Research Article

Cancer Epidemiology, Biomarkers & Prevention

Development of a Breast Cancer Risk Prediction Model for Women in Nigeria

Shengfeng Wang[1,2], Temidayo Ogundiran[3], Adeyinka Ademola[3], Oluwasola A. Olayiwola[4], Adewunmi Adeoye[4], Adenike Sofoluwe[5], Imran Morhason-Bello[6], Stella Odedina[6], Imaria Agwai[6], Clement Adebamowo[7], Millicent Obajimi[5], Oladosu Ojengbede[6], Olufunmilayo I. Olopade[2], and Dezheng Huo[2,8]

Abstract

Background: Risk prediction models have been widely used to identify women at higher risk of breast cancer. We aimed to develop a model for absolute breast cancer risk prediction for Nigerian women.

Methods: A total of 1,811 breast cancer cases and 2,225 controls from the Nigerian Breast Cancer Study (NBCS, 1998–2015) were included. Subjects were randomly divided into the training and validation sets. Incorporating local incidence rates, multivariable logistic regressions were used to develop the model.

Results: The NBCS model included age, age at menarche, parity, duration of breastfeeding, family history of breast cancer, height, body mass index, benign breast diseases, and alcohol consumption. The model developed in the training set performed well in the validation set. The discriminating accuracy of the NBCS model [area under ROC curve (AUC) = 0.703, 95% confidence interval (CI), 0.687–0.719] was better than the Black Women's Health Study (BWHS) model (AUC = 0.605; 95% CI, 0.586–0.624), Gail model for white population (AUC = 0.551; 95% CI, 0.531–0.571), and Gail model for black population (AUC = 0.545; 95% CI, 0.525–0.565). Compared with the BWHS and two Gail models, the net reclassification improvement of the NBCS model were 8.26%, 13.45%, and 14.19%, respectively.

Conclusions: We have developed a breast cancer risk prediction model specific to women in Nigeria, which provides a promising and indispensable tool to identify women in need of breast cancer early detection in Sub-Saharan Africa populations.

Impact: Our model is the first breast cancer risk prediction model in Africa. It can be used to identify women at high risk for breast cancer screening. *Cancer Epidemiol Biomarkers Prev; 27(6); 636–43. ©2018 AACR.*

Introduction

Breast cancer is the most common cancer in women, with nearly 1.7 million new cases diagnosed worldwide and 0.88 million in low- to middle-income countries in 2012 (1). Ideally, women at increased risk for breast cancer should be accurately identified so that appropriate prevention strategies can be offered, (e.g., control of specific modifiable risk factors and effective integrated prevention of noncommunicable diseases). This risk-stratified screening strategy is particularly important for countries with low incidence where cancer often present at late stage due to low awareness. To our knowledge, no breast cancer screening program is generally offered in Sub-Saharan Africa (SSA), and there is no risk prediction model for women in this region.

The Breast Cancer Risk Assessment Tool, also known as the Gail model, has been widely used and was validated in white women (2). The Gail model was modified for African Americans using data from the Women's Contraceptive and Reproductive Experiences study (3). Researchers from the Black Women's Health Study (BWHS) developed a new breast cancer risk prediction model using a prospective cohort of African American women ages 30 to 69 years (4, 5). However, to our knowledge, none of the existing breast cancer risk prediction models has been tested for application in indigenous women in SSA. The end users of a risk prediction model need to know whether the model is applicable to their population.

Compared with women in high-income countries, women in SSA have different features and risk factor profiles, including later menarche age, and higher parity (6–8). In addition, the incidence rates of breast cancer in SSA are lower than those in high-income countries (1, 9, 10). Biased risk projections could result in women receiving misleading counseling even missing the best time to intervene. Therefore, it is very important to develop accurate risk prediction model for SSA women based on data collected from indigenous populations in this region. The aims of this study were (i) to develop and validate an absolute breast cancer risk

[1]Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China. [2]Center for Clinical Cancer Genetics and Global Health, Department of Medicine, University of Chicago, Chicago, Illinois. [3]Department of Surgery, College of Medicine, University of Ibadan, Ibadan, Nigeria. [4]Department of Pathology, College of Medicine, University of Ibadan, Ibadan, Nigeria. [5]Department of Radiology, University College Hospital, Ibadan, Nigeria. [6]Center for Population and Reproductive Health, College of Medicine, University of Ibadan, Ibadan, Nigeria. [7]Department of Epidemiology and Preventive Medicine, University of Maryland, Baltimore, Maryland. [8]Department of Public Health Sciences, University of Chicago, Chicago, Illinois.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Corrected online July 30, 2018.

Corresponding Authors: Dezheng Huo, University of Chicago, 5841 S. Maryland Avenue, MC 2000, Chicago, IL 60637. Phone: 773-834-0843; E-mail: dhuo@health.bsd.uchicago.edu; Olufunmilayo I. Olopade, Center for Clinical Cancer Genetics and Global Health, University of Chicago, 5841 S. Maryland Avenue, MC 2020, Chicago, IL 60637. Phone: 773-702-1632; E-mail: folopade@medicine.bsd.uchicago.edu

doi: 10.1158/1055-9965.EPI-17-1128

prediction model specifically for SSA using data from the Nigerian Breast Cancer Study (NBCS), and (ii) to compare the performance of the new model with the Gail model for white population, Gail model for black population, and BWHS models among SSA women.

Materials and Methods

Study sample

The NBCS is a case–control study initiated in March 1998 and conducted in Ibadan, Nigeria. This study was conducted in accordance with recognized ethical guidelines, including Belmont Report and U.S. Common Rule, and the study protocol was approved by the institutional review boards of the University of Chicago and the University of Ibadan. The study setting and design have been described elsewhere (11–14). Briefly, cases were identified at the University College Hospital (UCH) in Ibadan. Serving a population in southwest Nigeria, UCH is the main tertiary referral center for other hospitals and thus treats the majority of breast cancer cases in Ibadan. Cases were defined as women who were at least 18 years old with a histologic or clinical diagnosis of invasive breast cancer. Controls were initially recruited from the communities that represent the diversity of UCH patients in terms of ethnicity and socioeconomic status. Field interviewers approached households in these communities and invited eligible women to visit community centers for the study. Additional controls were recruited through general medical outpatient clinic and ophthalmology clinic in UCH, and they were unselected for their medical conditions. Because risk prediction models using two types of controls were similar, they were pooled together in the analysis. All study participants gave written informed consent, and recruitment was highly successful with a response rate of >90%. By November 2015, 4,368 participants were recruited. A total of 332 participants (7.6%) were excluded from the analysis, due to recurrent breast cancer ($n = 34$) or missing key variables, including age at menarche ($n = 185$), benign breast diseases ($n = 20$), parity ($n = 19$), duration of breastfeeding ($n = 27$), body mass index (BMI; $n = 79$), and alcohol consumption ($n = 10$).

Relative risk prediction model

Based on literature on breast cancer risk factors and previous findings from NBCS (6, 11–13, 15–17), we considered the following factors as potential predictors for breast cancer risk: current age, age at menarche, benign breast diseases, family history of breast cancer in first degree relatives, parity, age at first live birth, total duration of breastfeeding, height, BMI, waist–hip ratio, and alcohol consumption. Alcohol consumption was defined as consumption of alcoholic beverages at least once a week for 6 months or longer. Status of benign breast diseases was asked by the question "Has a doctor ever told you that you had benign breast disease, such as a noncancerous cyst or a breast lump?"

Subjects were randomly divided into the training set (2/3 of the data) and validation set (1/3 of the data) to examine the overfitting issue in model building, and multivariable logistic regressions were used to derive the model. Predictive factors identified in age-adjusted logistic analysis were subjected to backward stepwise logistic regression analysis (details in Supplementary Table S1). We also tested for interaction terms in the models and explored the appropriate form of continuous variables. Three indices were used to compare different models, including likelihood ratio test, Akaike information criterion, and Bayesian information criterion (18). Parity was finally modeled using a linear spline function with a knot at 1 child, separating into two variables: first live birth (yes, no), each additional live birth (continuous).

Discrimination performance of the relative risk prediction models was assessed using concordance index (C-index), which is also known as area under the receiver operating characteristic curve (AUC), with 1 indicating perfect discrimination and 0.5 indicating no discriminating value. As a case–control study cannot fully examine model calibration, we can only assess model refinement, which describes the fitness of the model given corrected calibration (19). Refinement performance was examined using the expected/observed ratio (E/O), with E/O = 1 indicating perfect refinement. The expected number of cases was calculated by summing the individual projected probabilities based on model developed from training set. The 95% confidence intervals (CIs) for E/O ratios were calculated assuming a Poisson distribution (3). For a robust evaluation of our method, we also used 10-fold cross-validation procedure, which is less sensitive to parameter tuning. Logistic regression models were also used to estimate the odds ratios for breast cancer by percentile of the predicted chance of cases.

Absolute risk prediction model

After validating the relative risk prediction model, we developed an absolute risk model by updating the age-specific intercepts of the logistic regression using breast cancer incidence rates (2006–2009) from Ibadan Cancer Registry (Drs. Olufemi Ogunbiyi and Maxwell Parkin, personal communication; Supplementary Table S2). Ibadan Cancer Registry is a population-based cancer registry that covers the same catchment area as the NBCS. Both cases and controls in NBCS were selected independent of breast cancer risk factors other than age. In particular, we applied Bayes's rule to derive an adjustment term, $-log[\frac{d_i}{c_i} \times \frac{(1-Incidence_{age_i})}{Incidence_{age_i}}]$, to account for different selection probabilities of cases and controls from general population into the study (details in Supplementary Methods). In the adjustment term, d_i stands for number of cases, c_i stands for number of controls, and $Incidence_{age_i}$ stands for incidence rate in each age group i.

To remove competing risk, we obtained female Nigerian mortality rates in 2013 from World Health Organization (http://apps.who.int/gho/data/view.main.61200?lang=en, accessed on March 21, 2016; Supplementary Table S3). We calculated competing mortality rates as total mortality rate minus breast cancer–specific mortality rate. Because breast cancer–specific mortality in Nigeria is not available, we used breast cancer mortality-to-incidence rate ratio (MR:IR) for African countries, 0.69 (20). MR:IR, as an indirect measure of cancer survival, was calculated by dividing the mortality rate by the incidence rate. For each study participant who was younger than age 80, we calculated projected cumulative risks up to age 80 using the strategy specified in the Gail models (2, 3, 21).

Comparison with existing models

The two Gail models and the BWHS model (2, 3, 5) were applied to the study sample (Supplementary Fig. S1), except that we applied age-specific incidence rates and mortality rates from Nigeria to ensure consistence with NBCS model. Receiver-operating characteristic (ROC) analyses were performed, and area under ROC curve, or C-index, was used to indicate discriminating capacity. We also

Table 1. Distribution of breast cancer risk factors in cases and controls in the NBCS (1998–2015)

Characteristics	Case, N (%)	Control, N (%)	Age-adjusted odds ratio (95% confidence interval)	*P*	P_{trend}
N	1811	2225			
Age group				<0.001	<0.001
<25	9 (0.5)	153 (6.9)	0.06 (0.03–0.13)		
25–29.9	59 (3.3)	212 (9.5)	0.30 (0.21–0.41)		
30–34.9	151 (8.3)	314 (14.1)	0.51 (0.40–0.66)		
35–39.9	273 (15.1)	327 (14.7)	0.89 (0.71–1.12)		
40–44.9	297 (16.4)	317 (14.3)	1.00		
45–49.9	306 (16.9)	263 (11.8)	1.24 (0.99–1.56)		
50–54.9	254 (14.0)	222 (10.0)	1.22 (0.96–1.55)		
55–59.9	176 (9.7)	145 (6.5)	1.30 (0.99–1.70)		
60–64.9	132 (7.3)	139 (6.3)	1.01 (0.76–1.35)		
65–69.9	75 (4.1)	67 (3.0)	1.19 (0.83–1.72)		
70–74.9	50 (2.8)	43 (1.9)	1.24 (0.80–1.92)		
≥75	29 (1.6)	23 (1.0)	1.35 (0.76–2.38)		
Age at menarche				0.017	0.013
≥17	421 (23.3)	564 (25.4)	1.00		
16–16.9	255 (14.1)	343 (15.4)	1.02 (0.83–1.26)		
15–15.9	515 (28.4)	538 (24.2)	1.33 (1.11–1.59)		
14–14.9	255 (14.1)	341 (15.3)	1.09 (0.88–1.35)		
13–13.9	186 (10.3)	219 (9.8)	1.27 (1.00–1.62)		
12–12.9	151 (8.3)	171 (7.7)	1.30 (1.00–1.69)		
8–11.9	28 (1.6)	49 (2.2)	0.86 (0.52–1.42)		
Benign breast diseases	207 (11.4)	144 (6.5)	1.80 (1.43–2.26)	<0.001	<0.001
Family history of breast cancer	95 (5.3)	68 (3.1)	1.96 (1.40–2.74)	<0.001	<0.001
Parity				<0.001	0.001
0	155 (8.6)	246 (11.1)	1.00		
1	126 (7.0)	209 (9.4)	0.57 (0.40–0.80)		
2	198 (10.9)	268 (12.0)	0.48 (0.35–0.66)		
3	287 (15.8)	326 (14.7)	0.49 (0.36–0.67)		
4	297 (16.4)	393 (17.7)	0.36 (0.26–0.49)		
5	306 (16.9)	325 (14.6)	0.41 (0.30–0.56)		
6	213 (11.8)	206 (9.3)	0.41 (0.29–0.58)		
7	119 (6.6)	124 (5.6)	0.37 (0.25–0.54)		
≥8	110 (6.1)	128 (5.8)	0.32 (0.22–0.47)		
Age of first live birth				<0.001	<0.001
<20	352 (19.6)	428 (19.3)	1.00		
20–24.9	679 (37.7)	827 (37.2)	1.00 (0.83–1.19)		
25–29.9	419 (23.3)	542 (24.4)	0.98 (0.80–1.19)		
≥30	195 (10.8)	178 (8.0)	1.25 (0.97–1.60)		
No birth	155 (8.6)	246 (11.1)	2.25 (1.67–3.04)		
Age at menopause				0.536	0.384
<35	26 (3.23)	12 (1.90)	0.57 (0.20–1.63)		
35–44.9	162 (20.15)	96 (15.19)	1.04 (0.75–1.45)		
45–54.9	521 (64.80)	444 (70.25)	1.00		
≥55	95 (11.82)	80 (12.66)	1.19 (0.84–1.69)		
Total months of breastfeeding				<0.001	<0.001
<12	225 (12.4)	342 (15.4)	1.00		
12–23.9	155 (8.6)	193 (8.7)	0.71 (0.53–0.96)		
24–35.9	224 (12.4)	263 (11.8)	0.56 (0.42–0.71)		
36–47.9	207 (11.5)	224 (10.1)	0.57 (0.43–0.77)		
48–59.9	215 (11.9)	263 (11.8)	0.47 (0.35–0.63)		
60–71.9	163 (9.0)	225 (10.1)	0.37 (0.40–0.72)		
72–83.9	215 (11.9)	203 (9.1)	0.53 (0.40–0.72)		
84–95.9	82 (4.5)	103 (4.6)	0.38 (0.26–0.55)		
≥96	325 (18.0)	409 (18.4)	0.35 (0.26–0.46)		
Height in cm				<0.001	<0.001
<150	85 (4.7)	136 (6.1)	0.91 (0.68–1.23)		
150–159	654 (36.1)	1,067 (48.0)	1.00		
160–169	819 (45.2)	885 (39.8)	1.58 (1.37–1.82)		
≥170	253 (14.0)	137 (6.2)	3.26 (2.57–4.15)		
Body mass index in kg/m^2					
Before menopause				0.001	0.004
18.5–24.9	425 (44.1)	700 (44.8)	1.00		
<18.5	67 (7.0)	95 (6.1)	1.43 (1.00–2.05)		
25–29.9	276 (28.6)	432 (27.6)	0.89 (0.73–1.09)		
≥30	196 (20.3)	338 (21.6)	0.71 (0.57–0.89)		

(Continued on the following page)

Table 1. Distribution of breast cancer risk factors in cases and controls in the NBCS (1998–2015) (Cont'd)

Characteristics	Case, N (%)	Control, N (%)	Age-adjusted odds ratio (95% confidence interval)	*P*	P_{trend}
After menopause				<0.001	<0.001
18.5–24.9	327 (38.6)	191 (28.9)	1.00		
<18.5	55 (6.5)	28 (4.2)	1.12 (0.68–1.85)		
25–29.9	242 (28.6)	248 (37.6)	0.57 (0.44–0.74)		
≥30	223 (26.3)	193 (29.2)	0.68 (0.52–0.89)		
Waist–hip ratio				0.061	0.055
<0.80	453 (25.4)	599 (27.5)	1.00		
0.80–0.84	458 (25.7)	528 (24.3)	1.04 (0.87–1.25)		
≥0.85	871 (48.9)	1,050 (48.2)	0.87 (0.74–1.02)		
Oral contraceptive	512 (28.3)	741 (33.4)	0.75 (0.65–0.86)	<0.001	<0.001
Alcohol consumption	179 (9.9)	109 (4.9)	1.95 (1.52–2.52)	<0.001	<0.001
Study periods				0.009	0.389
1998–2004	619 (34.2)	732 (32.9)	1.00		
2005–2010	588 (32.5)	770 (34.6)	0.79 (0.67–0.92)		
2011–2015	604 (33.4)	723 (32.5)	0.93 (0.80–1.09)		

NOTE: There were 15, 71, 77, and 10 women with missing in age of first birth, age at menopause, waist–hip ratio, and oral contraceptive, respectively.
P value and P_{trend} value were all adjusted by age.

calculated age-adjusted C-index to remove effect of age. The Hosmer–Lemeshow test was applied to assess the fitness of model (22). We also evaluated classification accuracy using reclassification tables and quantified the differences in classification by net reclassification improvement (NRI; ref. 23). The threshold for enrolling in breast cancer prevention trials (a 5-year predicted risk of at least 1.66%) and recommended by the updated American Cancer Society guidelines for breast cancer screening (5-year projected risk for 45–49 years old was 0.9%) were used to determine how the new model would affect eligibility (24, 25).

All *P* values reported are two-sided. Statistical analyses were conducted with SAS 9.4 (SAS Institute Inc.) and Stata 15.0 (StataCorp). We developed both online risk calculator (http://bcrisktool.uchicago.edu/) and a standalone Windows package for the NBCS model to do individual risk projection and counseling.

Results

This study included a sample of 4,036 subjects: 1,811 cases and 2,225 controls recruited from March 1998 to November 2015 in Nigeria. The mean age at diagnosis (or at interview) was 47.4 years for cases and 42.5 years for controls ($P < 0.001$). As shown in Table 1, Nigerian women with breast cancer were characterized by late age at menarche, higher rates of a positive family history, multiple parity, and longer duration of breastfeeding. A total of 2,692 participants were randomly assigned to the training set (1,208 cases and 1,484 controls) and the remaining 1,344 were assigned to the validation set. The final model built using the training set includes age, age at menarche, number of live birth, duration of breastfeeding, benign breast diseases, family history of breast cancer in first degree relatives, height, BMI, and alcohol consumption (Supplementary Table S4). No significant interactions between risk factors were found.

The odds ratios for developing breast cancer by percentiles of the predicted chance of cases showed a monotonic increasing trend in the training set (Table 2). In the validation set, the trend was quite similar, and there was a 45-fold difference between the top 5 percentile and bottom 5 percentile. Discriminating accuracy (C-index) of the model applied to the validation set was 0.694 (95% CI, 0.666–0.721), similar to that in the training set (C-index = 0.720; 95% CI, 0.701–0.739). A more reliable evaluation of discriminating accuracy is from 10-fold cross-validation (C-index = 0.703; 95% CI, 0.687–0.719). Furthermore, the expected and observed numbers of breast cancer by percentiles of the predicted risk were almost the same except for the bottom 5 percentile group, with the overall E/O ratio being 1.01 (95% CI, 0.93–1.09). Taken together, the model has good refinement and discrimination capacity, without overfitting problem.

By pooling the training and validation sets, we reestimated regression coefficients without changing the form of each variable using a logistic regression. These new sets of coefficients (Table 3) were actually very similar to those from the training dataset. Supplementary Table S5 shows the absolute risks of developing breast cancer to age 80 years old.

Figure 1 shows the ROC curves of 5-year absolute risks projected from four models. The C-index of the NBCS model was 0.703 (95% CI, 0.687–0.719), which is significantly greater than the other three models (all $P < 0.001$). The C-index of the BWHS model was greater than that of two Gail models (both $P < 0.001$), and there was no difference between two Gail models ($P = 0.37$). In the age-specific ROC analysis, the NBCS model performed quite stably with C-indexes of 0.613 to 0.734, which were uniformly better than the other three models in each age group, especially in the youngest age categories (Table 4). The age-adjusted C-index of the NBCS model was 0.662 (95% CI, 0.641–0.682). We found the discriminating accuracy of the BWHS, two Gail models were relatively low after removing the effect of age (age-adjusted C-indexes of 0.574, 0.529, and 0.493).

Supplementary Fig. S2 shows the refinement of the models, and the NBCS model presented a good agreement between observed and predicted 5-year risk, over the whole range of probabilities. The other three models did not perform well in the Nigerian sample.

The 5-year predicted risk from the NBCS model was moderately correlated with those from the BWHS model (r = 0.66), the Gail model for white population (r = 0.55), and the Gail model for black population models (r = 0.54, all $P < 0.001$). Figure 2 showed the distribution of 5-year predicted risks for each model, only a small group of participants have 5-year risk greater than predefined thresholds. The distinguishing ability of the NBCS model was significantly better at the thresholds of 0.9% and 1.66%. According to the NBCS model, 95 (5.3%) cases had 5-year risk ≥1.66%, compared with 19 (0.9%) controls, yielding a 6-fold difference. At the threshold of 0.9%, 370 (20.7%) cases could be detected at the expense of 6.8% controls being screened. The BWHS model gave higher estimates of 5-year risk, but these risk estimates cannot distinguish cases and controls very well at

Table 2. The performance of the relative risk prediction model in the validation set

	Refinement in validation set					Discrimination	
Percentile of predicted risk[a]	No. of participants	Observed no. of cases (O)	Expected no. of case (E)	E/O	95% CI	Training set OR (95% CI)	Validation set OR (95% CI)
<5%	53	5	2.07	0.41	0.17–0.99	0.06 (0.03–0.14)	0.11 (0.04–0.29)
5%–20%	215	50	47.08	0.94	0.71–1.24	0.31 (0.23–0.42)	0.32 (0.22–0.48)
20%–40%	277	92	98.82	1.07	0.87–1.31	0.61 (0.47–0.77)	0.53 (0.38–0.75)
40%–60%	263	127	120.30	0.95	0.80–1.13	1.00 (referent)	1.00 (referent)
60%–80%	264	149	146.52	0.98	0.83–1.15	1.55 (1.22–1.97)	1.39 (0.98–1.96)
80%–95%	210	129	141.30	1.10	0.93–1.31	2.26 (1.73–2.95)	1.71 (1.18–2.47)
95%–	62	51	50.25	0.99	0.75–1.30	6.31 (3.85–10.36)	4.96 (2.48–9.95)
Total	1344	603	606.33	1.01	0.93–1.09		
C-index (95% CI)						0.720 (0.701–0.739)	0.694 (0.666–0.721)
Age-adjusted C-index (95% CI)						0.670 (0.645–0.695)	0.641 (0.604–0.679)

NOTE: Expected breast cancers = No. of participants × exp(OR)/(1+ exp(OR)).
Abbreviations: C-index, concordance index; OR, odds ratio.
[a]Percentile of risk from the logistic model developed in the training set.

threshold of 1.66%: 2.3% of cases and 1.0% of controls had a 5-year risk ≥1.66%. Both Gail models (using Nigerian incidence rates) gave an overall low estimate of risk, and neither can distinguish cases and controls very well at threshold of 1.66%: no women were predicted to have 5-year risk ≥1.66%. Compared with the BWHS model, the NBCS model was able to correctly change risk categories for a net of 8.26% of women. The NRI for the NBCS model was 13.45% compared with the Gail model for the white population and 14.19% compared with the Gail model for the black population (Supplementary Table S6).

Discussion

By integrating a large case–control study in Nigeria with breast cancer incidence rates, we developed an absolute risk prediction model for breast cancer for Nigerian women ages 20 to 79 years. Among Nigerian women, our model performed better than the BWHS and the two Gail models. The Gail model for the white population contains five variables: age, age at menarche, number of previous breast biopsy, age at first live birth, and number of first-degree relatives with breast cancer (2), whereas the Gail model for the black population uses the same set of predictors with some modifications (3). The BWHS model adds five extra factors, namely, BMI at age 18 years, oral contraceptive use, bilateral oophorectomy, estrogen plus progestin use, and height (5). Of the 9 risk factors used in the NBCS model, 6 are already used in aforementioned models, and 3 additional variables, alcohol intake, parity, and duration of breastfeeding, are included based on findings from our previous studies (6, 13, 17). Age at first live birth was not in our model due to its insignificant association with breast cancer after adjusting for other covariates (13), whereas bilateral oophorectomy and hormone replacement therapy use were very rare in the Nigerian population. Oral contraceptive use was excluded due to its inconsistent result with previous study to avoid misleading in practice, which might due to unmeasurable confounders (5). Furthermore, categorization of shared variables

Table 3. Parameters for breast cancer used in the absolute risk prediction model in the NBCS (1998–2015, $N = 4{,}036$)

Variables	Coefficient	χ^2	*P*	OR	95% CIs
Intercept[a]	−5.934			—	—
Age group[a] (ref: 40–44.9)		258.01	<0.001		
<25	−4.037			0.03	0.01–0.06
25–29.9	−2.568			0.20	0.14–0.29
30–34.9	−1.117			0.42	0.32–0.55
35–39.9	−0.412			0.86	0.68–1.09
45–49.9	0.675			1.42	1.12–1.80
50–54.9	0.596			1.49	1.16–1.93
55–59.9	0.987			1.86	1.39–2.49
60–64.9	0.621			1.57	1.15–2.14
65–69.9	0.813			1.93	1.30–2.87
70–74.9	0.513			2.08	1.30–3.33
≥75	0.085			2.53	1.37–4.70
Age at menarche (per year)	−0.032	3.74	0.053	0.97	0.94–1.00
First live birth	−0.611	15.28	<0.001	0.54	0.40–0.74
Each additional live birth	0.092	8.20	0.004	1.10	1.03–1.17
Breastfeeding (per 12 months)	−0.120	41.65	<0.001	0.89	0.85–0.92
Benign breast diseases	0.553	19.48	<0.001	1.74	1.36–2.22
Family history of breast cancer	0.430	5.57	0.018	1.54	1.08–2.20
Height (per 10 cm, centered at 160)	0.519	101.93	<0.001	1.68	1.52–1.86
Body mass index (ref: 18.5–24.9 kg/m^2)		31.92	<0.001		
<18.5	0.241			1.27	0.94–1.72
25–29.9	−0.333			0.72	0.61–0.84
≥30	−0.387			0.68	0.57–0.81
Alcohol consumption	0.529	14.60	<0.001	1.70	1.29–2.22

[a]Intercept and regression coefficients (odds ratios) for each age group have been adjusted using breast cancer incidence rates from the Ibadan Cancer Registry.

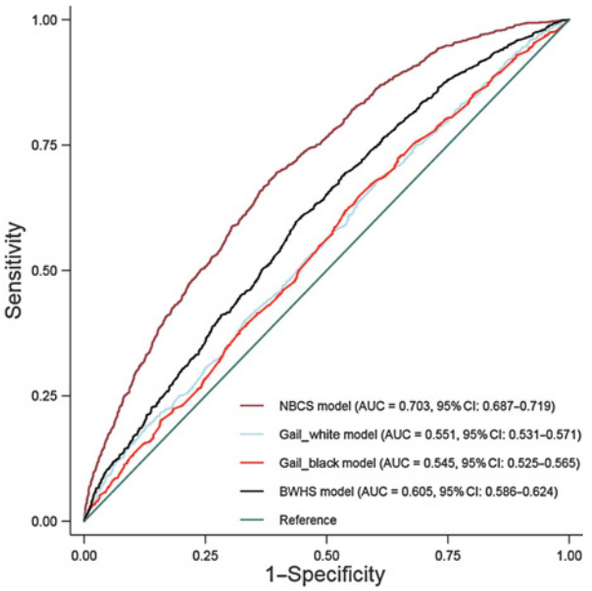

Figure 1.

ROC analysis and corresponding area under the curve (AUC) for the Gail model for the white population, Gail model for the black population, BWHS model, and NBCS model of prediction of invasive breast cancer on the NBCS. The discriminating accuracy of the four absolute risk prediction models was evaluated in their applicable age range.

across the four models is not the same, reflecting risk factor profiles in the samples used for model development.

We found that the age-adjusted C-index of the BWHS model was 0.57 in Nigerian women, which was close to that reported in the original study in African Americans (0.59; ref. 5). This is probably due to several risk factors, including height, BMI, family history, and benign breast disease in the BWHS being applicable to Nigerian women (Supplementary Table S7). We found the discriminating accuracy of the two Gail models was low after removing the effect of age. As shown in Supplementary Table S8, the risk factor categories of the Gail models are not applicable to our study participants: Less than 2% of our participants attained menarche at age 12 years or younger, few participants had breast biopsies. As a result, neither of the Gail models can differentiate our participants' risk, although age-specific incidence rates and mortality rates in Nigerian population were used in the Gail models.

The development of our risk prediction model has important public health implications for breast cancer control and prevention in SSA. First, early detection to improve outcome and survival remains the cornerstone of breast cancer control and it is more urgent in SSA countries because many breast cancers present with advanced stage. Most SSA countries face resource constraints that limit the capacity to universally screen breast cancer using mammography, let alone magnetic resonance imaging (MRI; ref. 26). In addition, the mean age at diagnosis of the SSA women was more than 10 years younger than American women (14, 27), so screening guidelines in the general population based on age alone, such as 45 years old by the American Cancer Society (24) or 50 years old by U.S. Prevention Service Task Force (28), will miss the majority of breast cancers in SSA countries. As a low-cost screening approach based on questionnaires and physical examination, our prediction model could be implemented in limited resource settings to identify the high-risk population for breast cancer screening. For example, if intensive surveillance with clinical breast examinations and ultrasono-mammographic screening only targets women with 5-year risk of 0.9% or higher.

Second, a woman's decision to accept prophylactic mastectomy or other interventions depends on her individualized risk estimate of developing breast cancer in a defined period (29, 30). For example, if a 5-year predicted risk of at least 1.66% were used for enrollment in breast cancer prevention trials, the NBCS model could help to enroll just top 1% of SSA women who would have high risk of breast cancer. Thirdly, we included potentially modifiable risk factors, such as alcohol consumption and BMI in the NBCS model, so that health professionals and education counselors can use the NBCS risk calculator to illustrate the change in projected risks in 5-year or up to age 80 for scenario that the counselee change her behaviors. Lastly, because of the common ancestry shared by sub-Saharan Africans and African American women, our work may provide some inspiration or clues to improve the existing model for African Americans, such as integrating reproductive factors and genomic risk factors (31).

Table 4. Discriminating accuracy of the four absolute risk prediction models in NBCS participants (1998–2015, $N = 3{,}963$[a])

		NBCS model		BWHS model		Gail-white model		Gail-black model	
Age group	**Number of participants**	**C-index**	**95% CI**	**C-index**	**95% CI**	**C-index**	**95% CI**	**C-index**	**95% CI**
20–29	412	0.734	0.664–0.803						
30–34	465	0.684	0.633–0.735	0.573	0.518–0.627				
35–39	600	0.682	0.640–0.725	0.626	0.582–0.671	0.604	0.558–0.649	0.580	0.534–0.626
40–44	614	0.662	0.619–0.705	0.577	0.532–0.622	0.563	0.518–0.609	0.555	0.510–0.600
45–49	569	0.650	0.605–0.695	0.577	0.529–0.623	0.533	0.485–0.580	0.509	0.462–0.556
50–54	476	0.664	0.615–0.712	0.590	0.539–0.641	0.523	0.471–0.575	0.525	0.473–0.576
55–59	321	0.613	0.551–0.674	0.501	0.438–0.564	0.483	0.419–0.546	0.442	0.379–0.504
60–64	271	0.648	0.583–0.713	0.575	0.507–0.643	0.503	0.433–0.572	0.516	0.447–0.585
65–69	142	0.670	0.612–0.782	0.573	0.479–0.668	0.519	0.429–0.614	0.557	0.462–0.652
70–74	93	0.697	0.570–0.789			0.505	0.385–0.625	0.427	0.308–0.546
Total[b]	3,963	0.703	0.687–0.719	0.605	0.586–0.624	0.551	0.531–0.571	0.545	0.525–0.565
Total[c] (ages 35–69)	2,993	0.658	0.639–0.678	0.580	0.560–0.601	0.555	0.534–0.575	0.549	0.528–0.569
Age-adjusted	—	0.662	0.641–0.683	0.574	0.550–0.598	0.529	0.503–0.556	0.493	0.466–0.521

[a]73 individuals were either younger than 20 years old ($n = 21$) or older than 75 years old ($n = 52$).

[b]BWHS model was evaluated only in participants ages 30 to 70 years old, while two Gail models were evaluated only in participants aged $\geq$ 35 years old.

[c]The discriminating accuracy of the four absolute risk prediction models was only evaluated in the overlap of their applicable age range (35–69 years old, $N = 2{,}993$).

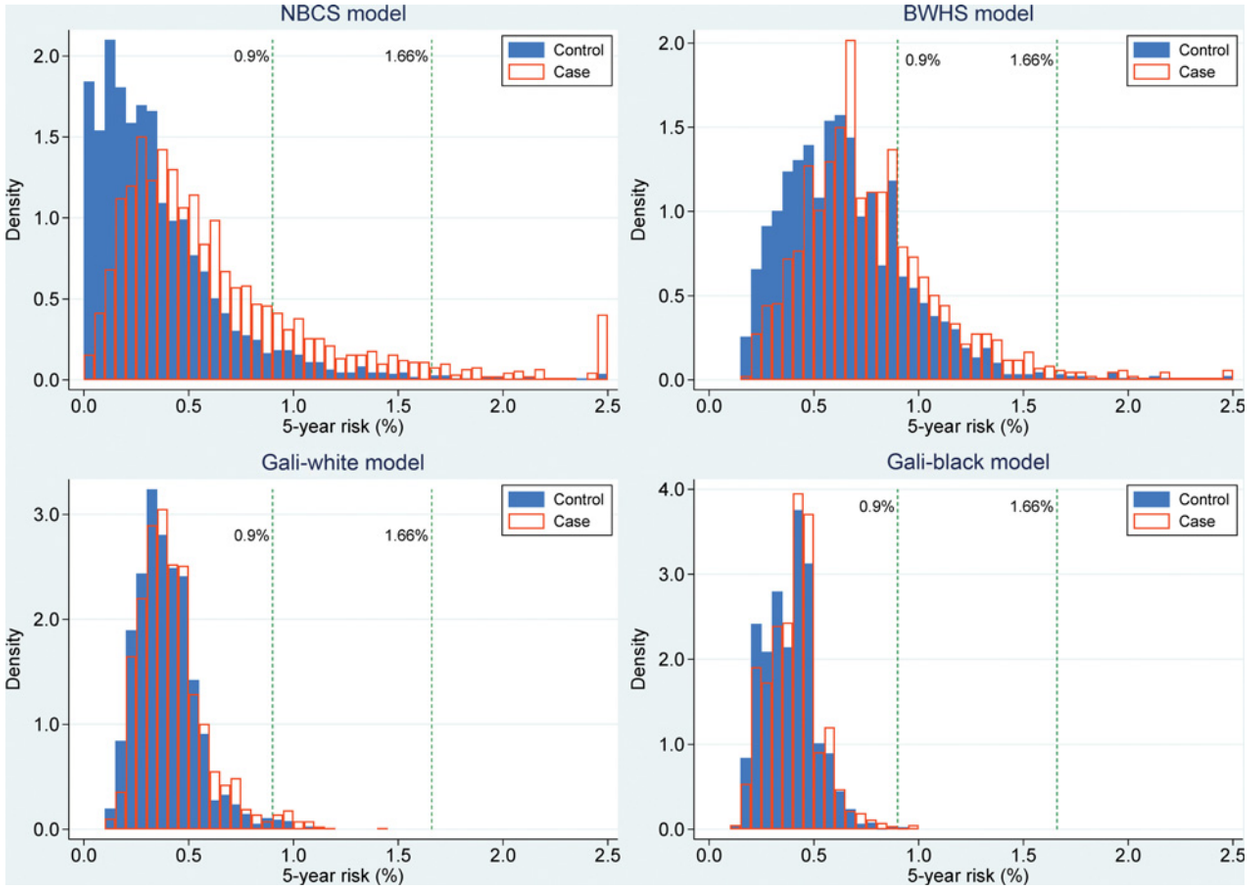

Figure 2.
The distribution of 5-year predicted risks for both breast cancer cases and controls in NBCS with the Gail models for the white population and black population, BWHS model, and NBCS model. The BWHS model was evaluated only in participants ages 30 to 70 years old, while two Gail models were evaluated only in participants age ≥35 years old. Risk was truncated at 2.5%.

To our knowledge, our model represents the first breast cancer risk prediction model for SSA women, a population in which breast cancer risk has not been fully appreciated. We acknowledge several limitations to this work. First, the model was developed and validated in the same population, which may have resulted in optimistic model performance in internal validation. Our model may not perform well in other African populations. Further replication study of our model in other African countries would be highly desirable to assess its applicability to SSA populations and to obtain more reliable estimates of the associations of established risk factors. Additionally, we lacked information on several other predictors, such as mammographic density (32) and exact histology of benign breast diseases (33), although the relatively high cost of obtaining these factors could restrict the model's application in SSA countries. Third, the incompleteness in case reporting in the Ibadan Cancer Registry might lead to underestimate of absolute risks in our NBCS model, although the registry has contributed the years 2006 to 2009 data to the GLOBOCAN database (http://globocan.iarc.fr), which suggests the data quality is acceptable. Moreover, we applied the same incidence and mortality in all four models, and the quality of above information would not affect the relative performance of these four models. The main purpose of comparing our model with the three existing models developed for other populations is to illustrate potential misleading results if applying the existing models in African population directly.

In summary, we developed an absolute risk prediction model for breast cancer that is specific to SSA women. It performed better than existing models and can be used to identify individuals at high risk of breast cancer to appropriately tailor surveillance and risk reduction strategies. Future attempts will include validating the model in other SSA countries and integrating genetic information (e.g., BRCA1/2 and low penetrance common variants) to further improve model's performance.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: S. Wang, A. Ademola, O.A. Olayiwola, O.I. Olopade, D. Huo

Development of methodology: S. Wang, O.A. Olayiwola, A. Adeoye, I. Morhason-Bello, O.I. Olopade, D. Huo

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): T. Ogundiran, A. Ademola, O.A. Olayiwola, A. Adeoye,

I. Morhason-Bello, S. Odedina, I. Agwai, C. Adebamowo, O. Ojengbede, O.I. Olopade, D. Huo
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S. Wang, A. Adeoye, O.I. Olopade, D. Huo
Writing, review, and/or revision of the manuscript: S. Wang, T. Ogundiran, A. Ademola, O.A. Olayiwola, A. Adeoye, I. Morhason-Bello, C. Adebamowo, O. Ojengbede, A. Sofoluwe, I. Agwai, M. Obajimi, O.I. Olopade, D. Huo
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Adeoye, C. Adebamowo, O. Ojengbede, O.I. Olopade, D. Huo
Study supervision: T. Ogundiran, A. Ademola, I. Morhason-Bello, C. Adebamowo, O. Ojengbede, O.I. Olopade, D. Huo
Other (obtained funding for the study): O.I. Olopade

Acknowledgments

We would like to thank the more than 4,000 study participants who were enrolled in the Nigerian Breast Cancer Study for their time and contribution. We are also immensely grateful to Professor Gladys A. Falusi (Institute for Medical Research and Training, University of Ibadan, Nigeria) and Professor Christopher O. Olopade (University of Chicago) for the ground-breaking ethical training work they did, which benefited this and other research projects in Nigeria.

This study was supported by the NCI R01CA89085 (to O.I. Olopade), P50CA125183 (to O.I. Olopade), and U01CA161032 (to D. Huo and O.I. Olopade); American Cancer Society MRSG-13-063-01-TBG (to D. Huo), CRP-10-119-01-CCE (to O.I. Olopade); Breast Cancer Research Foundation (to D. Huo and O.I. Olopade); and Susan G. Komen Foundation SAC110026 (to O.I. Olopade).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 7, 2017; revised February 8, 2018; accepted April 2, 2018; published first April 20, 2018.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87–108.
2. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989; 81:1879–86.
3. Gail MH, Costantino JP, Pee D, Bondy M, Newman L, Selvan M, et al. Projecting individualized absolute invasive breast cancer risk in African American women. J Natl Cancer Inst 2007;99:1782–92.
4. Boggs DA, Rosenberg L, Pencina MJ, Adams-Campbell LL, Palmer JR. Validation of a breast cancer risk prediction model developed for black women. J Natl Cancer Inst 2013;105:361–7.
5. Boggs DA, Rosenberg L, Adams-Campbell LL, Palmer JR. Prospective approach to breast cancer risk prediction in African American women: the Black Women's Health Study model. J Clin Oncol 2015; 33:1038–44.
6. Sighoko D, Ogundiran T, Ademola A, Adebamowo C, Chen L, Odedina S, et al. Breast cancer risk after full-term pregnancies among African women from Nigeria, Cameroon, and Uganda. Cancer 2015;121:2237–43.
7. Corbex M, Bouzbid S, Boffetta P. Features of breast cancer in developing countries, examples from North-Africa. Eur J Cancer 2014;50: 1808–18.
8. Brinton LA, Figueroa JD, Awuah B, Yarney J, Wiafe S, Wood SN, et al. Breast cancer in Sub-Saharan Africa: opportunities for prevention. Breast Cancer Res Treat 2014;144:467–78.
9. Jedy-Agba E, Curado MP, Ogunbiyi O, Oga E, Fabowale T, Igbinoba F, et al. Cancer incidence in Nigeria: a report from population-based cancer registries. Cancer Epidemiol 2012;36:e271–8.
10. Wabinga HR, Nambooze S, Amulen PM, Okello C, Mbus L, Parkin DM. Trends in the incidence of cancer in Kampala, Uganda 1991–2010. Int J Cancer 2014;135:432–9.
11. Hou N, Ogundiran T, Ojengbede O, Morhason-Bello I, Zheng Y, Fackenthal J, et al. Risk factors for pregnancy-associated breast cancer: a report from the Nigerian Breast Cancer Study. Ann Epidemiol 2013; 23:551–7.
12. Ogundiran TO, Huo D, Adenipekun A, Campbell O, Oyesegun R, Akang E, et al. Case-control study of body size and breast cancer risk in Nigerian women. Am J Epidemiol 2010;172:682–90.
13. Huo D, Adebamowo CA, Ogundiran TO, Akang EE, Campbell O, Adenipekun A, et al. Parity and breastfeeding are protective against breast cancer in Nigerian women. Br J Cancer 2008;98:992–6.
14. Hou N, Ndom P, Jombwe J, Ogundiran T, Ademola A, Morhason-Bello I, et al. An epidemiologic investigation of physical activity and breast cancer risk in Africa. Cancer Epidemiol Biomarkers Prev 2014;23:2748–56.
15. Adebamowo CA, Ogundiran TO, Adenipekun AA, Oyesegun RA, Campbell OB, Akang EU, et al. Obesity and height in urban Nigerian women with breast cancer. Ann Epidemiol 2003;13:455–61.
16. Ogundiran TO, Huo D, Adenipekun A, Campbell O, Oyesegun R, Akang E, et al. Body fat distribution and breast cancer risk: findings from the Nigerian breast cancer study. Cancer Causes Control 2012;23:565–74.
17. Qian F, Ogundiran T, Hou N, Ndom P, Gakwaya A, Jombwe J, et al. Alcohol consumption and breast cancer risk among women in three sub-Saharan African countries. PLoS One 2014;9:e106908.
18. Shmueli G. To explain or to predict? Stat Sci 2010;25:289–310.
19. Miller ME, Hui SL, Tierney WM. Validation techniques for logistic regression models. Stat Med 1991;10:1213–26.
20. Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. J Clin Oncol 2006;24:2137–50.
21. Benichou J, Gail MH. Methods of inference for estimates of absolute risk derived from population-based case–control studies. Biometrics 1995;51: 182–94.
22. Hosmer DW, Lemesbow S. Goodness of fit tests for the multiple logistic regression model. Commun Stat A. Theor 1980;9:1043–69.
23. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157–72; discussion 207–12.
24. Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA 2015;314:1599–614.
25. Vogel VG, Costantino JP, Wickerham DL, Cronin WM, Cecchini RS, Atkins JN, et al. Effects of tamoxifen vs. raloxifene on the risk of developing invasive breast cancer and other disease outcomes: the NSABP study of tamoxifen and raloxifene (STAR) P-2 trial. JAMA 2006;295:2727–41.
26. Formenti SC, Arslan AA, Love SM. Global breast cancer: the lessons to bring home. Int J Breast Cancer 2011;2012:7.
27. Anderson WF, Rosenberg PS, Menashe I, Mitani A, Pfeiffer RM. Age-related crossover in breast cancer incidence rates between black and white ethnic groups. J Natl Cancer Inst 2008;100:1804–14.
28. Siu AL, U.S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force Recommendation statement. Ann Intern Med 2016;164:279–96.
29. Howell A, Anderson AS, Clarke RB, Duffy SW, Evans DG, Garcia-Closas M, et al. Risk determination and prevention of breast cancer. Breast Cancer Res 2014;16:446.
30. Van Ravesteyn NT, Miglioretti DL, Stout NK, Lee SJ, Schechter CB, Buist DS, et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: a comparative modeling study of risk. Ann Intern Med 2012;156:609–17.
31. Fregene A, Newman LA. Breast cancer in sub-Saharan Africa: How does it relate to breast cancer in African-American women? Cancer 2005;103: 1540–50.
32. Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. J Natl Cancer Inst 2006;98:1215–26.
33. Pankratz VS, Degnim AC, Frank RD, Frost MH, Visscher DW, Vierkant RA, et al. Model for individualized prediction of breast cancer risk after a benign breast biopsy. J Clin Oncol 2015;33:923–9.