

ARTICLE

Practical Problems With Clinical Guidelines for Breast Cancer Prevention Based on Remaining Lifetime Risk

Anne S. Quante, Alice S. Whittemore, Tom Shriver, John L. Hopper, Konstantin Strauch, Mary Beth Terry

Affiliations of authors: Joseph L. Mailman School of Public Health, Columbia University, New York, NY (ASQ, TS, MBT); Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany (ASQ, KS); Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany (ASQ, KS); Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA (ASW); Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Carlton, Victoria, Australia (JLH); Department of Epidemiology and Institute of Health and Environment, School of Public Health, Seoul National University, Seoul, Korea (JLH); Herbert Irving Comprehensive Cancer Center, Columbia Medical Center, New York, NY (MBT).

Corresponding Author: Mary Beth Terry, PhD, Department of Epidemiology, Joseph L. Mailman School of Public Health, Columbia University, 722 West 168th Street, 724A, New York, NY 10032 (e-mail: mt146@columbia.edu).

Abstract

Background: Clinical guidelines for breast cancer chemoprevention and MRI screening involve estimates of remaining lifetime risk (RLR); in the United States, women with an RLR of 20% or higher meet “high-risk” criteria for MRI screening.

Methods: We prospectively followed 1764 women without breast cancer to compare the RLRs and 10-year risks assigned by the risk models International Breast Cancer Intervention Study (IBIS) and Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) and to compare both sets of model-assigned 10-year risks to subsequent incidence of breast cancer in the cohort. We used chi-square statistics to assess calibration and the area under the receiver operating characteristic curve (AUC) to assess discrimination. All statistical tests are two-sided.

Results: The models classified different proportions of women as high-risk (IBIS = 59.3% vs BOADICEA = 20.1%) using the RLR threshold of 20%. The difference was smaller (IBIS = 52.9% vs BOADICEA = 43.2%) using a 10-year risk threshold of 3.34%. IBIS risks (mean = 4.9%) were better calibrated to observed breast cancer incidence (5.2%, 95% confidence interval (CI) = 4.2% to 6.4%) than were those of BOADICEA (mean = 3.7%) overall and within quartiles of model risk ($P = .20$ by IBIS and $P = .07$ by BOADICEA). Both models gave similar discrimination, with AUCs of 0.67 (95% CI = 0.61 to 0.73) using IBIS and 0.68 (95% CI = 0.62 to 0.74) using BOADICEA. Model sensitivities at thresholds for a 20% false-positive rate were also similar, with 41.8% using IBIS and 38.0% using BOADICEA.

Conclusion: RLR-based guidelines for high-risk women are limited by discordance between commonly used risk models. Guidelines based on short-term risks would be more useful, as models are generally developed and validated under a short fixed time horizon (≤ 10 years).

Breast cancer risk models, which estimate a woman’s absolute risk of developing breast cancer either for a fixed horizon (eg, five or 10 years) or for a woman’s remaining lifetime, are used in clinical guidelines for decisions about MRI screening and risk-reducing surgeries. For example, the US National Comprehensive Cancer Network (NCCN) guidelines (1) recommend consideration of risk-reducing strategies for women

over the age of 35 years whose five-year invasive breast cancer risk as determined by the Breast Cancer Risk Assessment Tool (BCRAT) (2–4) is 1.67% or higher. Furthermore, consideration of annual mammograms and MRI starting at age 30 years is recommended for women with remaining lifetime risks (RLRs) of 20% or higher (as determined by risk models that are largely dependent on family history) (1). However, the clinical guidelines do not

Received: December 9, 2014; Revised: March 12, 2015; Accepted: April 8, 2015

© The Author 2015. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com.

suggest which risk model to use, and model predictions can differ depending on the risk factors they include and whether or not they consider the competing risk of death. In addition, using RLRs as basis for screening recommendations is problematic; for example, a young woman can have small short-term risk but large RLR—based on her age alone.

Breast cancer risk models (reviewed by Meads et al. [5]) differ in the risk factors they include and in the way they handle the competing risk of death. In general, the models are designed for two groups: 1) women without a predisposing mutation or strong family history and 2) women at higher risk because of personal or family history of breast or ovarian cancer (6). Models of the first type (e.g., BRCAT [2]) use only limited information on family history (e.g. number of first-degree relatives with breast cancer), while those of the second type use more detailed information (e.g. ages at onset of relatives' cancers, and/or carriage of specific breast cancer susceptibility alleles). Underlying assumptions about the nature of genetic risks differ among the models of the second type (e.g., the Claus model [7] assumes one risk locus, the International Breast Cancer Intervention Study (IBIS) model [8] and the BRCAPRO model [9] assume two risk loci, and the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm [10–12] assumes an additional familial/polygenic component).

The performance of a risk model when applied to a cohort of unaffected women is evaluated with respect to two characteristics: its calibration, which reflects how well the model's assigned risks agree with the actual observed incidence overall and within subgroups of the cohort, and its discrimination, which reflects its ability to distinguish between those who do and do not develop breast cancer (13). There are limited comparative evaluations of existing breast cancer risk models as applied to women at higher breast cancer risk (14). Amir and colleagues (15) compared the risks assigned by five models to observed incidence in a cohort of 1933 women at higher risk, 52 of whom developed breast cancer. They found that the IBIS model performed best with respect both to overall calibration, with an expected (E)-to-observed (O) ratio of 0.81 (95% confidence interval [CI] = 0.62 to 1.08) and with respect to discrimination, with an area under the receiver operating characteristic curve (AUC), of 0.76 (95% CI = 0.70 to 0.82) (15). Laitman et al. (16) compared the overall calibration of IBIS and BOADICEA predictions to incidence in a small cohort of 358 Israeli women without BRCA1/2 mutations, 15 of whom developed breast cancer, preventing precise estimates of performance measures (IBIS: O/E = 0.80, 95% CI = 0.48 to 1.33; BOADICEA: O/E = 0.52, 95% CI = 0.32 to 0.87). The authors did not evaluate discrimination. The BOADICEA model has been evaluated for Australian women and found to be well calibrated overall (E/O = 0.92, 95% CI = 0.76 to 1.10) and modestly discriminatory (AUC = 0.70, 95% CI = 0.66 to 0.75) (17). BOADICEA has also been evaluated for Swedish women where the ratio of observed to expected was 1.41 (95% CI = 0.91 to 2.08) and the AUC was 0.62 (95% CI = 0.52 to 0.73) (18). We have shown previously that the IBIS model outperformed the BCRAT model in our cohort (6). However, the IBIS and BOADICEA models have not been compared for US women at higher risk. As both models are commonly used in clinical practice to identify women eligible for MRI screening according to the NCCN guidelines (1), it is essential to understand how these models perform in high-risk cohorts. Therefore, using a cohort of higher risk women from New York City, we compared several measures of calibration and discrimination of IBIS and BOADICEA, as applied to the development of breast cancer within 10 years of risk assignment. We used a 10-year rather than lifetime horizon for two

reasons: 1) the two models define remaining lifetime differently (until age 80 years for BOADICEA and 85 years for IBIS) and thus are not comparable; and 2) existing cohorts lack the long-term observation needed to evaluate RLR predictions.

Methods

Study Sample

We studied women with no history of invasive or in situ breast cancer at the time of recruitment to the New York site of the Breast Cancer Family Registry (BCFR) (for details see [6,19]). We restricted eligibility to women with at least one subsequent update on cancer and vital status, and who at cohort entry were age 20 to 70 years and had no prior history of bilateral prophylactic mastectomy. We excluded women with a personal history of ovarian or pancreatic cancer.

At recruitment, each eligible patient completed a questionnaire that included information on demographics, lifestyle and environmental factors, past surgeries, and family history of cancer (19). We actively followed participants for subsequent information on cancer incidence and vital status and attempted to verify cancers through pathology reviews and reports and medical records. All cohort participants provided written informed consent, and the study was approved by the relevant local ethics committees.

Risk Models

Both the IBIS and BOADICEA models specify a woman's probability of developing breast cancer during a subsequent period after recruitment, called either her RLR or her 10-year risk. The two models define a woman's remaining lifetime as the years between her age at recruitment and either age 85 years (IBIS) or 80 years (BOADICEA). We used the software packages IBIS v7 (<http://www.ems-trials.org/riskevaluator/>) (8) and BOADICEA (<https://pluto.srl.cam.ac.uk/cgi-bin/bd3/v3/bd.cgi>) (10–12) to assign RLRs and 10-year risks. The IBIS and BOADICEA models assign risks without adjusting for mortality before breast cancer development, so they do not account for the competing risk of death. (A recent upgrade of the IBIS software now allows optional inclusion of population death rates; however, we used the default setting to make IBIS comparable with BOADICEA, which does not include death rates.)

The breast cancer hazard rate assumed by the IBIS model depends on several nongenetic risk factors, plus BRCA1 and BRCA2 mutation status, with residual familial clustering modeled as cosegregation of a single latent, dominantly acting gene. The nongenetic risk factors are age at menarche, parity, age at first live birth, age at menopause, prior use of hormone replacement therapy, history of hyperplasia/atypical hyperplasia, history of lobular carcinoma in situ, height, and body mass index (8). The breast cancer hazard rate assumed by the BOADICEA model includes BRCA1 and BRCA2 mutation status, with residual familial breast cancer clustering modeled as a polygenic component using the hypergeometric distribution, but does not include nongenetic risk factors (10–12).

Statistical Analysis

We evaluated the IBIS and BOADICEA model predictions by comparing their 10-year assigned risks to observed breast cancer incidence within 10 years of recruitment. To assess calibration,

we obtained Kaplan-Meier estimates of cumulative breast cancer incidence at 10 years postrecruitment for the entire cohort and within each quartile of risk predicted by each model. Because neither the IBIS v7 nor the BOADICEA model considers death without breast cancer to be a competing risk, we regarded women who died without breast cancer within 10 years of recruitment as censored (rather than unaffected with breast cancer at time of death [20]). To evaluate model calibration, we compared mean model-assigned risk to observed breast cancer incidence in the entire cohort and in each of the four assigned risk quartiles, using a chi-squared goodness-of-fit statistic (21). This goodness-of-fit statistic is based on a sum of squared differences between the estimated 10-year breast cancer probabilities and mean model-assigned risks within the quartiles of assigned risk (21). To evaluate discrimination, we computed the overall AUC and we plotted case risk percentiles (CRPs, also called standardized placement values) (21) for the subjects who developed breast cancer within 10 years of recruitment (case patients). (For a given model, a case patient's CRP is the percentile of her assigned risk in the distribution of all risks assigned to women free of breast cancer at 10 years postrecruitment.) In particular, we compared the two models' sensitivities when their thresholds were chosen to give the same specificity (eg, 80%). We estimated a model's sensitivity at its 80% specificity threshold as the proportion of case patients with CRPs exceeding the threshold. We estimated the concordance measure C that is applied to a binary outcome (positive, ie, breast cancer occurrence within 10 years, or negative, ie, 10-year survival without breast cancer). C is the probability that, given a randomly selected pair of outcome-discordant patients, the model assigns a higher 10-year outcome probability to the outcome-positive patient than to the outcome-negative one. To estimate C , we removed all subjects last observed without breast cancer before 10 years from baseline, which will give unbiased estimates of this concordance measure, provided the censoring mechanism is independent of

breast cancer risk (22). The C -statistic proposed by Harrell et al. (23) and estimated by Uno et al. (24) represents a different concordance measure C^* , defined as the probability that a randomly selected pair of patients who develop breast cancer at times $T_1 < T_2$ are assigned risks r_1 and r_2 with $r_1 > r_2$. This measure is inappropriate for evaluating the models needed for clinical guidelines, which are tied to outcome development within a specific time period.

Although the cohort contains pairs of first-degree relatives whose breast cancer risks are correlated because of unmeasured genetic and nongenetic familial factors, we ignored this in computing test statistics and confidence intervals, because the proportion of such pairs was small. All calibration and discrimination statistics were computed using the freely available software RMAP (<http://www.stanford.edu/~ggong/rmap/index.html>). All statistical tests were two-sided, and a P value of less than .05 was considered statistically significant.

Results

A total of 1764 subjects met the eligibility criteria, of whom 95 developed breast cancer subsequent to recruitment. Of these, 79 developed the disease within 10 years of enrollment, 1099 were known to still be unaffected by breast cancer after 10 years, 41 died without being diagnosed with breast cancer, and 545 were last observed before completion of 10 years of follow-up (for additional descriptive data on our cohort, see [6]).

RLR vs 10-Year Risks

Figure 1 shows scatterplots of patients' RLRs (Figure 1A) and 10-year risks (Figure 1B), as assigned by IBIS and BOADICEA, with the women who developed breast cancer within the risk period indicated in red. Figure 1A shows that IBIS risks generally exceeded those of

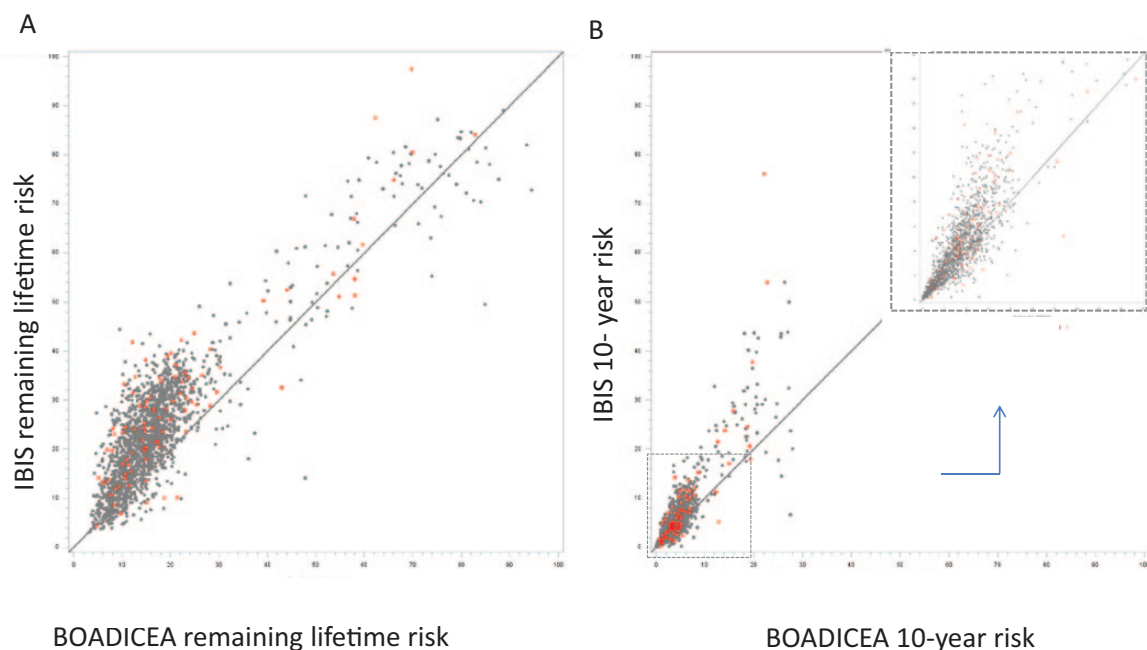


Figure 1. Remaining lifetime risks (RLR) vs 10-year risks. **A)** Scatterplot of International Breast Cancer Intervention Study (IBIS) and Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) RLRs. The points denote the 1764 patients' remaining lifetime risks as assigned by IBIS and BOADICEA. The women who developed breast cancer at any time are highlighted in red ($n = 95$). **B)** Scatterplot of IBIS and BOADICEA 10-year risks. The points denote the 1764 patients' 10-year risks as assigned by IBIS and BOADICEA. The women who developed breast cancer at within 10 years are highlighted in red ($n = 79$).

BOADICEA, with the mean difference in RLR between IBIS and BOADICEA being 6.7% (95% CI = 6.4% to 7.0%). In contrast, Figure 1B shows greater model similarity in the 10-year risks, with the mean IBIS-BOADICEA difference being 1.3% (95% CI = 1.1% to 1.4%).

Table 1 shows the joint distribution of women classified by the two models as high and low risk based on their remaining and 10-year assigned risks, using the NCCN guidelines to define “high risk” (1). Note that when high risk is defined as an RLR of 20% or higher, 705 women (40.0%) were classified discordantly as high risk by IBIS and low risk by BOADICEA (high IBIS/low BOADICEA), whereas only 13 (0.7%) were classified low IBIS/high BOADICEA. Overall, IBIS assigned 1047 (59.3%) patients as high risk, compared with 355 (20.1%) by BOADICEA. These differences could in part be because of the higher RLR upper age bound used by IBIS (85 years) compared with BOADICEA (80 years). We found the discordance was less when we defined high risk by a 10-year risk of 3.34% or higher (which is roughly equivalent to the NCCN five-year risk of 1.67%): Only 16.0% of women were discordantly classified, with 227 (12.9%) women classified as high IBIS/low BOADICEA and 55 (3.1%) classified as high BOADICEA/low IBIS. Overall, for this sample of women oversampled for having a family history of breast cancer, IBIS assigned 934 (52.9%) women as high risk, compared with 762 (43.2%) for BOADICEA.

Calibration of 10-Year Risks

We cannot address which of the two models gave a better fit (calibration) to the observed breast cancer incidence in the cohort using RLR because it would require observing subjects for their remaining lifetimes. We estimated the 10-year cumulative breast cancer probability to be 5.2% with a 95% confidence interval of 4.2% to 6.4%, compared with the mean of the predicted 10-year assigned risks of 4.9% by IBIS and 3.7% by BOADICEA. When we excluded BRCA1 and BRCA2 carriers ($n = 83$), we estimated the cumulative breast cancer probability to be 4.8% (95% CI = 3.8% to 6.0%), which exceeds the mean 10-year assigned risk of both models (3.9% for IBIS and 3.0% for BOADICEA (data not shown). We also evaluated model calibration by comparing observed incidence with mean model-assigned risks for subgroups of women defined by quartiles of assigned risk. The four points in each panel of Figure 2 give observed risk (vertical

coordinates) and mean assigned risk (horizontal coordinates) for each of the four quartiles of risk determined by IBIS (Figure 2A) and BOADICEA (Figure 2B). The vertical bars represent 95% confidence intervals for the observed risks. Because perfect agreement between observed and assigned risks corresponds to points on the diagonal line, the figure shows better agreement for IBIS than BOADICEA, all of whose assigned risks were lower than those observed. These graphical observations were confirmed by the models' goodness-of-fit statistics, which are based on the squared vertical distances from the points to the diagonal line (IBIS $\chi^2_4 = 6.0$, $P = .20$; BOADICEA $\chi^2_4 = 8.8$, $P = .07$) (Figure 2). When we restricted to BRCA1/2 noncarriers, we found an IBIS χ^2_4 of 4.2 ($P = .37$), BOADICEA χ^2_4 of 9.7 ($P = .04$) (data not shown). We did not observe any differences in calibration by age group (<50 years vs ≥ 50 years) (data not shown).

Discrimination of 10-Year Risks

We compared the models' receiver operating characteristic (ROC) curves and case risk percentiles (CRPs), omitting subjects who were last observed free of breast cancer before 10 years. Figure 3 shows similar discrimination by the two models, with area under the ROC (AUC) of 0.67 (95% CI = 0.61 to 0.73) for IBIS and 0.68 (95% CI = 0.62 to 0.74) for BOADICEA. IBIS AUC was little affected when nongenetic covariates were deleted from the model, with an AUC of 0.66 (95% CI = 0.61 to 0.72) (data not shown).

To supplement the AUC-based comparison of IBIS and BOADICEA discrimination, we also compared the IBIS and BOADICEA case risk percentiles (25) of the 79 women who developed breast cancer within 10 years of recruitment. Figure 4 shows the CRPs from the IBIS model compared with those from the BOADICEA model. Points above the diagonal line represent case patients more likely to be classified as high risk by IBIS than by BOADICEA. The vertical dashed line gives the threshold corresponding to 80% specificity for BOADICEA, while the horizontal line at 80% marks the 80th percentile of the risks assigned to non-case patients by BOADICEA, while the horizontal dashed line gives the corresponding 80% specificity threshold for IBIS. These lines allow visual comparison of the two model sensitivities when each model's high-risk threshold is determined to yield a common 80% specificity. As can be seen in the figure, this comparison gives a sensitivity of 33 of 79, or 41.8% for the IBIS model, and of 30 of 79, or 38.0% for the BOADICEA model. The figure also shows that seven case patients were correctly deemed high risk by IBIS but not by BOADICEA, while four cases were correctly identified as high risk by BOADICEA but not by IBIS.

We also assessed sensitivity and specificity associated with a common 10-year threshold of 3.34% (Table 2). IBIS was more sensitive than BOADICEA (77.2% vs 68.3% true-positive rate), but less specific (54.4% vs 43.7% false-positive rate) (Table 2). These differences also held for the subgroup of women younger than 50 years at recruitment (68.9% true-positive rate for IBIS vs 51.1% for BOADICEA and 39.5% false-positive rate for IBIS vs 25.3% for BOADICEA) (data not shown). In older women, however, the models showed similar sensitivity and specificity (88.2% vs 91.1% true-positive rate and 85.4% vs 81.8% false-positive rate) (data not shown).

Discussion

The NCCN guidelines (1) advise that high-risk women consider risk reduction strategies and annual MRI and mammography starting at the age of 30 years. The NCCN definition of high risk includes women who have a five-year risk of invasive breast cancer of 1.67% or more, or a remaining lifetime risk (RLR) of 20%

Table 1. Classification of 1764 study subjects into high- and low-risk groups by the IBIS and BOADICEA models, using remaining lifetime risk and 10-year risk thresholds

BOADICEA	Remaining lifetime risk*		
	Low†	High	Total
Low	704	705	1409
High	13	342	355
Total	717	1047	1764
	10-year risk		
	Low‡	High	Total
Low	775	227	1002
High	55	707	762
Total	830	934	1764

* Defined as risk from assessment until age 80 years (BOADICEA) or 85 years (IBIS). BOADICEA = Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; IBIS = International Breast Cancer Intervention Study.

† Remaining lifetime risks of 20% or greater.

‡ Ten-year risks of 3.34% or greater.

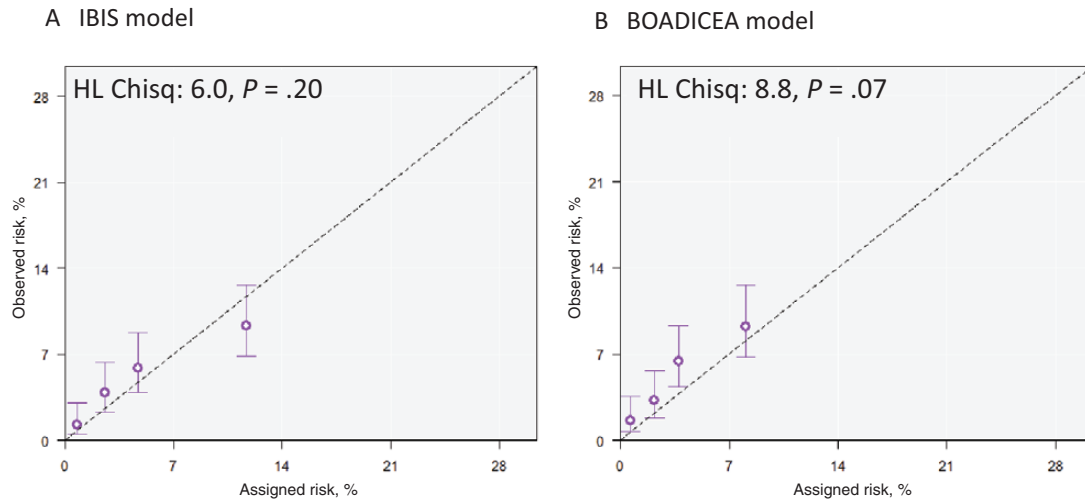


Figure 2. Calibration of International Breast Cancer Intervention Study (IBIS) and Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) models. The coordinates on the x-axis represent the mean 10-year assigned risks of the IBIS model (left panel) and the BOADICEA model (right panel) within quartiles of assigned risk. The coordinates on the y-axis represent quartile-specific estimates of 10-year breast cancer probabilities based on the women's observed breast cancer status, and the bars denote 95% confidence intervals for the observed risk. The P value was calculated using a chi-squared goodness-of-fit statistic.

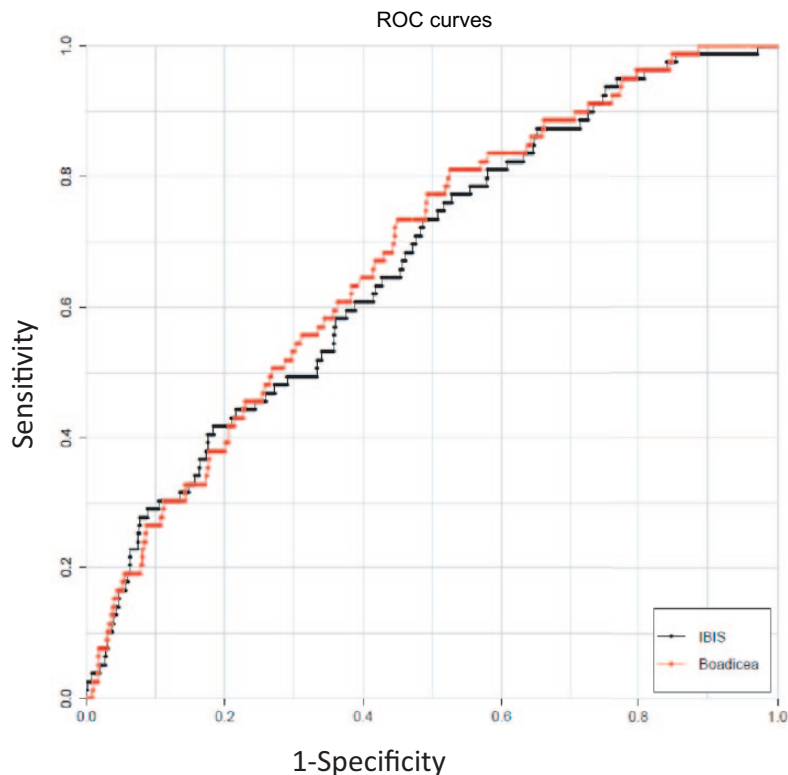


Figure 3. Discrimination of International Breast Cancer Intervention Study (IBIS) and Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) models. The black line shows the receiver operating characteristic (ROC) curves from IBIS, and the red line shows the receiver operating characteristic (ROC) curve from BOADICEA.

or more. However, we found that implementing these RLR-based guidelines is problematic for two reasons. First, the models disagree on the definition of a woman's remaining lifetime with IBIS using age 85 years and BOADICEA using age 80 years. We found that the RLRs assigned by IBIS and BOADICEA differ substantially and that these differences were reduced when comparing the models' predicted 10-year risks. Second, it is not feasible to

assess model RLR predictions, because we cannot observe breast cancer outcomes during the remaining lifetimes of cohort subjects. In fact, risk models are generally developed and validated over a fixed horizon (eg, five or 10 years of follow-up), making estimates of RLR less precise. In particular, we cannot assess model specificity, as we rarely have cohort patients followed and deemed breast cancer-free at age 80 or 85 years. When using an

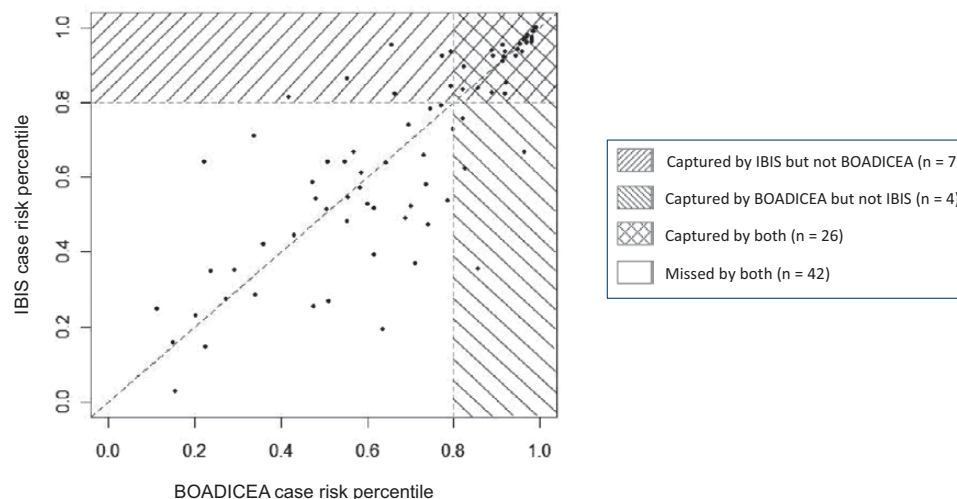


Figure 4. Percentiles of 10-year case patients (as measured by International Breast Cancer Intervention Study [IBIS] vs Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm [BOADICEA]) in the distribution of control patients' risks. A case patient's risk percentile is the proportion of all control patients with assigned risk less than hers. The vertical line at 80% marks the 80th percentile of the risks assigned to control patients by BOADICEA, and similarly the horizontal line marks the 80th percentile of the risks assigned to control patients by IBIS. By using as threshold the 80th percentile of a model's control risks, we are requiring both BOADICEA and IBIS to have the same false-positive rate of 20%.

Table 2. Comparing sensitivity and specificity between breast cancer prediction models

Model	Test result*	Cancer			Sensitivity	Specificity
		Yes	No	Total		
IBIS	Positive	61	598	659	77.2%	45.6%
	Negative	18	501	519		
	Total	79	1099	1178†		
BOADICEA	Positive	54	480	534	68.3%	56.3%
	Negative	25	619	644		
	Total	79	1099	1178†		

* Test result classifying women as positive/negative for 10-year risk of 3.34% or greater (the test classification). BOADICEA = Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; IBIS = International Breast Cancer Intervention Study.

† Women who had a follow-up time less than 10 years were excluded.

RLR threshold of 20%, the models classified different proportions of women as high risk (59.3% using IBIS, 20.1% using BOADICEA). Thus, if one wants to perform more MRIs, it seems that one should use the IBIS model. However, using the 10-year threshold of 3.34%, the difference was smaller (52.9% using IBIS, 43.2% using BOADICEA).

When using clinical risk thresholds to make decisions about screening intensity, it is important to have accurate assessment of women's absolute breast cancer risks. We compared the risk distribution, calibration and discrimination of the IBIS and BOADICEA models, both of which are used in cancer risk clinics and primary care settings to assess women's risks and decide preventive strategies. Both models use extensive information on family history as well as any known mutation carrier status for BRCA1 or BRCA2, but IBIS also uses information on nongenetic risk factors. While IBIS (15) and BOADICEA (17) have each been evaluated prospectively, there is a general paucity of prospective data comparing multiple risk models, and to our knowledge this is the first direct comparison of these

two models when applied to the same data from a prospective cohort of US high-risk women.

Our cohort is enriched for women with a family history and therefore has greater power than other cohorts to examine breast cancer risk across the spectrum. Nevertheless, our inferences are limited by the relatively small number of participants and outcomes. In particular, there were too few women to compare performance across racial and ethnic subgroups. We plan to extend the present analysis to larger cohorts and compare clinical model performance across different racial and ethnic subgroups.

BCRA1 and BRCA2 mutations confer a high risk of 40% to 65% for breast cancer, and an intensified surveillance or risk-reducing prophylactic surgery is highly recommended to mutation carriers. However, the discrepancy found between the two models was not due to the mutation carriers (as both models classify them as high risk and carriers are only a small percent of the cohort). Decisions for all noncarriers are based on individual risk profiles as determined by genetic and nongenetic risk factors. Therefore, the choice of a particular prediction model is an important aspect of risk assessment. We found that: 1) the two models yield vastly different estimates of which women would be high risk (40.0% of women were discordant based on the estimates of RLR of 20% or higher); 2) the difference was less when using the 10-year threshold of 3.34%; and 3) IBIS was more sensitive than BOADICEA but less specific when using the 10-year threshold of 3.34%. These findings indicate the limitations of clinical guidelines based on the concept of "remaining lifetime risk," which may be defined differently by different risk models, and which is difficult to evaluate empirically. For example, BOADICEA and IBIS have been evaluated only for five to 16 years and not over the lifetime (15–18). We recommend instead that practitioners use time periods of briefer duration, such as five- or 10-year risks. Nevertheless, both IBIS and BOADICEA models still underestimated 10-year breast cancer risks in our cohort, with the discrepancies larger for BOADICEA than IBIS. The data suggest that the improved IBIS calibration reflects its inclusion of nongenetic risk factors. These observations need replication

in other large cohorts spanning a broad range of risks, including a substantial number of women at high risk. Thus, a greater recognition of the limitations of RLRs estimates is needed in the clinic. Increased clinical use of shorter fixed time horizons when conveying risk will be particularly important to improve the validity of risk assessment.

Funding

This work was supported by the following grants UM1CA164920, R01 CA 159868, and R01 CA094069.

Notes

The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US government or the BCFR.

References

1. Bevers TB, Anderson BO, Bonaccio E, et al. NCCN clinical practice guidelines in oncology: breast cancer screening and diagnosis. *J Natl Compr Canc Netw*. 2009;7(10):1060–1096.
2. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81(24):1879–1886.
3. Gail MH, Costantino JP, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst*. 2007;99(23):1782–1792.
4. Matsuno RK, Costantino JP, Ziegler RG, et al. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. *J Natl Cancer Inst*. 2011;103(12):951–961.
5. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. *Breast Cancer Res Treat*. 2012;132(2):365–377.
6. Quante AS, Whittemore AS, Shriver T, et al. Breast cancer risk assessment across the risk continuum: genetic and nongenetic risk factors contributing to differential model performance. *Breast Cancer Res*. 2012;14(6):R144.
7. Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer*. 1994;73(3):643–651.
8. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med*. 2004;23(7):1111–1130.
9. Parmigiani G, Berry D, Aguilar O. Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *Am J Hum Genet*. 1998;62(1):145–158.
10. Antoniou AC, Cunningham AP, Peto J, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer*. 2008;98(8):1457–1466.
11. Antoniou AC, Pharoah PD, McMullan G, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer*. 2002;86(1):76–83.
12. Antoniou AC, Pharoah PP, Smith P, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br J Cancer*. 2004;91(8):1580–1590.
13. Whittemore AS. Evaluating health risk models. *Stat Med*. 2010;29(23):2438–2452.
14. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, et al. Risk prediction models of breast cancer: a systematic review of model performances. *Breast Cancer Res Treat*. 2012;133(1):1–10.
15. Amir E, Evans DG, Shenton A, et al. Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. *J Med Genet*. 2003;40(11):807–814.
16. Laitman Y, Simeonov M, Keinan-Boker L, et al. Breast cancer risk prediction accuracy in Jewish Israeli high-risk women using the BOADICEA and IBIS risk models. *Genet Res (Camb)*. 2013;95(6):174–177.
17. MacInnis RJ, Bickerstaffe A, Apicella C, et al. Prospective validation of the breast cancer risk prediction model BOADICEA and a batch-mode version BOADICEACentre. *Br J Cancer*. 2013;109(5):1296–1301.
18. Stahlbom AK, Johansson H, Liljegren A, et al. Evaluation of the BOADICEA risk assessment model in women with a family history of breast cancer. *Fam Cancer*. 2012;11(1):33–40.
19. John EM, Hopper JL, Beck JC, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res*. 2004;6(4):R375–R389.
20. Gong G, Quante AS, Terry MB, et al. Assessing the goodness of fit of personal risk models. *Stat Med*. 2014;33(18):3179–3190.
21. Whittemore AS, Halpern J. Two-stage sampling designs for external validation of personal risk models. *Stat Methods Med Res*. 2013; In press.
22. Heagerty PJ, Zeger SL. Multivariate continuation ratio models: connections and caveats. *Biometrics*. 2000;56(3):719–732.
23. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387.
24. Uno H, Cai T, Pencina MJ, et al. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011;30(10):1105–1117.
25. Pepe MS, Longton G. Standardizing diagnostic markers to evaluate and compare their performance. *Epidemiology*. 2005;16(5):598–603.