Journal of Universal Computer Science, vol. 27, no. 12 (2021), 1347-1370
submitted: 11/2/2021, accepted: 9/8/2021, appeared: 28/12/2021 CC BY-ND 4.0

Uncertainty Handling in Genetic Risk Assessment and Counseling

Ekaterina Auer
(University of Applied Sciences Wismar, Germany
https://orcid.org/0000-0003-4059-3982, ekaterina.auer@hs-wismar.de)

Wolfram Luther
(University of Duisburg-Essen, Germany
https://orcid.org/0000-0002-1245-7628, wolfram.luther@uni-due.de)

Abstract: In this paper, we consider genetic risk assessment and genetic counseling for breast cancer from the point of view of reliable uncertainty handling. In medical practice, there exist fairly accurate numerical tools predicting breast cancer (or gene mutation) probability based on such factors as the family history of a patient. However, they are too complex to be applied in normal doctors' offices, so that several simplified, questionnaire-type support tools appeared. This process is highly affected by uncertainty. At the same time, reliability of test interpretations and counseling conclusions is especially important since they have direct influence on humans and their decisions. We show how expert opinions on mutation probabilities can be combined using the Dempster-Shafer theory. Based on multi-criteria binary decision trees and interval analysis, we combine the referral screening tool designed to determine patients at risk of breast cancer (and recommend genetic counseling or testing for them) with three further risk assessment tools available for this purpose. A patient's confidence in the outcome of a genetic counseling session can be heightened by the proposed method since it combines different sources to provide score ranges leading to more information. Finally, based on this approach, a decision tree for assigning a risk category is proposed which enhances the existing methodology. The great impact of epistemic uncertainty is reflected through large overlapping intervals for the risk classes.

Keywords: Binary decision tree, *BRCA1/2* mutations, Dempster-Shafer theory, interval analysis, family history assessment model, genetic counseling verification and validation assessment, uncertainty in data
Categories: A.1, H.4.2, I.2.4, I.6.4, J.3
DOI: 10.3897/jucs.77103

1 Introduction

Imperfect or unknown information leading to uncertainty is an intrinsic part of both our everyday lives and science. Especially in the area of medicine, it is necessary to make qualified decisions which significantly affect human lives using highly uncertain, ambiguous, or unreliable data. Until approximately the 1960s, the major way to deal with uncertainty has been offered by the classical probability theory. Since then, various further approaches to uncertainty quantification and propagation have started to establish themselves, ranging from fully deterministic ones (e.g., interval analysis IA [Moore et al. 2009]) to those generalizing the classical probability measure (e.g., the Dempster-Shafer theory DST [Yager and Liu 2008]).

Uncertainty handling is an important, although at the moment still optional, part of the general verification and validation (V&V) assessment scheme [IEEE 2017]. V&V is a prerequisite in many areas of science or engineering, in particular, if safety critical applications are concerned. In our previous work, we employed it in the context of visual analytics [Auer et al. 2020] and biomechanics [Auer et al. 2013], with a focus on its reliability. In this contribution, we apply its parts in the context of *BRCA 1* or *2*[1] related cancer and *BRCA1/2* mutation probability prediction tools [Amir et al. 2010, Nelson et al. 2019] as a case study, with a special emphasis on risk assessment and decision making under epistemic uncertainty.

In Figure 1, the general treatment scheme for patients concerned about their *BRCA* related cancer risk is shown. Although most kinds of breast, ovarian, prostate, and pancreatic cancers are sporadic, a minority are actually caused by germline mutations in breast cancer susceptibility genes (*BRCA1/2*). Inherited mutations in these genes can lead to an increased risk of breast cancer (BC). That is why the first step might seem to be to test concerned patients for DNA mutations in these genes. However, the actual genetic testing (GT), while expensive, is not always worthwhile or even helpful (e.g., leading to an increased instead of decreased anxiety without being able to provide decisive results). In consequence, the patients are not referred for testing directly but undergo first a risk assessment (RA) procedure and then a genetic counseling session (GC). That is, mathematical models are applied first to assess the mutation probability (or BC risk) based on various factors such as the age of cancer onset of a given patient or their family history of cancer. Then, the pros and cons of GT and general results of RA are explained during a GC session. Finally, if an actual genetic test becomes necessary, its results are explained in a further GC session. Reliability of RA and GC stages in

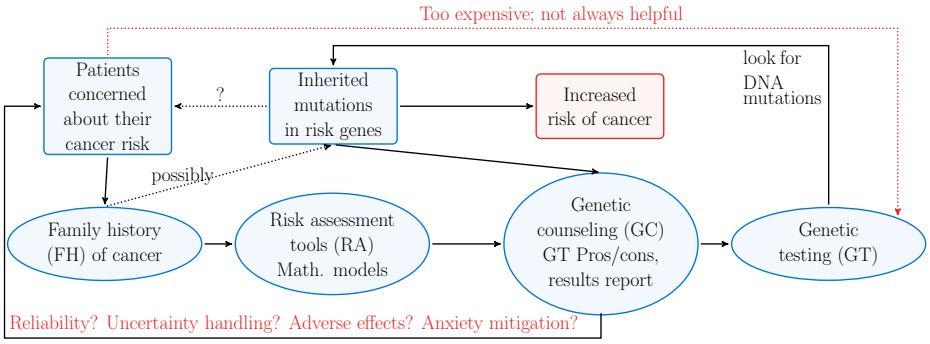

Figure 1: Genetic risk assessment, counseling and testing scheme

the process from Figure 1 is especially important because they are quite frequent and, possibly in combination with GT results, have a direct influence on humans and their decisions.

The process in Figure 1 is highly influenced by uncertainty, in particular, while interpreting its results. For example, it is known that mutations in *BRCA1/2* genes reduce their tumor-suppressor qualities essential for the repair of DNA double-strand breaks by homologous recombination (HR). However, the HR pathway for DNA repair is disrupted

[1] So-called tumor suppressor genes

by pathogenic mutations not only in *BRCA1/2*, but also in other involved genes [Stoppa-Lyonnet 2016], making it harder to pin the probability of getting cancer on *BRCA1/2* mutation alone. Other sources of uncertainty are pointed out later on.

An introduction to genetic screening, test performance measures, and clinical testing of BC genes can be found in [Struewing et al. 1995]. An overview of existing V&V standards as well as quality criteria and metrics for GC and clinical molecular testing is given in [Christian Jr. and Drilling 2009, Mattocks et al. 2010, Dequeker 2017]. In [Li et al. 2018, Ye et al. 2018], the authors emphasize RA as well as sense and decision making under various forms of uncertainty. At the RA stage, Claus tables [Claus et al. 1994], Frank tables [Frank et al. 1998, Frank et al. 2002], BRCAPRO[2], BOADICEA[3], and the Penn II risk model[4] are five important mathematical models for computing the probability of breast or ovarian cancer based on Mendelian genetics and the Bayes theorem [Evans et al. 2004]. However, their usage might be considered too complicated by a normal cancer specialist, so that several simplified, questionnaire-type support tools are usually employed for identifying candidates for whom GT is necessary, for example, the Family History (FH) Assessment Tool FHAT [Gilpin et al. 2000], the Referral Screening Tool RST [Bellcross et al. 2009], the Manchester Scoring System MSS [Evans et al. 2004], or the FH Screen FHS-7 [Ashton-Prolla et al. 2009].

At the RA stage, there are several influential studies which aim at validating various approaches for predicting *BRCA1/2* mutations. In [Parmigiani et al. 2007], the authors quantify the accuracy of seven publicly available models (including BRCAPRO, Penn II and FHAT) employing them for 3342 persons to predict the status of a mutation carrier. The study is based on three population-specific sample groups of participants from research and eight samples from GC clinics. As validation criteria, it relies on sensitivity and specificity of predictions as well as on how well a model discriminates between individuals testing positive for a *BRCA1/2* mutation and those testing negative using statistical methods as a metric. Conducting comprehensive statistical analysis with data of 9390 FH, [Kast et al. 2014] validates the recalibrated MSS model using pathology information on the histological subtype, on the grade of differentiation, on estrogen and other factors. Quite recently, [Himes et al. 2019] assesses five currently widely used screening tools for genetics referral, namely, FHS-7, Pedigree Assessment Tool [Teller et al. 2010], MSS, RST, and FHAT. Finally, the comprehensive and thorough meta-study in [Nelson et al. 2019] reviews 103 medical studies and 110 research articles (with 92712 patients overall) w.r.t. their methodology, scientific rigor, study parameters, relevance, quality criteria and metrics, performance, accuracy, limitations as well as adverse effects and benefits for the patients. A common drawback of such studies or meta-studies is that women with unknown *BRCA1/2* mutation status are not included, so that, for example, adverse GC effects and the suggested treatment options after testing cannot be evaluated appropriately. As stated in [Owens et al. 2019], there is fair evidence that RA, GC, GT, and interventions have (moderate) benefits only for patients whose family or personal history corroborates an increased risk for harmful mutations in the *BRCA1/2* genes. However, regardless of family or personal history, overall harms of RA, GC, GT, and interventions are small to moderate. It is therefore important to be able to reliably assign patients to risk classes. Although a person's assignment to a risk class w.r.t. a genetic mutation can be validated with reliable genetic tests, population groups must be stratified according to given criteria and test persons interviewed about their own illnesses and

[2] https://projects.iq.harvard.edu/bayesmendel/brcapro
[3] https://ccge.medschl.cam.ac.uk/boadicea/
[4] https://pennmodel2.pmacs.upenn.edu/penn2/

those in their family (i.e., first, second and third degree relatives on the maternal and paternal side) for the purpose of reliability.

Young women with *BRCA1/2* mutations and their families usually face conflicting healthcare decisions regarding family formation and risk management. They must decide whether they prioritize risk reducing interventions or family formation goals. Documented with 115 references, [Peshkin and Isaacs 2020] gives an up-to-date overview on GC and management of individuals at risk of hereditary BC and ovarian cancer (OC) syndromes covering topics from genetic criteria for risk evaluation and RA models to pre-test and post-test counseling to, finally, approaches to treat positive results as well as risk management for negative or uninformative results. In a prior study [Interrante et al. 2017], it is shown that patients subjected to telephonic post-test GC were at least non inferior to patients under usual care w.r.t. their GT decision, distress, quality of life and uptake of cancer risk management strategies. In [Li et al. 2019], a survey and evaluation tool Feelings About genomiC Testing Results (FACToR) is presented that measures the psychosocial impact of presenting genomic findings to patients in research and clinical practice. Finally, [Richardson et al. 2020] comes to the conclusion that oncology clinic-based GT using a multi-gene panel approach combined with post-test GC significantly reduces wait times and is acceptable for patients and healthcare providers.

From the studies and meta-studies on mutation prediction tools, it is evident that there is significant uncertainty both in data and in the processes/models. Aside from the usual, in this case less influential, sources of uncertainty due to the employed numerical methods and the modeling error, the major uncertainty factors, from the point of view of a FH, are the age of cancer onset in a patient's relative, the degree of kinship of the patient and the affected as well as the number of instances and kinds of cancer in the family tree. Moreover, interpretation/handling of the RA or GT results and prevailing strategies for advising people from various risk classes about appropriate behavior, treatments and preventive examinations need to be critically monitored and assessed according to a set of standardized rules for a longer period of time. The bottom line of all the above mentioned research is that it is important to standardize V&V management for RA and GC, to pose common requirements on the quality of heterogeneous data, to systematize dealing with a large number of parameters during test subjects' selection, to consider explicitly further kinds of epistemic uncertainty and to ensure interaction between experts during the evaluation, interpretation and conclusions for counseling, care and treatment of the subjects.

In this paper, we take a first step towards a consistent and reliable V&V framework for RA of GC stages by addressing the topic of uncertainty in the data. It is structured as follows. In Section 2, we discuss approaches to uncertainty handling and point out its role in the reliable V&V assessment procedure. Additionally, we overview briefly the two methods we rely on for this purpose, IA and DST. Having outlined the major tools serving as the basis for RA and GC in Section 3, we consider application of methods from Section 2 to these stages in Section 4. First, we show an example of how expert opinions on mutation probabilities can be combined using the DST in Subsection 4.1. To take into account uncertainty in patient data in combination with the advantages of established models and to ensure validation, we present a new multi-criteria categorical counseling test ERST in Section 4.2. It combines the binary decision structure of RST with the features of FHAT and MSS and uses an accumulated interval risk function. We propose to use interval scores and lower limits for probabilities of pathogenic variants since crisp scores do not reflect the available information correctly. The result of the original RST test combines the presence of different constellations of cancers in the FH. Now, a referral vector U for eight decision paths is produced to help assign participants to low,

moderate and high risk categories. Our validation for the scores is based on prevalence and frequency of mutations in *BRCA1/2* correlated with various combinations of personal and FH of cancer given by tables from [Frank et al. 2002]. In this way, patients can be assigned an individual, family and fused risk category (cf. [Bellcross et al. 2015]) characterizing the likelihood of a *BRCA1/2* mutation more accurately (cf. Section 4.3). Conclusions and an outlook on future work are in the last section.

This article is an updated and extended version of the contribution [Auer and Luther 2020] presented at the CODASSCA meeting on Collaborative Technologies and Data Science in Smart City Applications in Yerevan, Armenia, September 14–17, 2020 [Hajian et al. 2020].

2 Approaches to Uncertainty Treatment and Its Place in the General V&V Procedure

In this paper, we show how to apply parts of the general V&V procedure in the context of RA and GC[5] with the focus on uncertainty in Section 4. Therefore, in this section, we give a short overview of approaches to uncertainty handling, their role in the reliable V&V assessment scheme, and of two particular techniques for dealing with epistemic uncertainty (IA and DST).

2.1 Uncertainty Handling

As explained in the Introduction, uncertainty as a constituent part of almost any field of science needs to be considered a priori while developing a system model to obtain reliable simulation results. Therefore, uncertainty quantification dealing with characterization and reduction of uncertainty [Smith 2013] has become a cutting-edge research area. There are two directions for propagation of uncertainties through systems, forward (from uncertain inputs to the outcomes) and backward (answering the question of what uncertainty is allowed in the inputs if a given uncertainty in the outputs should not be exceeded).

In this paper, we rely on forward propagation. Three general types of methods for forward uncertainty propagation can be discerned [Ferson et al. 2003]: *rigor-preserving* (or with result verification, or outer enclosure: the result is guaranteed to enclose the uncertainty completely, if inputs enclose it completely); *best possible* (or inner enclosure: the enclosure of the output cannot get any tighter without more information); and *statistical*, providing a guarantee of the type "in x percent of the trials, the result is sure to enclose the uncertainty completely".

Although one of the prevalent directions to deal with uncertainty is through probability, other techniques emerged that generalize the classical probability measure. For various types of uncertainty, the probability-based approach is arguably not the best (cf. Figure 2). For example, methods with result verification (such as IA) were originally designed to provide a guarantee that the result obtained in a computer simulation using floating point instead of real arithmetic was correct. However, they can be used to represent and propagate bounded epistemic uncertainty in a rigor-preserving way using solely deterministic techniques. Moreover, such techniques as DST can provide more conclusive results than the traditional probability theory if both aleatory and epistemic

[5] Although V&V assessment in GT is not in the foreground, some of the results and conclusions of procedures existing there can also be applied here

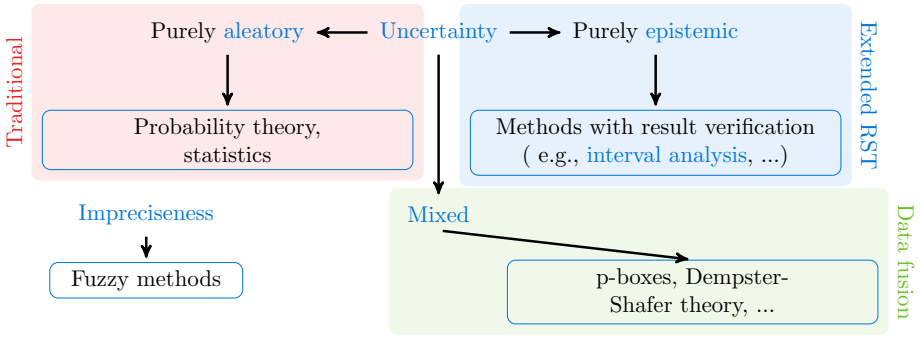

Figure 2: Common groups of approaches to uncertainty handling and the place of methods proposed in this paper there

(bounded) uncertainty is present in an application. Sometimes authors differentiate between uncertainty and impreciseness. To represent the latter kind of incertitude, fuzzy sets are developed [Zadeh 1996], which also can be combined with the classical probability techniques. In this paper, we rely on IA to combine the RST with FHAT and MSS in Section 4.2 and on DST to merge data on mutation probabilities in Section 4.1.

2.2 Reliable V&V Assessment

In accordance with the IEEE Std 1012TM-2016 norm [IEEE 2017], Figure 3 shows a broad approach to *reliable* V&V assessment advocated in [Auer et al. 2020]. It allows its users to relate the concept of reliability to data, design strategies, processes, software and outcome analysis; to define requirements, quality criteria and metrics for the result of the considered process or task; to analyze the process systematically (e.g., w.r.t. uncertainty) at the early stages in its development cycle; to choose the appropriately evaluated tools (e.g., for data mining, visualization, analysis, decision making and risk assessment) or get them recommended; and to facilitate interaction between experts during the decision-making process. The usual modeling and simulation cycle [Schwer 2007] on the outermost left of Figure 3 with verification at the level of implementation and validation at the level of simulation can be augmented with reliable computing techniques from the third column of the figure. At each stage of the cycle (which can be reiterated), certain methodologies and technologies (shown in Column 2) have to be used, complemented by optional features such as those from Column 4, of which uncertainty quantification, propagation and visualization play an especially important role [Markov and Akyildiz 1996, Weyers et al. 2019]. From the point of view of genetic RA and GC, the item in the right bottom corner of Figure 3 is also fairly significant. Methods and design criteria to support online collaborative help seeking actions and decision processes through the exchange of questions and answers in the area of healthcare are presented, for example, in [Santos et al. 2016].

As mentioned in the Introduction, there are many studies and meta-studies that review/evaluate RA and GC tools and validate GT and RA, for example, [Amir et al. 2010, Bellcross et al. 2015, Himes et al. 2019, Kast et al. 2014, Louro et al. 2019, Nelson et al. 2013, Nelson et al. 2019, Owens et al. 2019, Parmigiani et al. 2007] among a multitude of other publications. However, they focus on aleatory uncertainty mainly and consider the epistemic one rather indirectly through comparisons. From these studies,

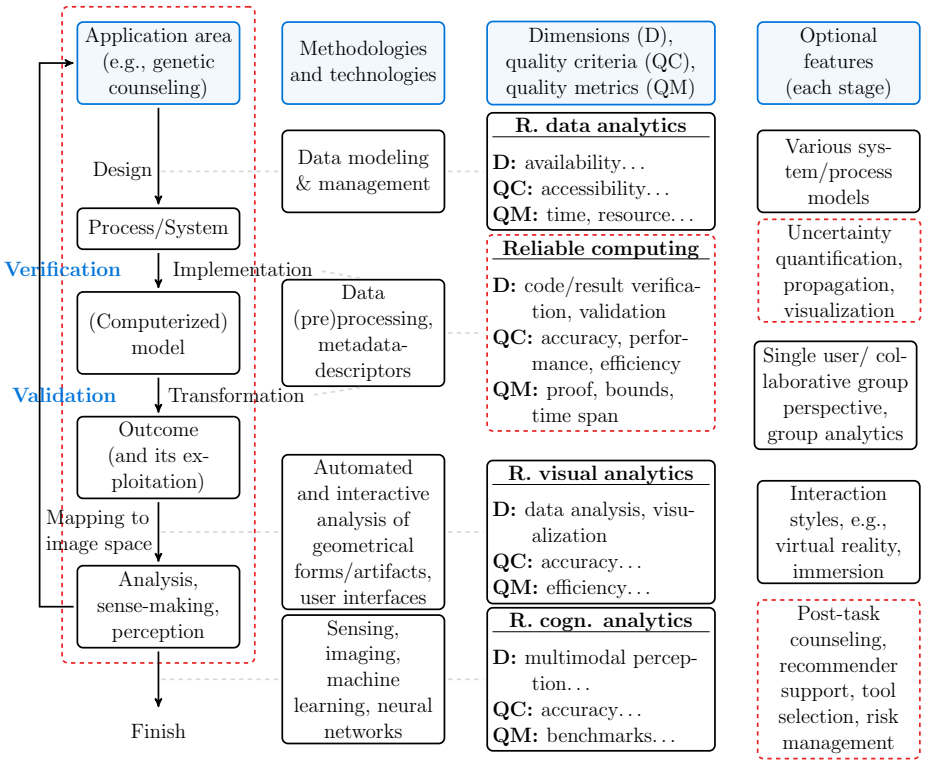

Figure 3: A scheme for a V&V approach to assess an environment from its modeling to its outcome analysis stage

it is clear that V&V assessment in GT primarily describes positive assurances, makes evident that specifications and quality criteria are met and the test achieves previsions and objectives [Nelson et al. 2019] as in "fit for the intended use" [ISO 2005]. The authors point out that negative results, damage, misinterpretation and uncertainty should also be highlighted. Additionally, V&V assessment should include implications of inaccurate risk assessment; inappropriate testing; false-positive and false-negative results; adverse effects on the patient's family relationships; overdiagnosis and overtreatment; false reassurance; incomplete testing; misinterpretation of test results; anxiety; cancer worry; and ethical, legal, and social implications. At the same time, these requirements constitute the corresponding specifications for (post-test) GC sessions.

As examples of quality metrics (cf. Column 3 in Figure 3) used in the area of RA and GT, sensitivity, specificity, and discriminatory accuracy can be named [Mattocks et al. 2010, Panchal et al. 2008]. They are usually described by a confidence interval around the mean result. With the abbreviations P (positive), N (negative), T (true), F (false), PR (positive rate), and LR (likelihood ratio), *sensitivity* (or true positive rate, or recall) is defined as $\frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FN})}$. The formula $\frac{\mathrm{TN}}{(\mathrm{TN}+\mathrm{FP})}$ is used for *specificity* (or the true negative rate) and $\frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FP})}$ for *precision*. Then, the *positive likelihood ratio* can

be defined as $\frac{\text{sensitivity}}{(1-\text{specificity})}$. *Discriminatory accuracy* is formalized as the power of a method (or risk factor/criterion) to predict true or false outcome of test cases. In [Panchal et al. 2008], a comprehensive graphical representation of the performance of the RA tools BRCAPRO, FHAT, MSS and Penn II mentioned in the Introduction is assessed according to these metrics.

At the moment, different studies use different criteria for choosing test persons, ages or degrees of kinship that often cannot be mapped to each other, which requires careful data fusion (or sometimes even prevents it). In particular, it is necessary to work with sets or confidence intervals while comparing approaches. In summary, V&V assessment in RA, GC, GT deals with determining accuracy in its various meanings, detection limit, specificity, linearity, repeatability, reproducibility, robustness and cross-sensitivity. It is always a balance between costs, risks and technical possibilities. There are many cases in which the range and uncertainty of the values can only be given in a simplified way due to lack of information [ISO 2005].

2.3 Interval Analysis

Interval analysis [Moore et al. 2009] is an approach for result verification with applications in many areas of engineering, medical science, (bio)mechanics and others. Methods based on IA ascertain formally that the outcome of a simulation implemented on a computer using them is correct despite such factors as numerical or discretization errors (assuming that the underlying implementation is correct). The results are intervals with bounds expressed by floating point numbers which with certainty contain the exact solution to the model. A common drawback of such rigor-preserving methods, caused by the dependency problem or the wrapping effect, is the possibility of too wide bounds for the solution sets (e.g., between $-\infty$ and $+\infty$).

An interval $[\underline{x}, \overline{x}]$, where $\underline{x}$ is the lower, $\overline{x}$ the upper bound, is defined as $[\underline{x}, \overline{x}] = \{x \in \mathbb{R} | \underline{x} \leq x \leq \overline{x}\}$. For an operation $\circ = \{+, -, \cdot, /\}$ and two intervals $[\underline{x}, \overline{x}]$, $[\underline{y}, \overline{y}]$, the corresponding interval operation can be defined as

$$[\underline{x}, \overline{x}] \circ [\underline{y}, \overline{y}] = [\min(\underline{x} \circ \underline{y}, \underline{x} \circ \overline{y}, \overline{x} \circ \underline{y}, \overline{x} \circ \overline{y}), \max(\underline{x} \circ \underline{y}, \underline{x} \circ \overline{y}, \overline{x} \circ \underline{y}, \overline{x} \circ \overline{y})],$$

that is, the result of an interval operation is also an interval. Every possible combination $x \circ y$ with $x \in [\underline{x}; \overline{x}]$ and $y \in [\underline{y}; \overline{y}]$ lies inside this interval. (For division of intervals, usually $0 \notin [\underline{y}, \overline{y}]$ is assumed.) The general formula can be simplified for a given $\circ$ (e.g., $[\underline{x}, \overline{x}] + [\underline{y}, \overline{y}] = [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$). Based on this interval arithmetic, higher-level interval methods can be defined, for example, those for solving systems of algebraic or differential equations.

2.4 The Dempster-Shafer Evidence Theory

The Dempster-Shafer theory described, for example, in [Ferson et al. 2003], makes it possible to combine evidence from different experts or other sources and to provide a measure of confidence that a given event occurs. A special feature of this theory is its ability to characterize uncertainties arising because of the lack of knowledge as discrete probability assignments associated with the power set of values from a given set Ω. Wheres a classical probability mass function produces as its result the probability that the random variable X is equal to a certain crisp value x_i, DST allows us to assign a

probability to the event that a realization of X belongs to a given set (e.g., $[\underline{x}_i, \overline{x}_i]$)[6]. That means that only lower and upper bounds (belief and plausibility) on the probability of a subset of Ω can be computed using the DST. A random DST variable can be characterized by its basic probability assignment (BPA) m. If $A_1, \ldots, A_n$ are the sets of interest where each $A_i \in 2^\Omega$, then

$$m : 2^\Omega \to [0,1], \quad m(A_i) = p_i,\ i = 1 \ldots n, \quad m(\emptyset) = 0, \quad \sum_{i=1}^{n} m(A_i) = 1 \ . \quad (1)$$

The mass of the impossible event $\emptyset$ is equal to zero. Every element A_i with a mass unequal zero is 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 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

$$Pl(Y) := \sum_{A_i \cap Y \neq \emptyset} m(A_i), \quad Bel(Y) := \sum_{A_i \subseteq Y} m(A_i). \quad (2)$$

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 [Ferson et al. 2003], of which Dempster's rule is used in this paper:

$$m_{DR}(A_i) = \frac{m_{1,2}(A_i)}{1 - K_{1,2}} = \frac{\sum\limits_{\text{all } A_j \cap A_k = A_i} m_1(A_j) m_2(A_k)}{1 - \sum\limits_{\text{all} A_j \cap A_k = \emptyset} m_1(A_j) m_2(A_k)},\ A_i \neq \emptyset,\ m_{1,2}(\emptyset) = 0 \ . \quad (3)$$

The constant $K_{1,2}$ in the denominator has the meaning of $m_{1,2}(\emptyset)$.

3 Gene Mutation Risk Scoring Tools and Probability Tables

One in four cancers diagnosed in women worldwide is a case of BC. Overall, an estimated 5–10% of all breast cancers is hereditary, that is, caused by a germline mutation in a high-risk gene. Among other BC predisposition genes, "germline mutations in the *BRCA1* and *BRCA2* genes are related to an increased risk for breast, ovarian and other cancers […]. Specific features in the family history may suggest the diagnosis of a hereditary breast cancer syndrome" [Ashton-Prolla et al. 2009], hence the need for the procedure shown in Figure 1. In this section, we give a brief overview of probability tables and simplified, questionnaire-type RA tools we rely on in this paper, the purpose of which is to identify individuals with high risk of BC. Additionally, where necessary, we reorganize the existing material to suit out later purposes (e.g., provide the corresponding intervals or combine data).

3.1 Probability Tables

Generally, RA goals can be either to provide the chances of developing BC over a given time span (including the lifetime) or the probability of a mutation in a high-risk gene

[6] A similar interpretation is possible for continuous random variables

(abbreviated as mp in the following). Tables from [Claus et al. 1994] give predictions for cumulative BC probability based on a survey considering primarily age-specific and FH risk factors and using a Bayesian model (with data on 4730 patients with confirmed BC matched against 4688 control subjects). In contrast, [Frank et al. 1998, Frank et al. 2002] provide predictions for mutations in *BRCA1/2* correlated with such risk factors as age of diagnosis, personal and FH, or ethnicity (compiled in tables, denoted Frank tables in the following). The first study considered 238 women with BC before age 50 or OC at any age and at least one first or second degree relative with either diagnosis. The participants underwent sequence analysis of *BRCA1* followed by analysis of *BRCA2* for those of them who agreed to it. Based on the results of this survey, the authors identified risk factors for *BRCA1/2* mutations (cf. Table 2, Column 1) and correlated them with rates for developing BC, BC with subsequent OC, and contralateral BC. This allowed them to model the probability of detecting a mutation in *BRCA1/2* with the help of logistic regression analysis (cf. Table 2, Column 2).

The second study [Frank et al. 2002] considered a bigger proband group of 2,233 [4,716] [Non]-Ashkenazi individuals. With the aim of accurately and reliably identifying different risk classes, this study describes the prevalence of mutations in *BRCA1* and *2* in correlation with the personal and FH taking into account the first and second degree relatives of the individual undergoing the test. We describe tables from it in some detail since we will use them for validation inside the proposed method ERST.

In Table 1, we reproduce Tables 1 and 2 from [Frank et al. 2002] in full for better understanding. This is a rectangular scheme of BC/OC disease constellations in the patient's FH and for the patient herself with the age as a parameter. The patient has the role of a child with her first degree relatives defined as parents, siblings, children. One of risk factors explicitly identified in this survey is the occurrence of BC at the age over 50 (possibly, in combination with other criteria). Although this criterion is not explicitly mentioned on the FH side, the entry in the upper left corner of the table shows mutations in 3.9% of cancer-free individuals. It is explained as follows: "This finding may result from individuals who were prompted to seek testing by a strong family history of breast cancer after age 50 years" [Frank et al. 2002]. That is, from the logical point of view, the entry should be interpreted as: "no breast cancer $<$50y or ovarian cancer is diagnosed in anyone in the FH of the patient; none, one (or two) cases of BC $\geq$50y occur in the FH; the patient has no BC or OC diagnosis at any age", which is not directly readable from the table. That is, we can assume that, although cases of BC$\geq$50y in FH are not explicitly reflected in the risk factors from the table, they are somehow considered in the study. The occurrence of genetic mutations seems to be rather probable for patients diagnosed with multiple cancers (BC and OC) at age under 50. However, the data set supporting this conclusion is quite small (cf. the right bottom half of the table).

A further example of how to interpret Table 1 is the entry in the 3rd row and 1st column: *BRCA1* or *2* mutations occur in 55 out of 579 individuals diagnosed with BC at age below 50 with no diagnosis of BC or OC for anyone below age 50 in the FH. If BC is diagnosed between 40 and 50 years of age, the mp is 5.6% (16/284). From this information, we can conclude that the corresponding mp for the probands at ages below 40 is 13.2% (55-16=39 out of 579-284=295). Another example is the entry in the same column, Row 4: the mp for a proband diagnosed with OC at any age with no BC and the same FH of cancer as in the first example is 6.5% (5/77), with mp for OC diagnosed at age over 50 being 6.7% (3/45); that is, under the same conditions, mp for OC diagnosed at age less than 50 is 6.9% (2/29).

	Family history (includes 1st/2nd degree relatives; excludes the proband)											
	No breast cancer <50y or ovarian cancer in anyone		Breast cancer <50y in one relative, no ovarian cancer in anyone		Breast cancer <50y in > one relative, no ovarian cancer in anyone		Ovarian cancer at any age in one relative, no breast cancer <50y in anyone		Ovarian cancer in > one relative, no breast cancer <50y in anyone		Breast cancer <50y and ovarian cancer at any age	
Proband	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
No breast cancer or ovarian cancer at any age	9/229	3.9	19/434	4.4	46/419	11	6/153	3.9	10/117	8.5	58/354	16.4
Breast cancer ≥50y	4/172	2.3	22/197	11.2	12/118	10.2	3/69	4.3	1/18	5.6	19/87	21.8
Breast cancer <50y (40-49y)	55/579 16/284	9.5	89/484 31/289	18.4	117/322 41/172	36.3	34/194 15/115	17.5	7/42 3/25	16.7	126/267 55/141	47.2
Ovarian cancer at any age (≥50y), no breast cancer	5/77 3/45	6.5	14/41	34.1	11/26	42.3	23/83	27.7	12/28	42.9	38/71	53.5
Breast cancer ≥50y and ovarian cancer at any age	5/27	18.5	1/9	11	4/11	36.4	1/6	17	1/3	33	3/6	50
Breast cancer <50y and ovarian cancer at any age	5/25	20	7/14	50	4/5	80	5/9	56	2/2	100	13/18	72.2

Table 1: Table 1 combined with Table 2 from [Frank et al. 2002]: Prevalence of mutations in BRCA1 and BRCA2 correlated with personal and family history of cancer in 4,716 Non-Ashkenazi individuals

3.2 Ontario Family History Assessment Tool (FHAT)

The FHAT test published by [Gilpin et al. 2000] is designed to select individuals for GC who are at approximately twice the population risk of BC or OC based on their FH. The underlying survey considered 184 patients with BC or OC who accepted the offer of *BRCA1/2* testing and answered a questionnaire on BC, OC, bilateral BC, BC and OC in the same person, male BC, colon and prostate cancer in first, second, and third degree relatives, and (broadly specified) age of diagnosis. Each of these factors is assigned an individual score from 1 to 10 points. These are summed up to a final score for a person undergoing the test. The cutoff threshold of 10 indicates referral for GC and corresponds to doubling of lifetime BC risk in comparison to general population (22% by BRCAPRO to carry a *BRCA1* or *BRCA2* mutation in their families). FHAT was validated using Claus tables and BRCAPRO, among others.

In Table 2, a comparison of different disease constellations of BC and OC both in personal and FH is given for Frank tables and FHAT. An interval score is computed for each criterion from [Frank et al. 1998] in Column 3 based on intervals for FHAT

reported in [Auer and Luther 2020]. The comparison[7] shows a strong correlation between *BRCA1/2* probabilities and FHAT scores. We can arrive at the approximately the FHAT score interval from the corresponding probabilities mp_{BRCA1}, mp_{BRCA2} by evaluating the formula $[-2, 2] + 0.25 \cdot [1, 2] \cdot (\text{mp}_{BRCA1} + \text{mp}_{BRCA2})$ using IA.

Diagnosis – The proband (Pr) has the role of a child in FH (risk factors from Frank tables)	**Mutation p. (%)** *BRCA1*	*BRCA2*	**FHAT score (f)**
Any relative with BC $<$ 50y	10.1	14.5	[4,10]
Any relative with OC	22.9	12.5	[5,13]
(BC$<$50y)&(Pr with BC (PrBC) $<$ 40y)	28.2	11.6	[11,19]
(BC$<$50y)&(OC)&(PrBC$<$40y)	50.9	7.9	[16,32]
(BC$<$50y)&(OC)&(Pr Bilateral BC or OC)	65.0	5.7	[15,35]
(BC$<$50y)&(OC)&(PrBilBC or OC)&(PrBC$<$40y)	86.7	2.2	[22,44]

Table 2: From probabilities to FHAT scores

3.3 Manchester Scoring System (MSS)

Another well-known questionnaire-type BC or OC genetics referral tool is MSS [Evans et al. 2004]. Its goal is to predict *BRCA1/2* mp in families suspected of having hereditary BC and OC. The underlying survey considered 422 patients from non-Jewish families who were tested for *BRCA1* mutation, with 318 of them tested subsequently w.r.t *BRCA2*. In this approach, 1 to 8 points are assigned to the factors of *BRCA1* or *2* mutation, degree of kinship, age and type of disease. The scores can be summed up for a person undergoing the test either individually for *BRCA1*, *BRCA2* or combined, with the limit for a referral being 10 points again in the individual case or 15 in the combined one. It corresponds to a 10% probability of a pathogenic mutation in *BRCA1* or *BRCA2* [Owens et al. 2019]. The similarities and differences of MSS and FHAT are highlighted in Table 3. Where necessary, interval scores are provided.

3.4 Referral Screening Tool (RST)

Similarly, the goal of the third tool RST [Bellcross et al. 2009] we consider in this paper is rapid identification of individuals at potential hereditary risk of BC/OC. The underlying survey covered 2464 women (without preselection) undergoing a screening mammography. This questionnaire consists of a single page used to record patient's yes/no responses to FH questions concerning Ashkenazi Jewish ancestry; occurrence of male BC at any age and in any relative; occurrence of two or more cases of BC after age 50 on the same side of the family; and occurrence of BC at or before age 50 or OC at any age (in the first and second degree relatives, that is, overall in the patient, her mother, sister, daughter, grandmother, or aunt at father's and mother's side). The assessment "scored positive" is given if there are two or more checks on the page and corresponds to a high risk of cancer. In comparison with the FH analyses using BRCAPRO, Myriad II[8], BOADICEA, FHAT and their high risk definition, RST proved an overall sensitivity of 81.2%, specificity of 91.9%, and discriminatory accuracy of 0.87 [Bellcross et al. 2009], which is good performance for such a simple, binary test.

[7] Female cancer only

[8] https://myriad.com/products-services/hereditary-cancers/bracanalysis/

Risk factor		**FHAT**	**MSS** *BRCA1 / BRCA2*
BC and OC	Mother/Sibling/2nd-3rd dr	10 / 7 / 5	0/0
BC relatives	Parent/Sibling/2nd-3rd dr	4 / 3 / 2	0/0
	Male	+2	[5,8]
BC onset age	20-29	6	6 / 5
	30-39	4	4 / 4
	40-49	2	3 / 3
	50-59	0	2 / 2
	$\geq$60	0	1 / 1
	Bilateral/multifocal	+3	$\times$2
OC relatives	Mother/Sibling/2nd-3rd dr	7 / 4 / 3	0
OC onset age	$<$40	6	8 / 5
	40-60(FHAT), 40-59(MSS)	4	
	$>$60(FHAT), $\geq$60(MSS)	2	5 / 5
Prostate C	Onset age $<$50	1	0 / 2
	Onset age $<$60	0	
	Onset age $\geq$60	0	0 / 1
Panrc.C	Any age	0	0 / 1
Colon C	Onset age $<$50	1	0

Table 3: Risk factors and scores of FHAT and MSS in comparison (dr="degree relative")

3.5 A Remark on Validation

In Subsection 4.2, we provide mp based on tables from [Frank et al. 2002] (cf. Table 1) for our extended RST approach. This in itself conclusive information is not always easy to use for validation of RA tools such as RST, FHAT and MSS we consider in this paper. In general, differences (or inconsistencies) in risk factors pose an information fusion problem. For example, the risk factor of age is not very finely grained in Frank tables (mostly just binary: over or below 50 years). Moreover, there is no information on the influence of multifocal or bilateral BC (which gives additional points in FHAT or MSS, cg. Table 3). Using MSS, it is not clear how to take into account male BC (e.g., diagnosed in a father). Moreover, there is an ambiguity while considering repeated cases of the same cancer category: it is not possible to say if they refer to the same person at different ages or several individuals (MSS, FHAT or Frank tables). A particular difficulty arises from the fact that the number of BC diseases after age 50 does not play a role in RST and is listed in Frank tables only for the proband's personal, but not for the family history.

4 Dealing with Uncertainty in Genetic BC Risk Assessment and Counseling

Whereas V&V assessment for GT in general and uncertainty handling in particular is covered fairly well (cf. [Mattocks et al. 2010, Panchal et al. 2008]), this is not quite so in the area of RA and GC. Having described in short the available RA tools in Section 3, we make a first step in this direction by addressing uncertainty in the data. First, we define unified, consistent risk factors (criteria) across the three RA tools we consider in this paper (RST, FHAT, MSS) and Frank tables. Data fusion for mutation probabilities is carried out using DST, see Subsection 4.1. We work with intervals for representing

epistemic uncertainty if the criteria does not map to each other in full. Moreover, we propagate this uncertainty using the extended RST method proposed in Subsection 4.2. Finally, patients can be assigned a more accurate risk category (cf. [Bellcross et al. 2015]) characterizing the likelihood of a *BRCA1/2* mutation as suggested by the decision tree in Subsection 4.3.

4.1 Data Fusion with DST

Based on [Tao et al. 2019, El-Mahassni and White 2015], we use Dempster's rule to combine data on mutation probabilities in *BRCA1* and *BRCA2* correlated with personal and family history of cancer. Let $\Omega = \{f_1, f_2, \ldots, f_n\}$ be the frame of discernment containing n distinct elements $f_i, i = 1, \ldots, n$. In the example of data fusion considered in this subsection, we assume that BPA m_1 is inspired by proband's mutation probabilities (personal risk) and BPA m_2 by those of her family members (family risk). Let the number n of elements in Ω equal 9. We define the elements corresponding to relevant risk factors which seem to be the major ones after a literature review (cf. Section 3) as follows: $f_1 = b_1$ (BC$\geq$50), $f_2 = b_2$ (BC$<$50), $f_3 = o_1$ (OC at any age), $f_4 = o_2$ (OC$\leq$50), $f_5 = ba$ (premature BC ($\leq$40)), $f_6 = oa$ (premature OC ($\leq$40)), $f_7 = nr$ (cancer in a near relative), $f_8 = bil$ (bilateral BC), and $f_9 = bm$ (male BC). Note that although the probabilities for o_1 and o_2 are very similar as given by Frank tables, both FHAT and MSS differentiate more finely w.r.t. to age on OC onset. Therefore, there is a necessity to have (at least 2) different OC types. Further important risk factors can be discerned as combinations of the basic f_i: two BC cases $\{b_1, b_2\}$, two OC cases $\{o_1, o_2\}$, and other combinations mentioned in Column 1 of Table 4. The corresponding probabilities for these risk factors in case of m_1 and m_2 are shown in Columns 2-3. They are inspired by the first column and the first row of Table 1 for m_1 and m_2, respectively. The mass of the subsets of Ω not mentioned in Table 4 is supposed to be zero. Using Dempster's rule from Eq. (3), the combined BPA for m_1 and m_2 can be computed (cf. Column 4). After that, applying the definition in Eq. (2) provides the corresponding values for the belief function of the combined BPA m_D (Column 6). It is possible to use intervals inside the resulting mass m_D (cf. Column 5). For example, we can capture a greater age span by using the interval $m_{ID}(\{ba\}) = [0.036, 0.076]$. Then, the mass of the last element Ω should be adjusted to $m_{ID}(\Omega) := [0.121, 0.161]$. For reference, the corresponding combined probabilities from the same Frank table are given in the last column where available.

To show how these results can be used, we consider the following examples. Suppose the patient's father was diagnosed with BC at less than 50 and the patient has BC at over 50, then the belief function corresponding to the modeled mp in the patient (the best case) can be computed from the combined BPA as $Bel_{m_D}(\{bm, nr, b_1, b_2\}) = m_D(\{b_1, b_2\}) + m_D(\{b_1\}) + m_D(\{b_2\}) + m_D(\{bm\}) + m_D(\{nr\}) = 0.257$ (cf. mp= 17%-23% given by the Penn II model for family risk in the same case). In the next example, we assume that the patient is diagnosed OC and BC at age 22 and her mother had bilateral BC at over 50. In this case, the corresponding mp can be computed as

$$Bel_{m_{ID}}(\{b_2, o_1, b_1, nr, oa, ba, bil\}) = m_{ID}(\{o_1\}) + m_{ID}(\{b_2\}) + m_{ID}(\{b_1\}) + m_{ID}(\{nr\}) + m_{ID}(\{bil\}) + m_{ID}(\{b_2, o_1\}) + m_{ID}(\{b_1, b_2\}) + m_{ID}(\{oa\}) + m_{ID}(\{ba\}) = 0.096 + 0.116 + 0.039 + 0.026 + 0.053 + 0.098 + 0.023 + 0.020 + [0.036, 0.76] = [0.507, 0.547]$$

compared to mp of 54% provided by Penn II (family risk).

Factor	m_1	m_2	m_D	m_{ID}	Bel_{m_D}	**[Frank et al. 2002]**
$\{b_1\}$	0.02	0.04	0.039	0.039	0.039	
$\{b_2\}$	0.1	0.04	0.116	0.116	0.116	
$\{o_1\}$	0.07	0.04	0.096	0.096	0.096	
$\{o_2\}$	0.07	0.04	0.096	0.096	0.096	
$\{ba\}$	0.06	0.04	0.056	[0.036,0.076]	0.056	
$\{oa\}$	0.02	0.02	0.02	0.02	0.02	
$\{nr\}$	0.02	0.04	0.026	0.026	0.026	
$\{bil\}$	0.05	0.05	0.053	0.053	0.053	
$\{bm\}$	0.05	0.05	0.053	0.053	0.053	
$\{b_1, b_2\}$	0.02	0.03	0.023	0.023	0.178	0.11
$\{b_1, o_2\}$	0.08	0.05	0.076	0.076	0.211	
$\{b_2, o_1\}$	0.1	0.06	0.098	0.098	0.312	0.341
$\{o_1, o_2\}$	0.12	0.01	0.090	0.090	0.2834	0.277
$\{b_2, o_1, o_2\}$	0.02	0.01	0.017	0.017	0.515	0.535
Ω	0.20	0.48	0.142	[0.121, 0.161]	1	

Table 4: DST based data fusion for nine risk factors

4.2 Merging RST Decision Rules with Multi-Criteria Binary Trees

In this subsection, we consider multi-criteria tests with binary and n-ary risk factor output and augment RST based on binary responses with it to arrive at an extended interval-based categorical counseling test. For this purpose, we rely on a binary decision tree (BDT) proposed in [Zhang and Varshney 1999], cf. Figure 4. There, X denotes an input (sensor) data vector; $U = (u_0, \dots, u_a, \dots)$ a decision/referral vector issued at the lowest BDT level in terminal nodes as a sequence of 0/1 decisions from root to leaf; and AF$(\cdot, \cdot)$ an (accumulated) individual risk function. Each u_a is in the format (interval risk score(s) rs_1,rs_2,..., mp%) and each index a represents the binary decision path as a decimal number, read left to right. At node t, $\Phi(t)$ denotes the set of (interval valued) features/conditions used by the BDT, $\Gamma(t)$ represents the decision rule as a function of the features/conditions with values $0, 1$, and the (interval) function $AF(t, \cdot)$ provides scores rs and mutation probabilities mp%. In this way, it is possible to replace complex one-stage decision procedures, for example, medical diagnoses or results of surveillance systems, by simpler yes/no decisions. We can exchange rules at the inner node level by modifying $\Phi(t)$ or $\Gamma(t)$ (alternative/new rules). At the leaves level, the entries of the decision vector can be combined to a final decision [Zhang and Varshney 1999] if required.

The developed extended referral screening tool ERST is shown in Figure 5. We combine the binary decision structure of RST with the features of FHAT, MSS and Frank tables in an accumulated interval risk function. The resulting vector U contains the eight components u_0, u_1, u_2, u_3, u_6, u_8, u_9 and u_{12} (with and without the alternative rules in orange).

Since the RST documentation acknowledges that the tool does not differentiate in its risk factor BC between cancers on the maternal and paternal side in the manual scoring system, we understand the criterion on the third level of the tree in Figure 5 in such a

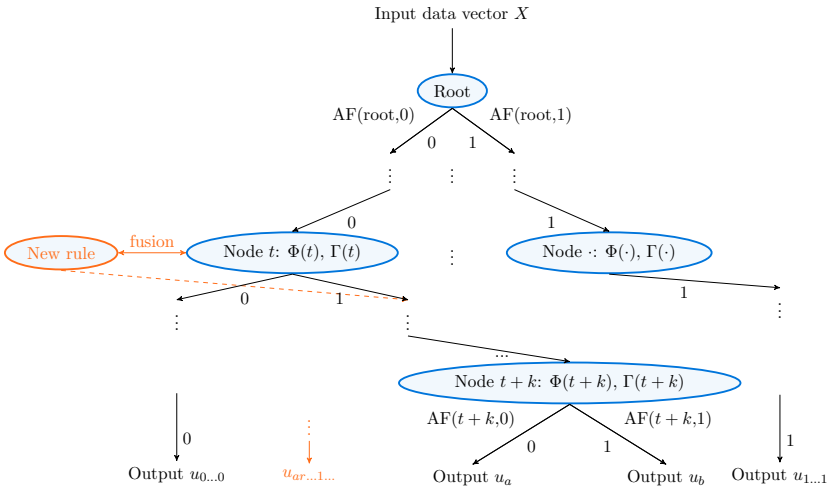

Figure 4: A multi-criteria binary decision tree

way that two and not more BC diseases happen after age 50. This is also the assumption used in our bound on the risk function.

The FHAT and MSS scores are validated using results given in Frank tables (cf. Section 3). The percentages of positively tested individuals in eight paths and five referral categories represent lower limits for the assigned risk, which would increase if further information on lower age time spans for the diseases were available.

To understand how the method works, let us consider the computation for the path 0001 from the ERST tree. The decision path corresponds to the following patient's responses: no Jewish ancestry, no male BC at any age in any relative, no cases of BC at age over 50, and either two cases of BC under 50 or two cases of OC or a case of BC under 50 combined with a case of OC. The corresponding mp from [Frank et al. 2002] are 46/419; 89/484; 5/25; 58/354; 12/118; 19/87; 34/194; 14/41; 1/9; 10/117; 1/18; 23/83; 1/6, so that the total amounts to 313/1955 or 16%. (From Table 1, the individual combinations of risk factors contributing to this total can be read.) It is also possible to consider the corresponding interval between the smallest (1/18) and the largest (19/87) percentage: in this case, it amounts to [5.5, 34.2]% using outwards rounding to the first place after the decimal point. The intervals [6,16] for MSS and [8,26] for FHAT are obtained by using interval operations at each node.

In principle, it is possible to compare the predictions derived from the FH of different families using Claus, Frank tables, and the Penn II model. Here, the Frank tables could be seen as corresponding to ground truth since they contain frequencies of observed mutations. Penn II is a mathematical model giving predictions on those frequencies. However, the entries for it (available online) are more detailed. That is, it is preferable to use all the available material for validation. As a rule, the corresponding percentages in the ERST tree nodes obtained using Frank tables are lower than those given by the Penn II model. Consider, for example, the decision path 0011, for which we can obtain that 90 out of 949 persons are positively tested according to the same principle as in the first example, that is, the mean percentage of 9.5% is recorded in [Frank et al. 2002]. However, if the OC in the last criterion is interpreted as "OC in the proband", the probability is

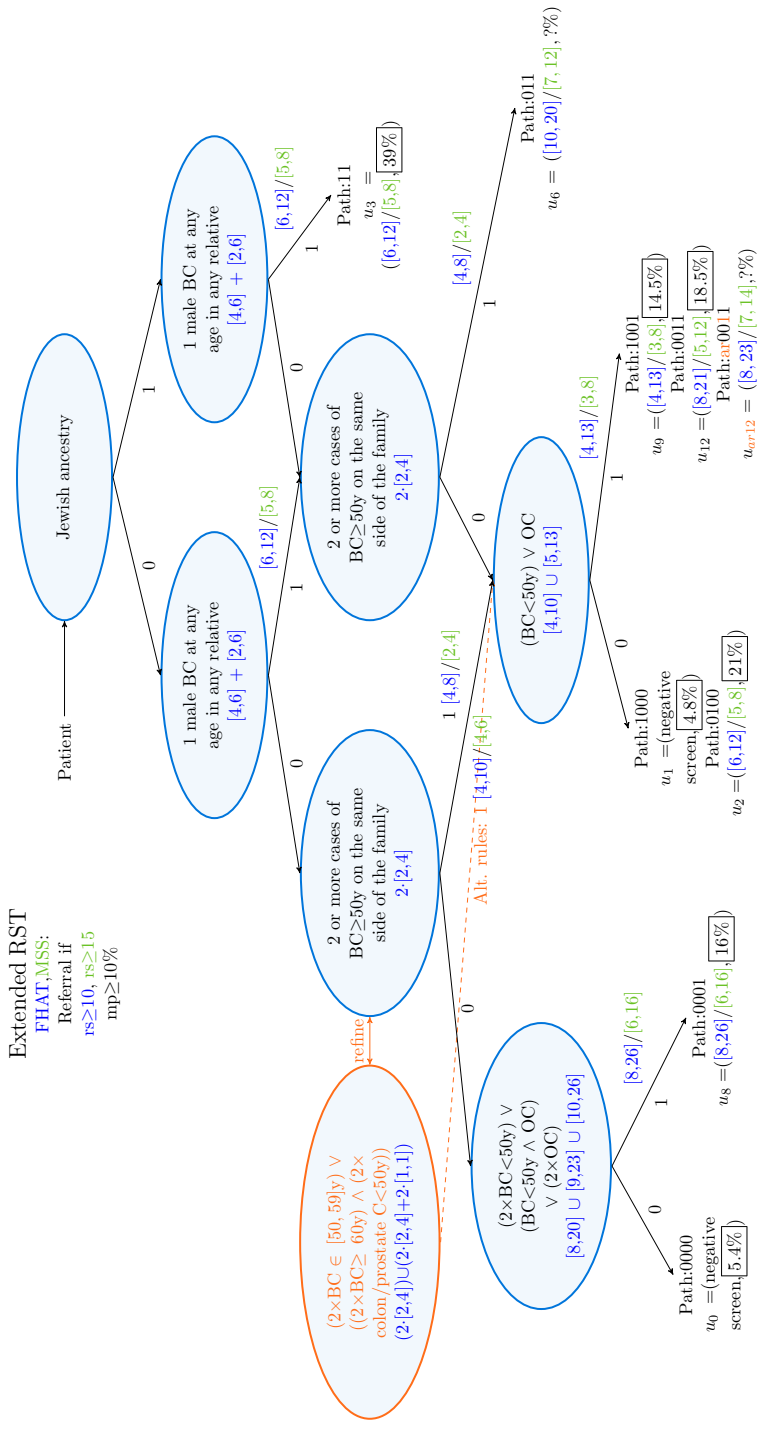

Figure 5: ERST: RST augmented with FHAT, MSS and Frank tables. (The notation ?% means that the corresponding mp cannot be given due to the lack of data). An alternative rule in orange replaces Rule 1 at level 3, the alternative result appearing in orange at the bottom right

18.5%.

Penn II provides higher probabilities for individual and family risk of *BRCA1/2* mutation: 21%/4%. Based on Mendelian genetics as realized in Penn II, the individual risk of a genetic defect for the proband without cancer diagnosis is reduced by half in comparison to the risk of a first degree relative (with a cancer diagnosis). The risk diminishes exponentially in comparison to that of more distant relatives.

4.3 Risk Category Assignment

In this section, we provide risk perception metrics which allow us to assign patients a suitable risk category based on FHAT scores using [Bellcross et al. 2015]. There, the authors randomly selected 3307 subjects from a group of 16720 eligible women within the Henry Ford Health System and interviewed 2524 of them for the purpose of discerning risk categories and subsequent genetics related healthcare activities. The participants were categorized into average, elevated/moderate, and high risk groups based on a series of personal and FH risk factors, US Preventive Services Task Force specifications and expert opinions. The test had again a fairly binary structure and used risk factors similar to those which we relied on in this paper (cf. Figure 6).

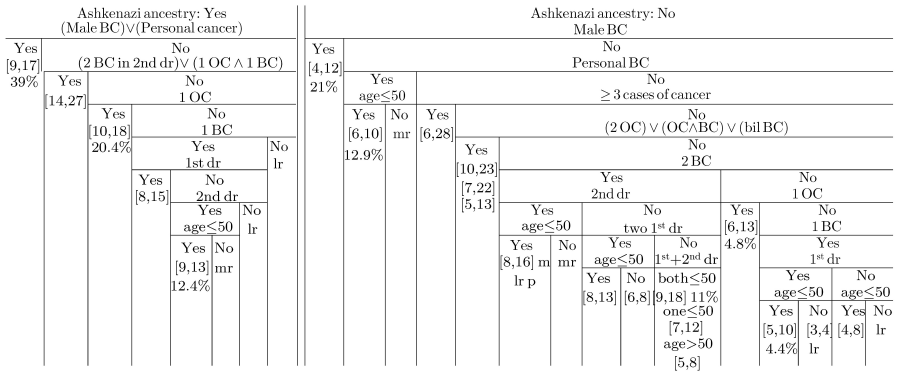

Figure 6: Decision tree to assign participants to low (lr, $[0, 6]$*), moderate (mr,* $[5, 13]$*) and high risk (*$[7, 16]$*,* $[9, 18]$*,* $[10, 23]$*,* $[14, 28]$*) categories based on FHAT scores (dr="degree relative", m="maternal", p="paternal")*

The risk categories proposed in [Bellcross et al. 2015] exhibit inconsistencies pointed out in [Owens et al. 2019]. Since there is no binding standard for determining the risk classes, the study in [Bellcross et al. 2015] cannot be expected to provide generally valid statements about them. However, it would be beneficial to replace the reference to a list of disease cases (criteria) by a comprehensible risk classification based on one or more validated RA tools and using actual probability intervals for *BRCA1/2* mutations. In Figure 6, we attempt to systematize the findings from [Bellcross et al. 2015] according to these principles and compute the corresponding FHAT scores using IA at the terminal leaves of the decision tree similarly to Section 4.2. (Where available, we also cite the corresponding mp% from Frank tables.) It can be seen from the computed FHAT intervals

that the three high risk categories H1/H2/H3 from the study are essentially the same as the elevated moderate risk categories M4/M5/M6 employed there. Moreover, several categories can be merged, for example, the lowest moderate risk category M1 and the low risk LR. Since there is no standard at the moment, we propose to take the classes $[0, 6]$ for low, $[5, 13]$ for moderate and the classes $[7, 16]$, $[9, 18]$, $[10, 23]$, $[14, 28]$ for high risk. A person belongs to a given risk class if the interval FHAT score computed for her according to Figure 6 is inside the interval assigned to this class.

4.4 Summary

In Section 4, we augmented known RA and GC methods by taking into account uncertainty, validated them using patient data and mathematical risk models, and identified accurate risk classes. Combining RST decision rules with multi-criteria binary trees allowed us to merge different RA proposals into a simple decision logic and to compare/validate the resulting risk classes. The application of the DST balanced the epistemic uncertainty in the Frank tables.

In particular, we demonstrated first how to combine data on probands' mutation probabilities (personal risks) and those of their family members (family risks) to obtain the joint mutation probability associated with each of the considered major risk factors with the help of the DST in Section 4.1. The uncertainty in mutation probabilities could be taken into account by using intervals. Based on that, a belief function (a lower bound) on the probability of any combination of this risk factors could be constructed for each individual patient's situation, which could augment the data from literature. Where possible, the corresponding probabilities from Frank tables were supplied as validation. Second, we combined the information from the three popular tools for risk assessment used in genetic counseling FHAT, MSS and RST based on intervals in order to ensure that the epistemic uncertainty present in the methods was understood by the patient (cf. Section 4.2). The provided intervals were supplemented by the data on the mp from Frank tables, if possible, also in this case. Finally, we made a suggestion for an improvement of the risk classification categories that took into account the underlying uncertainty in Section 4.3. The impact of epistemic uncertainty was found to be rather dramatic which necessitated using large overlapping intervals for the risk classes.

As can be seen from Figure 5, not all mp can be derived directly from the available data (e.g., Frank tables). One possibility to improve the situation is to use the Dempster-Shafer theory to augment the available information, which is feasible only if the basic mass probability can be obtained from somewhere. Additionally, the manual derivation of the intervals in the ERST can be error-prone. Therefore, a very important direction for the future work is to automate the procedure from the points of view of both data acquisition (are there any freely accessible data bases? are the data properly cleaned up and standardized?) and data synthesis (can we automate the data propagation through the rules given data on risk factors?).

5 Conclusions and Future Work

In this paper, we extended well-established tools from the field of genetic risk assessment and counseling by an additional (epistemic) uncertainty handling. To model uncertainty in data ranges and its propagation, we used IA in combination with DST and multi-criteria decision trees and paid attention to ways of merging information across different available RA tools and probability tables. This simplifies qualified genetic counseling and

handling of results from genetic testing since complex one-stage decision procedures can be replaced by several simpler yes/no stages in this way. In addition, the extended referral screening tool ERST proposed here takes into account scores from several current RA tools and supplements them by data on mutation probabilities from Frank tables which can heighten a patient's confidence in the outcome of a genetic counseling session. Finally, we demonstrated how missing or conflicting information on mutation probabilities could be completed through the application of DST.

While working on this paper, we came to the conviction that requirements for genetic risk evaluation in the counseling and testing process for individuals and families affected by breast and ovarian cancer caused by pathogenic mutations should address the following key points.

1. Each involved discipline should recognize the need for standardized evaluation and V&V of processes and their models as well as for fusion of conclusions from their outcome. For this, appropriate quality criteria and their metrics as well as standardized procedures need to be used.

2. Validation of numerical results according to robust and comparable criteria is impeded by the fact that existing methods usually account only for aleatory uncertainty and tend to disregard other kinds (e.g., epistemic or mixed uncertainty).

3. Gathering of data is not standardized at the moment, for example, participants might have different ethnicities or data might be incomplete w.r.t. patient's origin, age, type of cancer and genetic mutation, first/repeated occurrence, count (bilateral, multifocal), and family history. Moreover, different studies rely on mismatched numbers of participants or take place at different times. That is, calibration is necessary.

4. FHAT and MSS are difficult to assess with data from the Frank tables (e.g., because there are very small numbers of subjects for certain disease constellations in the tables). We observe that breast cancer for age over 50 (e.g., in the family history) is often not considered explicitly although used somehow inside the studies.

From this point of view, our paper reveals the great impact of epistemic uncertainty reflected through very large overlapping intervals for the risk classes. This uncertainty can only be reduced by employing finer scales for the parameters, their completeness, and appropriate standardized data collection.

The goal of further work should be to clarify whether such standardized data already exist and how to access it. Moreover, it is interesting to investigate existing data w.r.t. whether and how it can be cleaned up automatically, so that the entries can again be subdivided by genetic defects and age groups, and to make it available to the public research. After that, better models can be built, calibrated, and validated using ground truth; results can be properly assessed according to standardized criteria and metrics; and the recommendations of the experts can be feasibly merged into proposals for risk management strategies that are understandable and not harmful for the patients as suggested in [Peshkin and Isaacs 2020].

References

[Amir et al. 2010] Amir, E., Freedman, O. C., Seruga, B., & Evans, D. Gareth. 2010. Assessing Women at High Risk of Breast Cancer: A Review of Risk Assessment Models. J. Natl. Cancer Inst., 102(10), 680–691.

[Ashton-Prolla et al. 2009] Ashton-Prolla, P., Giacomazzi, J., Schmidt, A. V., Roth, F. L., Palmero, E. I., K., Luciane, A., Ernestina S., Moreira, S. M., Batassini, E., Belo-Reyes, V., Schuler-Faccini, L., Giugliani, R., Caleffi, M., & Camey, S. Alves. 2009. Development and validation of a simple questionnaire for the identification of hereditary breast cancer in primary care. BMC Cancer, 9(1), 283.

[Auer and Luther 2020] Auer, E., & Luther, W. 2020. Validation and Verification Assessment of Genetic Counseling and Testing. Pages 61–72 of: Hajian, A., Baloian, N., Inoue, T., & Luther, W. (eds), Collaborative Technologies and Data Science in Artificial Intelligence Applications. Logos Verlag Berlin. Extended abstract.

[Auer et al. 2013] Auer, E., Chuev, A., Cuypers, R., Kiel, S., & Luther, W. 2013. Relevance of Accurate and Verified Numerical Algorithms for Verification and Validation in Biomechanics. Pages 391–398 of: Deodatis, G., Ellingwood, B. R., & Frangopol, D. M. (eds), Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures.

[Auer et al. 2020] Auer, E., Luther, W., & Weyers, B. 2020. Reliable Visual Analytics, a Prerequisite for Outcome Assessment of Engineering Systems. Acta Cybernetica, 24(3), 287–314.

[Bellcross et al. 2009] Bellcross, Cecelia A., Lemke, Amy A., Pape, Laura S., Tess, Angela L., & Meisner, Lorraine T. 2009. Evaluation of a breast/ovarian cancer genetics referral screening tool in a mammography population. Genetics in Medicine, 11(11), 783–789.

[Bellcross et al. 2015] Bellcross, Cecelia A., Peipins, Lucy A., McCarty, Frances A., Rodriguez, Juan L., Hawkins, Nikki A., Hensley Alford, Sharon, & Leadbetter, Steven. 2015. Characteristics associated with genetic counseling referral and BRCA1/2 testing among women in a large integrated health system. Genetics in Medicine, 17(1), 43–50.

[Christian Jr. and Drilling 2009] Christian Jr., D. R., & Drilling, Stephanie. 2009. Implementing Quality in Laboratory Policies and Processes: Using Templates, Project Management, and Six Sigma. CRC Press.

[Claus et al. 1994] Claus, Elizabeth B., Risch, Neil, & Thompson, W. Douglas. 1994. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. Cancer, 73(3), 643–651.

[Dequeker 2017] Dequeker, E. 2017. Quality Assurance in Genetic Laboratories. Pages 493–500 of: Molecular Diagnostics. Elsevier.

[El-Mahassni and White 2015] El-Mahassni, E., & White, K. 2015 (Aug). A Discussion of Dempster-Shafer Theory and its Application to Identification Fusion Standard Form 298 (Rev. 8-98) Prescribed by ANSI Std Z39-18.

[Evans et al. 2004] Evans, D. G. R., Eccles, D. M., Rahman, N, Young, K, Bulman, M, Amir, E, Shenton, A, Howell, A, & Lalloo, F. 2004. A new scoring system for the chances of identifying a BRCA1/2 mutation outperforms existing models including BRCAPRO. Journal of Medical Genetics, 41(6), 474–480.

[Ferson et al. 2003] Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D. S., & Sentz, K.. 2003. Constructing Probability Boxes and Dempster-Shafer Structures. Washington, D.C: United States. Dept. of Energy.

[Frank et al. 1998] Frank, Th. S, Manley, S. A, Olopade, O. I, Cummings, S., Garber, J. E., Bernhardt, B., Antman, K., Russo, D., Wood, M. E., Mullineau, L., et al. 1998. Sequence analysis of BRCA1 and BRCA2: correlation of mutations with family history and ovarian cancer risk. J. Clin. Oncol., 16(7), 2417–2425.

[Frank et al. 2002] Frank, Th. S., Deffenbaugh, A. M., Reid, J. E., Hulick, M., Ward, B. E., Lingenfelter, B., Gumpper, K. L., Scholl, Th., Tavtigian, S. V., Pruss, D. R., & Critchfield, G. C. 2002. Clinical Characteristics of Individuals With Germline Mutations in BRCA1 and BRCA2: Analysis of 10,000 Individuals. J. Clin. Oncol., 20(6), 1480–1490.

[Gilpin et al. 2000] Gilpin, C. A., Carson, N., & Hunter, A. G. W. 2000. A preliminary validation of a family history assessment form to select women at risk for breast or ovarian cancer for referral to a genetics center. Clinical Genetics, 58(4), 299–308.

[Hajian et al. 2020] Hajian, A., Baloian, N., Inoue, T., & Luther, W. 2020. Collaborative Technologies and Data Science in Artificial Intelligence Applications (Proceedings of CODASSCA 2020). Logos Verlag Berlin.

[Himes et al. 2019] Himes, Deborah O., Zaro, Maren L., Williams, Mary, Freeborn, Donna, Eggett, Dennis L., & Kinney, Anita Y. 2019. Breast cancer risk assessment: Evaluation of screening tools for genetics referral. Journal of the American Association of Nurse Practitioners, 31(10).

[IEEE 2017] IEEE. 2017. IEEE Standard for System, Software, and Hardware Verification and Validation. IEEE Std 1012-2016 (Revision of IEEE Std 1012-2012/ Incorporates IEEE Std 1012-2016/Cor1-2017), Sept, 1–260.

[Interrante et al. 2017] Interrante, M. K., Segal, H., Peshkin, B. N., Valdimarsdottir, H. B., Nusbaum, R., Similuk, M., DeMarco, T., Hooker, G., Graves, K., Isaacs, C., Wood, M., McKinnon, W., Garber, J., McCormick, Sh., Heinzmann, J., Kinney, A. Y., & Schwartz, M. D. 2017. Randomized Noninferiority Trial of Telephone vs In-Person Genetic Counseling for Hereditary Breast and Ovarian Cancer: A 12-Month Follow-Up. JNCI Cancer Spectrum, 1(1).

[ISO 2005] ISO. 2005. ISO 17025:2005 General requirements for the competence of testing and calibration laboratories.

[Kast et al. 2014] Kast, K., Schmutzler, R. K., Rhiem, K., Kiechle, M., Fischer, C., Niederacher, D., Arnold, N., Grimm, T., Speiser, D., Schlegelberger, B., Varga, D., Horvath, J., Beer, M., Briest, S., Meindl, A., & Engel, C. 2014. Validation of the Manchester scoring system for predicting BRCA1/2 mutations in 9,390 families suspected of having hereditary breast and ovarian cancer. Int J Cancer, 135(10), 2352–2361.

[Li et al. 2018] Li, D., Li, Y., Ye, Zh., Ibrahim, M., Ryu, K. H., & Jeong, S. Ph. 2018. An Effective Risk Factor Detection and Disease Prediction (RFD-DP) Model Applied to Hypertension. J. Univers. Comput. Sci., 24(9), 1192–1216.

[Li et al. 2019] Li, M., Bennette, C. S., Amendola, L. M., Ragan Hart, M., Heagerty, P., Comstock, B., Tarczy-Hornoch, P., Fullerton, S. M., Regier, D. A., Burke, W., Trinidad, S. B., Jarvik, G. P., Veenstra, D. L., & Patrick, D. L. 2019. The Feelings About genomiC Testing Results (FACToR) Questionnaire: Development and Preliminary Validation. Journal of Genetic Counseling, 28(2), 477–490.

[Louro et al. 2019] Louro, J., Posso, M., Hilton Boon, M., RomÃ¡n, M., Domingo, L., Castells, X., & Sala, M.. 2019. A systematic review and quality assessment of individualised breast cancer risk prediction models. British Journal of Cancer, 121(1), 76–85.

[Markov and Akyildiz 1996] Markov, S. M., & Akyildiz, Y. 1996. Curve Fitting and Interpolation of Biological Data Under Uncertainties. J. Univers. Comput. Sci., 2(2), 58–69.

[Mattocks et al. 2010] Mattocks, Ch. J., Morris, M., Matthijs, G., Swinnen, E., Corveleyn, A., Dequeker, E., Mueller, C., Pratt, V., & Wallace, A. 2010. A standardized framework for the validation and verification of clinical molecular genetic tests. Eur. J. Hum. Genet., 18, 1276–1288.

[Moore et al. 2009] Moore, R. E., Kearfott, R. B., & Cloud, M. J. 2009. Introduction to Interval Analysis. Philadelphia: Society for Industrial and Applied Mathematics.

[Nelson et al. 2013] Nelson, H. D., Fu, R., Goddard, K., Mitchell, J. P., Okinaka-Hu, L., Pappas, M., & Zakher, B.. 2013. Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer: Systematic Review to Update the U.S. Preventive Services Task Force Recommendation. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US).

[Nelson et al. 2019] Nelson, H. D., Pappas, M., Cantor, A., Haney, E., & Holmes, R. 2019. Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer in Women: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA, 322(7), 666–685.

[Owens et al. 2019] Owens, D. K., Davidson, K. W., Krist, A. H., Barry, M. J., Cabana, M., Caughey, A. B., Doubeni, Ch. A., Epling Jr., J. W., Kubik, M., Landefeld, C. S., Mangione, C. M., Pbert, L., Silverstein, M., Simon, M. A., Tseng, Ch.-W., & Wong, J. B. 2019. Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer: US Preventive Services Task Force Recommendation Statement. JAMA, 322(7), 652–665.

[Panchal et al. 2008] Panchal, S. M., Ennis, M., Canon, S., & Bordeleau, L. J. 2008. Selecting a BRCA risk assessment model for use in a familial cancer clinic. BMC Medical Genet., 9(1), 116.

[Parmigiani et al. 2007] Parmigiani, G., Chen, S., Iversen Jr., E. S., Friebel, T. M., Finkelstein, D. M., Anton-Culver, H., Ziogas, A., Weber, B. L., Eisen, A., Malone, K. E., Daling, J. R., Hsu, L., Ostrander, E. A., Peterson, L. E., Schildkraut, J. M., Isaacs, C., Corio, C., Leondaridis, L., Tomlinson, G., Amos, Ch. I., Strong, L. C., Berry, D. A., Weitzel, J. N., Sand, Sh., Dutson, D., Kerber, R., Peshkin, B. N., & Euhus, D. M. 2007. Validity of models for predicting BRCA1 and BRCA2 mutations. Ann. Intern. Med., 147(7), 441–450.

[Peshkin and Isaacs 2020] Peshkin, B. N., & Isaacs, C. 2020. Genetic testing and management of individuals at risk of hereditary breast and ovarian cancer syndromes. Wolters Kluwer.

[Richardson et al. 2020] Richardson, M., Min, H. J., Hong, Q., Compton, K., Mung, S. W., Lohn, Z., Nuk, J., McCullum, M., Portigal-Todd, Ch., Karsan, A., Regier, D., Brotto, L. A., Sun, S., & Schrader, K. A. 2020. Oncology Clinic-Based Hereditary Cancer Genetic Testing in a Population-Based Health Care System. Cancers, 12(2).

[Santos et al. 2016] Santos, P., Dennerlein, Se., Theiler, D., Cook, J., Treasure-Jones, T., Holley, D., Kerr, M., Attwell, G., Kowald, D., & Lex, E. 2016. Going beyond your personal learning network, using recommendations and trust through a multimedia question-answering service for decision-support: A case study in the healthcare. J. Univers. Comput. Sci., 22, 340–359.

[Schwer 2007] Schwer, L. E. 2007. An Overview of the PTC 60/V&V 10: Guide for Verification and Validation in Computational Solid Mechanics. Eng. Comput. (Lond.), 23(4), 245–252.

[Smith 2013] Smith, R. C. 2013. Uncertainty Quantification: Theory, Implementation, and Applications. USA: Society for Industrial and Applied Mathematics.

[Stoppa-Lyonnet 2016] Stoppa-Lyonnet, L. 2016. The biological effects and clinical implications of BRCA mutations: Where do we go from here? Eur. J. Hum. Genet., 3–9.

[Struewing et al. 1995] Struewing, J. P., Abeliovich, D., Peretz, T., Avishai, N., Kaback, M. M., Collins, F. S., & Brody, L. C. 1995. The carrier frequency of the BRCA1 185delAG mutation is approximately 1 percent in Ashkenazi Jewish individuals. Nature Genetics, 11(2), 198–200.

[Tao et al. 2019] Tao, X., Liu, L., Zhao, F., Huang, Y., Zhu, S., & Liang, Y. 2019. Ontology and Weighted D-S Evidence Theory-Based Vulnerability Data Fusion Method. J. Univers. Comput. Sci., 25(3), 203–221.

[Teller et al. 2010] Teller, P., Hoskins, K. F., Zwaagstra, A., Stanislaw, C., Iyengar, R., Green, V. L., & Gabram, S. G. A. 2010. Validation of the Pedigree Assessment Tool (PAT) in Families with BRCA1 and BRCA2 Mutations. Ann. Surg. Oncol., 17(1), 240–246.

[Weyers et al. 2019] Weyers, B., Auer, E., & Luther, W.. 2019. The Role of Verification and Validation Techniques within Visual Analytics. J. Univers. Comput. Sci., 25(8), 967–987.

[Yager and Liu 2008] Yager, R. R., & Liu, L. (eds). 2008. Classic Works of the Dempster-Shafer Theory of Belief Functions. Studies in Fuzziness and Soft Computing, vol. 219. Springer.

[Ye et al. 2018] Ye, L., Lee, T.-Sh., & Chi, R.. 2018. A Hybrid Machine Learning Scheme to Analyze the Risk Factors of Breast Cancer Outcome in Patients with Diabetes Mellitus. J. Univers. Comput. Sci., 24(6), 665–681.

[Zadeh 1996] Zadeh, L. A. 1996. Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh. USA: World Scientific Publishing Co., Inc.

[Zhang and Varshney 1999] Zhang, Q., & Varshney, P. 1999. Towards the Fusion of Distributed Binary Decision Tree Classifiers.