>0.02</td>
<td>0.02</td>
<td>0.022</td>
<td>0.022</td>
</tr>
<tr>
<td>( {bil} )</td>
<td>0.05</td>
<td>0.05</td>
<td>0.053</td>
<td>0.053</td>
<td>0.02</td>
<td>0.02</td>
<td>0.022</td>
<td>0.022</td>
</tr>
<tr>
<td>( {bm} )</td>
<td>0.05</td>
<td>0.05</td>
<td>0.053</td>
<td>0.053</td>
<td>0.02</td>
<td>0.02</td>
<td>0.022</td>
<td>0.022</td>
</tr>
<tr>
<td>( {b_1, o_2} )</td>
<td>0.02</td>
<td>0.03</td>
<td>0.023</td>
<td>0.178</td>
<td>0.11</td>
<td>0.03</td>
<td>0.024</td>
<td>0.228</td>
</tr>
<tr>
<td>( {b_2, o_1} )</td>
<td>0.08</td>
<td>0.05</td>
<td>0.076</td>
<td>0.211</td>
<td>0.05</td>
<td>0.03</td>
<td>0.048</td>
<td>0.189</td>
</tr>
<tr>
<td>( {o_1, o_2} )</td>
<td>0.1</td>
<td>0.06</td>
<td>0.098</td>
<td>0.312</td>
<td>0.341</td>
<td>0.01</td>
<td>0.021</td>
<td>0.303</td>
</tr>
<tr>
<td>( {b_2, o_1, o_2} )</td>
<td>0.12</td>
<td>0.01</td>
<td>0.090</td>
<td>0.2834</td>
<td>0.277</td>
<td>0.01</td>
<td>0.012</td>
<td>0.286</td>
</tr>
<tr>
<td>( \Omega )</td>
<td>0.20</td>
<td>0.48</td>
<td>0.139</td>
<td>1</td>
<td>0.32</td>
<td>0.50</td>
<td>0.212</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3. The \( l_{BRCAm} \) in dependence on FAO for different risk factors.

<table>
<thead>
<tr>
<th>(60,70)</th>
<th>(50,60)</th>
<th>(40,50)</th>
<th>(30,40)</th>
<th>(20,30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>OC</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>BCOCsp</td>
<td>0.08</td>
<td>0.08</td>
<td>0.09</td>
<td>0.1</td>
</tr>
<tr>
<td>TNBC</td>
<td>0.07</td>
<td>0.04</td>
<td>0.05</td>
<td>0.08</td>
</tr>
</tbody>
</table>

\(^{12}\)The patient’s relatives have a lower risk (according to the Mendelian inheritance principles)
Further estimations are 24.87% (compared to 24%) for [9], and, finally, in case of [11], 12.8% (12.1% is given) for the Western European group and 16.9% (15.6%) for the African group. The results show a good agreement.

4. Conclusions
We proposed two possibilities for modeling the IBRCAm after an extensive literature research that allowed us to identify important risk factors. The fist model was entirely DST based. The second one used cumulative risk functions to compare the studies and their results with each other. Moreover, it is even possible to complete missing results or compare the studies and their results with each other. Better model the dependency on FAO. In this way, it is possible to predict the IBRCAm for different disease patterns or compare the studies and their results with each other. Moreover, it is even possible to complete missing results for the FH of a cohort from one study with those from another study if the numbers for BC, OC and BCOC are appropriately scaled for similar ethnicities, age intervals and risk classes, which is a subject for our future work.

References