Comparing Mammography Abnormality Features to Genetic Variants in the Prediction of Breast Cancer in Women Recommended for Breast Biopsy

Elizabeth S. Burnside, MD, MPH, MS, Jie Liu, PhD, Yirong Wu, PhD, Adedayo A. Onitilo, MD, MSCR, FACP, Catherine A. McCarty, PhD, C. David Page, PhD, Peggy L. Peissig, PhD, Amy Trentham-Dietz, PhD, Terrie Kitchner, Jun Fan, PhD, Ming Yuan, PhD

Rationale and Objectives: The discovery of germline genetic variants associated with breast cancer has engendered interest in risk stratification for improved, targeted detection and diagnosis. However, there has yet to be a comparison of the predictive ability of these genetic variants with mammography abnormality descriptors.

Materials and Methods: Our institutional review board-approved, Health Insurance Portability and Accountability Act-compliant study utilized a personalized medicine registry in which participants consented to provide a DNA sample and to participate in longitudinal follow-up. In our retrospective, age-matched, case-controlled study of 373 cases and 395 controls who underwent breast biopsy, we collected risk factors selected a priori based on the literature, including demographic variables based on the Gail model, common germline genetic variants, and diagnostic mammography findings according to Breast Imaging Reporting and Data System (BI-RADS). We developed predictive models using logistic regression to determine the predictive ability of (1) demographic variables, (2) 10 selected genetic variants, or (3) mammography BI-RADS features. We evaluated each model in turn by calculating a risk score for each patient using 10-fold cross-validation, used this risk estimate to construct Receiver Operator Characteristic Curve (ROC) curves, and compared the area under the ROC curve (AUC) of each using the DeLong method.

Results: The performance of the regression model using demographic risk factors was not statistically different from the model using genetic variants ($P = 0.9$). The model using mammography features (AUC = 0.689) was superior to both the demographic model (AUC = .598; $P < 0.001$) and the genetic model (AUC = .601; $P < 0.001$).

Conclusions: BI-RADS features exceeded the ability of demographic and 10 selected germline genetic variants to predict breast cancer in women recommended for biopsy.

Key Words: Mammography; Genetic variants; BI-RADS; Risk estimation; Predictive value.

Acad Radiol 2016; 23:62–69

From the Department of Radiology, University of Wisconsin School of Medicine and Public Health, E3/311 Clinical Science Center, 600 Highland Ave., Madison, WI 53792-3252 (E.S.B., Y.W.); Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin (E.S.B., J.L., C.D.P.); Marshfield Clinic Research Foundation, Marshfield, Wisconsin (A.A.O., P.P., T.K.); Department of Hematology/Oncology, Marshfield Clinic Weston Center, Weston, Wisconsin (A.A.O.); Essentia Institute of Rural Health, Duluth, Minnesota (C.MC.); Department of Population Health Sciences, University of Wisconsin (A.T-D.); Department of Statistics, University of Wisconsin (J.F., M.Y.), Madison, Wisconsin; Morgridge Institute for Research (J.F., M.Y.). Received May 15, 2015; revised September 15, 2015; accepted September 28, 2015. **Address correspondence to:** E.S.B. e-mail: eburnside@uwhealth.org

http://dx.doi.org/10.1016/j.acra.2015.09.007

INTRODUCTION

Over the last several decades, predictive variables have been discovered and incorporated into risk prediction models (1–3) with the goal of personalizing breast cancer screening and diagnosis. One highly predictive source of information is abnormality level feature descriptors observed on mammography as described in the Breast Imaging Reporting and Data System (BI-RADS) (4–8). Other emerging sources are the ever-growing genome-wide association studies (GWAS) that identify genetic variants (single nucleotide polymorphisms [SNPs]). The SNPs discovered via recent GWAS are distinct from mutations in the BRCA1 and BRCA2

tumor suppressor genes (9). Although both are germline genetic risk factors (inherited from parental lineage), SNPs discovered in recent GWAS are single base pair DNA sequence variations conferring modest risk (low penetrance) but occurring commonly (high frequency) within the human population. Expansion of genetic risk prediction may depend on polygenic risk stratification, that is, weighing many high-frequency, low-penetrance SNPs at once (3,10). Early attempts to use such SNPs to predict breast cancer risk have demonstrated only modest improvements over conventional demographic risk factors, like those in the Gail model (11–13).

Breast cancer risk is determined by a combination of genetic and environmental factors. Intermediate phenotypes like imaging (14) can capture and convey these interactions of these risk factors and provide biomarkers that can augment comprehensive risk prediction. Because demographic risk factors, genetic variants, and imaging features will all likely have some level of predictive value, determining which variables provide the best predictive power in any given setting becomes extremely important. Investing limited resources in collection of the best predictive variables will provide the most benefit. Prior literature evaluated risk prediction with genetics and breast density (15,16) and one paper added BI-RADS assessment category (17). Despite the proven predictive ability of abnormality-level features described in the BI-RADS lexicon (4–8) (e.g. mass and calcification descriptors as well as associated findings like architectural distortion), comparison to demographic or genetic risk has been limited. To estimate breast cancer risk in women recommended for breast biopsy, we compare the performance of predictive models using distinct data elements: demographic risk factors, germline genetic variants, or mammography abnormality features.

MATERIALS AND METHODS

Subjects

The source of subjects for this project was the population-based anonymized personalized medicine registry (APMR), details of which have been published previously (18). Briefly, Marshfield Clinic patients aged 18 years and older residing in one of 19 zip codes surrounding Marshfield, Wisconsin were invited to participate. After giving written informed consent, participants provided a blood sample from which DNA, plasma, and serum were extracted and stored. Permission was given to link the biological samples with medical records and a brief questionnaire was completed.

We selected subjects from the APMR using a retrospective case-control design. Women with available DNA sample, a diagnostic mammogram, and a breast biopsy within 12 months after the mammogram were included. Cases were defined as women having a confirmed diagnosis of invasive breast cancer or ductal carcinoma in situ (DCIS) obtained from the Marshfield Clinic institutional cancer registry. Controls were confirmed through the electronic medical records (and absence from the cancer registry) as having a benign biopsy result and never having had a breast cancer diagnosis.

To ensure a similar age distribution, we selected a control whose age was within 5 years of the age of each case. A total of 35 subjects were excluded from the statistical analysis. We excluded three cases with known BRCA1 mutation and three cases with known BRCA2 mutation (because these mutations would likely dominate all other predictive variables). We excluded eight nonwhite women from our study because GWAS variants can differ between races, and we did not have an appropriate number or distribution of nonwhites to effectively consider race or to race-match cases and controls. Finally, we excluded all instances in which BI-RADS features and breast density were all missing (21 cases). Some of the excluded subjects met more than one exclusion criterion.

All epidemiological, genetic, and mammographic risk factors were chosen a priori based on the literature (11–13,19) to represent the variables most likely to influence breast cancer risk, and these factors were included in analysis regardless of subsequent statistical significance.

Epidemiological Risk Factors

Variables used in the current study that were collected at the time of enrollment into the APMR included age and gender. Medical records were manually abstracted for the following information based on Gail risk factors: family history of breast cancer, age at menarche, and number of biopsies (prior to the index biopsy qualifying each subject for inclusion). Age at first live birth was not available in our cohort, so parity was instead used in our "DEMOGRAPHIC" model because of known association with breast cancer risk and correlation with age at first birth (20).

Genetic Variants

The APMR was one of five initial biobanks in the eMERGE Network funded by the National Human Genome Research Institute (21). We identified 10 genetic variants shown to predict breast cancer in large GWAS studies (22,23) and tested for breast cancer risk prediction (Table 1) (11–13). We sequenced these 10 SNPs on the Sequenom MassARRAY system. Because humans have two paired chromosomes with two chances to inherit the higher risk ("risky") allele, there are several accepted methods to quantify risky alleles for analysis. We enumerated the SNPs using two methods previously described in the literature (11). The "variant count" method quantifies the number of risky alleles (one allele per SNP for heterozygotes and two alleles per SNP for homozygotes) aggregating them into categories of ≤6, 7 or 8, 9 or 10, 11 or 12, and ≥13. The "individual count" method, which we used in our "GENETIC" model, quantifies how many risky alleles are present (0, 1, or 2 risky alleles) for each individual SNP resulting in possible values of 0–30 inclusive.

Mammography Features

To capture mammography abnormality level data, the biopsies of both cases and controls were matched with one diagnostic