ADAPTATION OF THE GENETIC RISK PREDICTION MODEL BRCAPRO FOR PRIMARY CARE SETTINGS

by

Philamer M. Atienza, B.S., M.S.

APPROVED:

_____
Dr. Sumihiro Suzuki, Co-Chair

_____
Dr. Swati Biswas, Co-Chair

_____
Dr. Karan Singh, Committee Member

_____
Dr. Sumihiro Suzuki, Department Chair

_____
Dr. Dennis L. Thombs, Dean, School of Public Health

ADAPTATION OF THE GENETIC RISK PREDICTION MODEL BRCAPRO FOR PRIMARY CARE SETTINGS

DISSERTATION

Presented to the School of Public Health

University of North Texas
Health Science Center at Fort Worth

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

By

Philamer M. Atienza, B.S., M.S.

Fort Worth, Texas

May 20, 2017

ACKNOWLEDGEMENTS

The completion of this research was funded in part by Susan G Komen grant KG081303, National Cancer Institute grants 1R03CA173834-01, 1R03CA173834-02 and 2P30CA006516-47, and the Dana Farber Cancer Institute.

I would like to express my deepest gratitude to Professors Sumihiro Suzuki and Swati Biswas for their invaluable contribution towards the completion of this research and dissertation work.

To Jonathan Chipman, Amanda L. Blackford, Drs. Kevin Hughes, and Giovanni Parmigiani, thank you for your guidance and assistance, and for allowing me the opportunity to work with your team on the enhancements to the BRCAPRO model.

And to all my family, colleagues, and peers, thank you for all the support and willingness to help in every way possible.

Philamer M. Atienza

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Atienza, Philamer M., Adaptation of the Genetic Risk Prediction Model BRCAPRO for Primary Care Settings. Doctor of Philosophy (Biostatistics), May 20, 2017, 100 pp., 12 tables, 8 illustrations, bibliography, 94 titles.

Identifying women at high risks of carrying the breast cancer susceptibility genes is crucial for providing timely surveillance and necessary health management interventions. BRCAPRO is one of the most widely used statistical models for breast cancer risk prediction in genetic counseling. It provides carrier probabilities of BRCA1/2 mutations and calculates the risks of developing breast and ovarian cancers. This calculation requires extensive personal and family history information, which makes it difficult to use in primary care where a wider population could be reached. Thus, we developed a two-stage approach for the genetic risk prediction of BRCA1/2 mutation. In the first stage, limited information on the counselee and her family history of cancer are used in simplified versions of BRCAPRO. If the risk at this stage is found to be high, the full BRCAPRO model utilizing the complete family history is implemented in the second stage. We aimed to balance the tradeoff between the amount of information used and the accuracy of the predictions. We explored several first stage tools. BRCAPROLYTE uses information on the affected relatives up to the second degree only. BRCAPROLYTE-Plus additionally includes unaffected relatives by imputing their ages. BRCAPROLYTE-Simple eliminates the need to collect information on the numbers and

types of unaffected relatives and imputes them and their ages instead. The study cohorts include 1,917 families mostly at high risk from the Cancer Genetics Network, 796 high-risk families from MD Anderson Cancer Center, and 1,344 population-based families from Newton-Wellesley Hospital. To evaluate the models, we used sensitivity, specificity, area under the curve, and observed versus expected number of carriers. We also considered clinical criteria of number of referrals made by each model. We found the proposed two-stage approach (with BRCAPROLYTE, BRCAPROLYTE-Plus, and BRCAPROLYTE-Simple at the first stage) has very limited loss of discrimination and comparable calibration with BRCAPRO. It identifies a similar number of carriers without requiring a full family history evaluation on all probands. Thus, our two-stage approach allows for practical large-scale genetic risk assessment in primary care.

CHAPTER 1

INTRODUCTION

According to the Centers for Disease Control and Prevention (CDC), breast cancer is the most common cancer diagnosis in women in the United States (U.S.) (CDC, 2016), and it is one of the leading cause of cancer deaths among women of all races (United States Cancer Statistics (USCS) Working Group, 2016), aside from non-melanoma skin cancer (CDC, 2016; National Cancer Institute (NCI), 2016). In 2016, approximately 249,260 men and women are expected to be diagnosed with breast cancer from which 40,890 are expected to die (American Cancer Society, 2016). Research has found several risk factors that increase susceptibility to breast cancer such as age, family history of breast cancer, early menarche, early menopause, using combination hormone therapy, having dense breast tissue, taking birth control pills, never having given birth, not being physically active, being overweight, alcohol intake, as well as having mutations in the breast cancer related genes BRCA1 or BRCA2 (CDC, 2016). It is estimated that about 7% of all women in the U.S. will get breast cancer by age 70, and most will occur after the age of 50 (CDC, 2014; CDC, 2016). Among those who develop breast cancer, approximately 5% to 10% will have harmful mutations in the breast cancer susceptibility genes BRCA1 and BRCA2 (BreastCancer,org, 2016; Campeau, Foulkes, and Tischkowitz, 2008; CDC, 2015). For

women with BRCA1/2 mutations, the risks for lifetime and early breast and ovarian cancers are greatly increased (Lynch et al., 2008; Miki et al., 1994; Parmigiani, Berry, and Aguilar, 1998; Wooster et al., 1995). The lifetime breast cancer risk in women in the general population is about 12% (NCI, 2015), while for those who have inherited a harmful mutation in BRCA1 or BRCA2, the risks can be as high as 80% (BreastCancer.org, 2016). The prevalence of some deleterious mutations of BRCA1 or BRCA2 among some ethnicities such as Ashkenazi Jews (AJ) is higher, and hence, they are at even greater risk of breast cancer development (BreastCancer.org, 2015). The two mutations in BRCA1 (185delAG and 5382insC) and one mutation in BRCA2 (6174delT) substantially increase the risk for breast and ovarian cancer in AJ families (Hartge et al., 1999). However, no long-term general population study to directly compare the cancer risks of women who have and do not have the harmful BRCA1 or BRCA2 mutations has been performed (NCI, 2015). Thus, the risk of breast cancer in the general population may not be accurately reflected.

It is important to detect who has harmful BRCA1/2 mutations early to manage their risk of cancer. Interventions such as mastectomy, oophorectomy or chemoprevention can be performed. Frequent monitoring and health management can help reduce the woman's chance of breast cancer and/or increase her survival rate (NCI, 2015). Hence, different models for genetic risk prediction are currently used in research and academic centers as well as in high-risk clinics or breast cancer centers that provide advanced diagnosis and detection capabilities. One set of genetic risk prediction models, the Mendelian models, use the known inheritance pattern of the genes given by the Mendelian laws, i.e., laws describing how a mutated gene is being

passed from a parent to an offspring. An example of a Mendelian genetic risk prediction model is the BRCAPRO (Parmigiani, Berry, and Aguilar, 1998). In general, a woman's risk for breast cancer is greater when she has more close relatives who have had breast or ovarian cancer, and additionally, if their ages of diagnosis are younger (CDC, 2014). In BRCAPRO, the family history of breast and ovarian cancers as well as information on age, type, and number of relatives are considered in calculating the risks. Ethnicity, such as being of AJ descent, breast tumor marker status, and oophorectomy status of each family member are also included as risk factors (Biswas et al., 2012; Chen, Blackford, and Parmigiani, 2009; Chen et al., 2004; Katki, 2007; Mazzola et al., 2014; Parmigiani et al., 2011; Tai et al., 2008). Given all of this information, BRCAPRO calculates an individual's probability of carrying a mutation of BRCA1 or BRCA2 under the assumption that the information on all relatives is accurate. It uses the Bayes' rule to find this probability using known (estimated) prevalence and penetrance of the mutations; that is, BRCAPRO calculates the probability of an individual's being a BRCA1 or BRCA2 mutation carrier based on estimates of mutation prevalence and penetrance found in the research literature (Chen et al., 2004). BRCAPRO also provides the absolute risk of developing breast and ovarian cancer for the counselee at 5-year increments by default, which can be modified to specific age intervals up to a maximum age (Chen et al., 2014).

The prediction models can potentially reap preventative benefits for more women when used in primary clinical settings like mammography clinics because it will allow for large-scale risk assessments, and can in turn lead to managing the health of high-risk individuals before they develop breast cancer or are at an early stage of the disease.

However, these models have limited availability to the general population. In primary care settings, as opposed to genetic counseling settings, collecting a lot of information needed for prediction on extended family history such as the age of onset and affection status can become a burden due to limited time and resource constraints (Drohan, Ozanne, and Hughes, 2009). The NCI recommends that individuals concerned about their risks consult directly with genetics professionals, yet there is a limited availability of such professionals (Drohan, Ozanne, and Hughes, 2009). Thus, having a simpler model to implement in primary care may be more useful and practical and could promote widespread use in a wider range of clinical settings.

In this dissertation, to bring the benefits of the genetic risk prediction to the general population, many of whom remain unaware of their mutation carrier status, we propose a two-stage approach. In the first stage, intended for primary care settings, a simpler version of BRCAPRO is used, which requires limited information about the patient's family history. If the risk at this stage is sufficiently high, the patient is referred to a more exhaustive genetic risk assessment (stage two) utilizing the full BRCAPRO version that uses extensive family history information. Such an approach allows balancing of the tradeoff between the accuracy versus the practicality of using the BRCAPRO model in predicting hereditary risks.

The aims of this dissertation are as follows:

1. To propose and assess the performance of simplified versions of BRCAPRO for estimating the probabilities of being a BRCA1 or BRCA2 mutation carrier, using limited family history information.

2. To develop and assess the performance of a two-stage approach for genetic risk prediction of BRCA1 or BRCA2 carrier mutation. If the risk is sufficiently high at the first stage, the full BRCAPRO model using data on the more extensive family history will be implemented at the second stage.

It is expected that the results of this study will promote the use of the proposed two-stage approach and lead to more women being screened for the genetic predisposition to breast cancer. Our goal is to select optimal model(s) that will be most useful in a primary care setting, thereby potentially screening and identifying more women who are at high risks and can benefit from early diagnosis. We hope that our research findings will help modify the current research practice of limiting the benefits of genetic counseling and risk prediction models in high-risk clinics only and bring them to the general (wider) population. Additionally, the two-stage approach of genetic risk prediction of breast cancer may help reduce the genetic testing burden for individuals who have lower risks of carrying a BRCA1 or BRCA2 mutation, and increase the identification and management of individuals who have higher risks of carrying a BRCA1 or BRCA2 mutation without increasing the genetic counseling burden in the second stage.

1.1. ORGANIZATION OF THE DISSERTATION

This chapter covered the background information and rationale for the two-stage approach to BRCA1/2 genetic risk prediction based on the BRCAPRO model. Chapter 2 discusses the relevant works of literature. In Chapter 3, simpler versions of BRCAPRO are proposed and evaluated. Chapter 4 focuses on evaluation of the overall

two-stage approach. Chapters 3 and 4 are based on published manuscripts (Biswas et al., 2013; Biswas et al., 2016), and have their own sections for the introduction, methods, results, and discussion. Thus, there is an overlap of their content with this Introduction chapter. Chapter 5 summarizes the main findings of this dissertation and highlights the public health importance of the research. Some future directions for research are also discussed.

CHAPTER 2

LITERATURE REVIEW

Before offering genetic testing for the BRCA1/2 mutations, patients are usually assessed on their estimated probability of carrying the mutation. This allows genetic testing to be used where it is most appropriate and is also required by insurance companies to determine coverage (Parmigiani et al., 2007). Several types of prediction models have been developed to assess the risks of carrying a BRCA1/2 mutation and of developing breast and ovarian cancer. These models can be broadly categorized as the expert-based approach, empirical models, and Mendelian models (Parmigiani et al., 2007). Development of some models that fall into these three categories is described in Sections 2.1.1 to 2.1.3.

The expert-based models consist of guidelines in identifying high-risk patients which were developed based on clinical judgment. Empirical models are statistical models that stratify families according to family history whereby regression or other approaches are used to describe the relationship between the family history data (predictor variables) and the genotyping results. The Mendelian models are based on Mendelian inheritance patterns of genes passed on from one generation to the other. These models use explicit assumptions about genetic parameters such as allele frequencies, cancer penetrances, and gene transmission to calculate the probability of

the proband/counselee being a mutation carrier (Parmigiani et al., 2007).

2.1.1. EXPERT BASED

As there was no clear and validated guideline on referrals for genetic testing, Gilpin, Carson, and Hunter (2000) developed a family history assessment tool (FHAT) in 1996 that can be used for genetic counseling. Weights (scores) assigned to the different characteristics of the cancers in the affected patients were adjusted using trial and error until the pedigrees (i.e., families) that should be referred to genetic counseling (using the target cutoff of double the general population lifetime risks, 0.11 for breast cancer and 0.016 for ovarian cancer, as benchmarks) got a total score of ≥10, and <10 for the pedigrees not requiring a referral. In FHAT, a 17-question interview is used to produce a score ranging from 0 to 45 to represent the risk severity of the family history (Parmigiani et al., 2007). Separate scoring was done for each family member who has been diagnosed with breast, ovarian, early colon or prostate cancer, and the total family score assisted in determining whether genetic counseling should be referred. If both sides of the counselee's family have breast and/or ovarian cancer history, the side with the higher score was used for risk assessment. FHAT was effective in minimizing the number of referrals and the likelihood of missing mutation-positive women. It can be used for selective referral, but not for estimating the risks of developing breast or ovarian cancer, or the likelihood of being a BRCA1 or BRCA2 mutation carrier (Gilpin, Carson, and Hunter, 2000).

It is important to develop models for predicting whether a counselee is likely to have BRCA mutations so that genetic screening efforts can be directed to potential

carriers. In this regard, Evans et al. (2004) developed a simple scoring system called Manchester scoring system to identify the likelihood of carrying a BRCA1 or BRCA2 mutation. Similar to FHAT, this system assigns scores based on the cancer types and age at diagnosis, and the side of the family that had the highest score was used for prediction. A score of 10 was used separately for each gene, which was considered as equivalent to a 10% probability of BRCA1 or BRCA2 mutation. The system does not require inputting the family tree into the computer program, hence, practical for use in busy clinics to help in identifying families to refer for BRCA1/2 mutation testing. However, this model does not take into account AJ ancestry. In addition, it is based on the number of cases per family which can bias the estimates (Kang et al., 2006).

2.1.2. EMPIRICAL MODELS

Couch et al. (1997) aimed to use clinical information, family history, and deoxyribonucleic acid (DNA) analysis to define the incidence of BRCA1 mutations and create probability tables of finding the mutations in individual families. The study population was presumed to be representative of the kinds of patients seen in a breast cancer referral clinic. However, many of the families were too small (i.e., there was a small number of individual members per family) to predict mutation status from an inheritance pattern. Most families only had breast cancer cases, and only a few had ovarian cancer, resulting in large confidence intervals. Additionally, the study sample was almost entirely White, so the results cannot be generalized to women of other races or ethnicities. Contrary to expectations, results showed that bilateral breast cancer

incidence, the number of affected family members, and the average age at diagnosis of ovarian cancer were not associated with increased risk for a BRCA1 mutation.

Another prediction model for BRCA mutations was introduced by Vahteristo et al. (2001) based on the study of Finnish breast cancer families and their clinical characteristics. Only families containing three or more first or second degree relatives with breast or ovarian cancer were included in the model. The model was compared to results from Shattuck-Eidens et al. (1997) and Couch et al. (1997), both of which were developed only for BRCA1. The simple criterion of a family in this study population having a breast cancer before age 40 or any ovarian cancer was effective in predicting mutation carrier families. The results were similar when BRCA1 and BRCA2 carriers were analyzed separately, so a common model was used. Results also showed that mutation screening in families with only two affected patients (moderately affected families) was not useful in the study population.

Apicella et al. (2003) also created a simple, reliable algorithm to estimate the probability that an AJ woman carries an ancestral mutation in BRCA1 or BRCA2 based on multiple predictive factors (personal or family history of breast or ovarian cancer). In this study, the authors used an unconditional multiple linear logistic regression to model the probability that each woman was a mutation carrier: starting with a baseline score of -3.75, a multiple of 0.5 based on the logistic regression estimates was added for each predictive feature. The sum of the scores is the estimated log odds ratio of being a carrier, converted to a probability using a reference table. The model developed in this study is called LAMBDA, which stands for “Log odds of carrying an Ancestral Mutation in BRCA1 or BRCA2 for a Defined personal and family history in an Ashkenazi Jewish

woman." One advantage of the model is that it integrates multiple datasets to provide estimates across a range of personal and family histories. Because of its simplicity, the LAMBDA model may have wide use in clinical settings.

These statistical models have their limitations. The Couch model was based on BRCA1 mutation only. It is based on logistic regression and requires the exact ages of all affected relatives (Lindor et al., 2007). The Vahteristo model is difficult to use in a typical clinic setting, and while Apicella's LAMBDA is an easy tool because it does not require a computer program (Lindor et al., 2007), it was designed only for use with AJ families.

2.1.3. MENDELIAN MODELS

Formal probabilistic models were found to be optimal in the selection of families that can be referred for genetic testing than clinical criteria or scoring methods (James et al., 2006). Moreover, Mendelian models were found to be more accurate and effective at estimating an individual's risk of being a BRCA mutation carrier and the overall number of mutations when compared to empirical models (Marroni et al., 2004). Development of two Mendelian models, the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) and BRCAPRO, are discussed below.

Antoniou et al. (2002) developed the BOADICEA model under which susceptibility to breast cancer is explained by BRCA1 and BRCA2 mutations as well as by the joint multiplicative effects of many genes (polygenic component) using combined data on high-risk families and population based series of breast cancer cases in the

United Kingdom. Models incorporating the simultaneous effects of BRCA1, BRCA2, BRCA3 (which was a hypothetical gene assumed to have increased risks of breast cancer), and a polygenic effect, i.e., the effect of several low-risk polymorphisms which act multiplicatively, were analyzed using segregation analysis. The model with separate relative risks assumed for ages grouped by decade fit significantly better than the one with a constant relative risk (p-value=0.03). BRCA1 mutation carriers were more likely to develop early onset breast cancer than BRCA2 mutation carriers. The results were consistent with the Breast Cancer Linkage Consortium estimates (Easton et al., 1993; Easton et al., 1995; Ford et al., 1998), where high-risk families are expected to have a much larger polygenic component than average. The results of the polygenic model, determined to be the most parsimonious, suggested that several genes that have small but multiplicative effects on risk can account for the non-BRCA1/BRCA2 familial clustering of breast cancer. Additionally, the risk modifiers in BRCA1 and BRCA2 carriers may explain the differences in the estimates for population based studies and high-risk families.

In 2008 (Antoniou et al.), BOADICEA was extended to include a larger number of mutation-carrying families in order to produce more reliable estimates for the BRCA mutation incidences. The polygenic component's variance was also updated to be age dependent. A birth cohort effect was implemented leading to cohort-specific incidences for carriers and non-carriers. The fitted model predicted that the average breast cancer risks in carriers increase in more recent birth cohorts. BOADICEA returns the probabilities of carrying a BRCA1 or BRCA2 mutation and of developing breast or ovarian cancer.

For BRCAPRO, initially, Berry et al. (1997) developed a mathematical model for computing the probability of a woman carrying a mutation of the BRCA1 gene based on her family history of breast and/or ovarian cancer. The model was based on the available estimates of BRCA1 mutation frequencies in the general population and the age-specific incidence of breast and ovarian cancers for women who were BRCA1 mutation carriers. For women with breast cancer at a young age, the authors found that there was strong evidence for being a mutation carrier, because breast cancer was much more common at younger ages among carriers than non-carriers. However, for those without breast cancer, there was very little evidence for being a carrier, and this evidence of being a non-carrier became stronger with age as long as the woman remained free of cancer. The number of relatives was also an important consideration because having many unaffected family members can substantially lower a carrier probability. Using the weights from the developed model, the cumulative incidence probabilities for carriers and non-carriers can be averaged to get the cumulative probability of breast cancer before a particular age. BRCA1 carriers were very likely to have bilateral breast cancer (Easton et al., 1995). The authors assumed the independence of the incidence of breast and ovarian cancers separately for carriers and non-carriers, which is not necessarily true about the general population when carrier status is ignored. At that time, the assumption was that all other breast cancer is sporadic, but this is not true because BRCA2 has been cloned since then.

Extending Berry et al.'s work, Parmigiani, Berry, and Aguilar (1998) used Bayesian methods for evaluating the probabilities that a woman is a BRCA1 or BRCA2 mutation carrier on the basis of the occurrence of breast and ovarian cancers in her first

and second degree relatives. As there were uncertainties in the age-specific incidence of the diseases and the overall prevalence of mutations, Monte Carlo simulation was used to incorporate these in the model. The late onset of breast cancer cases made it more likely that the mutation was due to BRCA2 than BRCA1. In addition, the presence of breast cancer in males gave a strong indication that there was BRCA2 mutation. When there was an early age of onset of the diseases, BRCA1 was more likely than BRCA2. However, with late onset, the overall probability of the individual being a mutation carrier decreased.

In 2009, in a correspondence to the editor of the Journal of Clinical Oncology, Chen, Blackford, and Parmigiani addressed the findings of Kurian et al. (2008) that the BRCAPRO model underpredicts the proportion of BRCA2 mutations in relation to the total number of mutations, and also underpredicts the total number of BRCA1 and BRCA2 mutations among Asian-Americans. The allele frequencies were updated in a later BRCAPRO model (in BayesMendel version 1.3-1) than what Kurian et al. used (version 1.2-1). Penetrance estimates for breast and ovarian cancers were updated in version 1.3-1 (BayesMendel Lab, 2015). Furthermore, BRCAPRO was enhanced in version 2.0-2 by incorporating Asian-specific phenocopy rates which was justified because Asians have a lower risk of developing breast cancer than Whites (8% vs. 12%) according to Surveillance, Epidemiology, and End Results (SEER) registry data. The more recent BRCAPRO versions allow user specification for race as Asian, Black, Hispanic, Native American or White; and for ethnicity of each family member as AJ, non-AJ, Italian or Other (Mazzola et al., 2015).

Another factor that should be considered in genetic risk prediction models is medical interventions. Katki (2007) argued that medical interventions given to at-risk people should be included in Mendelian models. This is because interventions may imply that relatives are at high-risk and could be carriers, and because interventions alter mutation penetrances. For example, if a relative who is thought to be high-risk has an intervention at a young age which prevents her from developing cancer, her non-cancer status will affect the prediction of the patient's risk because prediction models use family history of cancers in calculating the risks. If the family member's intervention is not incorporated into the model, there is the untenable assumption that the penetrances for those who have and who have not had an intervention are the same. Thus, Katki incorporated the hazard ratio of having oophorectomy into the BRCAPRO model. Oophorectomy (removal of ovaries and fallopian tubes) halves the risk of breast cancer and removes the risk for ovarian cancer entirely, therefore it is a common intervention for women at high-risk. This study showed that the importance of accounting for interventions increases when mutation carriers and non-carriers differ more in terms of the benefits of intervention. Accounting for the oophorectomy was most important for older family members, and ignoring the procedure leads to carrier probabilities that are underestimated to a varied degree, especially if the post-oophorectomy disease is informative about carrier status. When the hazard ratio of the reduction due to intervention between carriers and non-carriers is more than one (that is, when there is more benefit for non-carriers than carriers), interventions are most informative about the carrier status; getting cancer after the procedure is more evidence for a mutation. However, ultimately, if the full family history and all their genetic test

results are known to the consultand, there is no carrier status information gained by the knowledge that a relative has chosen to have an intervention because each individual's carrier status, disease history, and associated penetrances should already be factored into the model. Having mastectomy has also been incorporated into the BRCAPRO model starting from BayesMendel version 2.1 to improve the carrier probability calculation (Mazzola et al., 2015).

Tai et al. (2008) also added improvements to the BRCAPRO cancer risk prediction model by directly integrating a patient's pathological subtype information into the estimation methods. Including estrogen receptor (ER), progesterone receptor (PR), cytokeratin (CK)5/6, and CK14 marker information improved the BRCAPRO carrier probabilities, thereby discriminating between BRCA1 and non-BRCA1 breast tumors better. In 2012, Biswas et al. added the possibility of incorporating the human epidermal growth factor receptor 2 (Her-2/neu) into BRCAPRO to help improve the risk prediction for BRCA mutation. These improvements in risk estimation could lead to different clinical recommendations for genetic testing and cancer prevention.

We focus on BRCAPRO, which is widely used and distributed free of charge as part of the BayesMendel computer software package (Chen et al., 2004; http://bayesmendel.dfci.harvard.edu/risk/), and in the genetic counseling packages CancerGene (https://www4.utsouthwestern.edu/breasthealth/cagene/) and Hughes RiskApps (HRA; http://bcb.dfci.harvard.edu/bayesmendel/riskservice.php). In particular, CancerGene is licensed to more than 6,000 sites including high risk clinics in large academic medical centers and providers. Literature shows that BRCAPRO is one of the better-performing models for assessing an individual's risks for carrying a BRCA

mutation and for developing breast and/or ovarian cancer, with the AUCs (area under the receiver operating characteristic (ROC) curves) ranging from 80% or above (James et al., 2006; Parmigiani et al., 2007), i.e., in about 80% of the time, BRCAPRO assigns high probability to a positive BRCA1 or BRCA2 mutation carrier than to a negative mutation carrier. Further discussions about model comparisons can be found in Section 2.3.

2.2. BRCAPRO VALIDATION STUDIES

Genetic testing for BRCA1 and BRCA2 mutations is commonly warranted from risk assessments based on family history of breast and/or ovarian cancer. Genetic tests have both health and cost implications, stressing the importance of both the accuracy of risk assessments and suitability for testing (Berry et al., 2002). In this section, validation studies of the BRCAPRO prediction model are discussed.

Berry et al. (2002) compared the probability estimates for genetic mutation produced by BRCAPRO with genetic test results as well as assess genetic testing sensitivity and the relevance of other breast and ovarian cancer susceptibility genes. Six high-risk counseling clinics provided data for families in which at least one member had undergone testing for both BRCA1 and BRCA2. Ages of onset, types of cancers, numbers and ages of unaffected family members, and the exact relationships among all family members should be considered as factors as what the BRCAPRO uses. There was a tendency for probands with a large number of affected family members with breast and ovarian cancers to have higher carrier probabilities, but there were unexpected results as well indicating that factors aside from the number of familial

cancers should also be included in genetic risk assessment. BRCAPRO may overestimate the probability of mutation for those with the highest risk and underestimate the probability for those with the lowest risk. However, the overall correspondence between the probabilities produced by BRCAPRO and the results of the genetic testing was good. The sensitivity of the genetic test was high at 85.4% [78.7%-90.5%; 95% Confidence Interval (CI)]. The study showed that BRCAPRO was an accurate tool upon which patients can be counseled about their probability of carrying the BRCA1 and BRCA2 mutations. BRCAPRO was determined to be particularly effective in predicting positive tests when the carrier probability is less than 70%.

Euhus et al. (2002) performed an analysis to compare BRCAPRO with eight cancer risk counselors in their identification of families likely to carry the mutation. The study was also used to assess the impact that having BRCAPRO probabilities alongside the pedigree information had on the genetic risk counselor's probability estimations. Pedigrees were chosen for probability estimation on the basis of a proband affected by either breast or ovarian cancer and who had a conclusive positive or negative result after full sequence genetic testing. All eight counselors and BRCAPRO provided probability estimations for each family in a supervised environment to ensure that the probabilities from the counselors were not biased from the use of BRCAPRO or other models. Comparisons with BRCAPRO were made using a >10% mutation probability threshold, because genetic testing is usually recommended for individuals with a probability estimate above 10%. Each of the pedigrees was presented twice to the counselors, once without BRCAPRO data and once with BRCAPRO estimates

alongside the pedigree information. The differences in their probability estimates were compared for each family. Six of the 8 showed improvement in the AUC estimates, but only one counselor's improvement was statistically significant with p-value=0.05 when the BRCAPRO information was added to the pedigrees. It was concluded that while sensitivity was comparable between BRCAPRO and the counselors, the computer modeling was consistently better in terms of specificity. Overall, BRCAPRO was slightly better than the counselors at identifying the genetic mutation carriers.

BRCA1 and BRCA2 genetic mutations are well documented as susceptibility genes for breast cancer, but the variety of the mutations and risk modifiers among ethnic minorities were still unclear. The BRCAPRO model was originally designed using mutation frequencies of White (non-AJ and AJ) populations to be able to predict a family´s likelihood of being a BRCA1 or BRCA2 mutation carrier. Vogel et al. (2007) applied BRCAPRO to the Hispanic population while Nanda et al. (2005) used it for Whites, African Americans, and AJ families in high-risk clinics to obtain information about the usefulness of the model in these populations. In both studies, the results showed that the probabilities were similar for the same level of specificity. The AUCs for both the Hispanic and African American families were around 0.77 which were comparable to the White and AJ populations. A founder mutation in the AJ population, 187delAG, was the most frequent in the Hispanic sample with BRCA1 mutations. These studies provided evidence that it is appropriate to use the BRCAPRO model in risk assessment of Hispanic families as well as high-risk African American families, as its performance in these ethnic groups was equally well with Whites.

Katki et al. (2008) studied the prediction accuracy of an extension of the BRCAPRO model. The extended model took into account the survival to other non-breast/ovary cancers as a single outcome and assessed the effect of excluding mutation-unrelated cancers. They also considered the effects of informative censoring caused by ignoring certain mutation-related diseases such as pancreatic or prostate cancers for BRCAPRO. Even if no family member gets the disease, the authors showed that the carrier probability will be inflated if the disease occurs more often in carriers, or deflated if it occurs less often in carriers, if it is not accounted for in the model. In the case of a disease unrelated to mutation that is dependent on other mutation-related diseases, a person's survival could be informative about the carrier status. Ignoring the member's disease which has a higher prevalence among mutation carriers will cause the carrier probability to be deflated. The extended model was shown to help extract more useful information from male relatives and from families that have many older relatives with cancer.

Biswas et al. (2012) validated the use of the three breast cancer markers ER, PR, and Her-2/neu in the BRCAPRO model. Results revealed that most BRCA1 carriers were ER negative but both non-carriers and BRCA2 carriers were ER positive, consistent with literature. With women who are Her-2/neu negative and ER positive, Her-2/neu could help distinguish between carriers versus non-carriers. ER and PR helped improve the discrimination between carriers of BRCA1 versus BRCA2, which is useful because these genes confer different risks (e.g., BRCA1 carriers are more likely to have ovarian cancer while BRCA2 carriers are more likely to have pancreatic cancer). The results revealed that the updated BRCAPRO model, which accounted for

these breast tumor markers, improved the BRCA1 or BRCA2 mutation status prediction and is beneficial to use in clinical settings whenever possible.

2.3. MODEL COMPARISONS

Each statistical model has its strengths and limitations in different scenarios. A summary of the common methods that several authors have used in their studies to evaluate the performance of each BRCA genetic risk prediction model is shown in Table 2.1. A careful examination of the results from each of these models in light of known strengths and limitations of one over the other is important when comparing the performance of the risk prediction models. Ideally, we want to select models that give higher predictive power and accuracy. Among these comparison methods, it is of greater interest to choose models that give higher sensitivity and AUC values without sacrificing the associated loss incurred on the other measures when adjusting thresholds. A model that gives better calibration, i.e., an observed over expected ratio closer to one, is also preferred.

Table 2.1. Criteria used in evaluating and comparing the performance of prediction models

Comparison Methods	Specific Purpose (if any)	References
positive and negative predictive values		Barcenas et al., 2006; Kang et al., 2006; Parmigiani et al., 2007
sensitivity and specificity		Barcenas et al., 2006; James et al., 2006; Kang et al., 2006; Marroni et al., 2004; Parmigiani et al., 2007
ROC curves analysis	• to determine the cutoff value that provides the best discrimination between BRCA mutation carriers and non-carriers	James et al., 2006; Marroni et al., 2004
AUC or c-statistic (the	• to measure the level of	Barcenas et al., 2006;

probability that a randomly-chosen positive case (carrier) has a higher probability of a BRCA mutation than a randomly-chosen negative case (non-carrier))	discrimination between carriers and non-carriers, or to discriminate between individuals testing positive and negative • to examine accuracy with an AUC of 1.0 being a perfect discrimination between BRCA mutation carriers and non-carriers, and 0.5 meaning that the model is completely unable to distinguish between carriers and non-carriers	James et al., 2006; Kang et al., 2006; Lindor et al., 2007; Parmigiani et al., 2007
comparison of the observed and expected number of mutations (χ^2 and z-distribution)	• to test whether predicted probabilities are systematically too high (over-estimation) or too low (under-estimation); i.e., assessment of accuracy • to test whether the distributions are too variable or not; i.e., assessment of predicted carrier probabilities for over- or under-dispersion, which happens when a large number of probands with low predicted carrier probabilities or when a very small number with higher predicted carrier probabilities are carriers	Lindor et al., 2007; Marroni et al., 2004
Wilcoxon rank sum test	• to assess the significance of trends when comparing all models	Parmigiani et al., 2007
ease of use as an office tool		Barcenas et al., 2006; Kang et al., 2006; Lindor et al., 2007

Kang et al. (2006) evaluated four models BRCAPRO, Manchester (Evans et al., 2004), University of Pennsylvania (Penn) (Couch et al., 1997), and Myriad-Frank (Frank et al., 2002) on families who had been tested for mutations in BRCA1 and BRCA2. These models could not rule in or rule out BRCA carrier status as indicated by the positive and negative likelihood ratios for those who had a model probability >10% and <10%, respectively. Each model performed equally well when mutations in both genes were considered, with the AUCs being around 0.75. Each of the Manchester,

BRCAPRO, and Penn models performed similarly for BRCA1, but the Penn model was better at, and the BRCAPRO model was worst at, distinguishing BRCA2 carriers. It was recommended that the use of these models routinely in the clinical setting was not justified. Future models should be less reliant on clinical history because of its inaccuracies, often due to the inability to verify cancer and diagnosis age. In addition, the choice of an individual from a high-risk family who tests negative for BRCA mutation does not necessarily mean that the family is mutation negative.

Marroni et al. (2004) compared the performance of eight models for predicting BRCA1 and BRCA2 mutations: the Penn model, the Myriad-1 model, the Myriad Tables, the Spanish model, the Finnish model, the Yale University (Yale) model, the BRCAPRO model, and the Italian Consortium (IC) model. The Mendelian models were more accurate because they calculate individualized probabilities. While the Mendelian models BRCAPRO and IC performed the best overall, this work showed that there was substantial room for improvement in model performance. In particular, the adjustment of genetic parameters for families at low risk and the discrimination between the two genes is expected to improve the ability of models to predict carrier status.

Barcenas et al. (2006) looked specifically at BOADICEA and compared its performance to five other mathematical models: BRCAPRO, Myriad I, Myriad II, Couch and Manchester Scoring System. They used a 10% threshold (as initially recommended by the American Society of Clinical Oncology (1996)) to assess family data where at least one member had genetic testing. BOADICEA, BRCAPRO, and Myriad II had comparable performance and all moderately effective at estimating the

risk of carrying a BRCA gene mutation. For non-AJ families, BRCAPRO was found to be the best for predicting BRCA1 mutations.

James et al. (2006) wanted to find the optimal method for determining which families in the moderate- to high-risk group should be tested for BRCA1 and BRCA2 mutations for use at standard familial cancer centers. They compared six methods for estimating the probability of being a carrier: Frank (empirical), Couch (logistic regression), BRCAPRO (probabilistic), Adelaide (clinical criteria), FHAT (clinical scoring), and Manchester Score (clinical scoring). The model which best discriminated between mutation-positive and negative families was the BRCAPRO score for any BRCA mutation (AUC=0.78 [0.72-0.85, 95% CI]). Overall, the highest accuracy was achieved with a combination of pathology data (grade and ER or PR status) and the combined BRCAPRO score, and particularly for BRCA1 families.

Parmigiani et al. (2007) compared seven models for estimating the probability of having a mutation in BRCA1 or BRCA2: Finnish, Myriad, NCI, and Penn which are empirical models; BRCAPRO and Yale which are Mendelian models; and FHAT which is expert-based. BRCAPRO performed best overall in all but two subgroups, but the range of c-statistics across different models was not large, making it difficult to point out a clear best model. In general, the models performed worse in cancer-free individuals and for younger people. Also, discrimination between mutation carriers and non-carriers was better in population-based studies than in high-risk samples (p-value=0.036), suggesting that the models were better at coping with broadly representative settings than high-risk centers.

The relative performance of the models LAMBDA, Myriad II, BRCAPRO, and Couch was compared in Lindor et al. (2007). Results showed that the AUCs were similar for the four models, ranging from 0.71 to 0.76, with BRCAPRO having the highest AUC, but no evidence of statistical difference was found (p-value=0.3). A test of observed versus predicted values found that the overprediction of carriers by LAMBDA was not statistically significant (p-value=0.3), but the underprediction of carriers by BRCAPRO, Couch, and Myriad II were all statistically significant (p-values are 0.01, 0.01, and <0.001, respectively). BRCAPRO was the best discriminator between carriers and non-carriers.

Fischer et al. (2013) evaluated the performance of four widely used genetic models BRCAPRO, BOADICEA, IBIS, and extended Claus (eCLAUS) using a large sample from central Europe. They estimated the BRCA1/2 mutation carrier probabilities under each model, and assessed the models' diagnostic accuracy via ROC analysis. Model calibration was compared via the ratio of observed to expected numbers of carriers at various categories of predicted mutation carrier probabilities. BRCAPRO and BOADICEA discriminated well between carriers and non-carriers with their AUCs significantly larger than those of IBIS and eCLAUS (0.796, 0.791, 0.749 and 0.745, respectively; p-value<0.001). At 10% cutoff, the sensitivities were also higher for BRCAPRO and BOADICEA than IBIS: 84.3% and 82.1% vs. 77%. eCLAUS had the highest sensitivity at 98%; however, its specificity was very low: 9.6% compared to 55%, 56.8%, and 56.5% for BRCAPRO, BOADICEA, IBIS, respectively. BOADICEA was best in calibration overall, while BRCAPRO overpredicted the mutations, especially the BRCA2 carriers. The study also showed that incorporating breast tumor marker

information on ER, PR, Her-2, (CK)5/6, and CK14 improved the predictive ability of BOADICEA, supporting the results from BRCAPRO studies (Biswas et al., 2012; James et al., 2006). The use of BRCAPRO and BOADICEA models for estimating BRCA1/2 mutations in clinical settings was recommended.

In terms of ease of use in an office, in the study of Barcenas et al. (2006), Myriad II was found to be the easiest to implement in actual clinical settings because of the table of probabilities that is provided. However, this model does not differentiate between BRCA1 and BRCA2 mutations. Mendelian models such as BRCAPRO, on the other hand, are capable of doing this. However, both Kang et al. (2006) and Lindor et al. (2007) indicated that the time needed for inputting all variables into the computer and amount of data required in implementation was a hindrance for BRCAPRO.

Overall, it might be beneficial for genetic counselors to consider several predictive models to get a more reliable result in estimating the probability of having BRCA mutation (Parmigiani et al., 2007). In light of similar performance among some models, the review of cancer family histories by an experienced clinician to supplement the model's predictions was recommended so as not to overlook critical elements, especially at the lowest or highest ends of the probability scale (Lindor et al., 2007).

2.4. CLINICAL DECISION SUPPORT SOFTWARES

Despite all the models and guidelines that were developed, there are many barriers to large-scale screening for women at high risk of having hereditary breast or ovarian cancer (HBOC). Such barriers include lack of clinicians required in taking and recording detailed family histories, lack of knowledge by clinicians to identify hereditary

cancer syndromes, and time-intensive or difficult-to-use risk assessment programs (Ozanne et al., 2009). Many clinicians also lack the skills and education necessary to analyze family history information in a clinic setting. It is unreasonable to expect that medical practitioners will remember all the details of the 188 hereditary syndromes associated with common adult chronic diseases listed in the Online Mendelian Inheritance in Man database (Drohan, Ozanne, and Hughes, 2009). Identifying patients with high risks of being a carrier of mutation for HBOC is challenging especially if medical professionals will only rely on memory to classify such patients (Drohan et al., 2012). Many physicians are not adept with linking family history data and genetic testing referrals to risk interventions (Drohan, Ozanne, and Hughes, 2009). Family history information such as the number of blood relatives affected with cancer, early age at diagnosis, multiple primary cancers in an individual, or male breast cancer are just some of the factors suggestive of HBOC that clinicians need to consider. It is, however, important to discover these high risk patients in the general population in order to prevent breast and ovarian cancer or to diagnose them at an early age so that the disease can be effectively managed (Drohan et al., 2012).

The need for a Clinical Decision Support (CDS) tool to effectively manage patient family history data was recommended by Drohan et al. (2012). CDS tools aide clinicians in choosing the correct approach to managing patient information and improving the quality of medical care. These systems must have the ability to evolve over time as new analysis techniques become available. Results should be presented in a manner that the patients can understand to be able to make an informed decision in managing their risks. Drohan, Ozanne, and Hughes (2009) discussed that the

approaches to CDS for HBOC had been unsuccessful due to several factors like data collection, data entry, data analysis, and interoperability issues. Collecting family history information is time consuming and often lacks pertinent information such as the age of diagnosis of the relatives affected with breast or ovarian cancer. The ideal situation is to have the family history updated at every patient visit but this is not a common practice. In busy clinics, many medical practitioners see little or no value in entering patients' family history data due to time and cost constraints. Most clinicians record data in free text form but the unstructured data format is not useful for computer programs in drawing pedigrees or running analyses. Also, different software applications require separate data entry, which only highlight the importance of interoperability. As suggested by the Family Health History Multi-Stakeholder Workgroup (Feero et al., 2008), having individual vendors develop their own complete family history pedigree drawing and risk assessment tool is not practical as compared to the modular approach. Top notch software that solves problems specific to small groups of clinicians should be integrated with electronic health records (EHRs) to reach a wider audience in a faster pace. However, the challenge of having various tools linked is the heterogeneity of data, i.e., the various data structures that each module application requires. Thus, standards were designed and developed by different organizations such as the Clinical Genomics Special Interest Group under HL7, an international group which specializes in standardizing health related data, and the American Health Information Community.

2.4.1. BAYESMENDEL

One computer tool designed to help practitioners in predicting the risks of being a BRCA1/2 mutation carrier and some other cancer-related genes is BayesMendel. Chen at al. (2004) developed the library BayesMendel, which is implemented in the R statistical language. R is an object-oriented structure language, freely distributed, and open source. BayesMendel follows a Mendelian model, which uses genotype and marginal distributions of genotype (penetrance and prevalence, respectively) as information to build conditional distributions of phenotype in order to predict the carrier status for genetic counseling. It includes the BRCAPRO model for breast and ovarian cancer to predict BRCA1 and BRCA2 carrier predictions, as well as other models to predict mutations associated with some other cancers such as pancreatic, melanoma, and colorectal and endometrial cancers (BayesMendel Lab, 2015). The library models can also be easily modified for different diseases and subpopulations without having to customize the library code. The backbone of the software is the calculation of the carrier probability, i.e., calculating the probability distribution of the counselee genotype given the family history, covariates, and pedigree structure. Within BayesMendel, there are three major object classes: pedigree objects, penetrance objects, and prediction objects. A pedigree object is in matrix format and includes the phenotype information of the family and the pedigree structure. The row represents each family member, and for each member, information such as member ID, relation to the counselee, gender, father and mother IDs, disease status, age of onset or current age, genetic testing result, etc., is in 12 or more columns. Penetrance objects include information on literature-based net penetrance by age, gender, phenotye, and mutation status; default values are

available in the software. Separate penetrance estimations are used for the AJ and non-AJ ancestries. Prediction objects include the joint probability of an inherited deleterious mutation in the two genes in matrix form. The rows or columns in the matrix signify the homozygous carrier, heterozygous carrier, and wild-type genotypes at BRCA1/BRCA2. For example, the matrix results:

	BRCA20	BRCA21
BRCA10	0.89051874	0.0729811285
BRCA11	0.03631815	0.0001819807

indicate that the probabilities of the proband being either a BRCA1 or BRCA2 mutation carrier is 0.10948126 (1 minus 0.89051874), being a BRCA1 mutation carrier is 0.03631815, being a BRCA2 mutation carrier is 0.0729811285, and being both a BRCA1 and BRCA2 mutation carrier is 0.0001819807. If the counselee is unaffected (without the disease), the net and crude cumulative risk of developing the disease is also included in the objects.

One caveat is that the use of the BayesMendel package by itself would require some computer or programming knowledge for data entry and risks prediction. In actual clinical settings, BayesMendel is used through the genetic counseling software packages CancerGene (https://www4.utsouthwestern.edu/breasthealth/cagene/) and HRA (http://bcb.dfci.harvard.edu/bayesmendel/riskservice.php). Both of these user-friendly tools do not require programming knowledge.

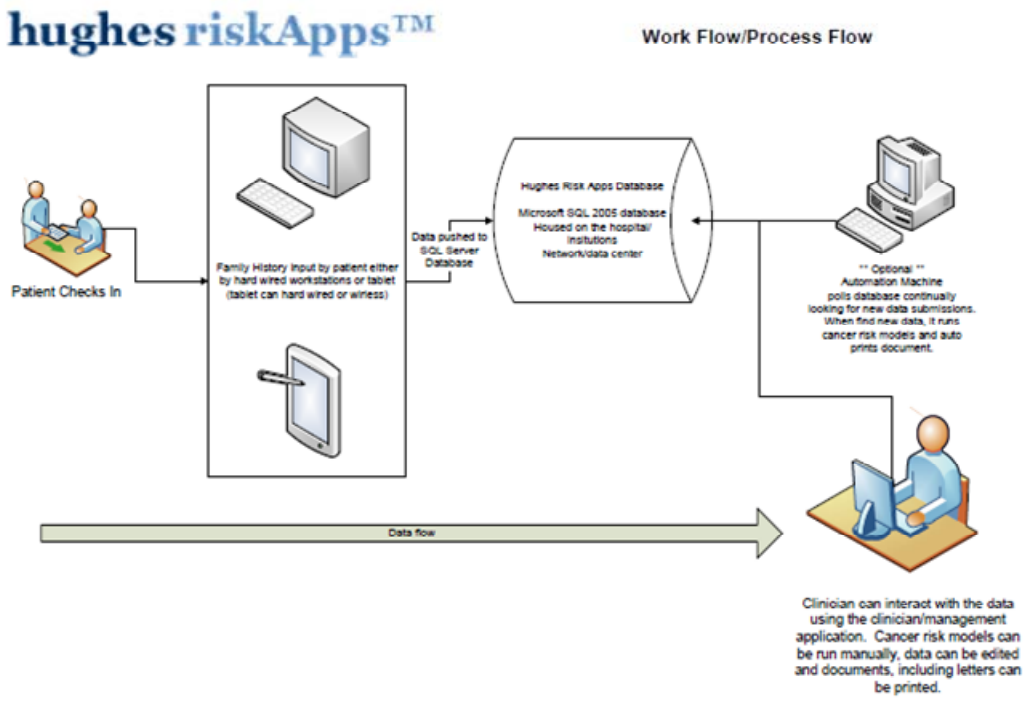
2.4.2. HUGHES RISKAPPS (HRA) as available currently from the Web Service

Drohan et al. (2012) offered a technological solution to help address the gap between the number of mutation carriers in the U.S. versus those who have been tested

and identified as BRCA1 or BRCA2 mutation carriers by developing HRA. To estimate the gap and help increase the awareness of genetic testing and its benefits, they used data available on the internet and made assumptions about the types and results of genetic testing. From their estimates of the U.S. population in 2000, there were 348,274 mutation carriers among women who are 20 years and above. Out of these, they estimate that only 29.3% carriers with cancer and 5.5% carriers without cancer have been identified; i.e., only a small proportion of BRCA1 or BRCA2 carriers have been identified even after 14 years of BRCA testing availability.

HRA (open-source family history collection, risk assessment, and CDS) software system was developed for use at primary care clinics and cancer centers (Ozanne et al., 2009). Figure 2.1 shows the HRA Work Flow and described briefly below.

Figure 2.1. Hughes RiskApps (HRA) Work Flow (http://www.hughesriskapps.net)



Due to time and resource constraints, only a limited amount of family history is asked from a patient when he/she checks in to the clinic. The process begins with a survey (either a Standard Survey or a Risk Clinic Survey) on a tablet, desktop computer, or kiosk, which the patient (or staff at the clinic) has to use to fill in her family information. The entered data including the numbers and types of affected first and second degree relatives, and their ages of diagnosis are analyzed in real time by risk prediction models such as BRCAPRO (or a simpler version BRCAPROLYTE) and Myriad models to assess the risk of carrying a BRCA1 or BRCA2 mutation. The results of the survey (risk factors, family history, risk of mutation, and lifetime risk of breast and ovarian cancers) are centrally stored and are available immediately to the clinicians involved with a patient, who can modify or add data to a patient's file. Family history information collected when the patient first enters the clinic will not have to be entered more than once if the patient is predicted with a higher BRCA1 or BRCA2 mutation carrier risk. For all patients with greater than 10% risk of carrying a mutation, an information sheet is given or mailed which explains HBOC and recommends a genetic counseling appointment. Finally, CDS is given to the clinician, which is composed of recommendations about genetic testing, surveillance, and prevention options, and then provides a list of relatives who should also be prioritized for genetic testing. The electronic nature of data collection and evaluation creates a work flow that is highly efficient and potentially less prone to errors. The auditing feature is also a desirable property as patient data and family history information can change over time and must be captured into the risk prediction model.

The HRA software was implemented and first installed in April 2007 at the Newton-Wellesley Hospital (NWH) (Drohan et al., 2012). Nine hundred fifteen out of 25,763 individuals (3.6%) were eligible for risk assessment and possible genetic testing based on a 10% risk of mutation threshold, representing a dramatic increase over previous numbers identified (from 1 to 2 per month to 5 to 6 per week) (Ozanne et al., 2009). This resulted in an increase in the number of patients sent for genetic testing, but a decrease in the workload of genetic counselors due to centralized data storage and sharing across clinics. Given the success of implementation, cost benefits by several departments, and increase in high-volume screening and risk prediction for HBOC, this system is likely to improve the efficiency and effectiveness of care as well as increase the number of at-risk relatives who can be identified and prioritized for assessment.

The successful implementation of HRA at the NWH illustrates how large-scale identification and management of women at high-risks can be conducted in an efficient manner in primary care settings, which can play a significant role in the prevention or early detection of cancer for many women who may be otherwise unaware of their risk. In 2012, Drohan et al. reported that 2,255 women had been identified by HRA to be at high risks. However, even though BRCAPROLYTE has been successfully implemented in clinical settings, its effectiveness is unknown as there has been no formal statistical evaluation and validation performed to-date, as is with other statistical models for predicting an individual's lifetime risk (Amir et al., 2010; Gail & Mai, 2010). This process which can critically affect the health of millions of women must be formally validated and enhanced for maximum efficacy and efficiency.

We formalize the idea of BRCAPROLYTE by developing simplified versions of BRCAPRO. However, to bring BRCAPRO to this simplified level and yet be helpful, we need to balance the tradeoff between simplicity (with regard to the amount of information entered) and the overall accuracy achieved. This motivation naturally leads to a two-stage approach. The first stage is intended for primary care settings wherein individuals provide limited information about their family history of cancers, which is then used to calculate carrier probabilities using a simplified version of BRCAPRO or other simpler models such as FHAT. If this probability is sufficiently high, the patient will be referred for a more exhaustive risk assessment, the second stage, wherein the full version of BRCAPRO is utilized. This two-stage approach could play a central role in identifying potential BRCA mutation carriers so that they can be referred for further risk assessment and genetic testing.

CHAPTER 3

SIMPLIFYING CLINICAL USE OF THE GENETIC RISK PREDICTION MODEL BRCAPRO[1]

3.1. INTRODUCTION

Carriers of deleterious mutations of the *BRCA1* and *BRCA2* genes are at a much higher lifetime risk of developing breast and ovarian cancers than the general population (Antoniou et al., 2003; King, Marks, and Mandell, 2003), and may benefit from more intensive screening, prophylactic surgery, and/or chemoprevention (Schwartz et al., 2008). Yet a majority of mutation carriers remain unaware of their status and risk, and are not managed in a way that might mitigate their risk (Drohan et al., 2012). This is partly because health care providers lack tools that can help them efficiently identify high-risk patients within the time and resource constraints of a busy practice. Genetic risk prediction models used currently in genetic counseling could help fill this gap if adapted and incorporated into Electronic Medical Records (EMRs) or other Health Information Technology (HIT) solutions (Drohan, Ozanne, and Hughes, 2009). Such adaptation could play a central role in identifying potential carriers so that they can be referred for risk assessment, genetic testing, and appropriate management.

The BRCAPRO genetic risk prediction model (Parmigiani, Berry, and Aguilar, 1998) is widely used in genetic counseling and is freely available through the

[1]Originally appearing in Breast Cancer and Research Treatment

BayesMendel R package (Chen et al., 2004), CancerGene genetic counseling package (http://www.utsouthwestern.edu/utsw/cda/dept47829/files/65844.html), HughesRiskApps (HRA; http://www.HughesRiskApps.net), and through a web-based risk service (http://bayesmendel.dfci.harvard.edu/risk/). It estimates the probability that a counselee carries a deleterious mutation of *BRCA1* or *BRCA2* as well as his/her risk of developing cancer. BRCAPRO is improved continually and currently can utilize a wealth of relevant information on proband and family history (Biswas et al., 2012; Chen, Blackford, and Parmigiani, 2009; Katki, 2007; Katki et al., 2008; Tai et al., 2008). However, in many health care settings, collecting exhaustively the family history used by BRCAPRO is not practical. Thus, it is useful to develop simplified adaptations of BRCAPRO. An example, named BRCAPROLYTE, is implemented in HRA (Ozanne et al., 2009).

HRA is a freeware program designed to manage high-risk clinic data as well as to identify high-risk women within the framework of a breast imaging center, a breast surgery practice, or an obstetrics practice. HRA collects family history via a tablet-based, patient self-administered questionnaire and assesses risk fully electronically (Ozanne et al., 2009). To address time and resource constraints, the basic HRA survey only collects a limited family history including the numbers and types of affected first- and second-degree relatives, and their ages of diagnosis. The data are analyzed in real time by BRCAPROLYTE and other models to assess the risk of carrying a *BRCA1/2* mutation. If the risk is high (10 % or greater in many clinical applications), the patient is informed that counseling is advised. The electronic nature of the process makes it

highly efficient. HRA and hence BRCAPROLYTE are currently in use in many clinical settings; however, BRCAPROLYTE has not been evaluated and tested.

In this article, we evaluate BRCAPROLYTE and three other simplified versions of BRCAPRO that we refer to as BRCAPROLYTE-Plus, BRCAPROLYTE-Simple, and BRCAPRO-1Degree. BRCAPROLYTE-Plus takes the same information as BRCAPROLYTE plus the family structure and additionally imputes ages of unaffected relatives. BRCAPROLYTE-Simple does not require knowledge of family structure and imputes both the family structure and the ages of unaffected relatives. BRCAPRO-1Degree is the same as BRCAPROLYTE but uses information on first-degree affected relatives only. In addition, we also consider the Family History Assessment Tool (FHAT) (Gilpin, Carson, and Hunter, 2000), another tool designed to rapidly identify high-risk individuals for testing. We compare the performances of these five tools and also investigate the clinical implications of using them.

3.2. METHODS

3.2.1. DATA

We use data originally collected for the Cancer Genetics Network (CGN) Carrier Probability Validation project (Parmigiani et al., 2007), and additional data from the MD Anderson Cancer Center (MDA), as summarized in Table 3.1. While MDA was one of the sites in Parmigiani et al. (2007), we excluded it from this analysis to avoid overlap with larger and more up-to-date data we have available. Data from Baylor College of Medicine are a population-based sample of Ashkenazi Jews and include a much

smaller proportion of BRCA mutation carriers (2.1 %) compared to other sites, wherein families were selected for participation in high-risk clinics. In total, we consider 2,713 probands with family history information and genetic test results. A total of 576 (21.2 %) probands are BRCA mutation carriers. Three probands are carriers of both *BRCA1* and *BRCA2* mutations. The median family size ranges from 10 to 35, which highlights the practical difficulty of collecting complete family information in many health care settings. We ran BRCAPRO, BRCAPROLYTE, BRCAPROLYTE-Plus, BRCAPROLYTE-Simple, BRCAPRO-1Degree, and FHAT on all probands. In the following, we discuss these tools and the evaluation strategy.

3.2.2. BRCAPRO

BRCAPRO is a Mendelian model utilizing detailed information on all available relatives (of any degree) including relationships between members, ethnicity, ages of breast and ovarian cancer diagnosis, and current age/age of death for unaffected members. We used the version implemented in BayesMendel 2.0-8, which also incorporates breast tumor marker (ER, PR, and Her-2/neu) information for members affected with breast cancer (Biswas et al., 2012; Tai et al., 2008). However, this information was only available for MDA, and so for those families, we evaluated BRCAPRO both with and without using marker information. To be consistent across all sites while combining the results, the latter is the one we used in our summaries. We present differences resulting from including tumor marker in the Discussion section. None of the following simpler tools use marker information.

Table 3.1 Pedigree characteristics by sites

	All sites	MDA	GT	Penn	Duke	JHU	Baylor	UTSW	HCI
Total pedigrees	2713	796	248	773	277	106	282	115	116
Pro. AJ descent	744 (27.4)	80 (10.1)	89 (35.9)	194 (25.1)	26 (9.4)	48 (45.3)	282 (100)	22 (19.1)	3 (2.6)
Pro. BRCA1+	377 (13.9)	107 (13.4)	54 (21.8)	131 (16.9)	37 (13.4)	10 (9.4)	6 (2.1)	17 (14.8)	15 (12.9)
Pro. BRCA2+	202 (7.4)	80 (10.1)	22 (8.9)	57 (7.4)	17 (6.1)	5 (4.7)	0 (0)	9 (7.8)	12 (10.3)
Pedigree size**	20 (15)	35 (19)	18 (8)	16 (10)	19 (8)	15 (7)	15 (7)	10 (6)	23 (16.5)
Age of pro.**	49 (17)	46 (16)	51.5 (14)	49 (19)	48 (13)	50 (12.75)	52 (14)	47 (16)	61.5 (24.25)
Males tested	87 (3.2)	9 (1.1)	2 (0.8)	68 (8.8)	0 (0)	2 (1.9)	0 (0)	6 (5.2)	0 (0)
Males tested and BC	48 (1.8)	5 (0.6)	1 (0.4)	39 (5)	0 (0)	1 (0.9)	0 (0)	2 (1.7)	0 (0)
Pro. unilateral BC	1628 (60)	517 (64.9)	198 (79.8)	500 (64.7)	193 (69.7)	49 (46.2)	33 (11.7)	51 (44.3)	87 (75)
Pro. bilateral BC	244 (9)	94 (11.8)	34 (13.7)	51 (6.6)	46 (16.6)	7 (6.6)	0 (0)	12 (10.4)	0 (0)
Pro. with OC	245 (9)	87 (10.9)	21 (8.5)	86 (11.1)	27 (9.7)	5 (4.7)	2 (0.7)	10 (8.7)	7 (6)
Pro. with BC and OC	88 (3.2)	27 (3.4)	8 (3.2)	35 (4.5)	11 (4)	1 (0.9)	0 (0)	6 (5.2)	0 (0)
BC age for pro.**	43 (14)	42 (13)	44 (12)	42 (15)	42 (12)	46 (10.25)	48 (13)	42 (13.5)	47 (20.5)
OC age for pro.**	51 (14)	54 (17)	47 (8)	53.5 (14.75)	49 (7)	50 (14)	58 (20)	49 (11.75)	52 (8.5)

All data except for the MDA site are from Parmigiani et al. (2007). Entries are numbers followed by percents in parentheses except for rows denoted by ** where entries are median followed by Inter-quartile range (IQR) in parenthesis

Pro. Proband. *BC* Breast Cancer, *OC* Ovarian Cancer, *MDA* MD Anderson Cancer Center, *GT* Georgetown University, *Penn* University of Pennsylvania, *Duke* Duke University, *JHU* Johns Hopkins University, *Baylor* Baylor College of Medicine, *UTSW* University of Texas Southwestern Medical Center, *HCI* Huntsman Cancer Institute

3.2.3. BRCAPROLYTE

BRCAPROLYTE evaluates BRCAPRO using age of the proband and ages of diagnosis for *affected* first- and second-degree relatives. A proband is asked about the numbers and types of first- and second-degree relatives (including maternal/paternal side information), and if any of those relatives are affected with cancer. If the proband indicates that a relative has cancer, BRCAPROLYTE further requires the age of diagnosis. For unaffected relatives, no additional information is collected. Also, the AJ status of the proband is collected and utilized in calculations. We evaluated BRCAPROLYTE using BRCAPRO, by setting the current age/age at death of unaffected relatives as missing.

3.2.4. BRCAPROLYTE-PLUS

As BRCAPROLYTE ignores unaffected relatives, its carrier probabilities are generally inflated. However, it is not excessively onerous to collect information on the *numbers* of first- and second-degree relatives, as HRA does currently. Using these, in BRCAPROLYTE-Plus, we impute the ages of *unaffected* relatives to compensate for this inflation, and thereby reduce false positives.

For imputation purposes, “age” refers to current age or age at death. BRCAPROLYTE-Plus imputes ages by utilizing an external independent dataset of unaffected relatives from families collected in colorectal cancer high-risk clinics (Chen et al., 2006). In Table 3.2, we list the median and inter-quartile range of ages for different first- and second-degree relative types in this dataset, stratified by the number of relatives of that type (1, 2–4, and ≥5). For imputation, we use the median age from this

table. For example, if a proband has 3 maternal aunts, we impute their current ages using the median age of maternal aunts of probands who have 2–4 maternal aunts (67.5 years). We do not impute family structure in BRCAPROLYTE-Plus, so BRCAPROLYTE-Plus requires the same information as BRCAPROLYTE.

Table 3.2 Median and interquartile range of ages of various relative types stratified by the number of relatives obtained from the colorectal data (Chen et al., 2006)

Relative type	Number of relatives		
	1	2–4	≥5
Sister	48 (20)	48.5 (18.5)	54 (19.4)
Brother	49 (19)	48 (16)	57.1 (19.4)
Daughter	28 (23)	33.5 (21.5)	48.2 (13.1)
Son	28 (21)	30 (23.2)	46.4 (8.9)
Maternal aunt	70 (17)	67.5 (19)	68.2 (16.3)
Maternal uncle	67 (20.5)	65.7 (19.2)	61.8 (13.5)
Paternal aunt	70 (28)	68.5 (18.5)	69.5 (14.3)
Paternal uncle	67 (21.8)	66 (16)	67.5 (15.3)
Mother	70 (20)	-	-
Father	69 (19)	-	-
Paternal grandmother	76 (20)	-	-
Paternal grandfather	70 (25)	-	-
Maternal grandmother	75 (22)	-	-
Maternal grandfather	70 (24)	-	-

BRCAPROLYTE-Plus and BRCAPROLYTE-Simple impute the median ages for ages of unaffected relatives

3.2.5. BRCAPROLYTE-SIMPLE

BRCAPROLYTE-Plus requires that the family structure be known. To explore whether the burden for data collection can be further reduced, we examine BRCAPROLYTE-Simple, which only requires information on the numbers and types of *affected* relatives and their ages of diagnosis. Unlike BRCAPROLYTE-Plus, this does not need knowledge of the total number of relatives of each type. BRCAPROLYTE-Simple imputes the number of relatives using the median number of relatives for each

relative type from the same colorectal data used for imputing ages in BRCAPROLYTE-Plus (Chen et al., 2006). The median number of relatives is one for each relative type that is listed in Table 3.2. So, if a proband does not have an affected relative of a particular type, a single unaffected relative of that type is created. Imputation of ages for the newly created unaffected relatives proceeds as in BRCAPROLYTE-Plus.

3.2.6. BRCAPRO-1DEGREE

This tool is similar to BRCAPROLYTE but only uses affected relatives up to the first degree. So, to run it, we set information on all relatives beyond the first degree and all unaffected first-degree relatives as missing.

3.2.7. FHAT

FHAT uses a 17-question interview about affected relatives to produce a quantitative score. Any relative affected with breast, ovarian, prostate, or colon cancer up to 3rd degree contribute to the score. A score of 10 or higher is typically considered as indicative of high risk. So, for FHAT, we use this cutoff for calculating sensitivity, specificity, and predictive values as described below.

3.2.8. EVALUATION STRATEGY

We use scatterplots to visually compare the probability of carrying any BRCA mutation, as generated by each of the simpler tools to those obtained using BRCAPRO. Next, we evaluate the clinical impact of using a simplified tool in place of BRCAPRO. For this, first we compare various tools in terms of the numbers of probands whose

carrier probabilities exceed or are equal to different cutoffs (i.e., the number of referrals) by each tool, and the number of carriers captured among those referred. With this information, we investigate what cutoffs may be appropriate for simpler tools to clinically perform similar to how BRCAPRO performs at 10 %, the most commonly used cutoff. To further assess clinical impact, we consider the additional numbers of probands who are classified correctly or incorrectly as high or low risk (i.e., referred or not referred) using a simpler version as compared to BRCAPRO. Here the classification is considered correct if a carrier is classified as high risk or a non-carrier is classified as low risk. Thus, for such comparison, four numbers are of interest – two each for correct and incorrect classification. These are combined in a measure called Net Reclassification Improvement (NRI) (Pencina et al., 2008), which we report along with its four components. Next, we plot the Receiver Operating Characteristic (ROC) curve and report the Area under the ROC curve (AUC) for all tools. We also report the sensitivity, specificity, predictive value (PV) positive (PVP) and negative (PVN) at various cutoffs. To assess calibration, we compare the observed number of carriers to the number of carriers expected according to each method. For FHAT, we do not evaluate calibration and NRI as the FHAT score is not in the probability scale. We find 95 % confidence interval (CI) obtained using the bootstrap method (Efron and Tibshirani, 1994) for each of the reported statistics. We used the statistical software R 2.15.2 for all computations.

3.3. RESULTS

Figure 3.1 shows scatterplots of carrier probabilities from the five simpler tools plotted against those from BRCAPRO. The BRCAPROLYTE probabilities are, in general, larger than the corresponding BRCAPRO probabilities. This is expected as BRCAPROLYTE only uses information on affected relatives, leading to inflation of the probability. BRCAPROLYTE-Plus, by imputing the ages for those relatives, decreases the probabilities across the range as seen from the fact that the points in its plot are closer to the diagonal line of equality with BRCAPRO. BRCAPROLYTE-Simple shows an intermediate pattern between those of BRCAPROLYTE and BRCAPROLYTE-Plus. The probabilities from BRCAPRO-1Degree seem to have the least correlation with those from BRCAPRO. This demonstrates that information on first-degree relatives only is not generally enough to capture family history for counseling purposes. Finally, FHAT scores are positively correlated with BRCAPRO probabilities.

In clinical applications it is common to consider a specific threshold of risk as a trigger for differential clinical management, as in this case, referral to counseling. In Table 3.3, we report the total number of referrals and the number of carriers captured in those referrals. We would ideally like to have fewer referrals (and hence reduced burden of following up the referred patients) but capture more carriers in those referrals. At the most commonly used cutoff of 10 %, BRCAPRO captures 413 carriers out of 1,031 referrals giving a percentage of carriers per referral (or predictive value positive) of 40 %. In other words, among probands whose carrier probability is 10 % or higher, 40 % are actually carriers. The corresponding percentages for other tools are 30 %

(BRCAPROLYTE), 42 % (BRCAPROLYTE-Plus), 32 % (BRCAPRO-1Degree), 36 %

(BRCAPROLYTE-Simple), and 30 % (FHAT). Thus, BRCAPROLYTE-Plus is closest to

Figure 3.1 Probabilities of carrying any BRCA mutation as computed by the five simpler tools plotted against those from BRCAPRO

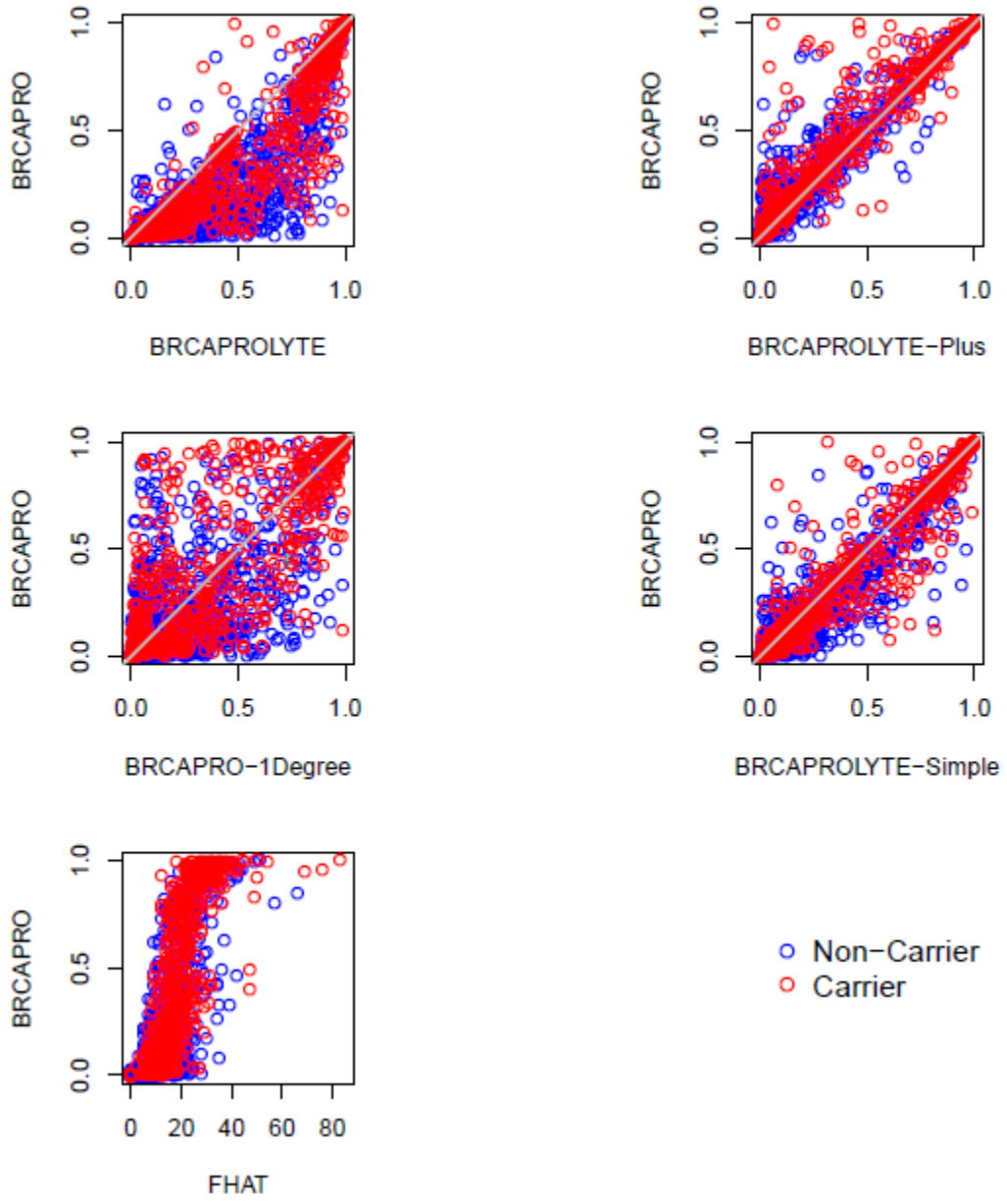


Table 3.3 Numbers of referrals (denominator) made by each tool at different cutoffs and the number of carriers (numerator) out of those referrals

Cutoff	BRCAPRO	LYTE	LYTE-Plus	BRCAPRO-1Degree	LYTE-Simple	FHAT
0.3	300/580 = 0.52	**387/997 = 0.39**	276/527 = 0.52	307/713 = 0.43	316/648 = 0.49	
0.25	320/649 = 0.49	**407/1105 = 0.37**	298/590 = 0.51	329/815 = 0.40	334/721 = 0.46	
0.2	345/734 = 0.47	436/1250 = 0.35	319/652 = 0.49	**352/922 = 0.38**	**357/818 = 0.44**	
0.1	**413/1031 = 0.40**	493/1631 = 0.30	**379/902 = 0.42**	**418/1301 = 0.32**	**426/1166 = 0.37**	
0.05	457/1358 = 0.34	531/2027 = 0.26	**438/1243 = 0.35**	475/1726 = 0.28	486/1521 = 0.32	
0.03	497/1620 = 0.31	552/2254 = 0.24	471/1476 = 0.32	516/2060 = 0.25	511/1788 = 0.29	
0.01	547/2069 = 0.26	568/2583 = 0.22	535/2002 = 0.27	560/2542 = 0.22	556/2299 = 0.24	
10						488/1625 = 0.30

The bold numbers correspond to the commonly used threshold of 10 % for referral by BRCAPRO and for simpler tools, they correspond to the modified thresholds at which the respective tools perform closest to BRCAPRO

LYTE represents BRCAPROLYTE

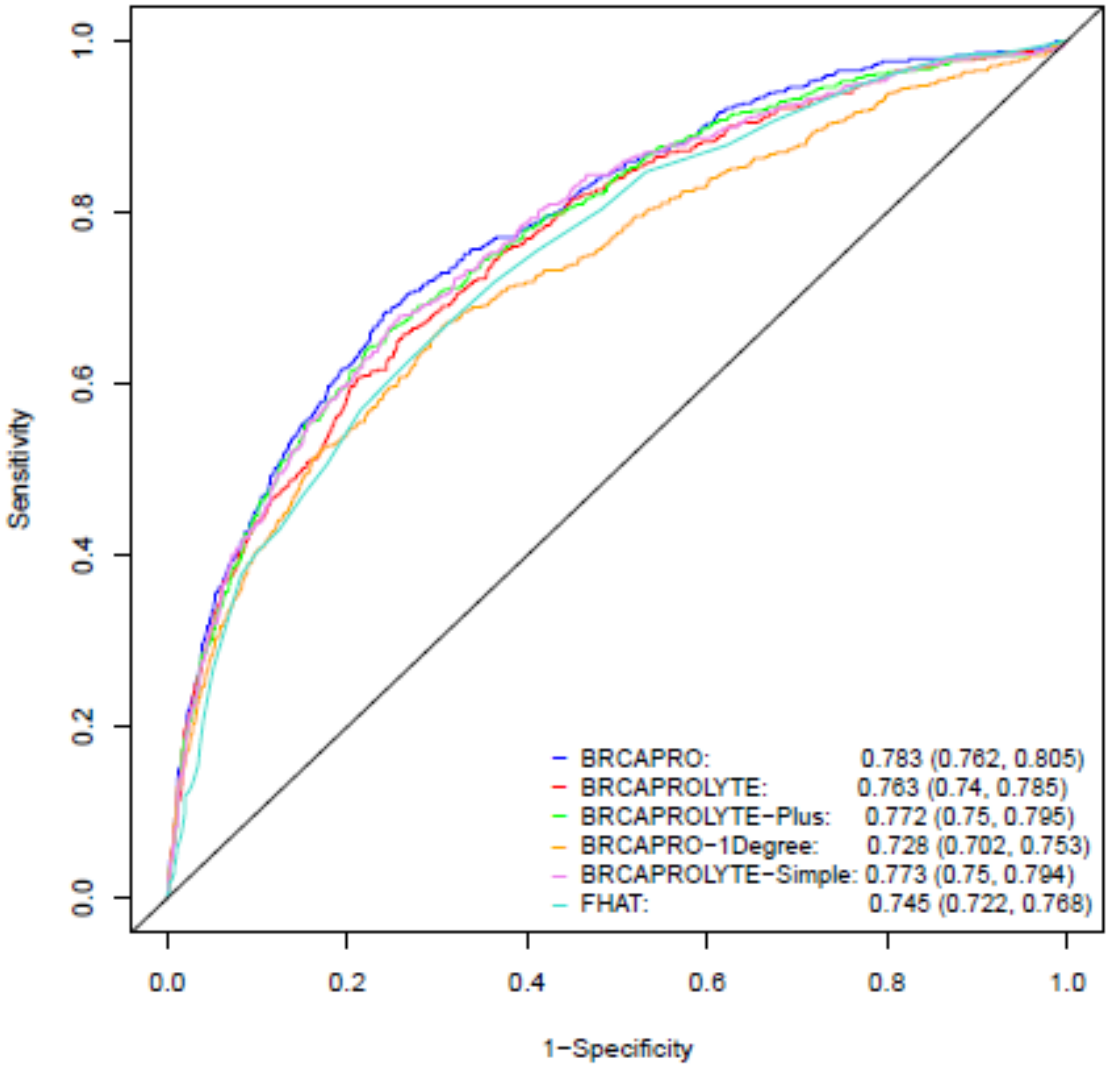
BRCAPRO in this regard followed by BRCAPROLYTE-Simple. The tools compare similarly at other cutoffs. Table 3.3 can also be used to find cutoffs at which simpler tools perform closest to what BRCAPRO provides at the 10 %, or other, cutoff. For example, if we want the number of referrals to be similar to that of BRCAPRO at 10 % (1,031), the cutoff to be used for BRCAPROLYTE should be slightly less than 30 %; however, if we want to capture similar number of carriers (413) we might set the threshold between 20 and 25 %. For BRCAPROLYTE-Plus and BRCAPROLYTE-Simple, the respective cutoffs should be slightly lower and higher than 10 %, and they will capture comparable numbers of carriers as BRCAPRO. For BRCAPRO-1Degree, the cutoff should be close to 20 %, and it will capture fewer carriers than BRCAPRO. Overall, it is clear that BRCAPROLYTE-Plus performs best among all simpler tools. Finally, FHAT at cutoff 10 has a comparable number of referrals as do BRCAPRO at cutoff 3 % and BRCAPROLYTE at cutoff 10 %, and has

also comparable percentage of carriers captured among those referred as BRCAPROLYTE.

To further evaluate the clinical implications of using these simpler versions, we consider how many additional probands would be reclassified if one was to switch from BRCAPRO to a simplified implementation. At the threshold of 10 % carrier probability, BRCAPROLYTE reclassifies to high risk (i.e., the carrier probability moves above the threshold), 14 % of carriers (a desirable reclassification) and 25 % of non-carriers (an undesirable reclassification), while it does not reclassify any carrier or non-carrier to low risk. This is summarized in NRI (Table 3.4), which is calculated as (0.14 – 0) – (0.25 – 0) = –0.11; the negative value in this case reflects the fact that BRCAPROLYTE is worse in classification than BRCAPRO. BRCAPROLYTE-Plus, at the same cutoff, reclassifies fewer families than BRCAPROLYTE, and the difference with BRCAPRO is not statistically significant on the NRI scale. In Table 3.4, we report the NRI and its four components for the tools studied here, at three different clinically relevant thresholds. BRCAPROLYTE-Plus is closest to BRCAPRO with its CI including 0 for each threshold. The next best is BRCAPROLYTE-Simple.

Figure 3.2 shows ROC curves and the corresponding AUC for the combined sample. Among the simplified tools, BRCAPROLYTE-Plus and BRCAPROLYTE-Simple have the highest AUC while BRCAPRO-1Degree performs worst. The AUCs vary by sites as reported in Table 3.5. In general, of the simpler tools, BRCAPROLYTE-Plus performs the best followed closely by BRCAPROLYTE-Simple.

Figure 3.2 ROC curves with AUC and their 95 % CI



Next, in Table 3.6, we report sensitivity, specificity, and PVs. These statistics vary across different tools at the same cutoff as reflected earlier in the varying numbers of referral and carriers captured. In fact, PVP is equivalent to the % of carriers captured per referral as reported in Table 3.3. At 10 % cutoff, BRCAPROLYTE has the highest sensitivity, even higher than BRCAPRO, but has the lowest specificity while BRCAPROLYTE-Plus has the lowest sensitivity and highest specificity. If one is interested in comparing specificities of different tools for a fixed sensitivity, Figure 3.3

may be used. For example, at the 80 % sensitivity value denoted by the horizontal gray line, the specificity values for different tools can be found by drawing a vertical line from the 80 % sensitivity point to the corresponding specificity curve. BRCAPROLYTE-Plus and BRCAPROLYTE-Simple give slightly higher specificity than BRCAPROLYTE for similar values of sensitivity. Also, from Table 3.6, the specificity, sensitivity, and PVs of FHAT are similar to that of BRCAPROLYTE at 10 % cutoff.

With regard to calibration, the average number of carriers estimated by each tool is 517.30 (BRCAPRO), 820.83 (BRCAPROLYTE), 604.34 (BRCAPRO-1Degree), 464.71 (BRCAPROLYTE-Plus), and 577.91 (BRCAPROLYTE-Simple). By comparing these to the observed number of carriers, 576, we see that BRCAPROLYTE overestimates the overall number of carriers while BRCAPROLYTE-Plus underestimates. BRCAPROLYTE-Simple is best and even slightly better than BRCAPRO.

3.4. DISCUSSION

We have developed and evaluated simplified versions of BRCAPRO. Of these, BRCAPROLYTE has been in use in clinical settings, though this is the first time that it is empirically evaluated. Our results show that it has high sensitivity but it overestimates carrier probabilities by a potentially large extent as it relies only on the affected relatives. Thus, we proposed BRCAPROLYTE-Plus wherein ages for unaffected relatives are imputed. This attempts to correct for the overestimation without increasing the burden of data collection. BRCAPROLYTE-Plus does balance the overestimation to some extent and thus gives higher specificity than BRCAPROLYTE for similar values of

Table 3.4 NRI statistic and its four components representing the proportions of carriers (C) and non-carriers (NC) who got reclassified as high risk (moved up) or low risk (moved down) when a simplified tool is used in place of BRCAPRO at the same cutoff

Cutoff	Tool	C.up	C.down	NC.up	NC.down	NRI	95 % CI
0.01	BRCAPROLYTE	0.04	0	0.23	0	–0.19	(–0.22, –0.17)
	BRCAPRO-1Degree	0.03	0.01	0.22	0.01	–0.19	(–0.22, –0.17)
	BRCAPROLYTE-Plus	0.01	0.03	0.04	0.07	0	(–0.02, 0.03)
	BRCAPROLYTE-Simple	0.02	0	0.11	0.01	–0.09	(–0.11, –0.07)
0.05	BRCAPROLYTE	0.13	0	0.28	0	–0.15	(–0.18, –0.12)
	BRCAPRO-1Degree	0.1	0.07	0.21	0.05	–0.13	(–0.17, –0.1)
	BRCAPROLYTE-Plus	0.02	0.05	0.02	0.07	0.01	(–0.01, 0.04)
	BRCAPROLYTE-Simple	0.06	0.01	0.09	0.02	–0.01	(–0.04, 0.01)
0.1	BRCAPROLYTE	0.14	0	0.25	0	–0.1	(–0.14, –0.07)
	BRCAPRO-1Degree	0.1	0.09	0.18	0.05	–0.12	(–0.15, –0.07)
	BRCAPROLYTE-Plus	0.01	0.07	0.01	0.06	–0.01	(–0.04, 0.01)
	BRCAPROLYTE-Simple	0.05	0.02	0.08	0.02	–0.03	(–0.06, –0.01)

The four components are C.up, C.down, NC.up, and NC.down

sensitivity. We also showed that the burden of data collection can be further reduced by asking only about the affected relatives and using BRCAPROLYTE-Simple to impute the rest of the family members and their ages. BRCAPROLYTE-Simple performs slightly better than BRCAPROLYTE. FHAT at cutoff 10 performed similar to BRCAPROLYTE at cutoff 10 %; however, BRCAPROLYTE has larger AUC. BRCAPRO-1Degree performs worst, clearly demonstrating the need for collecting information on affected second-degree relatives for genetic risk prediction.

Table 3.5 AUC and its 95 % CI by site

	BRCAPRO	LYTE	LYTE-Plus	BRCAPRO-1Degree	LYTE-Simple	FHAT
MDA	0.774 (0.735, 0.812)	0.737 (0.693, 0.779)	0.768 (0.729, 0.808)	0.681 (0.631, 0.728)	0.759 (0.715, 0.802)	0.737 (0.694, 0.779)
GT	0.765 (0.7, 0.826)	0.743 (0.674, 0.807)	0.755 (0.689, 0.822)	0.722 (0.643, 0.793)	0.75 (0.68, 0.816)	0.716 (0.643, 0.785)
Penn	0.771 (0.732, 0.811)	0.765 (0.724, 0.806)	0.765 (0.724, 0.806)	0.745 (0.7, 0.784)	0.772 (0.732, 0.81)	0.716 (0.671, 0.759)
Duke	0.823 (0.763, 0.877)	0.816 (0.75, 0.874)	0.827 (0.768, 0.881)	0.781 (0.706, 0.85)	0.819 (0.76, 0.873)	0.754 (0.68, 0.821)
JHU	0.829 (0.722, 0.914)	0.81 (0.703, 0.906)	0.841 (0.744, 0.931)	0.7 (0.522, 0.856)	0.816 (0.709, 0.917)	0.83 (0.721, 0.921)
Baylor	0.699 (0.588, 0.816)	0.723 (0.61, 0.842)	0.727 (0.613, 0.836)	0.744 (0.61, 0.869)	0.717 (0.591, 0.844)	0.759 (0.661, 0.845)
UTSW	0.82 (0.72, 0.907)	0.808 (0.71, 0.892)	0.817 (0.72, 0.899)	0.744 (0.615, 0.86)	0.821 (0.726, 0.905)	0.772 (0.673, 0.859)
HCI	0.696 (0.569, 0.815)	0.641 (0.52, 0.754)	0.675 (0.539, 0.797)	0.618 (0.494, 0.734)	0.673 (0.546, 0.791)	0.599 (0.474, 0.725)

LYTE represents BRCAPROLYTE

MDA MD Anderson Cancer Center, *GT* Georgetown University, *Penn* University of Pennsylvania, *Duke* Duke University, *JHU* Johns Hopkins University, *Baylor* Baylor College of Medicine, *UTSW* University of Texas Southwestern Medical Center, *HCI* Huntsman Cancer Institute

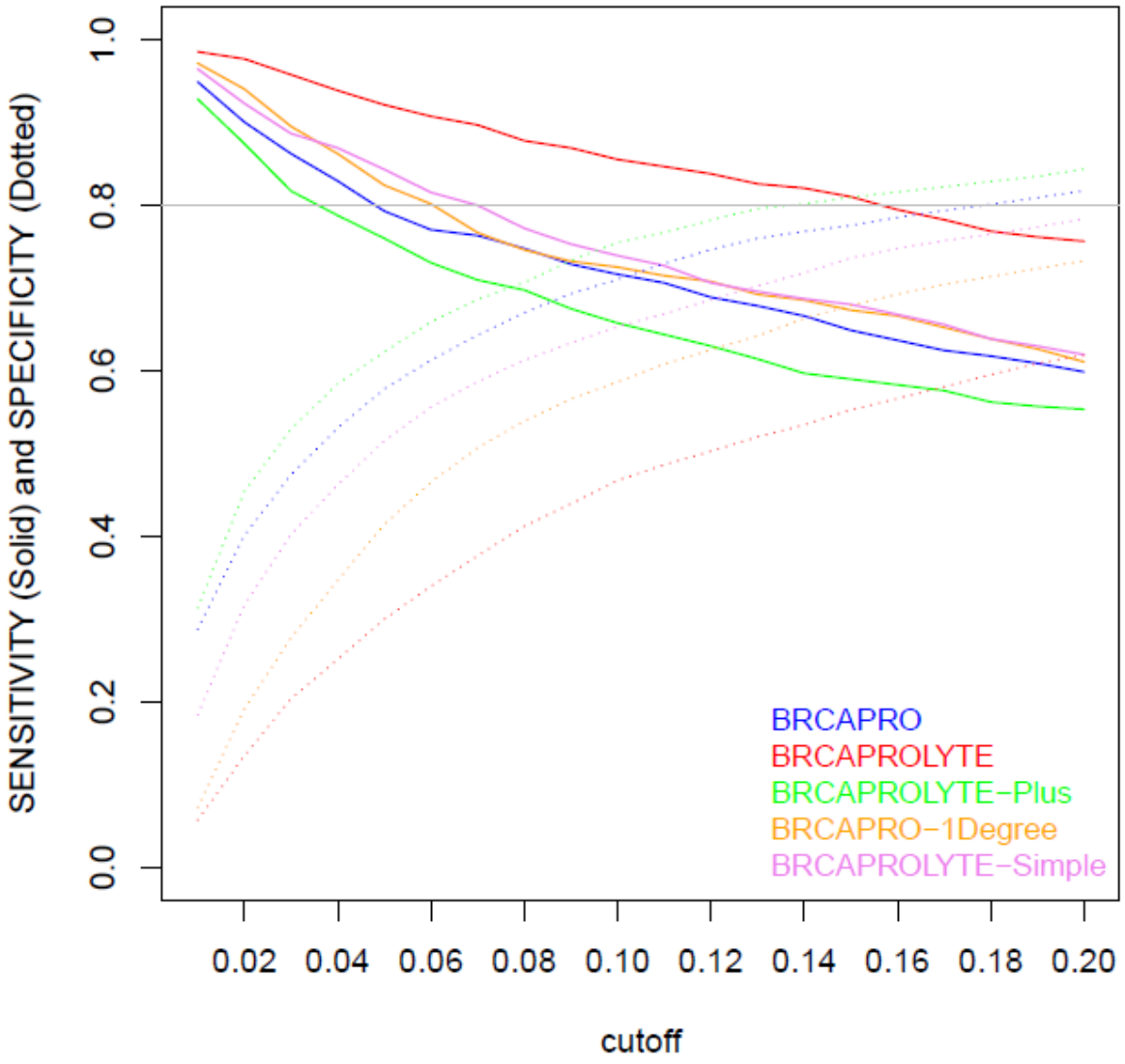
We also found that there is only modest loss in discrimination and calibration by BRCAPROLYTE-Plus and BRCAPROLYTE-Simple as compared to the complete BRCAPRO. From a practical point of view, as these simpler versions take limited amount of family information, they can be efficiently integrated into the EMR and other HIT solutions at the primary care or screening level, and thus can be routinely used to screen patients for their genetic risk. Nonetheless, it must be pointed out that BRCAPRO has additional features that are not included in simplified tools. It can utilize information on tumor markers, genetic test results, and medical interventions such as oophorectomy (Biswas et al., 2012; Katki, 2007; Tai et al., 2008). BRCAPRO must also be available for management beyond screening.

Table 3.6 Sensitivity, Specificity, PVP, and PVN, and their 95 % CI

	Cutoff	BRCAPRO	LYTE	LYTE-Plus	BRCAPRO-1Degree	LYTE-Simple	FHAT
Se	0.1	0.72 (0.68, 0.75)	0.86 (0.83, 0.88)	0.66 (0.62, 0.7)	0.73 (0.69, 0.76)	0.74 (0.7, 0.77)	
Sp		0.71 (0.69, 0.73)	0.47 (0.45, 0.49)	0.76 (0.74, 0.77)	0.59 (0.57, 0.61)	0.65 (0.63, 0.67)	
Se	0.05	0.79 (0.76, 0.83)	0.92 (0.9, 0.94)	0.76 (0.72, 0.79)	0.82 (0.79, 0.85)	0.84 (0.81, 0.87)	
Sp		0.58 (0.56, 0.6)	0.3 (0.28, 0.32)	0.62 (0.6, 0.64)	0.41 (0.39, 0.44)	0.52 (0.49, 0.54)	
Se	0.03	0.86 (0.83, 0.89)	0.96 (0.94, 0.97)	0.82 (0.79, 0.85)	0.9 (0.87, 0.92)	0.89 (0.86, 0.91)	
Sp		0.47 (0.45, 0.5)	0.2 (0.19, 0.22)	0.53 (0.51, 0.55)	0.28 (0.26, 0.3)	0.4 (0.38, 0.42)	
Se	0.01	0.95 (0.93, 0.97)	0.99 (0.98, 0.99)	0.93 (0.91, 0.95)	0.97 (0.96, 0.98)	0.97 (0.95, 0.98)	
Sp		0.29 (0.27, 0.31)	0.06 (0.05, 0.07)	0.31 (0.29, 0.33)	0.07 (0.06, 0.08)	0.18 (0.17, 0.2)	
Se	10						0.85 (0.82, 0.88)
Sp							0.47 (0.45, 0.49)
PVP	0.1	0.4 (0.37, 0.43)	0.3 (0.28, 0.32)	0.42 (0.39, 0.45)	0.32 (0.3, 0.35)	0.37 (0.34, 0.39)	
PVN		0.9 (0.89, 0.92)	0.92 (0.91, 0.94)	0.89 (0.88, 0.9)	0.89 (0.87, 0.9)	0.9 (0.89, 0.92)	
PVP	0.05	0.34 (0.31, 0.36)	0.26 (0.24, 0.28)	0.35 (0.33, 0.38)	0.28 (0.25, 0.3)	0.32 (0.3, 0.34)	
PVN		0.91 (0.9, 0.93)	0.93 (0.92, 0.95)	0.91 (0.89, 0.92)	0.9 (0.88, 0.92)	0.92 (0.91, 0.94)	
PVP	0.03	0.31 (0.28, 0.33)	0.24 (0.23, 0.26)	0.32 (0.3, 0.34)	0.25 (0.23, 0.27)	0.29 (0.26, 0.31)	
PVN		0.93 (0.91, 0.94)	0.95 (0.93, 0.97)	0.92 (0.9, 0.93)	0.91 (0.89, 0.93)	0.93 (0.91, 0.95)	
PVP	0.01	0.26 (0.25, 0.28)	0.22 (0.2, 0.24)	0.27 (0.25, 0.29)	0.22 (0.2, 0.24)	0.24 (0.22, 0.26)	
PVN		0.95 (0.94, 0.97)	0.94 (0.89, 0.98)	0.94 (0.92, 0.96)	0.91 (0.86, 0.95)	0.95 (0.93, 0.97)	
PVP	10						0.3 (0.28, 0.32)
PVN							0.92 (0.9, 0.94)

LYTE represents BRCAPROLYTE

Figure 3.3 Sensitivity and specificity for cutoffs ranging from 0.01 to 0.2 calculated at an increment of 0.01



In our data, MDA is the only site that had information on ER, PR, and Her-2/neu. These were not utilized to assess BRCAPRO in the results presented here, to facilitate comparison with other sites and to better focus on assessing the information loss from omitting the questions about unaffected relatives' age and/or family structure. If tumor marker information is available, the loss in using simplified versions of BRCAPRO is greater. For the MDA site, the AUC of BRCAPRO including marker information increases from 0.774 (in Table 3.5) to 0.802, and sensitivity/specificity at 10 % cutoff

increase from 0.64/0.78 to 0.66/0.79. The calibration is only slightly changed. If tumor marker information is readily available, the loss in using simpler tools is generally greater and must be weighed against the data collection burden associated with complete BRCAPRO.

For BRCAPROLYTE-Plus, we imputed missing ages using the median age, after stratifying by the number of relatives. We also carried out two sensitivity analyses by using the mean age in place of the median age and by using a coarser stratification by numbers of relatives. The results from both analyses are very close to what we have reported.

A practical issue is the choice of cutoff to be used for the simplified tools in clinical settings. We described how cutoffs may be chosen so that the number of referrals by a tool is comparable to that of BRCAPRO. One could also consider sensitivity and specificity. The trade-off between sensitivity, specificity, and burden of following up of referrals can be evaluated using Tables 3.3, 3.6, and Figure 3.3 together to choose a cutoff that suits specific needs. For example, a user of BRCAPROLYTE-Simple could achieve a sensitivity of 0.84 and a specificity of 0.52 using a cutoff of 0.05. This would lead to referral of about half of the patients (1521/2713). As seen in Figure 3.3, for achieving the same sensitivity, different tools require different cutoffs. As different clinical scenarios may require a different balance of specificity, sensitivity, and cost and benefit of genetic counseling, we recommend a careful weighing of cutoffs prior to implementation. One should also keep in mind that the widely used 10 % cutoff has different interpretations and implications depending on the tool used, as we have discussed in the Results section. For a specific clinical scenario, more formal statistical

analysis can be carried out to determine an optimal cutoff if the associated cost and benefit for genetic risk prediction can be quantified (Parmigiani, 2002).

A limitation of our study is that the data used here are mostly from high-risk families. For our one population-based sample, Baylor, we found that sensitivity dropped faster with increasing cutoff and so smaller cutoffs should be used for such a scenario, as expected. The performance of the proposed tools for this site is similar to the results we presented here for combined sample and this is consistent with earlier studies (Parmigiani et al., 2007). However, the sample size and the number of carriers for this site is small and so it would be useful to validate these approaches on a larger population-based sample.

In summary, we have shown that one can use modifications of BRCAPRO with limited collection of family history to construct simple and practical risk assessment tools whose performance is comparable to that of standard tools used in high-risk clinics. This limited data collection is feasible in a busy practice. Thus, these tools have formidable potential to bring the benefits of genetic counseling and testing to large sections of the population who are still unaware of the important prevention implications of inherited susceptibility.

CHAPTER 4

A TWO-STAGE APPROACH TO GENETIC RISK ASSESSMENT IN PRIMARY CARE[2]

4.1. INTRODUCTION

The risk of developing breast and ovarian cancers is high for the carriers of deleterious mutations of the BRCA1 and BRCA2 genes, and thus it is critical to identify carriers as early as possible (Antoniou et al., 2003; King, Marks, and Mandell, 2003). Yet there is a lack of streamlined procedures for identifying mutation carriers from a general population. As a result, many carriers remain unaware of their status (Drohan et al., 2012). The identification of potential carriers can be ideally initiated by primary care providers; however, they need tools that can help them in efficiently identifying high-risk patients within their constraints of limited time and resources. Genetic risk prediction models that are used currently in genetic counseling are effective but too complex for the primary care setting, unless they can be simplified and incorporated into the electronic medical records (EMR) or other health information technology (HIT) solutions (Drohan, Ozanne, and Hughes, 2009). With a simplified adaptation, potential carriers can be identified in primary care and referred for further risk assessment and genetic testing.

The BRCAPRO genetic risk prediction model (Parmigiani, Berry, and Aguilar, 1998) is used extensively in genetic counseling and is available in the BayesMendel R

[2]Originally appearing in Breast Cancer and Research Treatment

package (Chen et al., 2004), the CancerGene genetic counseling package (https://www4.utsouthwestern.edu/breasthealth/cagene/), the web-based risk service (http://bayesmendel.dfci.harvard.edu/risk/), the Hughes RiskApps (HRA) package (http://www.HughesRiskApps.com), and other computing environments. Based on the family history information provided by a counselee, BRCAPRO estimates the probability that she/he carries a BRCA1/2 mutation as well as her/his prospective risk of developing cancer. Several improvements to BRCAPRO in recent years allow it to use a variety of information (Biswas et al., 2012; Chen, Blackford, and Parmigiani, 2009; Katki, 2007; Katki et al., 2008; Mazzola et al., 2015; Tai et al., 2008). Primary care settings and breast imaging centers are ideal for fully reaping the preventative benefits of such risk prediction models at a large population level. However, in such settings, collecting and assembling the exhaustive family history used by BRCAPRO are not practical.

We have recently proposed three simplified versions of BRCAPRO: BRCAPROLYTE, BRCAPROLYTE-Plus, and BRCAPROLYTE-Simple (Biswas et al., 2013). We evaluated these tools on datasets collected in genetic counseling settings, and found that they entail only a modest loss of accuracy compared to BRCAPRO, especially BRCAPROLYTE-Plus and BRCAPROLYTE-Simple.

This suggests that we can use these tools to achieve a balance between simplicity (an issue of utmost importance in primary care) and accuracy by carrying out genetic risk prediction in two stages. In the first stage, intended for primary care, risk will be assessed using a simplified version of BRCAPRO. Those found to be at sufficiently high risk will be referred to the second stage (counseling), where the full

BRCAPRO will be used. A software implementation of such a two-stage approach is already available in HRA (Ozanne et al., 2009) which includes two sequentially administered surveys – "Short breast" and "Risk Clinic," using limited and more exhaustive family information. The two-stage procedure has not been evaluated in primary care settings so far.

Our aim is to formally develop and investigate the two-stage approach. We evaluate three versions of this approach, each with a different first-stage tool – BRCAPROLYTE, BRCAPROLYTE-Plus, and BRCAPROLYTE-Simple. In each case, the second stage uses BRCAPRO on all available information.

4.2. METHODS

4.2.1. COHORTS

We use retrospective data from three sources: Cancer Genetics Network (CGN), MD Anderson Cancer Center (MDA), and Newton-Wellesley Hospital (NWH). The first two have been described and analyzed earlier (Biswas et al., 2012; Biswas et al., 2013; Parmigiani et al., 2007); here we analyze them for the first time using two-stage approaches. In particular, the pedigree characteristics for each of seven sites in CGN and that of MDA can be found in Table 1 of Biswas et al. (2013). For completeness, in Table 4.1, we list the characteristics of the two datasets combined (referred as CGN+MDA and analyzed as a whole throughout). MDA as well as all CGN sites except one consists of high-risk families, i.e., the probands entered the study at least in part because of their personal and/or family history. The third cohort captures all probands

referred for genetic counseling at the NWH. The vast majority of these individuals enter the sample through a primary care encounter, typically a breast imaging visit. They are subsequently referred to counseling if they have a prior history of ovarian cancer, meet National Comprehensive Cancer Network (NCCN) guidelines (http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#detection), their BRCAPROLYTE probability exceeds 10 %, or are referred by a clinician based on his/her interpretation of risk. A fraction of this cohort sought genetic counseling after a relative had already been tested; they will be analyzed separately also. Pedigree characteristics of NWH data are listed in Table 4.1. Less than 10 % of NWH probands are BRCA mutation carriers compared to 21 % carriers in CGN+MDA data. The latter has also higher proportions of probands with breast and/or ovarian cancers along with younger affection ages. The median reported family size is about 20, highlighting the practical difficulty of collecting complete family information in primary care.

Table 4.1 Pedigree Characteristics

	CGN+MDA	NWH
Total pedigrees	2713	1344
Probands of AJ descent: *n* (%)	744 (27.4)	366 (27.2)
Probands tested BRCA1+: *n* (%)	377 (13.9)	49 (3.6)
Probands tested BRCA2+: *n* (%)	202 (7.4)	76 (5.7)
No. of members per pedigree: median (IQR)	20 (15)	19 (20)
Age of proband: median (IQR)	49 (17)	53 (16)
Males tested: *n* (%)	87 (3.2)	14 (1)
Males tested with BC: *n* (%)	48 (1.8)	6 (0.4)
Probands with unilateral BC: *n* (%)	1628 (60)	698 (51.9)
Probands with bilateral BC: *n* (%)	244 (9)	72 (5.4)
Probands with OC: *n* (%)	245 (9)	55 (4.1)
Probands with BC & OC: *n* (%)	88 (3.2)	17 (1.3)
BC age for proband: median (IQR)	43 (14)	47 (13.75)
OC age for proband: median (IQR)	51 (14)	53 (16.5)

IQR Inter-quartile range, *BC* breast cancer, *OC* ovarian cancer

4.2.2. TWO-STAGE APPROACH

In this approach, the first-stage tool is either BRCAPROLYTE, BRCAPROLYTE-Plus, or BRCAPROLYTE-Simple. We had earlier also considered the Family History Assessment Tool (FHAT) (Gilpin, Carson, and Hunter, 2000) and another simpler version of BRCAPRO based on first-degree relatives only. Their performance was inferior compared to the three selected and thus we do not pursue them here. In the second stage, the complete BRCAPRO is used on those probands whose first-stage probability exceeds a chosen cutoff.

4.2.3. BRCAPRO

BRCAPRO utilizes information on all available relatives including family structure, ages of diagnosis, current age/age at death for unaffected members, ethnicity, and additional information such as breast tumor markers and BRCA genetic test results, if available. We used the version implemented in BayesMendel 2.0-9.

4.2.4. BRCAPROLYTE

BRCAPROLYTE applies the BRCAPRO model using only information on the numbers and types of first- and second-degree relatives, which relatives are affected with breast and ovarian cancer, and their affection ages. We performed BRCAPROLYTE calculations using BRCAPRO after setting the current age/age at death of all relatives as missing to mimic the scenario when ages are not collected/known to the proband.

4.2.5. BRCAPROLYTE-PLUS

BRCAPROLYTE does not collect data on ages of unaffected relatives leading to inflated carrier probabilities in general (Biswas et al., 2013). As the numbers of first- and second-degree relatives are collected, we can impute the ages of unaffected relatives to compensate for the inflation of probabilities, and thereby offset false positives. This idea is implemented in BRCAPROLYTE-Plus. The imputation of ages is based on an external large dataset as described elsewhere (Biswas et al., 2013).

4.2.6. BRCAPROLYTE-SIMPLE

BRCAPROLYTE-Plus needs knowledge of the numbers of each type of relative. A further simplification can be achieved by imputing this information when it is unknown. BRCAPROLYTE-Simple does this through two levels of imputation: number of relatives of each type and ages of unaffected relatives, based on an external dataset (Biswas et al., 2013). The burden of data collection is therefore the least with BRCAPROLYTE-Simple.

4.2.7. EVALUATION STRATEGY

To establish a baseline, we apply each simplified tool and BRCAPRO separately to all probands and compare the results for NWH data. We earlier reported results for each of the simplified tools on the CGN and MDA data (Biswas et al., 2013). This is the first analysis addressing the two-stage approach.

We then evaluate the clinical impact of two-stage approaches by quantifying the reduction in genetic counseling burden as compared to applying BRCAPRO to all probands. We consider various combinations of cutoffs for the two stages (denoted as c_1 and c_2, respectively) and note the number of counselees whose carrier probabilities exceed c_1 and/or c_2. We compare scenarios constructed so that the numbers of carriers captured by the two-stage approaches and BRCAPRO are about the same.

The statistical evaluation of a two-stage approach is more involved than that of a single-stage tool as results from both stages as well as dependence of the second stage on the first-stage results must be considered. In Supplementary Methods, we show that the overall sensitivity (Se.O), specificity (Sp.O), Area Under ROC Curve (AUC.O), and predictive value positive and negative (PVP.O and PVN.O) can be written in terms of sensitivity and specificity of the first stage and of the second stage *given* the results of the first stage.

Next, we calculate the ratio of the observed (O) number of carriers to the expected (E) number (O/E). To compute E, we need one carrier probability per proband. We use the first-stage probability for counselees who undergo the first stage only (first-stage probability $<c_1$), while for the rest, we use their second-stage probability. We plot O/E, along with their 95 % confidence interval (CI), for a set of c_1 values ranging from 0.01 to 0.1, the cutoffs typically used in practice.

We also consider scenarios where the percentage of counselees to be followed-up in the second stage is fixed in advance in consideration of resource constraints. This requires setting a specific cutoff c_1 for the first stage. Depending on whether a proband was evaluated at the second stage, we use either the first- or second-stage probability

and calculate AUC referred as AUC.p, where “p” is the fixed percentage follow-up and is set to 25, 50, and 75 % in turn.

For all analyses, we also report the results when all probands are evaluated using BRCAPRO. We refer to each two-stage approach by the name of the corresponding first-stage tool, as BRCAPRO is always used in the second stage. We used the statistical software R 3.1.0 for all computations.

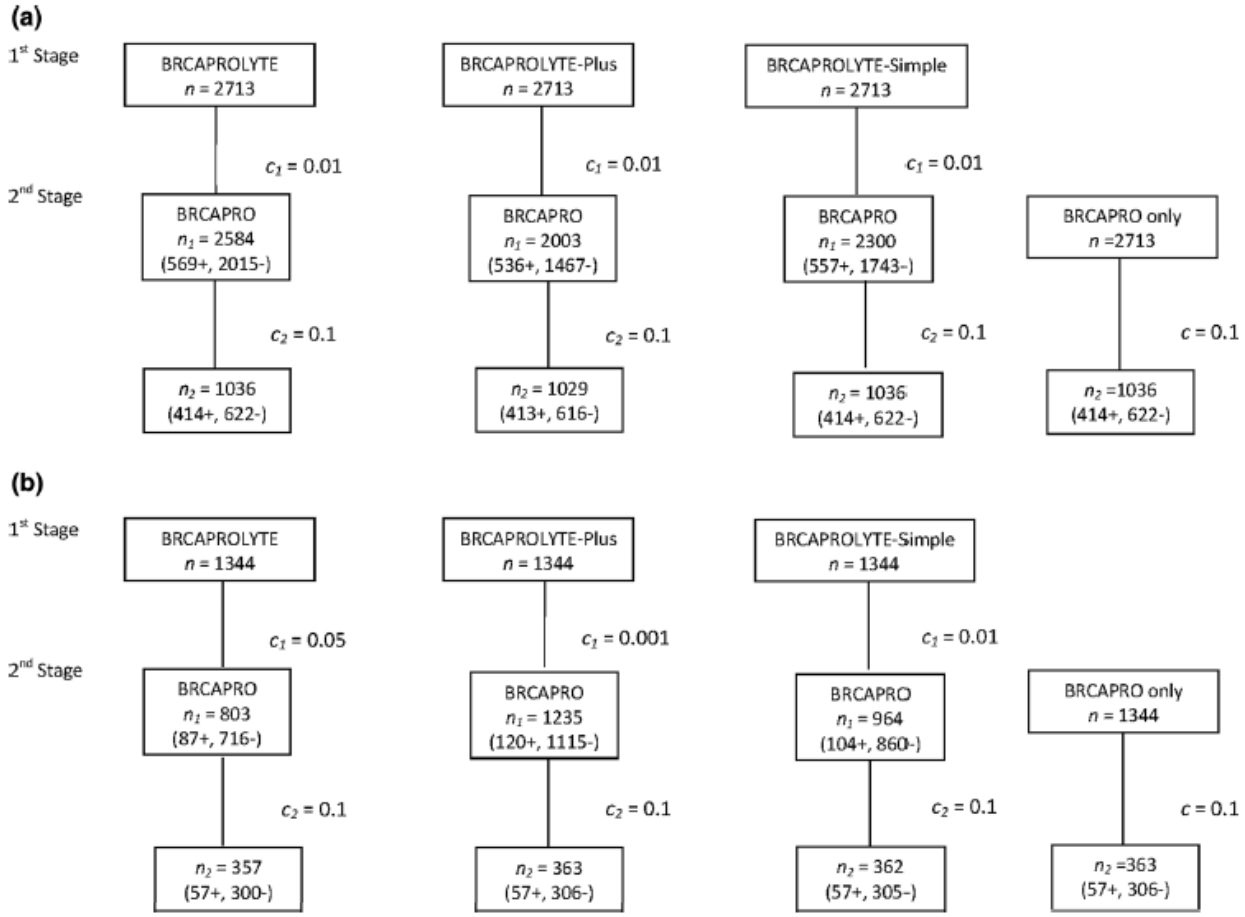
4.3. RESULTS

In NWH, BRCAPROLYTE (applied to all probands) overestimates carrier probabilities compared to BRCAPRO as reflected in O/E values smaller than 1 in Appendix 1 Table A.1. We also observe higher sensitivity and lower specificity of BRCAPROLYTE for a fixed cutoff. BRCAPROLYTE-Plus and BRCAPROLYTE-Simple are better calibrated with O/E values closer to 1, and show a trend toward better calibration than BRCAPRO. The AUCs of all three first-stage tools are close to 0.65, the AUC of BRCAPRO in this dataset. We reported similar results for CGN and MDA (Biswas et al., 2013).

Next, we quantify the consequences of using the two-stage approach on the clinical workflow. There are 576/2713 BRCA mutation carriers in CGN+MDA and 125/1344 in NWH. As shown in Figure 4.1, when BRCAPRO is applied to all probands, the carrier probabilities of 1036 and 363 probands exceed 10 %, a cutoff traditionally used in clinical practice, though not necessarily optimum (American Society of Clinical Oncology, 1996). Of these, only 414 and 57 are carriers. The genetic counseling burden with this single-stage approach is the totality of probands (2713 and 1344) and

genetic testing is done for 1036 and 363 probands. The sensitivity of BRCAPRO at this cutoff is 0.72 and 0.46 in CGN+MDA and NWH, respectively (Appendix 1 Table A.1). The corresponding specificities are 0.71 and 0.75. Now let us compare these numbers with those for the two-stage approaches.

Figure 4.1 Numbers of referrals made at each stage using a two-stage approach, as compared to using BRCAPRO only on all probands for **a** CGN+MDA and **b** NWH data

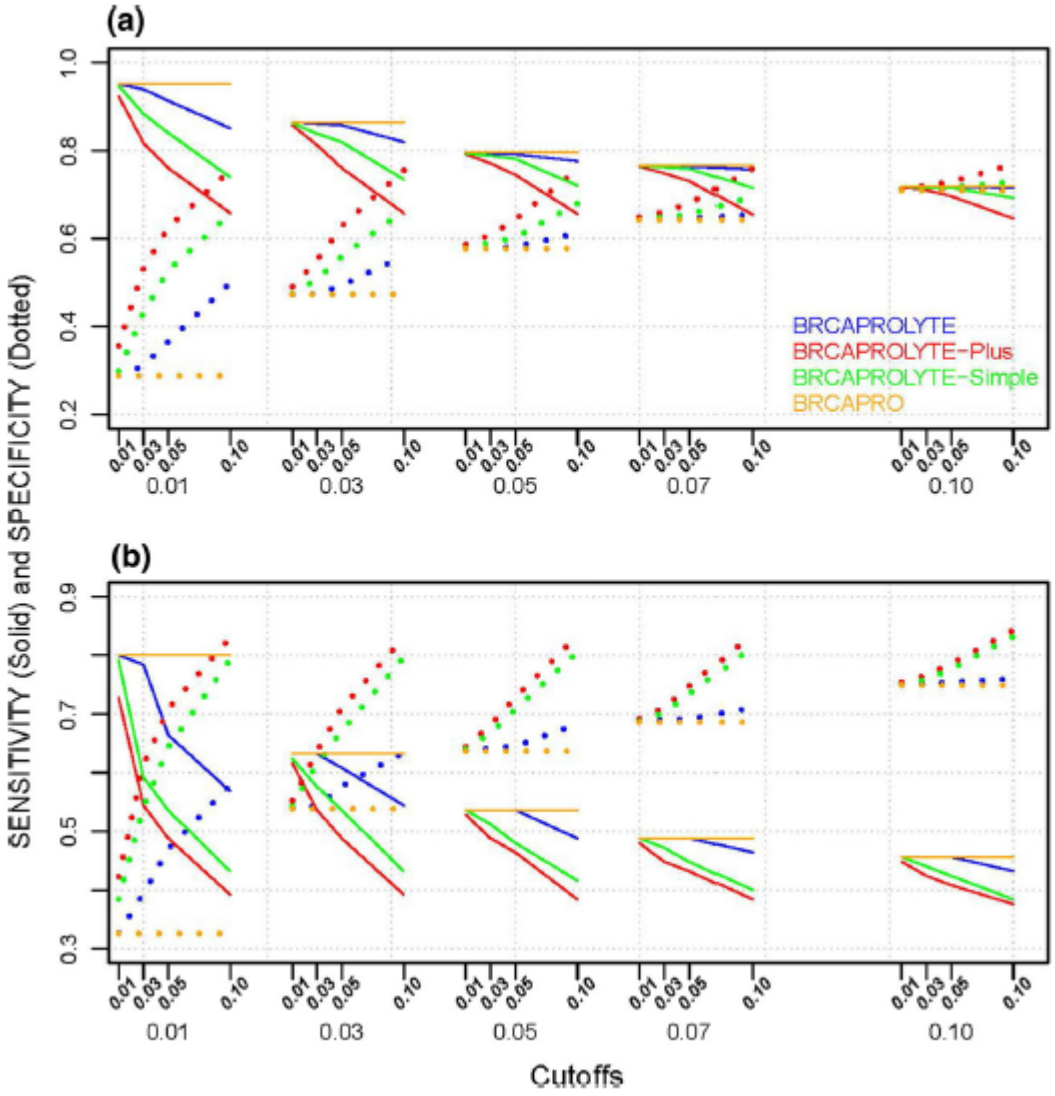


We denote by c_1 and c_2 the cutoffs used at the first and second stages, respectively. The number of probands whose carrier probability at the first stage exceeds c_1 is denoted by n_1, and out of n_1, the number of probands with second-stage carrier probability exceeding c_2 is denoted by n_2. When we evaluate BRCAPRO alone, the cutoff is labeled as c and the number of probands with carrier probability exceeding c is n_2. These numbers are further stratified by the carrier status in *parentheses*

In Figure 4.1b, for NWH, we see that out of a total of 1344 probands that go through the first stage, 803, 1235, and 964 are referred to the second stage by

BRCAPROLYTE, BRCAPROLYTE-Plus, and BRCAPROLYTE-Simple, respectively. Correspondingly, the reduction in genetic counseling burden as compared to direct counseling of all 1344 probands is 541 (40 %), 109 (8 %), and 380 (28 %) families. In the second stage, the numbers of probands referred for genetic testing are close to 363 as obtained using BRCAPRO only. The sensitivity and specificity for all two-stage

Figure 4.2 Sensitivity (Se.O) and specificity (Sp.O) of the two-stage approach and BRCAPRO for **a** CGN+MDA and **b** NWH data



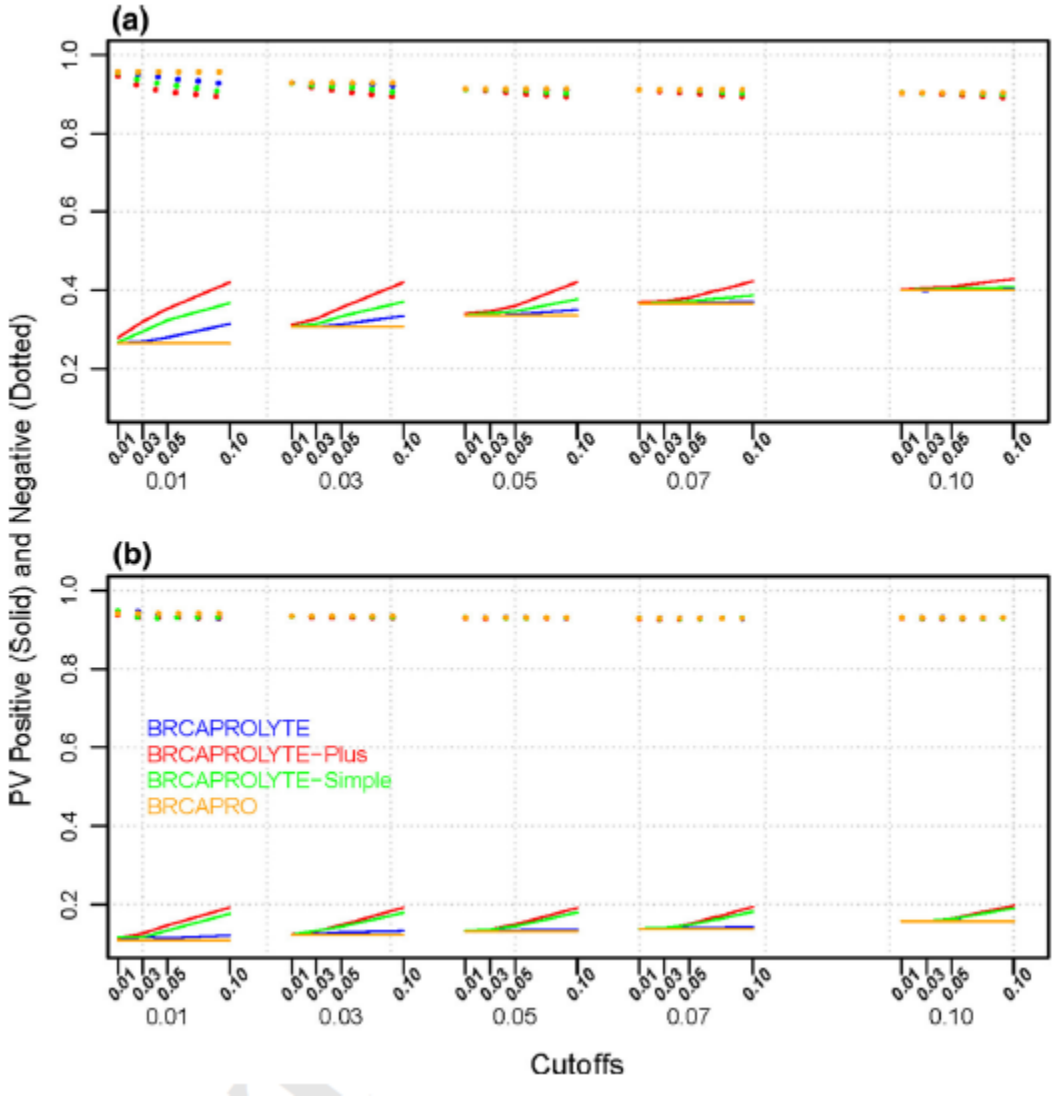
The *x-axis* has two sets of cutoffs, c_1 (first stage) followed by c_2 (second stage) below it. For BRCAPRO, only one cutoff (indicated by the second level c_2 values) is applicable

approaches are about the same as those of BRCAPRO (0.46 and 0.75; Figure 4.2 discussed below). The comparison on CGN+MDA is similar, as shown in Figure 4.1a. In particular, the reduction in genetic counseling is 129 (5 %), 710 (26 %), and 413 (15 %) by using BRCAPROLYTE, BRCAPROLYTE-Plus, and BRCAPROLYTE-Simple in the first stage, respectively. Thus, two-stage approaches are able to capture the same numbers of carriers and achieve same sensitivity and specificity as BRCAPRO with a reduction of genetic counseling burden. Additional scenarios are considered in Appendix 1 Tables A.1 and A.2. We see that similar numbers of carriers can be captured using other cutoff combinations as well. The genetic counseling burden can be further reduced through a higher first-stage cutoff, albeit with larger number of genetic tests.

Figure 4.2 plots Se.O and Sp.O for specific cutoff combinations for the two stages. We see that if we want a two-stage approach to achieve same Se.O value (e.g., 80 %) with a similar (or higher) Sp.O value as that of BRCAPRO, it is usually possible at different cutoff combinations for different models. In general, BRCAPROLYTE has high values of sensitivity, while BRCAPROLYTE-Plus and BRCAPROLYTE-Simple give slightly higher values of specificity for the same sensitivity. If we choose the same cutoff for BRCAPRO when applied by itself and when applied as the second stage (i.e., compare the curves within each column panel), then the former seems to have highest sensitivity and lowest specificity. However, by allowing the cutoffs to vary, the two-stage approaches can achieve similar values of sensitivity and specificity as BRCAPRO. Similar plots for PVP.O and PVN.O are shown in Figure 4.3, and the same trend is seen. Similar considerations apply to PVP.O and PVN.O.

Table 4.2 lists AUC.O, AUC.25, AUC.50, and AUC.75. AUC.O values for all two-stage approaches are practically same as that of BRCAPRO. In fact, AUCs remain comparable even when the percentage of follow-up in the second stage is restricted to 25, 50, or 75 %.

Figure 4.3 Predictive value positive (PVP.O) and negative (PVN.O) of the two-stage approach and BRCAPRO for **a** CGN+MDA and **b** NWH data



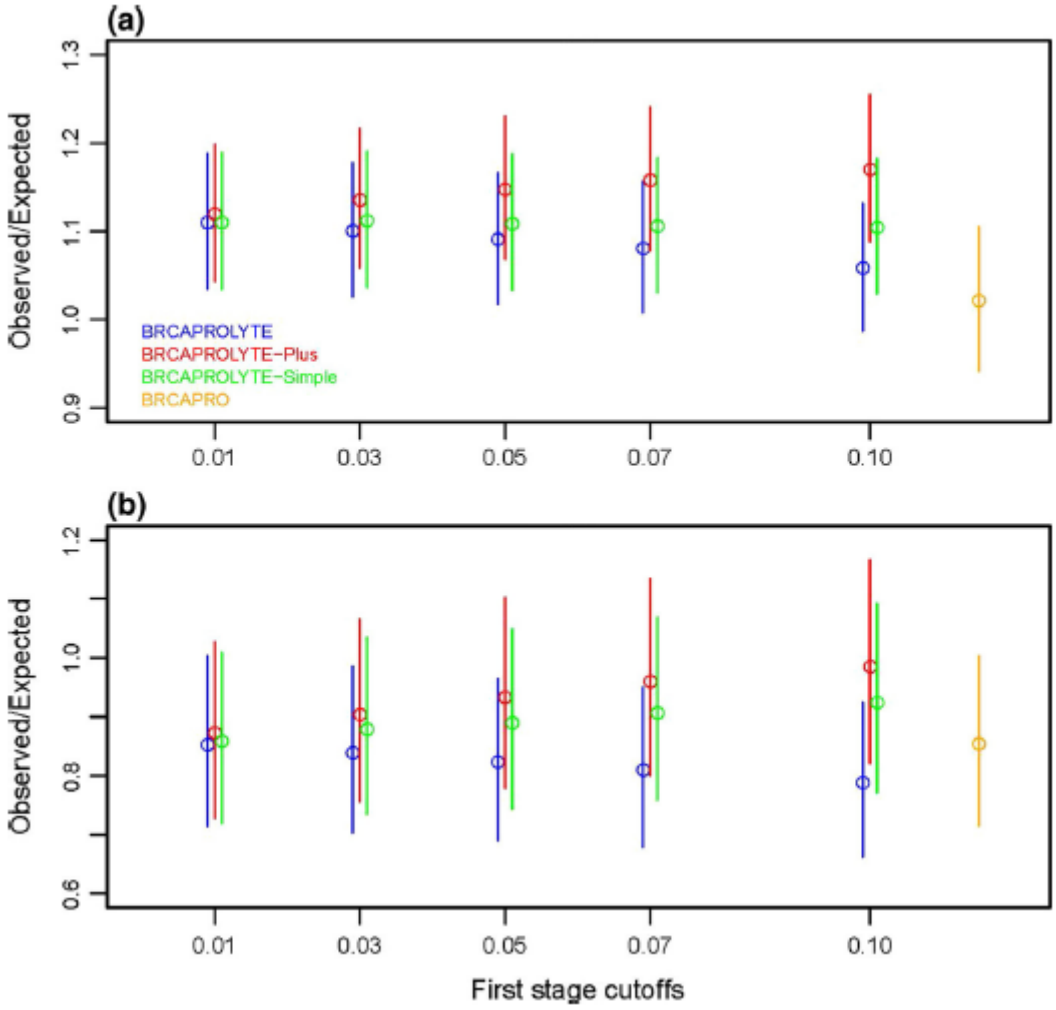
The *x-axis* has two sets of cutoffs, c_1 (first stage) followed by c_2 (second stage) below it. For BRCAPRO, only one cutoff (c_2) is applicable

Table 4.2 AUC (with CI) of the two-stage approach and BRCAPRO. For AUC.p, the first-stage cutoff corresponding to p (percentage followed-up in the second stage) is indicated as c_1

Dataset	BRCAPROLYTE	BRCAPROLYTE-Plus	BRCAPROLYTE-Simple	BRCAPRO
AUC.O				
CGN+MDA	0.79 (0.76, 0.81)	0.78 (0.76, 0.81)	0.79 (0.76, 0.81)	0.78 (0.76, 0.81)
NWH	0.66 (0.61 , 0.71)	0.65 (0.6, 0.71)	0.66 (0.61, 0.71)	0.65 (0.59, 0.70)
AUC.75 (cutoff)				
CGN+MDA	0.78 (0.76, 0.8), c_1=0.049	0.78 (0.76, 0.8), c_1=0.009	0.78 (0.76, 0.8), c_1=0.018	
NWH	0.65 (0.59, 0.7), c_1=0.027	0.65 (0.6, 0.7), c_1=0.005	0.65 (0.6, 0.7), c_1=0.009	
AUC.50 (cutoff)				
CGN+MDA	0.78 (0.75, 0.8), c_1=0.169	0.78 (0.76, 0.8), c_1=0.039	0.78 (0.76, 0.8), c_1=0.069	
NWH	0.65 (0.6, 0.71), c_1=0.08	0.64 (0.59, 0.7), c_1=0.018	0.65 (0.59, 0.7), c_1=0.028	
AUC.25 (cutoff)				
CGN+MDA	0.77 (0.74, 0.79), c_1=0.479	0.77 (0.75, 0.8), c_1=0.185	0.78 (0.75, 0.8), c_1=0.274	
NWH	0.63 (0.58, 0.69), c_1=0.217	0.64 (0.58, 0.69), c_1=0.068	0.64 (0.58, 0.7), c_1=0.092	

Next, we plot O/E values in Figure 4.4. In general, BRCAPROLYTE tends to overestimate the risk of mutation, BRCAPROLYTE-Plus tends to underestimate, and BRCAPROLYTE-Simple remains in between and is the most stable across varying cutoffs. When compared with BRCAPRO, the two datasets show somewhat different trends. In CGN+MDA, the two-stage approaches have somewhat worse calibration than BRCAPRO with BRCAPROLYTE performing the best (remains closest to 1), while in NWH, BRCAPROLYTE-Plus and BRCAPROLYTE-Simple have slightly better calibration than BRCAPRO. The CIs overlap substantially and so calibration of the two-stage approaches

Figure 4.4 Ratio of observed number of carriers to the expected number of carriers as predicted by the two-stage approach and BRCAPRO for **a** CGN+MDA and **b** NWH data



may not be very different from BRCAPRO, although there seems to be some differences between high-risk and community practice.

In summary, the two-stage approach has similar discrimination and calibration, and can achieve a similar clinical impact without requiring a full evaluation in primary care.

4.4. DISCUSSION

We proposed and evaluated two-stage approaches for genetic risk assessment. The first stage can be implemented in a primary care setting using limited family history information, and can be efficiently integrated into an EMR or other HIT solutions. Patients with sufficiently high risk can then be assessed in further detail, typically, though not necessarily, within a genetic counseling clinic. We showed that the overall performance of the two-stage approach is comparable to using the more complete assessment on all patients. Even though it is more complex than performing a single full evaluation of all patients, the clinical importance of this result lies in the fact that the latter is not currently scalable to primary care delivery. Moreover, testing everyone is currently not a practical option in the U.S. due to financial and other considerations. The two-stage approach makes it possible to screen the general population for risk of carrying BRCA mutations. It not only entails an increase in the burden of data collection in primary care, and a duplicate assessment on a relatively large subset of families, but also results in a reduction in genetic counseling activities, the most challenging and less easily scalable stage.

A practical issue is the choice of cutoffs for use in clinical settings. We illustrated the clinical implications using specific combinations of first- and second-stage cutoffs to quantify the genetic counseling and testing burden associated with using different first-stage tools in different populations. For practical applications, different clinical scenarios may require a different balance of specificity, sensitivity, and burden of data collection, and these considerations should guide the choice of appropriate combinations.

One of the strengths of our study is the use of high-risk as well as population-based data. Although specific assessments of discrimination and calibration for the two datasets were different, both analyses gave the same conclusion about the viability of the two-stage approach. Thus, our overall conclusions are likely to be applicable to the more general population, at least qualitatively. However, due to limitations of the data (as discussed below), these results may not be fully representative of the use of the two-stage approach in an unselected population.

At present, the paucity of medical environments where the potential for genetic testing is routinely incorporated into primary care workflows makes it challenging to carry out a population-based evaluation of the two-stage approach. Our analysis of the NWH data comes close, but still has limitations. The main limitation is that the NWH cohort is enriched for patients with BRCAPROLYTE probability exceeding 10 %, which makes it more difficult to generalize our conclusions on the operating characteristics of the two-stage approach involving BRCAPROLYTE as the first step. Generally, our analyses are likely to overestimate the number of high-risk families found in the first stage, when compared to an application of a two-stage approach to a completely unselected population.

The discrimination of the BRCAPRO model in the NWH cohort, as measured by AUC, is lower than reported in other datasets (Biswas et al., 2013; Parmigiani et al., 2007). This indicates that the carrier probability distributions for BRCA carriers and non-carriers are not well separated. For example, the median probabilities for carriers and non-carriers are 0.07 and 0.03, respectively, compared to 0.34 and 0.03 for CGN+MDA data. If we take each proband in CGN+MDA and find a matching proband

(with the closest probability) from NWH, the BRCAPRO AUC for this subset of NWH probands is 0.72, compared to 0.65 observed in the whole NWH data. Also, 175 probands in NWH have genetic test results available for at least one relative. Including their test results in BRCAPRO calculations increases the AUC from 0.65 to 0.79 showing the strong impact of genetic test results. However, if information on relatives' test results is not used (as it will be burdensome to collect this information in the primary care), the results of the two-stage approaches for this subset are similar to what we found for the whole NWH data.

Of the three two-stage approaches, BRCAPROLYTE tends to over-predict. BRCAPROLYTE-Plus gives less inflated estimates but appears to be affected by under-prediction of carrier probabilities. BRCAPROLYTE-Simple seems to provide a better balance between over- and under-prediction. This is somewhat unexpected as BRCAPROLYTE-Simple uses less information on family structure than BRCAPROLYTE-Plus. From a practical point of view, this is a useful result as the burden of data collection in the first stage is the least through use of BRCAPROLYTE-Simple. To provide a rough estimate of savings in time, consider that it takes about 10–30 minutes to collect the vital status and age of every relative. Compared to that, it may take only about 3 minutes to obtain the age and relationships of each person with cancer, which is sufficient for applying BRCAPROLYTE-Simple. Future work can focus on estimation of the burden of data collection as well as the acceptance rate of the simplified tools in clinical practice.

In summary, our work proposes a new paradigm of formally integrating genetic risk assessment in primary care. By implementing the process of risk assessment in

two stages, the proposed approach strikes a balance between two competing issues – identifying potential carriers among large populations who are currently not receiving adequate risk counseling and ensuring that the burden of exhaustive genetic counseling (second stage) remains manageable. By adjusting the cutoffs for the two stages, this approach allows identification of as many carriers as are practically possible. Although we focused on breast and ovarian cancer risk here, the approach is general and can be used for risk prediction for other cancers, for which well-established genetic models exist such as pancreas, colon, and melanoma (Chen et al., 2006; Wang et al., 2007; Wang et al., 2010).

CHAPTER 5

CONCLUSION AND PUBLIC HEALTH IMPLICATIONS

A woman who has a family history of breast and ovarian cancers may consider genetic testing. BRCAPRO is one of the most widely used statistical models for genetic risk prediction for breast cancer, and provides probabilities of being a carrier of the breast cancer susceptibility genes BRCA1 or BRCA2. Despite its widespread use in genetic counseling, a limitation is that it requires extensive information on the proband and her family history. This poses practical challenge for its use in many primary care clinics. As a further enhancement to the BRCAPRO model, we proposed the two-stage approach in order to bring its utility available to a more general population. In a primary care setting, simplified versions of BRCAPRO will be used as the first stage in estimating the risks of being a BRCA1 or BRCA2 mutation carrier using limited family history information. For those who are identified to be at potentially high risk, the full BRCAPRO model will be used in the second stage, typically in genetic counseling, to provide more accurate risk estimates.

BRCAPROLYTE relies only on first- and second-degree relatives who are affected with breast and/or ovarian cancer. BRCAPROLYTE-Plus additionally imputes the ages of the relatives who are unaffected with breast or ovarian cancer.

BRCAPROLYTE-Simple is a more simplified model which imputes the relatives who are unaffected, in addition to their ages. Overall, we have shown that among the five simpler tools considered at the first stage, BRCAPROLYTE, BRCAPROLYTE-Plus, and BRCAPROLYTE-Simple perform better and lead to only modest loss in accuracy compared to the results when the full BRCAPRO model is applied to all probands. The AUCs for BRCAPRO are 0.783 for CGN+MDA and 0.648 for NWH. At the first stage, the AUCs for CGN+MDA for BRCAPROLYTE, BRCAPROLYTE-Plus, and BRCAPROLYTE-Simple are 0.763, 0.772, and 0.773, while for NWH, the corresponding AUCs are 0.629, 0.644, and 0.641.

By identifying accurately women who are unlikely to have mutations, the developed approach has the potential to result in substantial savings in time, resources, and money. At the same time, more high risk patients can be identified for genetic testing who can benefit in time from preventative medicine. Knowledge of patient's increased susceptibility to breast cancer can help in medical management. These risk prediction models are used by insurance companies to determine whether the costs for genetic tests, which range from about $300 to $5,000 (BreastCancer.org, 2016), would be covered. Thus, these models crucially impact the degree, quality, and cost of care of millions of women in the U.S.

Different individuals will have different attitudes about testing, thus, a single probability cutoff so that testing is appropriate above and not appropriate below that cutoff cannot be recommended. The chance of carrying a mutation varies from person to person, thus, understanding and conveying risk information should be individualized. Great care is needed when assessing patient suitability for genetic testing of the BRCA

mutation. As well as a financial cost, there are legal, health, and ethical issues to be considered, stressing the importance of the accuracy of risk assessments. The possibility of an adverse psychological effect of receiving BRCA1/2 mutation test results must also be considered before undergoing genetic testing (Palma et al., 2006). In Schwartz et al.'s study (2002), participants who received negative test results showed significant decrease in perceived risk and distress compared with those who received positive test results. There are situations where individuals were perceived to have experienced "survivor's guilt" when test results showed they did not have the genetic mutations (Dudok deWit et al., 1998; Lodder, L.N., 2001; Wagner et al., 2000). In Croyle et al. (1997) and Dorval et al. (2000), carriers were found to have significantly higher levels of test-related psychological distress than non-carriers. In Hamilton, Lobel, and Moyer's (2009) meta-analysis, results indicate that BRCA1/2 testing have emotional consequences, with the most unfavorable outcome among carriers, but the distress levels appear to be minimal and returned to pre-test levels over time. All things considered, it is important to make personalized recommendations for cancer screening and prevention.

5.1. PUBLIC HEALTH IMPLICATIONS

Before actual genetic testing is conducted, statistical models are used to determine the probability of the patient being a mutation carrier through genetic counseling based on family history. If a patient has been identified as being high-risk, the psychosocial issues attached to the awareness of the mutation status such as the effect on the patient's relationship with the rest of the family, the effects on body image

and personal identity, cancer-related stress and worry, feelings of guilt for passing on the genetic risk to the offspring, etc. (NCI, 2015) can be tackled appropriately. The information gained from this research can be used for surveillance and management of women at a higher risk of an inherited predisposition to breast cancer. Women deemed to be at high risk can be tested to see whether they are carriers of mutations of BRCA1 and BRCA2 genes. Men also carry mutations of these genes and in rare cases may develop breast cancer, which is a strong indicator of BRCA2 mutation in the family (Komen, 2016). If genetic tests show presence of a BRCA1 or BRCA2 mutation, necessary preventative measures (such as oophorectomy, mastectomy, chemoprevention, etc.) can be taken. Alternatively, management strategies or frequent monitoring through mammograms, magnetic resonance imaging (MRI), ultrasounds, and other tests can be performed (NCI, 2015).

5.2. FUTURE WORK

As discussed in Chapter 4, one of the limitations of our study is that the NWH data used may not be entirely representative of the general population. Future work may involve validation of our proposed two-stage approach using a completely unselected population; i.e., patients will be selected randomly from those who go to primary care or breast clinics without consideration of their BRCA mutation risk probabilities. It is also of interest to investigate if the burden of data collection is indeed lessened, whether our first stage tools will be more accepted in primary care settings where it is intended, and how much that helps in identification of potential mutation carriers.

In addition, some known risk factors are not incorporated into the full BRCAPRO model. Body mass index, alcohol intake, age at menarche, age at first live birth, age at menopause, hormone replacement therapy use, and prostate and colon cancer in relatives (Amir et al., 2010; Gilpin, Carson, and Hunter, 2000) can be included to make the model more personalized and accurate. The full BRCAPRO model can also be enhanced to allow the inclusion of genes other than BRCA1/BRCA2 that are associated with breast/ovarian cancer cases found from genome-wide association studies of several single-nucleotide polymorphisms (SNPs) (Fanale et al., 2012; Foulkes, 2008). Among these SNPs, the mutations in serine-threonine protein kinase 11 (STK11), phosphatase and tensin homologue (PTEN), E-cadherin (CDH1), tumor protein 53 (TP53), BRCA1-interacting protein C-terminal helicase 1 (BRIP1, aka BACH1), partner and localizer of BRCA2 (PALB2), cell-cycle-checkpoint kinase (CHEK2), ataxia-telangiectasia mutated (ATM), fibroblast growth factor receptor 2 (FGFR2), tox high-mobility group box family member 3 (TOX3, aka TNRC9), mitogen-activated protein kinase kinase kinase 1 (MAP3K1, aka MEKK1), lymphocyte-specific protein 1 (LSP1, aka WP43), and caspase 8 (CASP8) genes have been identified to have different effect levels on breast cancer (Foulkes, 2008). Systematically incorporating the high-penetrance genes or genes that have been identified with high degree of certainty as risk factors for breast and ovarian cancers into the BRCAPRO model may help make risk predictions more accurate.

APPENDIX 1

SUPPLEMENTARY MATERIALS FOR CHAPTER 4

Supplementary Methods

Here we describe the methods for obtaining the overall sensitivity, specificity, AUC, PVP, and PVN of the two-stage approach. First, we obtain the overall sensitivity and specificity as follows. Denote the cutoffs used at the first and second stages as c_1 and c_2, and P(AnyBRCA) at these two stages as Pr_1 and Pr_2, respectively. Consider a specific first-stage tool and denote its sensitivity and specificity at c_1 as $Se.1 = P(Pr_1 > c_1|$ positive genetic test) and $Sp.1 = P(Pr_1 < c_1|$ negative genetic test). Further, denote the sensitivity and specificity of the second stage at c_2 given the results of the first stage as Se.2|1 and Sp.2|1, respectively. These can be estimated by evaluating the subsample of patients with $Pr_1 > c_1$ who undergo the second stage. A counselee will be considered at high risk overall, and referred for genetic testing, only if $Pr_1 > c_1$ and $Pr_2 > c_2$. Thus, the sensitivity Se.O and the specificity Sp.O of the overall two-stage procedure can be calculated as

$Se.O = P(Pr_2 > c_2|Pr_1 > c_1,$ positive genetic test) $*$ $P(Pr_1 > c_1|$ positive genetic test) = $Se.2|1 * Se.1$, and

$Sp.O = P(Pr_1 < c_1|$ negative genetic test) + $P(Pr_2 < c_2|Pr_1 > c_1,$ negative genetic test) $*$ $P(Pr_1 > c_1|$ negative genetic test) = $Sp.1 + Sp.2|1 * (1 - Sp.1)$.

By varying the values of c_1 and c_2, we can get a range of sensitivities and specificities for the two-stage approach. Next, we plot the Se.O versus 1-Sp.O over the range of c1 and c2 to obtain an empirical ROC curve, and estimate the AUC using trapezoidal rule. Also, we calculate the overall predictive values, PVP.O and PVN.O from Se.O and Sp.O values by using the Bayes rule. We obtain 95% CI for overall sensitivity, specificity, and predictive values using the bootstrap method (Efron and Tibshirani, 1994). For AUC, we use an asymptotic CI (Hanley and McNeil, 1982; Pepe, 2004). Details are provided in Appendix 3.

Table A.1 First Stage Results (with CI) for NWH data

	Cutoff	BRCAPRO	BRCAPROLYTE	BRCAPROLYTE-Plus	BRCAPRO-Simple
Sensitivity	0.001	0.97 (0.93, 0.99)	0.99 (0.97, 1)	0.96 (0.92, 0.99)	0.97 (0.93, 0.99)
Specificity		0.07 (0.05, 0.08)	0 (0, 0.01)	0.09 (0.07, 0.1)	0.06 (0.04, 0.07)
Sensitivity	0.003	0.92 (0.87, 0.96)	0.97 (0.93, 0.99)	0.91 (0.86, 0.96)	0.93 (0.88, 0.97)
Specificity		0.17 (0.15, 0.19)	0.05 (0.04, 0.07)	0.18 (0.15, 0.2)	0.1 (0.08, 0.12)
Sensitivity	0.005	0.89 (0.83, 0.94)	0.97 (0.93, 0.99)	0.86 (0.79, 0.92)	0.9 (0.85, 0.95)
Specificity		0.22 (0.2, 0.24)	0.07 (0.05, 0.08)	0.26 (0.24, 0.28)	0.14 (0.13, 0.16)
Sensitivity	0.007	0.84 (0.77, 0.9)	0.96 (0.92, 0.99)	0.8 (0.73, 0.87)	0.89 (0.83, 0.94)
Specificity		0.27 (0.25, 0.3)	0.08 (0.06, 0.09)	0.33 (0.3, 0.36)	0.21 (0.19, 0.23)
Sensitivity	0.01	0.8 (0.73, 0.87)	0.93 (0.88, 0.97)	0.76 (0.68, 0.83)	0.83 (0.76, 0.89)
Specificity		0.33 (0.3, 0.35)	0.1 (0.08, 0.12)	0.4 (0.37, 0.42)	0.29 (0.27, 0.32)
Sensitivity	0.03	0.63 (0.55, 0.71)	0.84 (0.77, 0.9)	0.54 (0.46, 0.63)	0.6 (0.51, 0.69)
Specificity		0.54 (0.51, 0.57)	0.29 (0.27, 0.32)	0.61 (0.58, 0.64)	0.53 (0.5, 0.55)
Sensitivity	0.05	0.54 (0.45, 0.62)	0.7 (0.61, 0.78)	0.49 (0.4, 0.58)	0.54 (0.45, 0.62)
Specificity		0.64 (0.61, 0.66)	0.41 (0.39, 0.44)	0.71 (0.68, 0.73)	0.64 (0.61, 0.67)
Sensitivity	0.1	0.46 (0.37, 0.54)	0.57 (0.48, 0.66)	0.39 (0.31, 0.48)	0.43 (0.34, 0.52)
Specificity		0.75 (0.72, 0.77)	0.56 (0.53, 0.59)	0.83 (0.81, 0.85)	0.79 (0.77, 0.81)
Sensitivity	0.2	0.39 (0.31, 0.48)	0.46 (0.37, 0.54)	0.32 (0.24, 0.4)	0.34 (0.25, 0.42)
Specificity		0.86 (0.84, 0.88)	0.74 (0.72, 0.77)	0.9 (0.88, 0.92)	0.88 (0.86, 0.9)
PVP	0.001	0.1 (0.08, 0.11)	0.09 (0.08, 0.11)	0.1 (0.08, 0.11)	0.1 (0.08, 0.11)
PVN		0.95 (0.9, 0.99)	0.83 (0.5, 1)	0.95 (0.91, 0.99)	0.95 (0.89, 0.99)
PVP	0.003	0.1 (0.08, 0.12)	0.09 (0.08, 0.11)	0.1 (0.08, 0.12)	0.1 (0.08, 0.11)
PVN		0.95 (0.93, 0.98)	0.94 (0.88, 0.99)	0.95 (0.92, 0.98)	0.93 (0.88, 0.97)
PVP	0.005	0.1 (0.09, 0.12)	0.1 (0.08, 0.11)	0.11 (0.09, 0.13)	0.1 (0.08, 0.12)
PVN		0.95 (0.92, 0.97)	0.95 (0.9, 0.99)	0.95 (0.92, 0.97)	0.94 (0.9, 0.97)
PVP	0.007	0.11 (0.09, 0.13)	0.1 (0.08, 0.11)	0.11 (0.09, 0.13)	0.1 (0.09, 0.12)
PVN		0.94 (0.92, 0.97)	0.95 (0.9, 0.99)	0.94 (0.92, 0.96)	0.95 (0.92, 0.97)
PVP	0.01	0.11 (0.09, 0.13)	0.1 (0.08, 0.11)	0.11 (0.09, 0.14)	0.11 (0.09, 0.13)

PVN		0.94 (0.92, 0.96)	0.93 (0.89, 0.97)	0.94 (0.92, 0.96)	0.94 (0.92, 0.97)
PVP	0.03	0.12 (0.1, 0.15)	0.11 (0.09, 0.13)	0.13 (0.1, 0.15)	0.12 (0.09, 0.14)
PVN		0.93 (0.92, 0.95)	0.95 (0.92, 0.97)	0.93 (0.91, 0.95)	0.93 (0.91, 0.95)
PVP	0.05	0.13 (0.1, 0.16)	0.11 (0.09, 0.13)	0.15 (0.11, 0.18)	0.13 (0.1, 0.16)
PVN		0.93 (0.91, 0.95)	0.93 (0.91, 0.95)	0.93 (0.91, 0.95)	0.93 (0.91, 0.95)
PVP	0.1	0.16 (0.12, 0.2)	0.12 (0.09, 0.14)	0.19 (0.14, 0.24)	0.17 (0.13, 0.22)
PVN		0.93 (0.91, 0.95)	0.93 (0.91, 0.94)	0.93 (0.91, 0.94)	0.93 (0.92, 0.95)
PVP	0.2	0.22 (0.17, 0.28)	0.15 (0.12, 0.19)	0.25 (0.18, 0.32)	0.22 (0.16, 0.28)
PVN		0.93 (0.92, 0.95)	0.93 (0.91, 0.95)	0.93 (0.91, 0.94)	0.93 (0.91, 0.94)
AUC		0.65 (0.59, 0.70)	0.63 (0.57, 0.68)	0.64 (0.59, 0.7)	0.64 (0.58, 0.7)
O/E		125/146.45 = 0.85 (0.71, 1)	125/228.94 = 0.55 (0.46, 0.64)	125/113.98 = 1.10 (0.91, 1.3)	125/135.37 = 0.92 (0.77, 1.09)

Table A.2 Numbers of referrals made at each stage using a two-stage approach, as compared to using BRCAPRO only on all probands for CGN+MDA data. For each combination of c_1 and c_2, three numbers are provided – number of probands with first stage probability exceeding c_1 (n_1), out of n_1, the number of probands with second stage probability exceeding c_2 (n_2), and out of n_2, the number of probands tested positive for BRCA mutation

BRCAPROLYTE

c_1 (%) \ c_2 (%)	1	3	5	10
1	2584	2584	2584	2584
	2070	1624	1361	1036
	548	498	458	414
3	2255	2255	2255	2255
	2012	1620	1359	1035
	541	496	456	413
5	2028	2028	2028	2028
	1881	1576	1350	1031
	526	494	456	413
10	1632	1632	1632	1632
	1559	1410	1276	1023
	490	472	447	412

BRCAPROLYTE-Plus

c_1 (%) \ c_2 (%)	1	3	5	10
1	2003	2003	2003	2003
	1909	1582	1340	1029
	532	494	456	413
3	1477	1477	1477	1477
	1470	1429	1280	1006
	471	468	444	409
5	1244	1244	1244	1244
	1240	1231	1191	980
	438	438	429	401
10	903	903	903	903
	902	902	898	867
	379	379	378	372

BRCAPROLYTE-Simple

c_1 (%) \ c_2 (%)	1	3	5	10
1	2300	2300	2300	2300
	2044	1619	1359	1036
	546	497	457	414
3	1789	1789	1789	1789
	1725	1536	1330	1024
	509	483	454	412

BRCAPRO

c_2 (%)			
1	**3**	**5**	**10**
2070	1624	1361	1036
548	498	458	414

5	1522	1522	1522	1522
	1494	1415	1299	1018
	484	472	450	412
10	1167	1167	1167	1167
	1160	1141	1100	978
	426	423	415	399

Table A.3 Numbers of referrals made at each stage using a two-stage approach, as compared to using BRCAPRO only on all probands for NWH data. For each combination of c_1 and c_2, three numbers are provided – number of probands with first stage probability exceeding c_1 (n_1), out of n_1, the number of probands with second stage probability exceeding c_2 (n_2), and out of n_2, the number of probands tested positive for BRCA mutation

BRCAPROLYTE

c_1 (%)	c_2 (%) 1	3	5	10
1	1212	1212	1212	1212
	921	641	509	363
	100	79	67	57
3	969	969	969	969
	836	636	504	360
	98	79	67	57
5	803	803	803	803
	729	592	500	357
	83	76	67	57
10	608	608	608	608
	585	513	448	347
	71	68	61	54

BRCAPROLYTE-Plus

c_1 (%)	c_2 (%) 1	3	5	10
1	831	831	831	831
	795	623	500	356
	91	77	66	56
3	540	540	540	540
	540	510	450	339
	68	67	61	53
5	416	416	416	416
	416	414	389	312
	61	61	58	51
10	256	256	256	256
	256	256	252	239
	49	49	48	47

BRCAPROLYTE-Simple

c_1 (%)	c_2 (%) 1	3	5	10
1	964	964	964	964
	849	635	506	362
	99	78	67	57
3	651	651	651	651
	634	542	470	349
	74	72	64	55

BRCAPRO

c_2 (%) 1	3	5	10
922	642	510	363
100	79	67	57

5	506	506	506	506
	502	468	417	326
	67	67	60	53
10	309	309	309	309
	307	303	290	253
	54	54	52	48

APPENDIX 2

THE BRCAPRO PROBABILITY MODEL

Denote R as the total number of relatives of the proband. The probability of the genotypes at BRCA1 and BRCA2 genes of the proband (γ_0), given the family history ($h_0, h_1, \ldots, h_R$), covariates, and pedigree structure ($X_0, X_1, \ldots, X_R$), is as follows:

$$P(\gamma_0 \mid h_0,h_1,\ldots,h_R;X_0,X_1,\ldots,X_R) = \frac{P(\gamma_0) * P(h_0,h_1,\ldots,h_R;X_0,X_1,\ldots,X_R \mid \gamma_0)}{\sum_\gamma P(\gamma) * P(h_0,h_1,\ldots,h_R;X_0,X_1,\ldots,X_R \mid \gamma)}$$

where γ is the set of all possible values of the genotypes of the individual family member. ($h_0, h_1, \ldots, h_R$) includes information on the relevant phenotypes and ages of onset of the proband and his/her relatives, if affected with cancer (or current age or age of death if unaffected). ($X_0, X_1, \ldots, X_R$) includes information on individual specific covariates such as being of AJ descent, each relatives' relationship to the proband, tumor marker information, and medical interventions. BRCA1 and BRCA2 mutations are assumed to be inherited independently of each other and all deleterious mutation variants are assumed to have the same phenotypic implications (Chen et al., 2004; Parmigiani, Berry, and Aguilar, 1998).

Using the law of total probability, the probability of the phenotypes for the entire family given the genotype of the proband is derived as:

$$P(h_0,h_1,...,h_R;X_0,X_1,...,X_R \mid \gamma_0) = \sum_{\gamma_1,...,\gamma_R}\left[\prod_{r=0}^{R} P(h_r;X_r \mid \gamma_r)\right] * P(\gamma_1,...,\gamma_R \mid \gamma_0)$$

BayesMendel uses the Elston-Stewart Algorithm (Elston and Stewart, 1971) to compute the above pedigree likelihood. Note that it is assumed that the histories of each family member (phenotypes) are conditionally independent given their genotypes. Also, the last term in the equation above, $P(\gamma_1,...,\gamma_R \mid \gamma_0)$, can be computed for all genotype configurations using Mendel's laws as long as the mode of inheritance is known (e.g., autosomal dominant for breast cancer) (Parmigiani, Berry, and Aguilar, 1998).

The overall prevalence of the mutation of each of these genes ($P(\boldsymbol{\gamma}_r)$) and the age-specific penetrance of breast and ovarian cancers resulting from carrying mutations ($P(h_r; X_r \mid \boldsymbol{\gamma}_r)$) used in BRCAPRO were based on a nine study meta-analysis (Chen et al., 2006), the CGN (Tai et al., 2007), Graeser et al. (2009), and Katki et al. (2008).

APPENDIX 3

AUC CONFIDENCE INTERVAL ESTIMATION

Bootstrap sampling to get the 95% CIs for AUC for the overall two-stage approach can be computationally intensive. Thus, as noted in Appendix 1, we calculated the asymptotic CIs for the overall measures of diagnostic accuracy. Let n_D = # subjects with positive test results, $n_{\bar{D}}$ = # subjects with negative test results, c_1 = cutoff at the first stage, and c_2 = cutoff at the second stage. For probands with positive test results, define Pr_{D_i} according to the following (reflecting the last carrier probability evaluation for each proband):

$$Pr_{D_i} = \begin{cases} Pr_1, & Pr_1 \leq c_1 \\ Pr_2, & Pr_1 > c_1 \end{cases}.$$

For probands with negative test results, define $Pr_{\bar{D}_j}$ similarly. The area under the empirical ROC curve is equivalent to the Mann-Whitney U-statistic as follows (Hanley and McNeil, 1982; Pepe, 2004):

$$\text{AUC} = \frac{\sum_{j=1}^{n_{\bar{D}}} \sum_{i=1}^{n_D} \left\{ I\left[Pr_{D_i} > Pr_{\bar{D}_j}\right] + \frac{1}{2} I\left[Pr_{D_i} = Pr_{\bar{D}_j}\right] \right\}}{n_D n_{\bar{D}}}.$$

Using large sample approximation, the variance of the AUC can be calculated as:

$$\text{var(AUC)} = \frac{\text{AUC(1-AUC)} + (n_D\text{-}1) * (Q_1\text{-AUC}^2) + (n_{\bar{D}}\text{-}1) * (Q_2\text{-AUC}^2)}{n_D n_{\bar{D}}},$$

where Q_1 = P($Pr_{D_i} \geq Pr_{\bar{D}_j}, Pr_{D_i'} \geq Pr_{\bar{D}_j}$),

Q_2 = P($Pr_{D_i} \geq Pr_{\bar{D}_j}, Pr_{D_i} \geq Pr_{\bar{D}_j'}$),

($Pr_{D_i}, Pr_{D_i'}$) = random pairs of observations from those with positive test results,

and ($Pr_{\bar{D}_j}, Pr_{\bar{D}_j'}$) = random pairs of observations from those with negative test results.

Therefore, the 95% asymptotic CI for the AUC can be calculated as $\text{AUC} \pm 1.96\sqrt{\text{var(AUC)}}$ (Pepe, 2004).

REFERENCES

American Cancer Society. (2016). Cancer facts and figures 2016. Atlanta: American Cancer Society. Retrieved from http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf.

American Society of Clinical Oncology. (1996). Statement of the American Society of Clinical Oncology: genetic testing for cancer susceptibility. *J Clin Oncol, 14*, 1730-1736.

Amir, E., Freedman, O.C., Seruga, B., & Evans, D.G. (2010). Assessing women at high risk of breast cancer: a review of risk assessment models. *J Natl Cancer Inst, 102*(10), 680-691.

Antoniou, A.C., Cunningham, A.P., Peto, J., Evans, D.G., Lalloo, F., Narod, S.A., Risch, H.A., Eyfjord, J.E., Hopper, J.L., Southey, M.C., Olsson, H., Johannsson, O., Borg, A., Passini, B., Radice, P., Manoukian, S., Eccles, D.M., Tang, N., Olah, E., Anton-Culver, H., Warner, E., Lubinski, J., Gronwald, J., Gorski, B., Tryggvadottir, L., Syrjakoski, K., Kallioniemi, O.-P., Eerola, H., Nevanlinna, H., Pharoah, P.D.P., & Easton, D.F. (2008). The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *British Journal of Cancer, 98*(8), 1457-1466.

Antoniou, A.C., Pharoah, P.D.P., McMullan, G., Day, N.E., Stratton, M.R., Peto, J., Ponder, B.J., & Easton, D.F. (2002). A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *British Journal of Cancer, 86*(1), 76-83.

Antoniou, A., Pharoah, P.D.P., Narod, S., Risch, H.A., Eyfjord, J.E., Hopper, J.L., Loman, N., Olsson, H., Johannsson, O., Borg, A., Pasini, B., Radice, P., Manoukian, S., Eccles, D.M., Tang, N., Olah, E., Anton-Culver, H., Warner, E., Lubinski, J., Gronwald, J., Gorski, B., Tulinius, H., Thorlacius, S., Eerola, H., Nevanlinna, H., Syrjakoski, K., Kallioniemi, O.-P., Thompson, D., Evans, C., Peto, J., Lalloo, F., Evans, D.G., & Easton, D.F. (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet., 72*, 1117-1130.

Apicella, C., Andrews, L., Hodgson, S.V., Fisher, S.A., Lewis, C.M., Solomon, E., Tucker, K., Friedlander, M., Bankier, A., Southey, M.C., Venter, D.J., & Hopper, J.L . (2003). Log odds of carrying an ancestral mutation in BRCA1 or BRCA2 for a defined personal and family history in an Ashkenazi Jewish woman (LAMBDA). *Breast Cancer Res, 5*(6), R206-R216.

Barcenas, C.H., Monawar Hosain, G.M., Arun, B., Zong, J., Zhou, X., Chen, J., Cortada, J.M., Mills, G.B., Tomlinson, G.E., Miller, A.R., Strong, L.C., & Amos, C.I. (2006). Assessing BRCA carrier probabilities in extended families. *J Clin Oncol, 24*(3), 354-360.

BayesMendel Lab. (2015). The BayesMendel R package archive. Retrieved from http://bcb.dfci.harvard.edu/BayesMendel/Rpackage.php.

BayesMendel. (2016). BRCAPRO Web Service Client. Retrieved from http://bayesmendel.dfci.harvard.edu/risk/.

Berry, D.A., Iversen, Jr., E.S., Gudbjartsson, D.F., Hiller, E.H., Garber, J.E., Peshkin, B.N., Lerman, C., Watson, P., Lynch, H.T., Hilsenbeck, S.G., Rubinstein, W.S., Hughes, K.S., & Parmigiani, G. (2002). BRCAPRO validation, sensitivity of genetic testing of BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes. *J Clin Oncol, 20*(11), 2701-2712.

Berry, D.A., Parmigiani, G., Sanchez, J., Schildkraut, J., & Winer, E. (1997). Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. *J Natl Cancer Inst, 89*(3), 227-237.

Biswas, S., Atienza, P., Chipman, J., Blackford, A.L., Arun, B., Hughes, K., & Parmigiani, G. (2016). A two-stage approach to genetic risk assessment in primary care. *Breast Cancer Res Treat, 155*(2), 375-383.

Biswas, S., Atienza, P., Chipman, J., Hughes, K., Gutierrez Barrera, A.M., Amos, C.I., Arun, B., & Parmigiani, G. (2013). Simplifying clinical use of the genetic risk prediction model BRCAPRO. *Breast Cancer Res Treat, 139*(2), 571-579.

Biswas, S., Tankhiwale, N., Blackford, A., Gutierrez Barrera, A.M., Ready, K., Lu, K., Amos, C.I., Parmigiani, G., & Arun, B. (2012). Assessing the added value of breast tumor markers in genetic risk prediction model BRCAPRO. *Breast Cancer Res Treat, 133*(1), 347-355.

BreastCancer.org. (2016). Genetic testing facilities and cost. Retrieved from http://www.breastcancer.org/symptoms/testing/genetic/facility_cost.

BreastCancer.org. (2016). Genetics. Retrieved from http://www.breastcancer.org/risk/factors/genetics.

BreastCancer.org. (2015). Risk of breast, ovarian cancer may vary depending on type and location of BRCA mutation. Retrieved from http://www.breastcancer.org/research-news/cancer-risk-may-vary-by-mutation-type-location.

Campeau, P.M., Foulkes, W.D., and Tischkowitz, M.D. (2008). Hereditary breast cancer: new genetic developments, new therapeutic avenues. *Human Genetics, 124*(1), 31-42.

CancerGene. (2016). UTSW CancerGene Software. Retrieved from https://www4.utsouthwestern.edu/breasthealth/cagene/.

Centers for Disease Control and Prevention. (2014). Breast and ovarian cancer and family health history. Retrieved from http://www.cdc.gov/genomics/resources/diseases/breast_ovarian_cancer.htm.

Centers for Disease Control and Prevention. (2016). Breast cancer. Retrieved from http://www.cdc.gov/cancer/breast/index.htm.

Centers for Disease Control and Prevention. (2016). Does breast or ovarian cancer run in your family? Retrieved from http://www.cdc.gov/features/hereditarycancer/.

Centers for Disease Control and Prevention. (2015). Hereditary breast cancer and BRCA genes. Retrieved from http://www.cdc.gov/cancer/breast/young_women/bringyourbrave/hereditary_breast_cancer/.

Centers for Disease Control and Prevention. (2016). What are the risk factors for breast cancer? Retrieved from http://www.cdc.gov/cancer/breast/basic_info/risk_factors.htm.

Chen, S., Blackford, A.L., and Parmigiani, G. (2009). Tailoring BRCAPRO to Asian-Americans. *J Clin Oncol, 27*(4), 642-643.

Chen, S., Iversen, E.S., Friebel, T., Finkelstein, D., Weber, B.L., Eisen, A., Peterson, L.E., Schildkraut, J.M., Isaacs, C., Peshkin, B.N., Corio, C., Leondaridis, L., Tomlinson, G., Dutson, D., Kerber, R., Amos, C.I., Strong, L.C., Berry, D.A., Euhus, D.M., & Parmigiani, G. (2006). Characterization of BRCA1 and BRCA2 mutations in a large United States sample. *J Clin Oncol, 24*(6), 863-871.

Chen, S., Wang, W., Broman, K.W., Katki, H.A., & Parmigiani, G. (2004). BayesMendel: an R environment for Mendelian risk prediction. *Stat Appl in Genet and Mol Biol, 3*, 21.

Chen, S., Wang, W., Lee, S., Nafa, K., Lee, J., Romans, K., Watson, P., Gruber, S.B., Euhus, D., Kinzler, K.W., Jass, J., Gallinger, S., Lindor, N.M., Casey, G., Ellis, N.,

Giardiello, F.M., Offit, K., & Parmigiani, G. (2006). Prediction of germline mutations and cancer risk in the Lynch syndrome. *JAMA, 296*(12), 1479-1487.

Chen, S., Wang, W., Tai, Y.C., Katki, H.A., & Blackford, A. (2014). Carrier probabilities for breast cancer susceptibility genes BRCA1 and BRCA2. *R Documentation, brcapro*(BayesMendel version 2.0-9).

Couch, F.J., DeShano, M.L., Blackwood, M.A., Calzone, K., Stopfer, J., Campeau, L., Ganguly, A., Rebbeck, T., & Weber, B.L. (1997). BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer. *N Engl J Med, 336*(20), 1409-1415.

Croyle, R.T., Smith, K.R., Botkin, J.R., Baty, B., & Nash, J. (1997). Psychological responses to BRCA1 mutation testing: preliminary findings. *Health Psychology, 16*(1), 63-72.

Dorval, M., Patenaude, A.F., Schneider, K.A., Kieffer, S.A., DiGianni, L., Kalkbrenner, K.J., Bromberg, J.I., Basili, L.A., Calzone, K., Stopfer, J., Weber, B.L., & Garber, J.E. (2000). Anticipated versus actual emotional reactions to disclosure of results of genetic tests for cancer susceptibility: findings from p53 and BRCA1 testing programs. *J Clin Oncol, 18*(10), 2135-2142.

Drohan, B., Ozanne, E.M., and Hughes, K.S. (2009). Electronic health records and the management of women at high risk of hereditary breast and ovarian cancer. *The Breast Journal, 15*(1), 46-55.

Drohan, B., Roche, C.A., Cusack, Jr., J.C., & Hughes, K.S. (2012). Hereditary breast and ovarian cancer and other hereditary syndromes: using technology to identify carriers. *Ann Surg Oncol, 19*, 1732-1737.

Dudok deWit, A.C., Duivenvoorden, H.J., Passchier, J., Niermeijer, M.F., & Tibben, A. (1998). Course of distress experienced by persons at risk for an autosomal dominant inheritable disorder participating in a predictive testing program: an explorative study. *Psychosom Med., 60*(5), 543-549.

Easton, D.F., Bishop, D.T., Ford, D., Crockford, G.P., & the Breast Cancer Linkage Consortium. (1993). Genetic linkage analysis in familial breast and ovarian cancer. *Am. J. Hum. Genet., 52*, 678-701.

Easton, D.F., Ford, D. , Bishop, D.T., & the Breast Cancer Linkage Consortium. (1995). Breast and ovarian cancer incidence in BRCA1 mutation carriers. *Am. J. Hum. Genet., 56*, 265-271.

Efron, B., and Tibshirani. (1994). R: an introduction to the bootstrap. *Chapman and Hall/CRC*.

Elston, R.C., and Stewart, J. (1971). A general model for the genetic analysis of pedigree data. Human Heredity, 21(6), 523-542.

Euhus, D.M., Smith, K.C., Robinson, L., Stucky, A., Olopade, O.I., Cummings, S., Garber, J.E., Chittenden, A., Mills, G.B., Rieger, P., Esserman, L., Crawford, B., Hughes, K.S., Roche, C.A., Ganz, P.A., Seldon, J., Fabian, C.J., Klemp, J., & Tomlinson, G. (2002). Pretest prediction of BRCA1 or BRCA2 mutation by risk counselors and the computer model BRCAPRO. *J Natl Cancer Inst, 94*(11), 844-851.

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. *J Med Genet, 41*, 474-480.

Fanale, D., Amodeo, V., Corsini, L.R., Rizzo, S., Bazan, V., & Russo, A. (2012). Breast cancer genome-wide association studies: there is strength in numbers. *Oncogene, 31*, 2121-2128.

Feero, W.G., Bigley, M.B., Brinner, K.M., & Family Health History Multi-Stakeholder Workgroup of the American Health Information Community. (2008). New standards and enhanced utility for family health history information in the electronic health record. *J Am Med Inform Assoc, 15*(6), 723-728.

Fischer, C., Kuchenbacker, K., Engel, C., Zachariae, S., Rhiem, K., Meindl, A., Rahner, N., Dikow, N., Plendl, H., Debatin, I., Grimm, T., Gadzicki, D., Flottmann, R., Horvath, J., Schrock, E., Stock, F., Schafer, D., Schwaab, I., Kartsonaki, C., Mavaddat, N., Schlegelberger, B., Antoniou, A.C., & Schmutzler, R., on behalf of the German Consortium for Hereditary Breast and Ovarian Cancer. (2013). Evaluating the performance of the breast cancer genetic risk models BOADICEA, IBIS, BRCAPRO and Claus for predicting BRCA1/2 mutation carrier probabilities: a study based on 7352 families from the German Hereditary Breast and Ovarian Cancer Consortium. *J Med Genet, 50*, 360-367.

Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M.D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T.R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B.A., Cayther, S.A., Zelada-Helman, M. (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *Am. J. Hum. Genet., 62*, 676-689.

Foulkes, W.D. (2008). Inherited susceptibility to common cancers. *N Engl J Med, 359*, 2143-2153.

Frank, T.S., Deffenbaugh, A.M., Reid, J.E., Hulick, M., Ward, B.E., Lingenfelter, B., Gumpper, K.L., Scholl, T., 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.

Gail, M.H., and Mai, P.L. (2010). Comparing breast cancer risk assessment models. *J Natl Cancer Inst, 102*(10), 665-8.

Gilpin, C.A., Carson, N., and 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. *Clin Genet, 58*, 299-308.

Graeser, M.K., Engel, C., Rhiem, K., Gadzicki, D., Bick, U., Kast, K., Froster, U.G., Schlehe, B., Bechtold, A., Arnold, N., Preisler-Adams, S., Nestle-Kraemling, C., Zaino, M., Loeffler, M., Kiechle, M., Meindl, A., Varga, D., & Schmutzler, R.K. (2009). Contralateral breast cancer risk in BRCA1 and BRCA2 mutation carriers. *J Clin Oncol, 27*(35), 5887-5892.

Hamilton, J.G., Lobel, M., and Moyer, A. (2009). Emotional distress following genetic testing for hereditary breast and ovarian cancer: a meta-analytic review. *Health Psychology, 28*(4), 510-518.

Hanley, J.A., and McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29-36.

Hartge, P., Struewing, J.P., Wacholder, S., Brody, L.C., & Tucker, M.A. (1999). The prevalence of common BRCA1 and BRCA2 mutations among Ashkenazi Jews. *Am. J. Hum. Genet., 64*(4), 963-970.

Hughes RiskApps. (2016). Retrieved from http://www.hughesriskapps.com/downloads.php.

Hughes RiskApps Work Flow. (2012). Retrieved from http://www.hughesriskapps.net.

James, P.A., Doherty, R., Harris, M., Mukesh, B.N., Milner, A., Young, M-A., & Scott, C. (2006). Optimal selection of individuals for BRCA mutation testing: a comparison of available methods. *J Clin Oncol, 24*(4), 707-715.

Kang, H.H., Williams, R., Leary, J., kConFab Investigators, Ringland, C., Kirk, J., & Ward, R. (2006). Evaluation of models to predict BRCA germline mutations. *British Journal of Cancer, 95*(7), 914-920.

Katki, H.A. (2007). Incorporating medical interventions into carrier probability estimation for genetic counseling. *BMC Medical Genetics, 8*, 13.

Katki, H.A., Blackford, A., Chen, S., & Parmigiani, G. (2008). Multiple diseases in carrier probability estimation: accounting for surviving all cancers other than breast and ovary in BRCAPRO. *Stat Med., 27*(22), 4532-4548.

King, M.C., Marks, J.H., and Mandell, J.B. (2003). Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science, 302*, 643-646.

Komen, S.G. (2016). Facts for Life. Breast cancer in men.

Kurian, A.W., Gong, G.D., Chun, N.M., Mills, M.A., Staton, A.D., Kingham, K.E., Crawford, B.B., Lee, R., Chan, S., Donlon, S.S., Ridge, Y., Panabaker, K., West, D.W., Whittemore, A.S., & Ford, J.M. (2008). Performance of BRCA1/2 mutation prediction models in Asian-Americans. *J Clin Oncol, 26*, 4752-4758.

Lindor, N.M., Lindor, R.A., Apicella, C., Dowty, J.G., Ashley, A., Hunt, K., Mincey, B.A., Wilson, M., Smith, M.C., & Hopper, J.L. (2007). Predicting BRCA1 and BRCA2 gene mutation carriers: comparison of LAMBDA, BRCAPRO, Myriad II, and modified Couch models. *Fam Cancer, 6*(4), 473-482.

Lodder, L.N. (2001). Dealing with the risk for hereditary breast and ovarian cancer: a prospective study on psychological consequences of choices on genetic testing, surveillance, and prophylactic surgery. *Erasmus University Rotterdam.*

Lynch, H.T., Silva, E., Snyder, C., & Lynch, J.F. (2008). Hereditary breast cancer: part I. diagnosing hereditary breast cancer syndromes. *The Breast Journal, 14*(1), 3-13.

Marroni, F., Aretini, P., D'Andrea, E., Caligo, M.A., Cortesi, L., Viel, A., Ricevuto, E., Montagna, M., Cipollini, G., Ferrari, S., Santarosa, M., Bisegna, R., Bailey-Wilson, J.E., Bevilacqua, G., Parmigiani, G., & Presciuttini, S. (2004). Evaluation of widely used models for predicting BRCA1 and BRCA2 mutations. *J Med Genet, 41*, 278-285.

Mazzola, E., Blackford, A., Parmigiani, G., & Biswas, S. (2015). Recent enhancements to the genetic risk prediction model BRCAPRO. *Cancer Inform, 14*(S2), 147-157.

Mazzola, E., Chipman, J., Cheng, S.-C., & Parmigiani, G. (2014). Recent BRCAPRO upgrades significantly improve calibration. *Cancer Epidemiol Biomarkers Prev, 23*(8), 1689-1695.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P.K., Norris, F.H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J.C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A., &

Skolnick, M.H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science, 266*(5182), 66-71.

Nanda, R., Schumm, L.P., Cummings, S., Fackenthal, J.D., Sveen, L., Ademuyiwa, F., Cobleigh, M., Esserman, L., Lindor, N.M., Neuhausen, S.L., & Olopade, O.I. (2005). Genetic testing in an ethnically diverse cohort of high-risk women: a comparative analysis of BRCA1 and BRCA2 mutations in American families of European and African ancestry. *JAMA, 294*(15), 1925-1933.

National Cancer Institute. (2015). BRCA1 and BRCA2: cancer risk and genetic testing. Retrieved from http://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet.

National Cancer Institute. (2016). Genetics of breast and gynecologic cancers (PDQ) – health professional version. Retrieved from http://www.cancer.gov/types/breast/hp/breast-ovarian-genetics-pdq#link/_2_toc.

NCCN Guidelines. (2015). Retrieved from http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#detection.

Ozanne, E.M., Loberg, A., Hughes, S., Lawrence, C., Drohan, B., Semine, A., Jellinek, M., Cronin, C., Milham, F., Dowd, D., Block, C., Lockhard, D., Sharko, J., Grinstein, G., & Hughes, K.S. (2009). Identification and management of women at high risk for hereditary breast/ovarian cancer syndrome. *The Breast Journal, 15*(2), 155-162.

Palma, M., Ristori, E., Ricevuto, E., Giannini, G., & Gulino, A. (2006). BRCA1 and BRCA2: the genetic testing and the current management options for mutation carriers. *Critical Reviews in Oncology/Hematology, 57*, 1-23.

Parmigiani, G. (2002). Modeling in medical decision making: a Bayesian approach. *Wiley*.

Parmigiani, G., Berry, D.A., and Aguilar, O. (1998). Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *Am. J. Hum. Genet., 62*, 145-158.

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, C.I., Strong, L.C., Berry, D.A., Weitzel, J.N., Sand, S., 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.

Parmigiani, G., Chen, S., Wang, W., Tai, Y.C., Katki, H.A., & Blackford, A. (2011). Determining carrier probabilities for cancer susceptibility genes. *R documentation, BayesMendel package version 2.0-6*.

Pencina, M.J., D'Agostino, Sr., R.B., D'Agostino, Jr., R.B., & Vasan, R.S. (2008). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med., 27*, 157-172.

Pepe, M.S. (2004). The statistical evaluation of medical tests for classification and prediction. *Oxford University Press*.

Schwartz, G.F., Hughes, K.S., Lynch, H.T., Fabian, C.J., Fentiman, I.S., Robson, M.E., Domchek, S.M., Hartmann, L.C., Holland, R., Winchester, D.J., & Consensus Conference Committee The International Consensus Conference Committee. (2008). Proceedings of the international consensus conference on breast cancer risk, genetics, & risk management, April, 2007. *Cancer, 113*, 2627-2637.

Schwartz, M.D., Peshkin, B.N., Hughes, C., Main, D., Isaacs, C., & Lerman, C. (2002). Impact of BRCA1/BRCA2 mutation testing on psychologic distress in a clinic-based sample. *J Clin Oncol, 20*, 514-520.

Shattuck-Eidens, D., Oliphant, A., McClure, M., McBride, C., Gupte, J., Rubano, T., Pruss, D., Tavtigian, S.V., Teng, D.H., Adey, N., Staebell, M., Gumpper, K., Lundstrom, R., Hulick, M., Kelly, M., Holmen, J., Lingenfelter, B., Manley, S., Fujimura, F., Luce, M., Ward, B., Cannon-Albright, L., Steele, L., Offit, K., Gilewski, T., Norton, L., Brown, K., Schulz, C., Hampel, H., Schluger, A., Giulotto, E., Zoli, W., Ravaioli, A., Nevanlinna, H., Pyrhonen, S., Rowley, P., Loader, S., Osborne, M.P., Daly, M., Tepler, I., Weinstein, P.L., Scalia, J.L., Michaelson, R., Scott, R.J., Radice, P., Pierotti, M.A., Garber, J.E., Isaacs, C., Peshkin, B., Lippman, M.E., Dosik, M.H., Caligo, M.A., Greenstein, R.M., Pilarski, R., Weber, B., Burgemeister, R., Frank, T.S., Skolnick, M.H., & Thomas, A. (1997). BRCA1 sequence analysis in women with high risk for susceptibility mutations: risk factor analysis and implications for genetic testing. *JAMA, 278*, 1242-1250.

Tai, Y.C., Chen, S., Parmigiani, G., & Klein, A.P. (2008). Incorporating tumor immunohistochemical markers in BRCA1 and BRCA2 carrier prediction. *Breast Cancer Res, 10*(2), 401.

Tai, Y.C., Domchek, S., Parmigiani, G., & Chen, S. (2007). Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. *J Natl Cancer Inst, 99*(23), 1811-1814.

U.S. Cancer Statistics Working Group. (2016). United States Cancer Statistics: 1999-2013 incidence and mortality web-based report. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Retrieved from www.cdc.gov/uscs.

Vahteristo, P., Eerola, H., Tamminen, A., Blomqvist, C., & Nevanlinna, H. (2001). A probability model for predicting BRCA1 and BRCA2 mutations in breast and breast-ovarian cancer families. *British Journal of Cancer, 84*(5), 704-708.

Vogel, K.J., Atchley, D.P., Erlichman, J., Broglio, K.R., Ready, K.J., Valero, V., Amos, C.I., Hortobagyi, G.N., Lu, K.H., & Arun, B. (2007). BRCA1 and BRCA2 genetic testing in Hispanic patients: mutation prevalence and evaluation of the BRCAPRO risk assessment model. *J Clin Oncol, 25*(29), 4635-4641.

Wagner, T.M.U., Moslinger, R., Langbauer, G., Ahner, R., the Austrian Hereditary Breast and Ovarian Cancer Group, Fleischmann, E., Auterith, A., Friedmann, A., Helbich, T., Zielinski, C., Pittermann, E., Seifert, M., & Oefner, P. (2000). Attitude towards prophylactic surgery and effects of genetic counselling in families with BRCA mutations. *British Journal of Cancer, 82*(7), 1249-1253.

Wang, W., Chen, S., Brune, K.A., Hruban, R.H., & Parmigiani, G.P. (2007). PancPRO: risk assessment for individuals with a family history of pancreatic cancer. *J Clin Oncol, 25*, 1417–1422.

Wang, W., Niendorf, K.B., Patel, D., Blackford, A., Marroni, F., Sober, A.J., Parmigiani, G., & Tsao, H. (2010). Estimating CDKN2A carrier probability and personalizing cancer risk assessments in hereditary melanoma using MelaPRO. *Cancer Res, 70*(2), 552–559.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., & Micklem, G. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature, 378*(6559), 789-92.