

Uncertainty quantification in breast cancer risk prediction models using self-reported family health history

Lance T. Pflieger¹, Clinton C. Mason² and Julio C. Facelli^{1*}

¹ Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

² Department of Pediatrics, University of Utah, Salt Lake City, UT, USA

Journal of Clinical and Translational Science (2016), 1, pp. 53–59 doi:10.1017/cts.2016.9

Introduction. Family health history (FHx) is an important factor in breast and ovarian cancer risk assessment. As such, multiple risk prediction models rely strongly on FHx data when identifying a patient's risk. These models were developed using verified information and when translated into a clinical setting assume that a patient's FHx is accurate and complete. However, FHx information collected in a typical clinical setting is known to be imprecise and it is not well understood how this uncertainty may affect predictions in clinical settings.

Methods. Using Monte Carlo simulations and existing measurements of uncertainty of self-reported FHx, we show how uncertainty in FHx information can alter risk classification when used in typical clinical settings.

Results. We found that various models ranged from 52% to 64% for correct tier-level classification of pedigrees under a set of contrived uncertain conditions, but that significant misclassification are not negligible.

Conclusions. Our work implies that (i) uncertainty quantification needs to be considered when transferring tools from a controlled research environment to a more uncertain environment (i.e. a health clinic) and (ii) better FHx collection methods are needed to reduce uncertainty in breast cancer risk prediction in clinical settings.

Received 16 September 2016; Accepted 11 October 2016; First published online 20 January 2017

Key words: Family health history, Risk prediction models, Breast and ovarian cancer, Monte Carlo simulations.

Introduction

Studies have shown a strong association between family health history (FHx) and disease susceptibility for a number of high-frequency chronic conditions such as diabetes, cardiovascular disease, and various cancers [1, 2]. FHx is comprised of information on a relative's demographics, disease diagnosis, and age of onset in a pedigree often tracing up to 3 generations [3]. The record helps to characterize a combination of shared environmental factors, genetic susceptibility, and common behaviors to provide an independent variable for risk analysis [4]. These associations can then facilitate patient risk stratification based solely on FHx. For example, patients with an FHx that meet certain disease-specific guidelines can be characterized as either moderate risk or high risk. Patients identified as high risk can then

receive earlier and/or more frequent screenings, recommendations for behavioral changes in health management, and other evidence-based measures for prevention of the identified disease [5].

The American Cancer Society (ACS) has published guidelines to identify patients at high risk for breast and ovarian cancers [6]. These guidelines also provide recommendations informing genetic testing and screening for early detection and prevention. Specifically, the ACS established breast cancer screening guidelines that include annual screening mammography and magnetic resonance imaging (MRI) for patients with known BRCA gene mutations and those with an approximate lifetime risk of 20% or greater. To calculate lifetime risk, the ACS recommends the use of risk assessment models that rely heavily on FHx [ie, Claus model [7], the Tyrer-Cuzick model [8], BRCAPRO [9], and the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) [10, 11] over those that do not. Although each of these models rely on FHx as input, they are derived using various methods and populations, resulting in each model stratifying a unique group of patients into each risk category [12].

These models have been developed and validated for highly controlled cohorts in which all data have been verified from original sources [13–15]. Hence, the models assume that FHx information provided for risk calculation is accurate. However, the clinical application of risk

* Address for correspondence: J. C. Facelli, Ph.D., Department of Biomedical Informatics, University of Utah, 421 Wakara Way, Suite 140, Salt Lake City, UT 84112, USA.
(Email: julio.facelli@utah.edu)

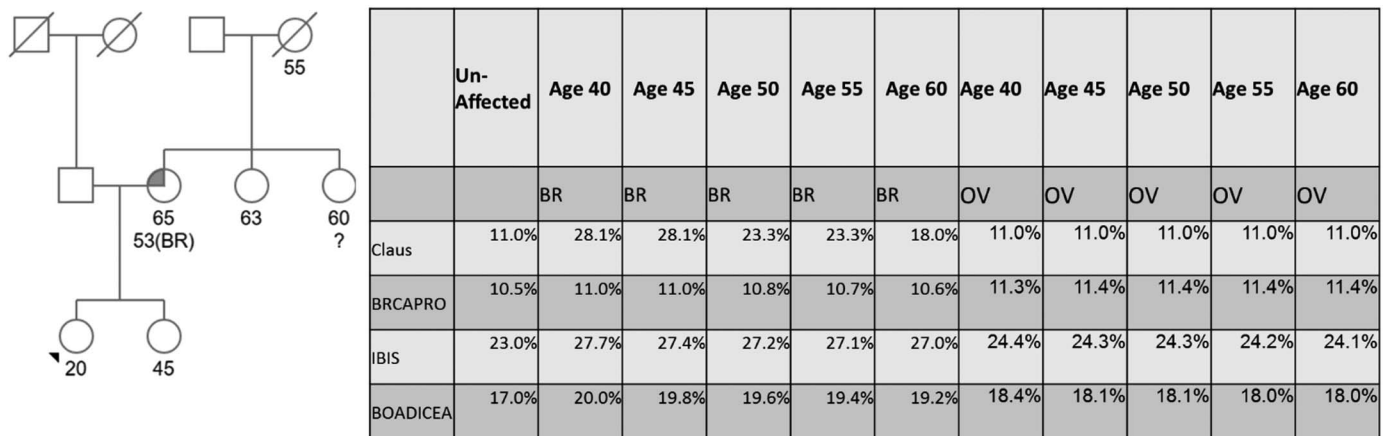


Fig. 1. Sensitivity in lifetime risk estimates of various models to uncertainty in an example pedigree. A hypothetical situation where a proband is assessing her lifetime risk for breast cancer (BR) based on her family knowledge. The proband is marked by the triangle and lifetime risk (risk of developing cancer from age 20 to age 80) is assessed by each model. The proband has a mother with BR with age at onset of 53 years and is uncertain of the cancer status of the 60-year-old aunt. The table on the right shows how each tool evaluates the proband's lifetime risk under the scenario of the 60-year-old maternal aunt being unaffected for BR or ovarian cancer (OV) in the first column, followed by proband's lifetime risk for BR under the perturbation that the aunt had affected status for BR or OV at various onset ages. IBIS, International Breast Intervention Study; BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm.

prediction models should allow for realistic data inaccuracies in a clinical setting. For instance, it has been observed that FHx information collected in a clinical setting is imprecise, with the most frequently used method to obtain FHx, self-reporting, as a common cause of error [16]. Therefore, the translational value of these models should be evaluated in light of such practical limitations when translating these population science discoveries from the research environment to the clinic.

Studies on self-reported family history have shown that inaccuracies in FHx of general diseases can range from 10% to 70%, depending on disease and degree of relatedness [17]. Self-reported FHx error originates from a lack of family history knowledge or recall bias during reporting. These errors, which may lead to uncertainty in the risk prediction, have been shown to vary based on the demographics of the patient, type of disease, and degree of the relative. For example, a false negative in the family history may result in an underestimation of risk, creating a missed opportunity for proper care [18]. Alternatively, a false positive in the family history may result in overestimation of risk, resulting in unnecessary, expensive and/or risky procedures, and unnecessary referrals for genetic testing or counseling (Fig. 1) [19]. For breast cancer, it has been previously shown that the effect of uncertain family history can seriously distort carrier probabilities and therefore lifetime risk estimates using the BRCAPRO model [20]. However, the effect of this distortion on clinical guidelines, as well as on other risk prediction models, has not been quantified.

Considering the uncertainty surrounding self-reported FHx, we focus our work in assessing its effect on breast cancer risk prediction models. It is important to quantify this effect as it aids in gaining a better understanding of how to transition complex prediction models from the research environment to the clinic, which can have an important impact on clinical guideline development and determine potential patient treatment. We present a general approach for this translational issue using a general Monte Carlo (MC) approach to quantitatively assess the uncertainty factors discussed above. The aim of this study was to develop a framework to evaluate commonly used risk assessment models heavily reliant on family history under uncertain conditions. The methods developed are quite general and can be applied to any biomedical predictive model including other cancer-related applications.

Methods

A comprehensive experimental design to estimate the effect of uncertainty in self-reported FHx on breast cancer risk classification

requires an adequately large number of pedigrees to be tested over a plausible range of uncertain conditions. For each pedigree considered in the analysis, we built a large number of derivative pedigrees or replicas with FHx input data modified according to the distributions expected for their uncertainties as derived from published data. In this study we use the self-reported FHx accuracy assessments from the recent analysis of Tehranifar *et al.* [21], with the understanding that our analysis will be restricted by the inherent limitations in their study. We used these accuracies to build probability distributions for age at onset and affected status of all the members of the pedigree under consideration and used the MC method to sample these distributions [22, 23]. MC simulations provide an effective method for estimating classification error and there is extensive literature for using this approach. Age at diagnosis was also considered because of its inclusion in all risk models considered here and its importance in determining hereditary cancer. As Tehranifar *et al.* did not include validation of age at diagnosis, we used data from Schneider *et al.* [24] showing 53% of reported ages at onset were within 5 years for relatives with hereditary breast or ovarian cancer syndrome.

Considering the wide variation in pedigree structures, we contrived simulated pedigrees for each MC simulation with the goal of achieving a wide range of familial risk which could be selected from to evaluate the effect of uncertainty across the risk spectrum [25]. The initial simulated pedigrees were classified by the ACS guidelines for MRI screening adjunct to mammography (Table 1). MC simulations were used for generating replicas of the original pedigrees perturbed to simulate uncertainty. Then, recalculation of the lifetime breast cancer risk using each of the models considered was performed to allow assessment of the effect of uncertainty upon final risk classification in contrast to the original risk classification strata of each pedigree. The effect of uncertainty was summarized for each pedigree by the total percentage of MC replicas whose risk tier was classified correctly (ie, same as

Table 1. American Cancer Society risk classification strata for breast screening

	Lifetime risk (%)	Recommendations
High risk	20–25	Annual MRI screening
Moderate risk	15–20	Insufficient evidence for screening
Low risk	<15	Recommend against MRI screening

MRI, magnetic resonance imaging.

initial pedigree) or misclassified (ie, change of risk-tier classification from initial pedigree).

Risk Prediction Models

As multiple models are available, and each model can potentially perform differently under uncertain conditions, we tested the 4 ACS recommended models that have widespread use: Claus, BRCAPRO, International Breast Intervention Study (IBIS), and BOADICEA [7–10]. For continuity across models we followed the BOADICEA guideline for lifetime risk. Lifetime risk for breast cancer was assessed by setting the age of the proband to 20 years and computing risk at 80 years of age, with the exception of the Claus model, which was measured as the cumulative probability of a woman at 79 years of age.

The Claus model was developed using data from the Cancer and Steroid Hormone Study, a large population-based study with histologically confirmed breast cancer cases and controls [13]. The model is based on segregation analysis for a rare single autosomal dominant allele. For this study we used the complete set of published risk tables, including the subsequently published ovarian cancer tables [26]. These risk tables include combinations of affected first-degree and second-degree relatives, however, some risk combinations are not provided (ie, mother and maternal grandmother). For the missing combinations we used a similar degree of relative combinations to extrapolate risk, comparable with what has been done in other studies [27]. A complete table can be found in online Supplementary Table S1. As the Claus table requires a maximum of 2 affected first-degree and second-degree relatives with age at onset, we found all possible affected combinations for a given pedigree and considered only the highest as the lifetime risk.

BRCAPRO is a Mendelian model that uses Bayesian statistics to obtain a probability of lifetime risk by combining the likelihood ratio that an individual carries a *BRCA1* or *BRCA2* mutation, extrapolated from family history information with mutation prevalence and penetrance data. The variables for BRCAPRO include a pedigree of any degree with ages of breast and ovarian cancer diagnosis, current age of family members, ethnicity with optional germline testing results (*BRCA1/2* positive, negative, or untested), and tumor marker status (Estrogen Receptor, Progesterone Receptor, and Human Epidermal Growth Factor Receptor 2 [Her2/neu]). For this study we used the implementation available in the R package BayesMendel 2.1.1 and excluded all optional variables [28, 29]. The model was used with default penetrance and risk objects.

IBIS, also known as the Tyrer-Cuzick model, was developed using results from the International Breast Intervention Study and a Swedish population study on Familial Breast and Ovarian Cancer [14]. The model is based on the assumption of an underlying gene that leads to breast cancer predisposition in addition to the *BRCA* genes. It uses family history in conjunction with Bayes theorem and Mendelian genetics to estimate the likelihood of a proband carrying any predisposing genes, as well as the likelihood of developing breast cancer. For this study we used version 7 of the windows desktop application. Model variables included affected status and age at onset for breast and ovarian cancer for first-degree and second-degree relatives in addition to half-siblings and affected cousins and nieces. Of note, IBIS also includes many personal health history variables. All other personal health history variables were entered as missing data including age at menarche, parity, age of first child, menopause, menopause age, hormone therapy use, and genetic testing.

The BOADICEA model was developed using complex segregation analysis on 2785 families collected through multiple population-based studies of breast cancer. The model uses probabilities for *BRCA1/2* mutations, as well as a polygenic component that represents the aggregative effects of a large number of genes, to generate risk.

BOADICEA has input values for first-degree, second-degree, and third-degree relatives with affected status for breast, ovarian, prostate, and pancreatic cancers. The model also includes values for genetic testing for *BRCA1* and *BRCA2*, age at onset of cancer, age of relatives, patient ethnicity, and tumor pathology (ER, TN, and basal markers). For the present study, we used a batch mode implementation of BOADICEA based on the version 7 release (personal communication). As with other models, all non-breast and ovarian cancer variables were entered as missing.

These models are fundamentally different as they require unique input, methodologies, and output [30]; however, no additional steps were taken to normalize risk estimates between models. For example, even though the risk estimates for the Claus and IBIS models include both invasive breast cancer and ductal carcinoma in situ, neither BRCAPRO nor BOADICEA include ductal carcinoma in situ in their risk estimates. Although this could potentially lead to outcome differences in terms of economic and health impact, the purpose of this study was to assess uncertainty on each model and not on model validity.

Pedigree Simulation

Custom python scripts were used to simulate pedigrees. Because each risk assessment program requires a unique input format, pedigrees were initially simulated using the BOADICEA format and then converted to the required input format for the other risk assessment programs. Pedigree simulation started with a backbone pedigree of proband (age 20, unaffected), parents, grandparents, and great-grandparents. All other first-degree, second-degree, and third-degree relatives were added randomly within the limitations of the risk assessment tools (eg, IBIS limits the proband to a maximum of 5 sisters, paternal aunts, and maternal aunts). Affected status for only breast cancer and ovarian cancer, along with age at onset, were then added randomly with a 10% chance of breast cancer and 5% incidence of ovarian cancer. Affected status was only applied to females as the accuracies of self-reported FHx for male breast cancer is, to our knowledge, unknown. In addition, breast and/or ovarian cancer was only applied to females over 30 years of age. The lifetime risk of the proband was then calculated using each risk assessment model. Pedigrees were then classified into low-risk, medium-risk, and high-risk categories with 50,000 pedigree simulations in each (50 pedigrees for each risk category with 1000 MC perturbation simulations for each pedigree). All pedigrees were assessed by each model. For the high-risk category, we used a cut-off of 35% in order to model more realistic pedigrees. Fig. 2 shows the distribution of initial risks in the simulated pedigrees used to assess the BOADICEA model. Similar distributions were used for each of the other models.

Sample Size Justification

Before performing the MC simulation, an analysis was done to justify the number of MC replicas needed per pedigree and the number of pedigrees per risk stratification classification that were necessary to achieve stable results. To start, 3 sample pedigrees were run through the simulation using increasingly larger number of replicas, up to 1 million. After each iteration the pedigree replicas were classified; it was found that the proportion of correctly classified simulations stabilized around 500–1000 iterations, depending on pedigree size. Using 1000 iterations, we calculated the 95% confidence interval (CI) around the percentage of correctly classified pedigrees. The CI around a binomial proportion population estimate (p) can be calculated (Equation 1) for a given critical value (z , $z = 1.96$ for a 95% CI), calculated proportion of correct classification (p , p is assessed for each risk tool), and sample size (n , $n = 1000$ in each of our simulations).

$$CI = p \pm z \sqrt{\frac{1}{n} p (1-p)}. \quad (1)$$

The average calculated CI across the 4 models was within ± 0.03 from each calculated proportion for each simulation with 1000 iterations, which is an acceptable value. Similarly, a 95% CI was calculated for each of the risk stratifications using a sample size of 50 pedigrees and the calculated standard deviation.

MC Simulations

MC simulations were performed using 3 independent input variables: affected status for breast cancer, affected status for ovarian cancer, and age of cancer onset. For affected status of each cancer type, binomial distributions were created based on the sensitivities (affected relatives) and specificity (unaffected relatives) from Tehranifar *et al.* [21]. Similar to pedigree generation, MC simulations for affected status was only applied to females over the age of 30 years. For onset age of cancer, a normal distribution was used with mean (μ) of age at onset in the original pedigree and a standard deviation corresponding to ~53% of generated ages being within ± 5 years (eg, $\mu = 70$ y of age and $\sigma = 6.92$). False positives for affected status were assigned an age at onset of their current age in the pedigree. In addition, we used the upper and lower bounds of 95% CI for sensitivities and specificities to generate “best-case” and “worse-case” scenarios, respectively.

The model for uncertainty quantification described above was used for each risk prediction model, using 50 pedigrees initially classified into low-risk, medium-risk, and high-risk categories as set by the ACS. In addition to the “average case” of uncertainty scenario (utilizing the mean sensitivity and specificity values from Tehranifar *et al.* [21]), best-case and worst-case uncertainty scenarios were also created, defined by the upper and lower bounds of the 95% CIs of sensitivity and specificity from that previously published analysis. For each risk prediction tool, a simulated family was tested by generating 1000 pedigree replicas using MC simulations, totaling 50,000 samples per initial ACS risk classification category (150,000 for each). This was then performed for each of the 3 (average, best, and worst) degrees of uncertainty scenarios, totaling 450,000 simulations for each of the 4 tools.

Results

Pedigree Simulation for Each Model

The average number of females in the 150 simulated pedigrees used in this study (all risk categories) was between 25 and 26 depending on the risk prediction model. This includes all first-degree, second-degree, and third-degree relatives with a high of 39 female members and a low of 7. The average number of affected individuals for each degree of relatedness, risk prediction model, and initial risk category is summarized in Table 2. As the initial goal was to simulate pedigrees which spanned the risk range of 10%–35% for each model (see Fig. 2), not all generated pedigrees were identical for each of the various risk prediction models (ie, some pedigrees were unique for a particular model’s assessment). Risk estimates were sometimes observed to vary widely across models for a given pedigree, whereas the effect of uncertainty of age in an affected relative often had only a minimal effect on a model’s overall risk (see Fig. 1).

Classification of Pedigree Replicas with Uncertainty

The resulting risk classification changes due to adding uncertainty to the simulated pedigrees by the initial risk strata are presented in Fig. 3 for the Claus, BRCAPRO, IBIS, and BOADICEA models, respectively. A colored bar represents no change in classification of the original pedigree with uncertainty added, whereas a gray bar represents a change in the classification due to the added uncertainty. The height of the bars indicates the percentage of the replicas from the original classification sorted into the

Table 2. Average number of affected individuals in the simulated pedigrees by model and degree of relatedness/cancer type

	Claus	IBIS	BRCAPRO	BOADICEA
Low risk				
BR-FDR	0.28	0	0.4	0
BR-SDR	0.9	0.26	0.76	0.48
BR-TDR		0.88	1.02	0.78
OV-FDR	0.16	0.2	0.36	0.08
OV-SDR		0.48	0.98	0.54
OV-TDR		0.76	0.88	0.4
Total affected	1.34	2.58	4.4	2.28
Moderate risk				
BR-FDR	0.58	0.08	0.56	0.56
BR-SDR	1.02	0.88	0.98	0.88
BR-TDR		1.12	0.98	1.12
OV-FDR	0.14	0.16	0.6	0.18
OV-SDR		0.48	1.14	0.62
OV-TDR		0.5	1.08	0.64
Total affected	1.74	3.22	5.34	4.00
High risk				
BR-FDR	0.68	0.62	0.66	0.76
BR-SDR	1.64	1.02	1.14	1.46
BR-TDR		0.88	1.12	1.06
OV-FDR	0.46	0.3	0.82	0.34
OV-SDR		0.7	1.54	0.9
OV-TDR		0.74	1.02	0.74
Total affected	2.78	4.26	6.3	5.26

IBIS, International Breast Intervention Study; BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; BR-FDR, first-degree relative with breast cancer; BR-SDR, second-degree relative with breast cancer; BR-TDR, third-degree relative with breast cancer; OV-FDR, first-degree relative with ovarian cancer; OV-SDR, second-degree relative with ovarian cancer; OV-TDR, third-degree relative with ovarian cancer.

The rows labeled “total affected” indicate average number of total affected with either cancer and across all degrees of relatedness.

category indicated in the x-axis under the average uncertainty scenario. In the absence of uncertainty there should be only 1 bar per x-axis category (green for low risk, yellow for moderate risk, and red for high risk) at 100%, that is, all of the pedigree simulations would classify into the same category as the original one. The bars in the figure represent the results for the best case and worst case of uncertainty scenario as defined in the Methods section. The values corresponding to these graphs are presented in online Supplementary Table S2.

Discussion

It is apparent from Fig. 3 that for all risk models considered here, the majority of the pedigree replicas with added uncertainty were classified into the original risk category. The percentage of pedigrees changing risk category was modest, with a 14% average of changed category. Extreme values of 25% reclassification for the high-risk class was observed with the BOADICEA model and a minimum misclassification of only 1% for the high-risk classification was seen with the Claus model. However, the changes did not appear to have any well-defined trend with respect to risk category of risk model, the lack of any extreme outlier changes may indicate that, for practical applications, the selection of the cut-off criteria in the uncertainty distributions of the input parameters is not critical. On the other hand, the results from Fig. 3 clearly indicate that misclassification of pedigrees with uncertainty was not uncommon and that uncertainty in the input parameters does have clinical implications. For instance, in our contrived population of high-risk patients, ~16% would not be

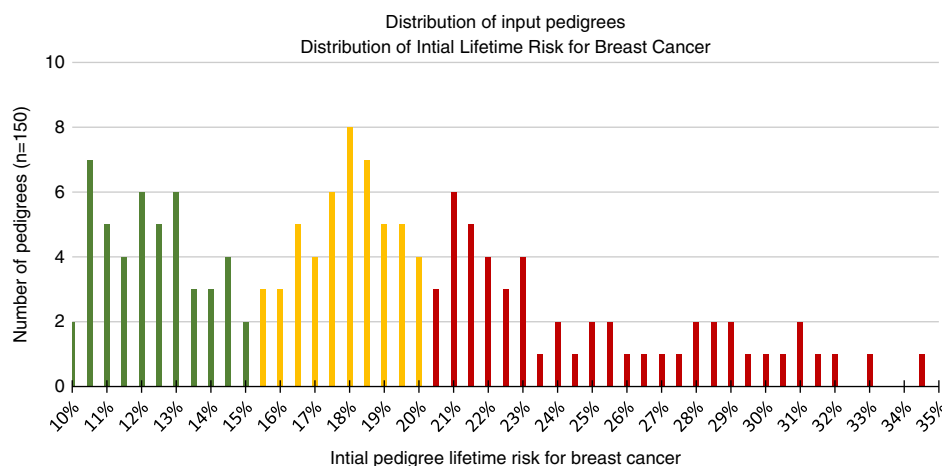


Fig. 2. Distribution of lifetime breast cancer risk for initial simulated pedigrees and risk estimates following uncertainty perturbation calculated by the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) risk prediction model. Distribution of lifetime breast cancer risk for initial simulated pedigrees utilized in assessing the BOADICEA model. Initial risk categories are noted by color (green = low; yellow = moderate; and red = high). Each risk category contains 50 pedigrees.

recommended for advanced breast cancer screening in a clinic using the BOADICEA model with average-case uncertainty. Moreover, our contrived data show that pedigrees with risk classifications closer to risk category cut-offs are more likely to be misclassified due to uncertainty (online Supplementary Fig. S1).

Consistently all the risk models misclassified pedigree replicas of moderate risk at much higher rates than the low-risk and high-risk categories, with the latter showing less misclassifications in most cases, with

BRCAPRO being the exception (Table 3). For the average misclassifications it appears that the Claus and BOADICEA risk models are less sensitive to uncertainty on the input parameter, but the level of misclassification for all the risk models considered here are non-negligible and their clinical consequences should be carefully considered.

Fig. 3 shows that uncertainty in self-reported family history has a non-negligible effect on risk classification regardless of the risk model used. Although this effect differs from model to model and a straight

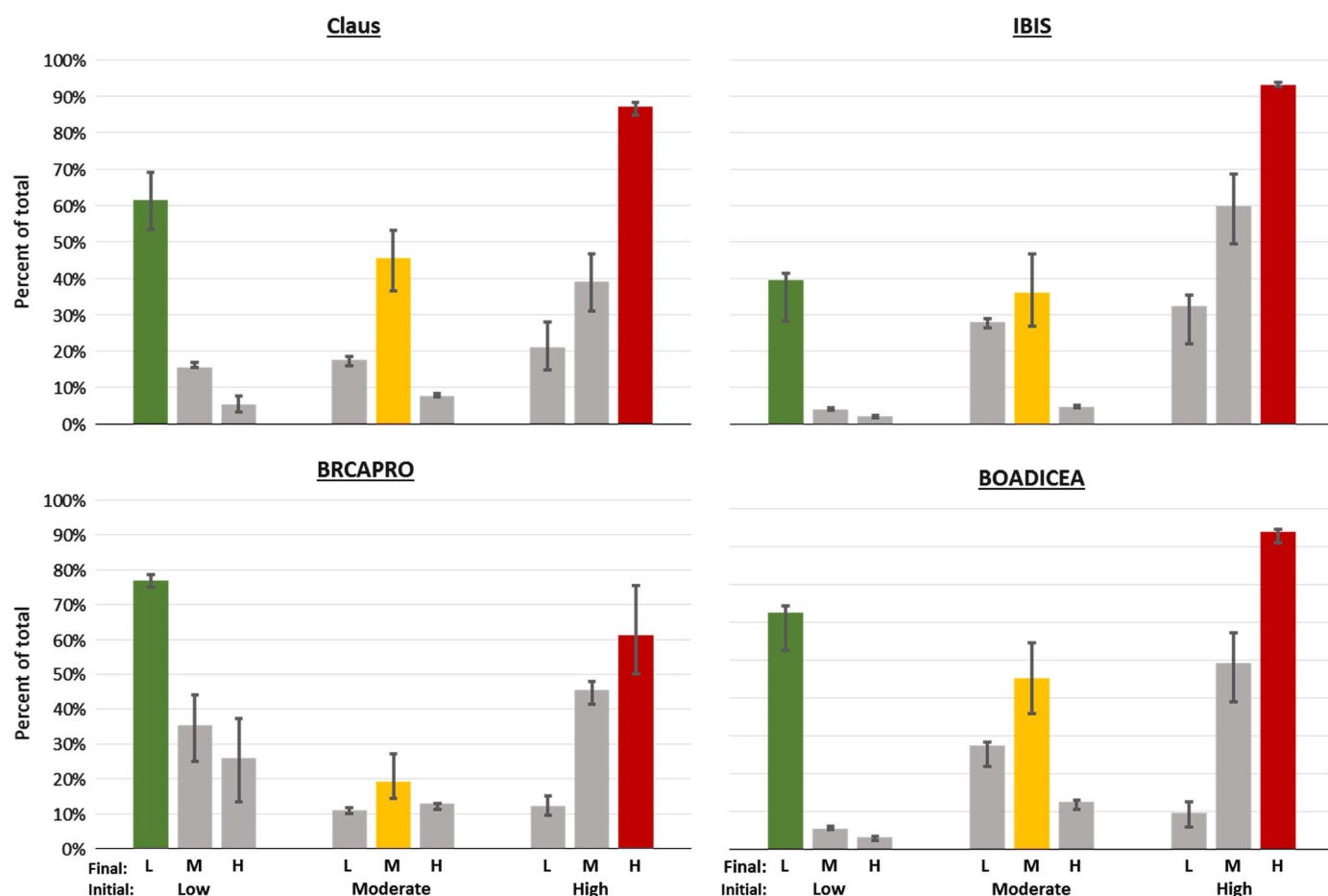


Fig. 3. Effect of uncertainty on initial versus final risk classification for each model. Bars show percentage of pedigree classification from the initial risk strata to the final risk strata (L, low; M, moderate; H, high) determined after adding uncertainty to the initial pedigree according to the average case of uncertainty scenario. Colored bars represent no change in classification, gray indicates a change in classification. Upper-bound and lower-bound bars show best-case and worst-case of uncertainty scenarios. IBIS, International Breast Intervention Study; BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm.

Table 3. Frequency of pedigree replicas that remained in the same risk-tier classification following perturbation with uncertainty

Risk models	Risk category			Average (%)
	Low risk (%)	Moderate risk (%)	High risk (%)	
Claus	61	45	87	64
BRCAPRO	77	19	61	52
IBIS	40	36	93	56
BOADICEA	63	45	84	64

BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; IBIS, International Breast Intervention Study.

across comparison is difficult to make as models take into account different variables for risk calculation and use differing methodologies, none of the models considered here appear to be unquestionably more robust to uncertainty of input parameters.

As mentioned in the Methods section, the IBIS model does not account for unaffected cousins and nieces. Considering that the specificity of breast cancer for third-degree relatives is 83% for the average-case scenario [21], it is possible that many false negatives will be included in a large pedigree, with multiple cousins. Overall, this tends to push lifetime risk estimates for the IBIS model into high-risk categories as can be observed in Fig. 3 as ~50% of low-risk pedigree replicas fall into the moderate-risk and high-risk categories. Conversely, IBIS has the highest average of high-risk pedigrees correctly classified at such high risk.

Although many studies have evaluated family history data for accuracy on self-reporting, the true sensitivity and specificity distributions are unknown. Studies tend to be limited by sampling bias—where the population under study has higher risk than the general population, a lack of gold standard—where self-reported family history is not compared against a gold standard such as a pathology report but instead collected through interviews, questionnaires, or death records, and lack of generalizable results due to the studies being held in 1 location. Nevertheless, the literature on self-reported accuracy shows an agreement of moderate-to-high accuracy for most cancers [17]. We have performed an MC simulation using values associated with the accuracy of self-reported family history to quantify the effect of uncertainty of breast cancer risk prediction. Our use of multiple risk prediction models, as recommended by the ACS, showed that uncertainty in family history can have a large effect on risk prediction. The effect of this uncertainty varies by model, but we show that it could ultimately affect prevention strategies for breast cancer, both in overuse of MRI screening to low-risk populations and in missed screening opportunities for high-risk patients.

We also show a highly generalizable method based on MC simulations to estimate the effect of self-reported family history on lifetime risk prediction tools for breast cancer. As medicine moves to a more personalized system, it is important to not only look to future research and discoveries, but also to better utilize and implement existing strategies to maximize their impact. Although we show that risk classification is subject to uncertain conditions, it is important not to dismiss the usefulness of these models. Instead, further research into methods of collection and storage of FHx are needed. We show that by decreasing uncertainty in FHx, existing tools become more effective, potentially saving cost, and, more importantly, lives.

Acknowledgments

The authors thank the Utah Center for High Performance Computing for computational resources (NLM Training Grant No. 5T15LM007124),

Richard A. Fay and Carol M. Fay Endowed Graduate Fellowship for the Department of Biomedical Informatics in Honor of Homer R. Warner (M.D., Ph.D.), and Andrew Lee for kindly providing the BOADICEA batch program. J.C.F. has been partially supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award No. 5UL1TR001067-03. C.C.M. gratefully acknowledges funding by the Pediatric Cancer Program supported by the Intermountain Healthcare and Primary Children's Hospital Foundations and the University of Utah, Department of Pediatrics, and Division of Hematology/Oncology.

Disclosures

None.

Declaration of Interest

None.

Supplementary Material

To view supplementary material for this article, please visit <https://doi.org/10.1017/cts.2016.9>

References

1. **Wilson BJ, et al.** Systematic review: family history in risk assessment for common diseases. *Annals of Internal Medicine* 2009; **151**: 878–885.
2. **Ziogas A, Anton-Culver H.** Validation of family history data in cancer family registries. *American Journal of Preventive Medicine* 2003; **24**: 190–198.
3. **Rich EC, et al.** Reconsidering the family history in primary care. *Journal of General Internal Medicine* 2004; **19**: 273–280.
4. **Yoon PW, et al.** Can family history be used as a tool for public health and preventive medicine? *Genetics in Medicine* 2002; **4**: 304–310.
5. **Ready K, Arun B.** Clinical assessment of breast cancer risk based on family history. *Journal of the National Comprehensive Cancer Network* 2010; **8**: 1148–1155.
6. **Saslow D, et al.** American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA: A Cancer Journal for Clinicians* 2007; **57**: 75–89.
7. **Claus EB, Risch N, Thompson WD.** Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer* 1994; **73**: 643–651. ([https://doi.org/10.1002/1097-0142\(19940201\)73:3<643::AID-CNCR2820730323>3.0.CO;2-5](https://doi.org/10.1002/1097-0142(19940201)73:3<643::AID-CNCR2820730323>3.0.CO;2-5))
8. **Tyrer J, Duffy SW, Cuzick J.** A breast cancer prediction model incorporating familial and personal risk factors. *Statistics in Medicine* 2004; **23**: 1111–1130.
9. **Parmigiani G, Berry D, Aguilar O.** Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *American Journal of Human Genetics* 1998; **62**: 145–158.
10. **Antoniou AC, et al.** The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *British Journal of Cancer* 2008; **98**: 1457–1466.
11. **Lee AJ, et al.** BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. *British Journal of Cancer* 2014; **110**: 535–545.
12. **Ozanne EM, et al.** Which risk model to use? Clinical implications of the ACS MRI screening guidelines. *Cancer Epidemiology, Biomarkers & Prevention* 2013; **22**: 146–149.
13. **Claus EB, Risch N, Thompson WD.** Genetic analysis of breast cancer in the cancer and steroid hormone study. *American Journal of Human Genetics* 1991; **48**: 232–242.
14. **Anderson H, et al.** Familial breast and ovarian cancer: a Swedish population-based register study. *American Journal of Epidemiology* 2000; **152**: 1154–1163.

15. **Mavaddat N, et al.** Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). *Cancer Epidemiology, Biomarkers & Prevention* 2012; **21**: 134–147.
16. **Ozanne EM, et al.** Bias in the reporting of family history: implications for clinical care. *Journal of Genetic Counseling* 2012; **21**: 547–556.
17. **Qureshi N, et al.** *Family History and Improving Health*. Washington D.C.: Agency for Healthcare Research and Quality (US), 2009.
18. **Severin MJ.** Hereditary cancer litigation: a status report. *Oncology (Williston Park, N.Y.)* 1996; **10**: 211–214; discussion 217.
19. **Murff HJ, Spigel DR, Syngal S.** Does this patient have a family history of cancer? An evidence-based analysis of the accuracy of family cancer history. *JAMA* 2004; **292**: 1480–1489.
20. **Katki HA.** Effect of misreported family history on Mendelian mutation prediction models. *Biometrics* 2006; **62**: 478–487.
21. **Tehranifar P, et al.** Validation of family cancer history data in high-risk families: the influence of cancer site, ethnicity, kinship degree, and multiple family reporters. *American Journal of Epidemiology* 2015; **181**: 204–212.
22. **Metropolis N.** The beginning of the Monte Carlo method. *Los Alamos Science* 1987; **15**: 125–130.
23. **Harrison RL.** Introduction to Monte Carlo simulation. *AIP Conference Proceedings* 2010; **1204**: 17–21.
24. **Schneider KA, et al.** Accuracy of cancer family histories: comparison of two breast cancer syndromes. *Genetic Testing* 2004; **8**: 222–228.
25. **Weitzel JN, et al.** Limited family structure and BRCA gene mutation status in single cases of breast cancer. *JAMA* 2007; **297**: 2587–2595.
26. **Claus EB, Risch N, Thompson WD.** The calculation of breast cancer risk for women with a first degree family history of ovarian cancer. *Breast Cancer Research and Treatment* 1993; **28**: 115–120.
27. **Evans DGR, Howell A.** Breast cancer risk-assessment models. *Breast Cancer Research* 2007; **9**: 213.
28. **Mazzola E, et al.** Recent BRCAPRO upgrades significantly improve calibration. *Cancer Epidemiology, Biomarkers & Prevention* 2014; **23**: 1689–1695.
29. **Mazzola BA, Parmigiani G, Biswas S.** Recent enhancements to the genetic risk prediction model BRCAPRO. *Cancer Informatics* 2015; **147**: 147–157. (<https://doi.org/10.4137/CIN.S17292>)
30. **Gail MH, Mai PL.** Comparing Breast Cancer Risk Assessment models. *Journal of the National Cancer Institute* 2010; **102**: 665–668.