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

Shengfeng Wang1,2, Temidayo Ogundiran3, Adeyinka Ademola3, Oluwasola A. Olayiwola4, Adewunmi Adeoye4, Adenike Sofoluwe5, Imran Morhason-Bello6, Stella Odedina6, Imaria Agwai6, Clement Adebamowo7, Millicent Obajimi5, Oladosu Ojengbede6, Olufunmilayo I. Olopade2, and Dezheng Huo2,8

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, especially when the incidence rates and survival rates of breast cancer are 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...
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).

**Absolute 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.

**Comparison with existing models**

The two Gail models and the BWHS model (2, 3, 21) were applied to the study sample (Supplementary Fig. S1), except that we applied age-specific incidence rates and mortality rates from Nigeria to ensure consistency 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
Wang et al.

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

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Case, N (%)</th>
<th>Control, N (%)</th>
<th>Age-adjusted odds ratio (95% confidence interval)</th>
<th>P</th>
<th>P_Brand</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1811</td>
<td>2225</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>25–29.9</td>
<td>59 (3.3)</td>
<td>212 (9.5)</td>
<td>0.30 (0.21–0.47)</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>30–34.9</td>
<td>151 (8.3)</td>
<td>314 (14.3)</td>
<td>0.50 (0.40–0.66)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>35–39.9</td>
<td>273 (15.1)</td>
<td>327 (14.7)</td>
<td>0.89 (0.71–1.12)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>40–44.9</td>
<td>297 (16.4)</td>
<td>317 (14.3)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45–49.9</td>
<td>306 (16.9)</td>
<td>263 (11.8)</td>
<td>1.24 (0.99–1.56)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>50–54.9</td>
<td>254 (14.0)</td>
<td>222 (10.0)</td>
<td>1.22 (0.96–1.55)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>55–59.9</td>
<td>176 (9.7)</td>
<td>145 (6.5)</td>
<td>1.30 (0.99–1.70)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Age at menarche</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥17</td>
<td>421 (23.3)</td>
<td>564 (25.4)</td>
<td>1.00</td>
<td>0.017</td>
<td>0.013</td>
</tr>
<tr>
<td>16–19.9</td>
<td>255 (14.3)</td>
<td>343 (15.4)</td>
<td>1.02 (0.83–1.26)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>15–19.9</td>
<td>515 (28.4)</td>
<td>538 (24.2)</td>
<td>1.33 (1.11–1.59)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>14–14.9</td>
<td>255 (14.3)</td>
<td>341 (15.3)</td>
<td>1.09 (0.88–1.35)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>13–13.9</td>
<td>186 (10.3)</td>
<td>219 (9.8)</td>
<td>1.27 (1.00–1.62)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>12–12.9</td>
<td>151 (8.3)</td>
<td>171 (7.7)</td>
<td>1.30 (1.00–1.69)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>8–11.9</td>
<td>28 (1.6)</td>
<td>49 (2.2)</td>
<td>0.86 (0.52–1.42)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Before first live birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>155 (8.6)</td>
<td>246 (11.1)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>126 (7.0)</td>
<td>209 (9.4)</td>
<td>0.57 (0.40–0.80)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>198 (10.9)</td>
<td>268 (12.0)</td>
<td>0.48 (0.35–0.66)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>287 (15.8)</td>
<td>326 (14.7)</td>
<td>0.49 (0.36–0.67)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>297 (16.4)</td>
<td>393 (17.7)</td>
<td>0.36 (0.26–0.49)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>306 (16.9)</td>
<td>325 (14.6)</td>
<td>0.41 (0.30–0.56)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>213 (11.8)</td>
<td>206 (9.3)</td>
<td>0.41 (0.29–0.58)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>119 (6.6)</td>
<td>124 (5.6)</td>
<td>0.37 (0.25–0.54)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>≥8</td>
<td>110 (6.1)</td>
<td>128 (5.8)</td>
<td>0.32 (0.22–0.47)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Age at menopause</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>26 (3.2)</td>
<td>12 (1.9)</td>
<td>0.57 (0.20–1.63)</td>
<td></td>
<td>0.536</td>
</tr>
<tr>
<td>35–44.9</td>
<td>162 (10.5)</td>
<td>96 (15.9)</td>
<td>1.04 (0.75–1.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45–54.9</td>
<td>521 (64.80)</td>
<td>444 (70.25)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥55</td>
<td>95 (11.82)</td>
<td>80 (12.66)</td>
<td>1.19 (0.84–1.69)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Total months of breastfeeding</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>&lt;12</td>
<td>225 (12.4)</td>
<td>342 (15.4)</td>
<td>1.00</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>12–23.9</td>
<td>155 (8.6)</td>
<td>193 (8.7)</td>
<td>0.71 (0.53–0.96)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>24–35.9</td>
<td>224 (12.4)</td>
<td>263 (11.8)</td>
<td>0.56 (0.42–0.71)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>36–47.9</td>
<td>207 (11.5)</td>
<td>224 (10.1)</td>
<td>0.57 (0.43–0.77)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>48–59.9</td>
<td>215 (11.9)</td>
<td>263 (11.8)</td>
<td>0.47 (0.35–0.63)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>60–71.9</td>
<td>163 (9.0)</td>
<td>225 (10.1)</td>
<td>0.37 (0.40–0.72)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>72–83.9</td>
<td>215 (11.9)</td>
<td>203 (9.1)</td>
<td>0.53 (0.40–0.72)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>84–95.9</td>
<td>82 (4.5)</td>
<td>103 (4.6)</td>
<td>0.38 (0.26–0.55)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>≥96</td>
<td>325 (18.0)</td>
<td>409 (18.4)</td>
<td>0.35 (0.26–0.46)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Height in cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>&lt;150</td>
<td>85 (4.7)</td>
<td>136 (6.1)</td>
<td>0.91 (0.68–1.23)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>150–159</td>
<td>654 (36.1)</td>
<td>1067 (48.0)</td>
<td>1.00</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>160–169</td>
<td>819 (45.2)</td>
<td>885 (39.8)</td>
<td>1.58 (1.37–1.82)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>≥170</td>
<td>253 (14.0)</td>
<td>137 (6.2)</td>
<td>3.26 (2.57–4.15)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Body mass index in kg/m²</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before menopause</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.001</td>
</tr>
<tr>
<td>18.5–24.9</td>
<td>425 (44.1)</td>
<td>700 (44.8)</td>
<td>1.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18.5</td>
<td>67 (7.0)</td>
<td>95 (6.1)</td>
<td>1.43 (1.00–2.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25–29.9</td>
<td>276 (28.6)</td>
<td>432 (27.6)</td>
<td>0.89 (0.73–1.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥30</td>
<td>156 (20.3)</td>
<td>358 (21.6)</td>
<td>0.71 (0.57–0.89)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on the following page)
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, in Table 1, Nigerian women with breast cancer were characterized by late age at menarche, higher rates of a positive family history, and predicted 5-year risk, over the whole range of probabilities. 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 = 0.53), 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 1. Distribution of breast cancer risk factors in cases and controls in the NBCS (1998-2015) (Cont’d)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Case, N (%)</th>
<th>Control, N (%)</th>
<th>Age-adjusted odds ratio (95% confidence interval)</th>
<th>P</th>
<th>P_Hurd</th>
</tr>
</thead>
<tbody>
<tr>
<td>After menopause</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.5–24.9</td>
<td>327 (58.6)</td>
<td>191 (28.9)</td>
<td>1.00</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>&lt;18.5</td>
<td>55 (6.5)</td>
<td>28 (4.2)</td>
<td>1.12 (0.68–1.85)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25–29.9</td>
<td>242 (28.6)</td>
<td>248 (37.6)</td>
<td>0.57 (0.44–0.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥30</td>
<td>223 (26.3)</td>
<td>193 (29.2)</td>
<td>0.68 (0.52–0.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waist–hip ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.80</td>
<td>453 (25.4)</td>
<td>599 (27.5)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.80–0.84</td>
<td>458 (25.7)</td>
<td>528 (24.3)</td>
<td>1.04 (0.87–1.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥0.85</td>
<td>871 (48.9)</td>
<td>1050 (48.2)</td>
<td>0.87 (0.74–1.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral contraceptive</td>
<td>512 (28.3)</td>
<td>741 (33.4)</td>
<td>0.75 (0.65–0.86)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>179 (9.9)</td>
<td>109 (4.9)</td>
<td>1.95 (1.52–2.52)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Study periods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998–2004</td>
<td>619 (34.2)</td>
<td>732 (32.9)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005–2010</td>
<td>588 (32.5)</td>
<td>770 (34.6)</td>
<td>0.79 (0.67–0.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011–2015</td>
<td>604 (33.4)</td>
<td>723 (32.5)</td>
<td>0.93 (0.80–1.09)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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_Hurd value were all adjusted by age.
Based on 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 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 progesterin use, and height (5). Of the 9 risk factors used in the BWHS 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 insignificance in our previous studies (6, 13, 17). Age at first live birth was not in our model due to its insignificance in our previous studies (6, 13, 17).

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 black population contains five variables: age, age at menarche, number of previous breast biopsies, age at first live birth, and number of first-degree relatives with breast cancer (2), whereas the Gail model for the black population contains five variables: age, age at menarche, number of previous breast biopsies, 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 progesterin use, and height (5). Of the 9 risk factors used in the BWHS 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 insignificance in our previous studies (6, 13, 17). Age at first live birth was not in our model due to its insignificance in our previous studies (6, 13, 17).
The discriminating accuracy of the four absolute risk prediction models was only evaluated in the overlap of their applicable age range (35–73 individuals were either younger than 20 years old (n = 21) or older than 75 years old (n = 52)).

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. To improve the existing model for African Americans, such as integrating reproductive factors and genomic risk factors (31). 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.
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
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): T. Ogundiran, A. Ademola, O.A. Olayiwola, A. Adeoye,

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. Mortahson-Bello, C. Adebamowo, O. Ojengbade, A. Sofielswe, 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. Ojengbade, O.I. Olopade, D. Huo

Study supervision: T. Ogundiran, A. Ademola, I. Mortahson-Bello, C. Adebamowo, O. Ojengbade, 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 NIH RO1CA9805 (to O.I. Olopade), PS0CA125183 (to O.I. Olopade), and U10CA161032 (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 (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

Cancer Epidemiology, Biomarkers & Prevention

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

Shengfeng Wang, Temidayo Ogundiran, Adeyinka Ademola, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-17-1128

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2018/05/04/1055-9965.EPI-17-1128.DC2

Cited articles
This article cites 33 articles, 4 of which you can access for free at:
http://cebp.aacrjournals.org/content/27/6/636.full#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cebp.aacrjournals.org/content/27/6/636.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.