Title

A Deep Learning Decision Support Tool to Improve Risk Stratification and Reduce Unnecessary Biopsies in BI-RADS 4 Mammograms.

Journal

Radiology: Artificial Intelligence, 5(6)

Authors

Ezeana, Chika
He, Tiancheng
Patel, Tejal
et al.

Publication Date

2023-11-01

DOI

10.1148/ryai.220259

Radiology: Artificial Intelligence

ORIGINAL RESEARCH

A Deep Learning Decision Support Tool to Improve Risk Stratification and Reduce Unnecessary Biopsies in BI-RADS 4 Mammograms

Chika F. Ezeana, MD, MS • Tiancheng He, PhD • Tejal A. Patel, MD • Virginia Kaklamani, MD, DSc • Maryam Elmi, MD • Erika Brigmon, MD • Pamela M. Otto, MD • Kenneth A. Kist, MD • Heather Speck, MPH • Lin Wang, PhD • Joe Ensor, PhD • Ya-Chen T. Shih, PhD • Bumyang Kim, PhD • I-Wen Pan, PhD • Adam L. Cohen, MD • Kristen Kelley, MD • David Spak, MD • Wei T. Yang, MD • Jenny C. Chang, MD • Stephen T. C. Wong, PhD, PE**

From the Department of Systems Medicine and Bioengineering, Houston Methodist Neal Cancer Center, Houston Methodist Hospital, Houston, Tex (C.F.E., T.H., L.W., S.T.C.W.); Houston Methodist Neal Cancer Center, Houston Methodist Hospital, Houston, Tex (J.E., J.C.C.); Departments of General Oncology (T.A.P.), Health Services Research (Y.C.T.S., B.K., I.W.P.), and Radiology (D.S., W.T.Y.), University of Texas MD Anderson Cancer Center, Houston, Tex; University of Texas Health Science Center, San Antonio, Tex (V.K., M.E., E.B., P.M.O., K.A.K.); University of the Incarnate Word School of Osteopathic Medicine, San Antonio, Tex (H.S.); Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah (A.L.C., K.K.); and Department of Radiology, Houston Methodist Hospital, Weill Cornell Medicine, 6670 Bertner Ave, Houston, TX 77030 (S.T.C.W.). Received November 25, 2022; revision requested December 23; revision received June 8, 2023; accepted July 7. **Address correspondence to** S.T.C.W. (email: *stwong@houstonmethodist.org*).

* J.C.C. and S.T.C.W. are co–senior authors.

Supported by the Ting Tsung & Wei Fong Chao Family Foundation, the John S. Dunn Research Foundation, the Breast Cancer Research Foundation, and The National Institutes of Health/National Cancer Institute grant no. 1R01CA251710.

Conflicts of interest are listed at the end of this article.

See also the commentary by McDonald and Conant in this issue.

Radiology: Artificial Intelligence 2023; 5(6):e220259 • https://doi.org/10.1148/ryai.220259 • Content codes: AI BR OI

Purpose: To evaluate the performance of a biopsy decision support algorithmic model, the intelligent-augmented breast cancer risk calculator (iBRISK), on a multicenter patient dataset.

Materials and Methods: iBRISK was previously developed by applying deep learning to clinical risk factors and mammographic descriptors from 9700 patient records at the primary institution and validated using another 1078 patients. All patients were seen from March 2006 to December 2016. In this multicenter study, iBRISK was further assessed on an independent, retrospective dataset (January 2015–June 2019) from three major health care institutions in Texas, with Breast Imaging Reporting and Data System (BI-RADS) category 4 lesions. Data were dichotomized and trichotomized to measure precision in risk stratification and probability of malignancy (POM) estimation. iBRISK score was also evaluated as a continuous predictor of malignancy, and cost savings analysis was performed.

Results: The iBRISK model's accuracy was 89.5%, area under the receiver operating characteristic curve (AUC) was 0.93 (95% CI: 0.92, 0.95), sensitivity was 100%, and specificity was 81%. A total of 4209 women (median age, 56 years [IQR, 45–65 years]) were included in the multicenter dataset. Only two of 1228 patients (0.16%) in the "low" POM group had malignant lesions, while in the "high" POM group, the malignancy rate was 85.9%. iBRISK score as a continuous predictor of malignancy yielded an AUC of 0.97 (95% CI: 0.97, 0.98). Estimated potential cost savings were more than $420 million.

Conclusion: iBRISK demonstrated high sensitivity in the malignancy prediction of BI-RADS 4 lesions. iBRISK may safely obviate biopsies in up to 50% of patients in low or moderate POM groups and reduce biopsy-associated costs.

Supplemental material is available for this article.

Screening mammography is performed for early detection of breast cancer before clinically detectable signs of disease manifest (1–4). A major limitation of mammography is that patients with suspicious mammographic findings, such as Breast Imaging Reporting and Data System (BI-RADS) category 4 lesions, are recommended for biopsies that often yield benign outcomes. Of the more than 1 million breast biopsies performed annually in the United States, up to 75% have benign findings (5,6). Despite substantial research over the decades, tissue biopsy-proven positive predictive value (PPV3) (7) has not improved much (8). Thus, radiologists would benefit from tools or adjuncts to the BI-RADS system to more precisely assess breast cancer probability in women with BI-RADS category 4 lesions.

The American College of Radiology developed the BI-RADS lexicon that standardizes mammographic reporting to facilitate cancer risk communication and biopsy recommendation (9). However, substantial inter- and intraobserver variability in the application of BI-RADS remains, resulting in variation in biopsy rates across the United States (10,11) More importantly, established risk factors associated with breast cancer such as personal or family history of cancers, hormone replacement therapy, obesity, diabetes, hypertension, and so forth are not incorporated into the clinical decision model. The inclusion

Abbreviations

AUC = area under the ROC curve, BI-RADS = Breast Imaging Reporting and Data System, FNR = false-negative rate, FPR = false-positive rate, HMH = Houston Methodist Hospital, iBRISK = intelligent-augmented breast cancer risk calculator, MDACC = MD Anderson Cancer Center, POM = probability of malignancy, PPV3 = biopsy-proven positive predictive value, ROC = receiver operating characteristic, UTHSCSA = University of Texas Health Science Center San Antonio

Summary

The intelligent-augmented breast cancer risk calculator (or, iBRISK) demonstrated potential to serve as an adjunct to Breast Imaging Reporting and Data System (BI-RADS) to improve risk stratification of BI-RADS category 4 lesions and reduce unnecessary biopsies in patients with lesions with low probability of malignancy.

Key Points

- The intelligent-augmented breast cancer risk calculator (iBRISK) was developed to assess probability of malignancy of Breast Imaging Reporting and Data System (BI-RADS) category 4 lesions.
- The iBRISK model achieved an accuracy of 89.5%, area under the receiver operating characteristic curve of 0.93 (95% CI: 0.92, 0.95), sensitivity of 100%, and specificity of 81%; only 0.16% of lesions determined to have a low probability of malignancy (POM) by the model were malignant, and lesions with high POM had a biopsy-proven predictive value of 85.9%.
- A cost savings analysis demonstrated that iBRISK can reduce unnecessary biopsies of BI-RADS category 4 lesions by up to 50% in patients with lesions classified as low or moderate POM and can reduce financial costs.

Keywords

Mammography, Breast, Oncology, Biopsy/Needle Aspiration, Radiomics, Precision Mammography, AI-augmented Biopsy Decision Support Tool, Breast Cancer Risk Calculator, BI-RADS 4 Mammography Risk Stratification, Overbiopsy Reduction, Probability of Malignancy (POM) Assessment, Biopsy-based Positive Predictive Value (PPV3)

of these factors could contribute to a more robust and holistic assessment of a suspicious mammographic finding (Table S1) (12–14). The BI-RADS category 4 subgroup assigns wide variability in the probability of malignancy (POM), ranging from 2% to 95%. Biopsies of BI-RADS category 4 lesions serve as a quality metric and performance standard (15–18). False-positive mammograms are estimated to cost around $4 billion in the United States yearly (19).

A recent report based on the National Mammography Database subcategorized the majority of BI-RADS category 4 cases into BI-RADS category 4A (55.6%) and BI-RADS category 4B (31.8%), with associated low PPV3s of 7.6% and 22%, respectively (8). These findings indicate the opportunity for exploring, developing, and validating precision diagnostic models that can appropriately downgrade low and moderate suspicious assessments to nonactionable levels. Supplemental tools and algorithms would have less impact on BI-RADS category 4C lesions because they are fewer (12.6%) and of much higher PPV3 (69.3%).

Several models (20–25) have been developed previously, but they differ from our work in terms of scope, model predictors, and performance accuracy. Some earlier models incorporate all BI-RADS categories or focus on either screening or diagnostic data only. Other models are strictly limited to imaging data or integrate few clinical parameters and are often trained on public breast cancer screening datasets only.

We aim to evaluate the performance of our improved biopsy decision support algorithmic model, the intelligent-augmented breast cancer risk calculator (26) (iBRISK), on a large patient dataset in a multicenter study.

Materials and Methods

Source of Data

The institutional review boards of the participating institutions approved this study, which was performed in strict compliance with Health Insurance Portability and Accountability Act guidelines, and granted waivers of informed consent. We used retrospective, multi-institutional datasets to assess the performance of our previously developed and now improved iBRISK, a decision support tool that characterizes breast lesions classified as BI-RADS category 4 on mammograms and stratifies women according to POM (26). Data for this validation study include patient clinical data and mammography reports, which were consecutively drawn from the systemwide data warehouse of our institution, Houston Methodist Hospital (HMH) (27) (the same data source for model development and improvement [March 2006–December 2016]). Data were also consecutively curated from the electronic medical records from MD Anderson Cancer Center (MDACC) (March 2016–September 2018) and the University of Texas Health Science Center San Antonio (UTHSCSA) (January 2015–June 2019). Our study evaluated only patients with BI-RADS category 4 lesions with mammographic abnormalities. To limit the number of input variables and keep the model user-friendly, mammography alone was used. Also, the final features of the iBRISK model were purely from mammographic descriptors and clinical factors.

Patients

The study included patients with lesions categorized as BI-RADS 4 at diagnostic mammography, including recalls from screening, who were seen consecutively in diverse clinical settings. Minimum inclusion criteria were data on age, height, weight, and calcification details at imaging, as well as a biopsy performed within 3 months after mammography. Patients with lesions classified into other BI-RADS categories or missing the aforementioned data were excluded (Fig 1). The study used only retrospective patient data; there was no direct patient contact, and patients did not receive any treatments.

Outcome

Each patient in the test dataset was evaluated using iBRISK by inputting the individual's set of 20 variables (Table S1) comprising the model, which were derived from clinical factors and mammographic descriptors (Table S1). Please see Appendix S1 for details on the iBRISK model (26), its im-

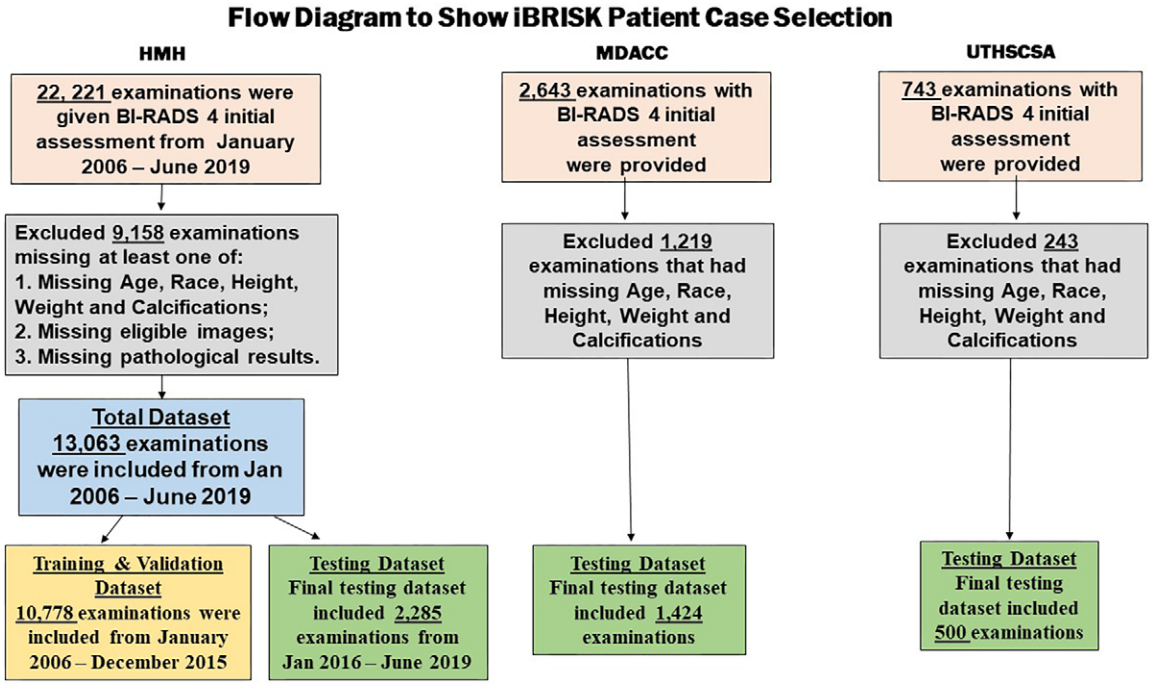

Figure 1: Flow diagram shows patient case selection for refined model training, validation, and multicenter testing. BI-RADS = Breast Imaging Reporting and Data System, iBRISK = intelligent-augmented breast cancer risk calculator.

provement and updating, comparison with other models, and predictors. The calculator provided a POM score between 0 and 1, as well as biopsy decision-making support recommendations based on these scores. Evaluations were performed while blinded to patients' pathologic results. Afterward, model results were compared with biopsy outcomes, which served as ground truth.

Missing Data

The validation data of HMH and MDACC were complete datasets without missing values. UTHSCSA, however, had 1.34% missing data. Thus, of 20 features each for all 500 cases (ie, 10 000 observations from this site), 134 observations were missing. We used one-hot encoding (28,29) to vectorize the input parameters, and after the parameters were vectorized, the binary values in the missing data became all zeroes (0,0), compared with (0,1) for "yes" and (1,0) for "no". Interestingly, continuous input variables like age, height, and weight were and would always be available in real-time workflow.

Statistical Analysis

An iBRISK score was derived for each patient (n = 4209). A single continuous predictor logistic regression model was fitted to the data to test the ability of the iBRISK scores to predict pathologic findings. The iBRISK scores were trichotomized into "low," "moderate," or "high" POM (cutoff points were determined based on the dynamics of our training data and model settings) and correlated with the pathologic findings (a dichotomous categorical factor: malignant or benign), which served as ground truth in a χ^2 test. Additionally, the model scores were dichotomized into low versus "not low" or "not high" versus high and then correlated with the same χ^2 analysis of the pathologic findings (benign or malignant). In either case, in the χ^2 analysis, we correlated the iBRISK trichotomized or dichotomized predictor with the pathologic finding. Receiver operating characteristic (ROC) curve analysis for all patients in the test set, as well as subdivision by race classifications in the data warehouse and electronic medical records, was performed, and the area under the ROC curve (AUC) was calculated. An assessment of the impact of missing variables on model accuracy was performed using data from MDACC (1424 patients); that is, we simulated states of missing features by progressively removing one feature at a time and assessing the impact on model accuracy and stability. We estimated possible iBRISK-assisted biopsy avoidance, and potential cost savings were calculated in a cost analysis.

To assess the importance of each of the 20 factors in the model, 20 unique sets of scores were derived in which each set of scores represented the effect of removing a different factor, that is, without imputation of said factor's value. Logistic regression was used to derive AUCs for each of the 20 sets of scores as a marker of model performance. Differences in performance (as measured by the areas under the empirical ROC curves) between the full and factor-restricted models were assessed using the method of DeLong et al (30). Decreases in AUC, along with their associated 95% Wald CIs, between the full model and each of the cluster-restricted models were used to measure cluster effect on model performance. The significance of POM differences between groups was determined using χ^2 tests. A P value of less than .05 was considered statistically significant. All analyses were conducted using SAS 9.4 software (SAS Institute) (31).

Table 1: Demographic Characteristics for Training, Validation, and Test Sets

Characteristic	Training Set	Validation Set	Multicenter Test Set
Sample size	9700	1078	4209
Age (y)	54 (45–63)	52 (43–62)	56 (45–65)
Height (cm)	162.6 (157.5–167.6)	161.5 (156.5–67.6)	162.6 (153.8–167.9)
Weight (kg)	73.4 (60.5–89.8)	72.6 (59.4–87.5)	85.73 (59.3–87.1)
BMI (kg/m^2)	28 (24–32)	27 (24–31)	32.08 (23–34)
Race or ethnicity			
African American	1562 (16.1)	140 (13.0)	592 (14.1)
Asian	718 (7.4)	43 (4.0)	273 (6.5)
Caucasian (non-Hispanic)	6557 (67.6)	852 (79.0)	2504 (59.5)
Hispanic	136 (1.4)	5 (0.5)	532 (12.6)
Native American	10 (0.1)	0 (0.0)	8 (0.2)
Other	630 (6.5)	32 (3.0)	256 (6.1)
Unknown	87 (0.9)	6 (0.6)	44 (1.1)
Insurance			
Commercial, self-pay, or other	7314 (75.4)	725 (67.3)	3006 (71.4)
Medicaid	252 (2.6)	31 (2.9)	155 (3.7)
Medicare (includes part B)	1950 (20.1)	300 (27.8)	934 (22.2)
Unknown	184 (1.9)	22 (2.0)	114 (2.7)
History			
Previous breast cancer	4161 (42.9)	498 (46.2)	1546 (36.7)
Previous other cancer	708 (7.3)	57 (5.3)	851 (20.2)
Family history of cancer	3550 (36.6)	390 (36.2)	303 (7.2)

Note.—Data are reported as medians, with IQRs in parentheses, or numbers of patients, with percentages in parentheses. Race category "other" includes Pacific Islanders, Alaskan Natives, and Native Hawaiians. BMI = body mass index.

Results

Patient Characteristics

The testing dataset (4209) comprised 2285 of 3887 (58.8%) patients from HMH, 1424 of 2643 (53.9%) from MDACC, and 500 of 743 (67.3%) from UTHSCSA (Fig 1). Median and IQR values for the multicenter validation dataset (n = 4209) were as follows: age, 56 years (IQR, 45–65 years); height, 162.6 cm (IQR, 153.8–167.9 cm); weight, 85.73 kg (IQR, 59.3–87.1 kg); and body mass index, 32.08 (IQR, 23–34). Non-Hispanic White individuals (2504 of 4209 [59.49%]) and commercial insurance and self-paying patients (3006 of 4209 [71.42%]) were the majority for race and insurance status, respectively. Descriptive statistics of other demographic predictors can be found in Tables 1 and 2.

Model Performance

Score as a continuous predictor.— Logistic regression was fitted to the data, assuming model score as a continuous predictor of malignancy, resulting in an AUC of 0.97 (95% CI: 0.96, 0.98) (Fig 2A). Figure 2B shows the graph for the logistic estimate. The model suggests the POM is zero for scores less than 0.4 and extremely high for scores above 0.7, as indicated by the trichotomized data. Figure S1 outlines the scores as a continuous predictor according to race and ethnicity distribution.

Trichotomized into low, moderate, and high POM.— iBRISK designated patients as having low POM for scores of less than 0.4, moderate POM for scores between 0.4 and 0.55, and high POM for scores greater than 0.55. Overall, 29.2% (1228 of 4209) of patients in the multicenter validation dataset had low POM, 42.8% (1788 of 4209) had moderate POM, and 28.1% (1193 of 4209) had high POM. While the distribution of patients in the three POM categories was significantly different between institutions (Fig S2), the calculator performed equally at all three sites in terms of sensitivity, accuracy, and when dichotomized as described below.

The proportion of benign lesions within these 4209 patients was significantly different between the POM groups ($P < .001$). When the model predicted a low POM, the likelihood of a benign biopsy finding was 99.8% (Fig 3), with a false-negative rate (FNR) of 0.16% (two of 1228 were malignant). Most patients in the moderate POM category also had benign biopsy findings (93.4%, 1670 of 1788), with a slightly higher malignancy rate of 6.6% (118 of 1788). The high POM group had a malignancy rate of 85.9% (1025 of 1193). The calculator designated only

Table 2: Distribution of Model Variables across Medical Centers in Test Set Grouped by iBRISK-determined Probability of Malignancy Categories

Variable	Low (*n* = 1228)	Intermediate (*n* = 1788)	High (*n* = 1193)	Overall (*n* = 4209)
Visit age (y)*	56 (18–91)	60 (27–96)	60 (22–94)	56 (18–96)
Race				
Asian	72	132	69	273
Black	147	282	163	592
Caucasian	712	1094	698	2504
Hispanic	171	210	151	532
Indian American	4	2	2	8
Other	122	68	110	300
Height (cm)*	162.6 (125.3–198.12)	165.1 (150–177.8)	162.6 (139.7–198.12)	162.6 (125.3–198.12)
Median weight (kg)	86	89.36	84.37	85.73
Median BMI (kg/m^2)	32.00	33.2	31.64	32.08
Insurance				
Commercial and self-pay	865 (70.44)	1335 (74.66)	806 (67.56)	3006 (71.42)
Medicare	289 (23.53)	379 (21.20)	266 (22.30)	934 (22.19)
Medicaid	38 (3.09)	64 (3.58)	53 (4.44)	155 (3.68)
Menopausal status				
Premenopausal	642 (52.28)	910 (50.89)	589 (49.37)	2141 (50.87)
Postmenopausal	586 (47.72)	878 (49.11)	604 (50.63)	2068 (49.13)
Left or right breast				
Left	600 (48.86)	1001 (55.98)	639 (53.56)	2240 (53.22)
Right	607 (49.43)	944 (52.80)	571 (47.86)	2122 (50.42)
Cancer history				
Family history of cancer	472	659	415	1546
Previous history of breast cancer	330	303	218	851
Personal history of other cancers	97	124	82	303
Skin change				
Nipple retraction	2	11	11	24
Trabecular thickening	3	6	13	22
Solitary dilated duct	12	41	24	77
Other skin changes	166	250	134	550
Presence of palpable lump	146	357	277	780
Presence of nipple discharge	23	42	18	83
Mammogram density				
Fatty	41	102	66	209
Scattered	393	755	438	1586
Heterogeneously dense	570	760	500	1830
Extremely dense	75	139	69	283
Presence of mass on mammogram	460 (37.46)	1289 (72.09)	777 (65.13)	2526 (60.01)
Mass appearance				
Oval	116	309	163	588
Round	40	134	78	252
Irregular	85	289	201	575
Spiculated	7	51	79	137
Circumscribed	90	228	91	409
Obscured	42	153	84	279
Microlobulated	11	37	17	65
Indistinct	21	129	104	254
Calcifications present	828 (67.43)	1244 (69.57)	776 (65.05)	2848 (67.66)
Calcification appearance				
Heterogeneous or coarse	259 (21.09)	283 (15.83)	195 (16.35)	737 (17.51)
Amorphous	348 (28.34)	211 (11.80)	137 (11.48)	696 (16.54)
Linear or pleomorphic	28 (2.28)	53 (2.96)	62 (5.20)	143 (3.40)

(Table 2 continues)

Table 2 *(continued)*: Distribution of Model Variables across Medical Centers in Test Set Grouped by iBRISK-determined Probability of Malignancy Categories

Variable	Low (n = 1228)	Intermediate (n = 1788)	High (n = 1193)	Overall (n = 4209)
Round or punctate	30 (2.44)	106 (5.93)	68 (5.70)	204 (4.85)
Calcification distribution				
Linear	45 (3.66)	19 (1.06)	28 (2.35)	92 (2.19)
Segmental	53 (4.32)	60 (3.36)	90 (7.54)	203 (4.82)
Clustered	560 (45.60)	583 (32.61)	377 (31.60)	1520 (36.11)
Regional or diffuse	47 (3.83)	76 (4.25)	56 (4.69)	179 (4.25)
Asymmetry or architectural distortion	51 (4.15)	131 (7.33)	85 (7.12)	267 (6.34)
Axillary adenopathy	33 (2.69)	139 (7.77)	127 (10.65)	299 (7.10)
Medications				
Blood pressure medication	238 (19.38)	318 (17.79)	205 (17.18)	761 (18.08)
Heart medication	140 (11.40)	131 (7.33)	97 (8.13)	368 (8.74)
Cholesterol medication	92 (7.49)	133 (7.44)	78 (6.54)	303 (7.20)
Diabetes medication	75 (6.11)	128 (7.16)	52 (4.36)	255 (6.06)

Note.—Unless otherwise noted, values are numbers, and values in parentheses are percentages. Race category "other" includes Pacific Islanders, Native Americans, Alaskan Natives, and Native Hawaiians. BMI = body mass index, iBRISK = intelligent-augmented breast cancer risk calculator.
* Values are medians, with ranges in parentheses.

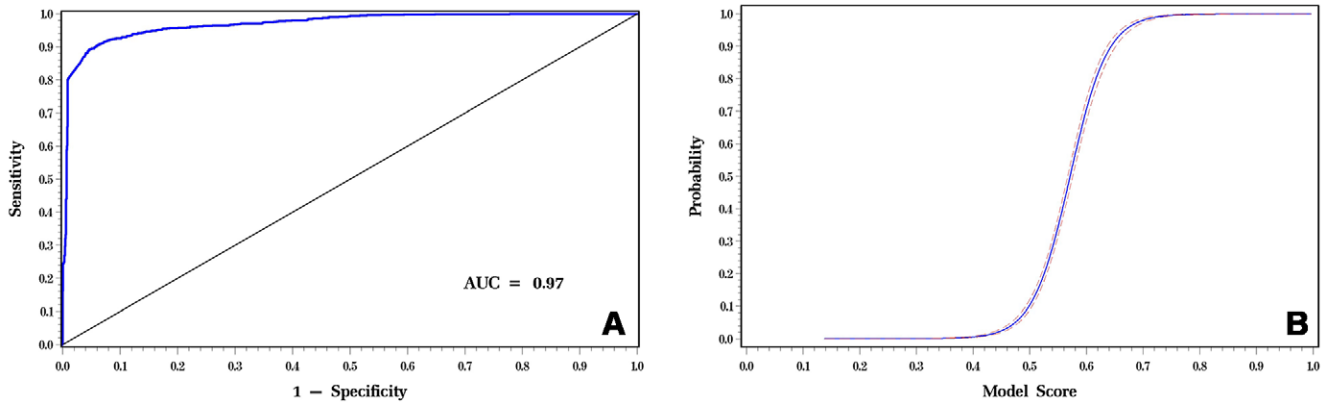

Figure 2: **(A)** Receiver operating characteristic curve of iBRISK score as a continuous estimate of the probability of breast lesion malignancy. **(B)** Graph shows probability of malignancy by model score (logistic estimate with confidence limits). AUC = area under the receiver operating characteristic curve, iBRISK = intelligent-augmented breast cancer risk calculator.

14.1% (168 of 1193) of benign biopsy findings as high POM (false-positive rate [FPR]) (Table 3, Fig 3).

Dichotomized into low versus not low POM.— Model scores were dichotomized into low versus not low POM following the clinical decision process in the 4209 patients. The proportion of benign lesions was significantly different between the two risk groups ($P < .001$). As previously mentioned, there were 1228 patients in the low POM group (FNR, 0.16%). There were 1838 patients in the not low POM group who would have undergone biopsy with benign results, with an FPR of 61.66%. Model sensitivity was 99.83% (ie, the model would detect malignant lesions 99.83% of the time) (Table 4). Performance metrics between groups per institution are shown in Table S2.

Dichotomized into high versus not high POM.— When model scores were dichotomized into high versus not high POM, the proportion of benign lesions was significantly different between the two risk groups ($P < .001$). Among the lesions categorized as not high POM, 116 were malignant (FNR, 3.86%), while 177 biopsies would have been conducted among patients with benign lesions (FPR, 14.68%). The model achieved 89.87% sensitivity, 94.22% specificity, and an AUC of 0.92 (Table 4; per institution, Table S2). Table S3 shows the percentage of benign and malignant biopsy findings and the FNRs and FPRs after model categorization of lesions as low, moderate, or high POM.

Contribution of factors in iBRISK, individually and in clusters.— The contribution of each of the 20 factors in iBRISK was calculated using simple logistic regression to estimate the AUC after removing each factor from the model. Each factor removal resulted in a small but statistically significant decrease in performance, reflecting its relative contribution.

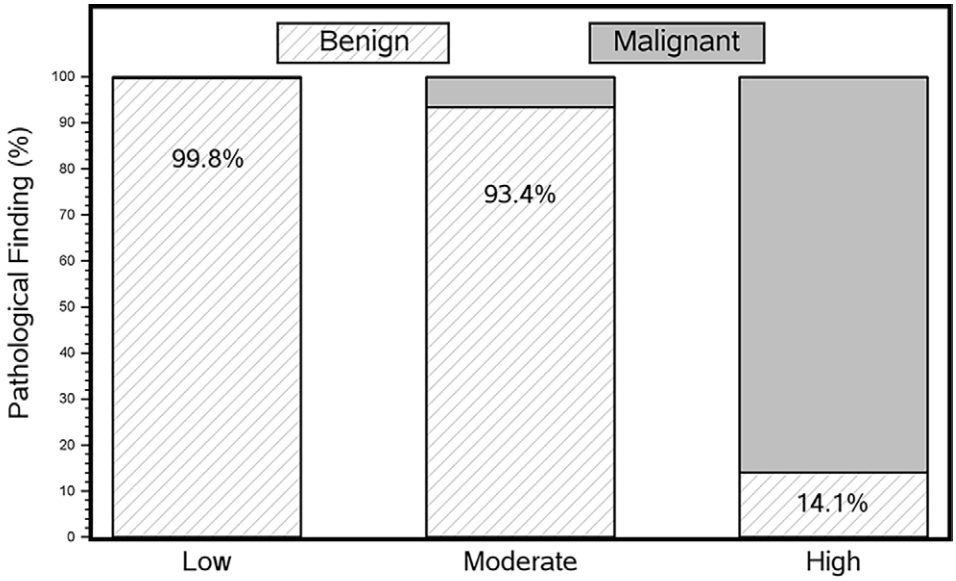

Figure 3: Percentage of benign and malignant pathologic findings after biopsy of breast lesions according to iBRISK probability of malignancy level. iBRISK = intelligent-augmented breast cancer risk calculator.

Table 3: Association between Trichotomized Model Probability of Malignancy and Pathologic Finding in Test Set

	Model Probability of Malignancy			
Pathologic Finding	Low	Moderate	High	Total
Benign				
No. of lesions	1226	1670	168	3064
Percentage	TNR: 99.84	TNR: 93.40	FPR: 14.08	…
Malignant				
No. of lesions	2	118	1025	1145
Percentage	FNR: 16.00	FNR: 6.60	TPR: 85.92	…
Total no. of lesions	1228	1788	1193	4209

Note.—Test set comprised 4209 patients. The distribution of pathologic findings was significantly different between the three model probability of malignancy groups on the basis of a χ^2 test ($P <$ .0001). FNR = false-negative rate, FPR = false-positive rate, TNR = true-negative rate, TPR = true-positive rate.

Menopausal status and mammographic mass had the largest contributions according to decrease in AUC (Fig 4A, Table 5). We grouped the final 20 factors in the model into the following six clusters: *(a)* demographics (age, race, menopausal status, and laterality), *(b)* metabolic factors (height, weight, insurance, and medications, including hormone replacement therapy), *(c)* history (personal history of breast or other cancers, family history of breast or other cancers), *(d)* physical signs (skin changes, nipple discharge, palpable lump, and lymphadenopathy), *(e)* mammographic density and mammographic mass, and *(f)* mammographic calcification features (vascular calcification, calcification morphology, and calcification distribution) and asymmetry and architectural distortion. Mammographic mass, metabolic factors, and mammographic calcification features showed the highest contributions (Fig 4B, Table 5).

Missing Feature Analysis

Using MDACC data (n = 1424) to assess the impact of missing features on the accuracy and stability of the iBRISK model, we observed progressively slighter declines with each additional missing feature and a statistically significant level drop in accuracy when the fourth feature was removed. Thus, the model can tolerate up to three missing features while retaining robustness and confidence in the accuracy of results generated (Fig 5).

Estimated Annual Cost Savings

Table S4 shows the median projected cost (based on the Medicare reimbursement rate) of biopsy ($380) and the average cost for each type of biopsy. The most common type of biopsy among MDACC patients was stereotactic biopsy (68%). The cost of biopsy ranged from $321 (stereotactic biopsy) to al-

Table 4: Association between Dichotomized Model Probability of Malignancy and Pathologic Finding in Test Set

A. Low versus Not Low

	Model Probability of Malignancy		
Pathologic Finding	Low	Not Low	Total
Benign			
No. of lesions	1226	1838	3064
Percentage	TNR: 99.84, specificity: 40.01	% FP: 61.66, FPR: 59.99	…
Malignant			
No. of lesions	2	1143	1143
Percentage	% FN: 0.16, FNR: 0.17	TPR: 38.34, sensitivity: 99.83	…
Total no. of lesions	1228	2981	4209

B. High versus Not High

	Model Probability of Malignancy		
Pathologic Finding	Not High	High	Total
Benign			
No. of lesions	2887	177	3064
Percentage	TNR: 96.15, specificity: 94.22	% FP: 14.68, FPR: 5.78	…
Malignant			
No. of lesions	116	1029	1145
Percentage	% FN: 3.86, FNR: 10.13	TPR: 85.32, sensitivity: 89.87	…
Total no. of lesions	3003	1206	4209

Note.—Test set comprised 4209 patients. The proportion of benign lesions was significantly different between each pair of probability of malignancy groups on the basis of a χ^2 test ($P < .0001$). FN = false negative, FNR = FN rate, FP = false positive, FPR = FP rate, TNR = true-negative rate, TPR = true-positive rate.

most $3600 (mammography-guided surgical biopsy). Table S5 provides the information used to derive our estimate of cost savings as a result of triaging patients classified as low risk by iBRISK to not undergo biopsy. As shown, biopsy can potentially be avoided for approximately 390 000 women, with cost savings of more than $420 million.

Discussion

We improved iBRISK and evaluated the model by using a retrospective multi-institutional dataset made up of patients from MDACC, UTHSCSA, and HMH. iBRISK demonstrated high sensitivity and specificity for the prediction of POM, resulting in improved risk stratification of BI-RADS category 4 lesions, such that only 0.16% (two of 1228) of lesions classified as low POM in women assessed by iBRISK were malignant, and PPV3 among the high POM group was 85.9% (1025 of 1193), which is close to that of the BI-RADS 5 risk category (80.3%–97.9%) and outperforms radiologists' BI-RADS 4 categorization accuracy (20,32,33). Thus, iBRISK can potentially obviate up to 50% of biopsies in patients with BI-RADS 4 mammograms.

The iBRISK calculator can assist physicians, primarily radiologists, in triaging patients to low POM groups to avoid biopsies of benign lesions, while high-risk groups can be treated as patients with BI-RADS category 5 lesions. A more precise stratification system that considers vital patient characteristics in addition to abnormal, suspicious imaging features is needed to enhance POM estimation to guide the safe management of such mammographic findings, prevent overbiopsy and associated costs, and reduce patient emotional distress (34,35). The goal is not to replace or modify the BI-RADS standards but to improve precision in predicting the malignancy of category 4 lesions, which are currently overbiopsied, when iBRISK is used alongside the BI-RADS system.

While various models have been proposed and several studies performed (20,21,24,25), a safe, pragmatic, and effective system that addresses these concerns has not been reported. A 2015 study included the Gail model, body mass index, and genetic marker information for breast cancer risk estimation in women with suspicious findings on BI-RADS 4 mammograms (21). Similar to our current study, this study considered clinical factors, albeit in a more limited fashion, but did not improve POM estimation precision within the BI-RADS 4 category. A 2019 study proposed a combined machine and deep learning approach applied to digital mammograms and electronic health records to identify false-negative findings in BI-RADS categories 1, 2, and 3 (24). The algorithm identified 34 of 71 (48%) of such findings on mammograms. Another study in 2021 developed a deep learning fusion network model using mammography imaging biomarkers and clinical features of BI-RADS category 3, 4, and 5 lesions to predict malignancy (25). However, their test cohort was relatively small (internal test cohort, 244 patients; external

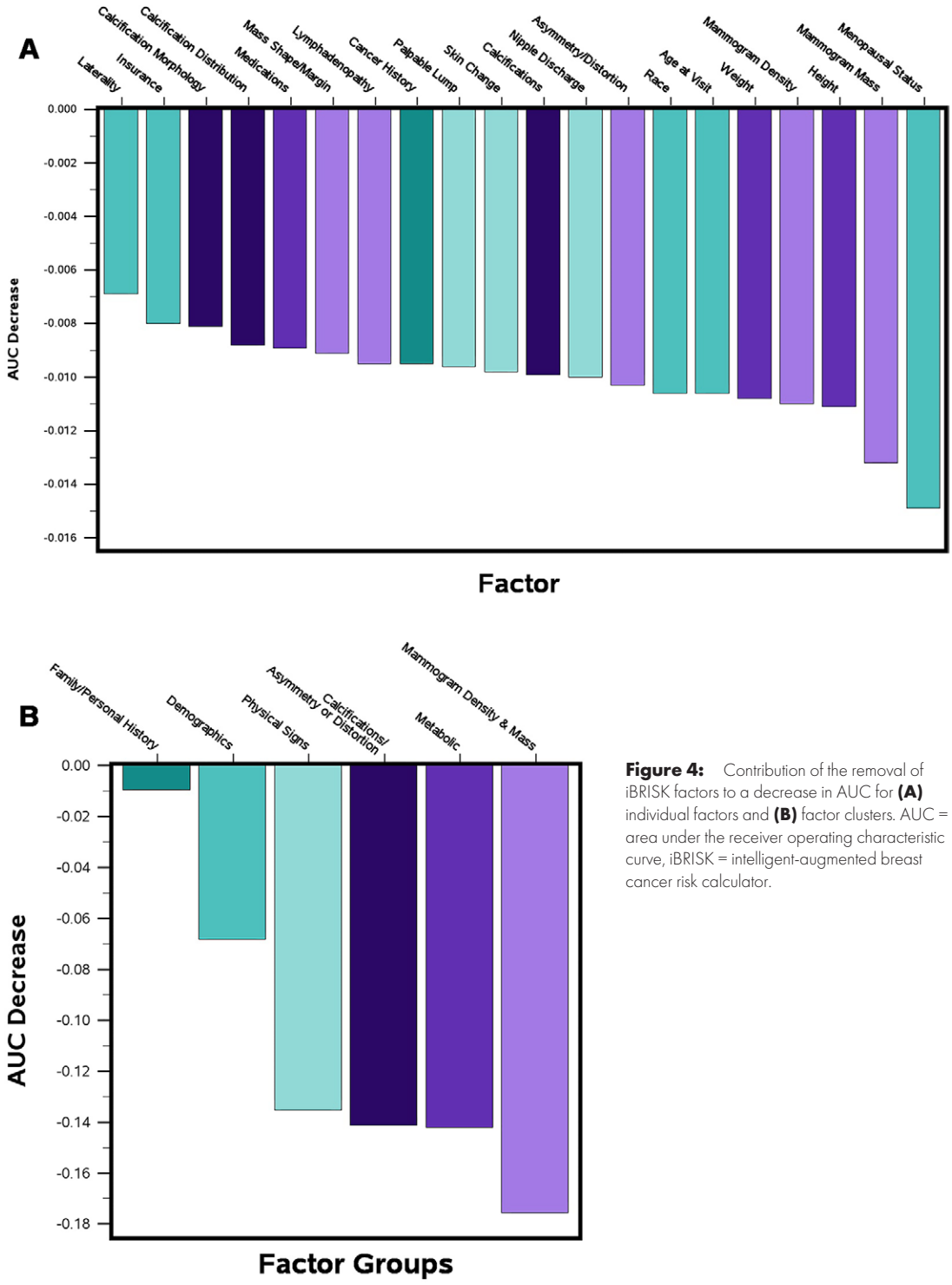

Figure 4: Contribution of the removal of iBRISK factors to a decrease in AUC for **(A)** individual factors and **(B)** factor clusters. AUC = area under the receiver operating characteristic curve, iBRISK = intelligent-augmented breast cancer risk calculator.

test cohort, 100 patients). Most current published works do not compare results with the BI-RADS guidelines, address the issue of precision POM estimation of BI-RADS category 4 mammogram suspicious findings, or address the issue of overbiopsy, and do not involve multiple nonimaging parameters. Further, current published artificial intelligence models for cancer probability estimation were developed using clinical images from a public breast cancer screening dataset (22,23,36) and do not incorporate the above parameters.

Because of variability in malignancy rates associated with BI-RADS category 4, the BI-RADS fifth edition proposes the following subcategories for likelihood of malignancy: 4A (2%–10%), 4B (11%–50%), and 4C (50%–95%). Reported PPV3 for BI-RADS category 4 ranged from 15% to 30% (37–40) and was recently reported as 21.1% (4A: 7.6%, 4B: 22.2%, 4C: 69.3%) (8) in the United States and between 21% and 27.1% in other countries (32,41). Subcategorization of BI-RADS 4 is not widely adopted, as the malignancy

Table 5: Contribution of Removal of iBRISK Model Factors to Decrease in Model Performance

A. Individual Factors

Scenario	Factor Removed	AUC		AUC Loss	*P* Value
		Estimate	95% CI		
0	None (full model)	0.97	0.96, 0.98	…	…
1	Age at visit	0.96	0.95, 0.97	0.0106	<.001
2	Race	0.96	0.95, 0.97	0.0106	<.001
3	Height	0.96	0.95, 0.97	0.0111	<.001
4	Weight	0.96	0.95, 0.97	0.0108	<.001
5	Insurance	0.96	0.96, 0.97	0.0080	<.001
6	Cancer history	0.96	0.95, 0.97	0.0095	<.001
7	Menopausal status	0.95	0.95, 0.96	0.0149	<.001
8	Laterality	0.96	0.96, 0.97	0.0069	<.001
9	Skin change	0.96	0.95, 0.97	0.0098	<.001
10	Palpable lump	0.96	0.95, 0.97	0.0096	<.001
11	Nipple discharge	0.96	0.95, 0.97	0.0100	<.001
12	Mammogram density	0.96	0.95, 0.97	0.0110	<.001
13	Mammogram mass presence	0.96	0.95, 0.96	0.0132	<.001
14	Mammogram mass shape or margin (appearance)	0.96	0.95, 0.97	0.0091	<.001
15	Calcifications	0.96	0.95, 0.97	0.0099	<.001
16	Calcification form	0.96	0.96, 0.97	0.0081	<.001
17	Calcification distribution	0.96	0.95, 0.97	0.0088	<.001
18	Mammogram lymphadenopathy	0.96	0.95, 0.97	0.0095	<.001
19	Mammographic asymmetry or distortion	0.96	0.95, 0.97	0.0103	<.001
20	Medications	0.96	0.95, 0.97	0.0089	<.001

B. Factor Groups

Scenario	Group Removed	AUC		AUC Loss	*P* Value
		Estimate	95% CI		
0	None (full model)	0.97	0.96, 0.98	…	…
1	Demographics	0.90	0.89, 0.91	0.0682	<.001
2	Metabolic	0.83	0.81, 0.84	0.1421	<.001
3	History	0.96	0.95, 0.97	0.0095	<.001
4	Physical signs	0.83	0.82, 0.85	0.1352	<.001
5	Mass features	0.79	0.78, 0.81	0.1756	<.001
6	Calcification features	0.83	0.81, 0.84	0.1411	<.001

Note.—AUC = area under the receiver operating characteristic curve.

rate in the 4A (low risk) category is up to 10% (20). Our study demonstrated that iBRISK outperforms BI-RADS subcategory recommendations, with an FNR of less than 1% in the low POM category. There were malignancy rates of 6.6% in the iBRISK moderate POM group, which is still lower than the published BI-RADS category 4A range, and 85.9% in the high POM group, close to the BI-RADS 5 category. Also, a recent published study found that essentially, in the BI-RADS 4 category, digital breast tomosynthesis had no comparative advantage over digital mammography in terms of PPV3 and cancer detection rate, which could answer questions on the impact of the broad implementation of digital breast tomosynthesis on precision (42).

Our study had certain limitations. First, the iBRISK model was built, refined, and internally validated with patient data from a major health system in the greater Houston area, and this reported multi-institutional testing is largely restricted to data from three leading hospitals in Texas. A larger multicenter study involving other states is being planned to further assess the model's performance in more diverse patient populations and breast imaging practices. Second, the study required complete retrospective data curation to incorporate both mammographic

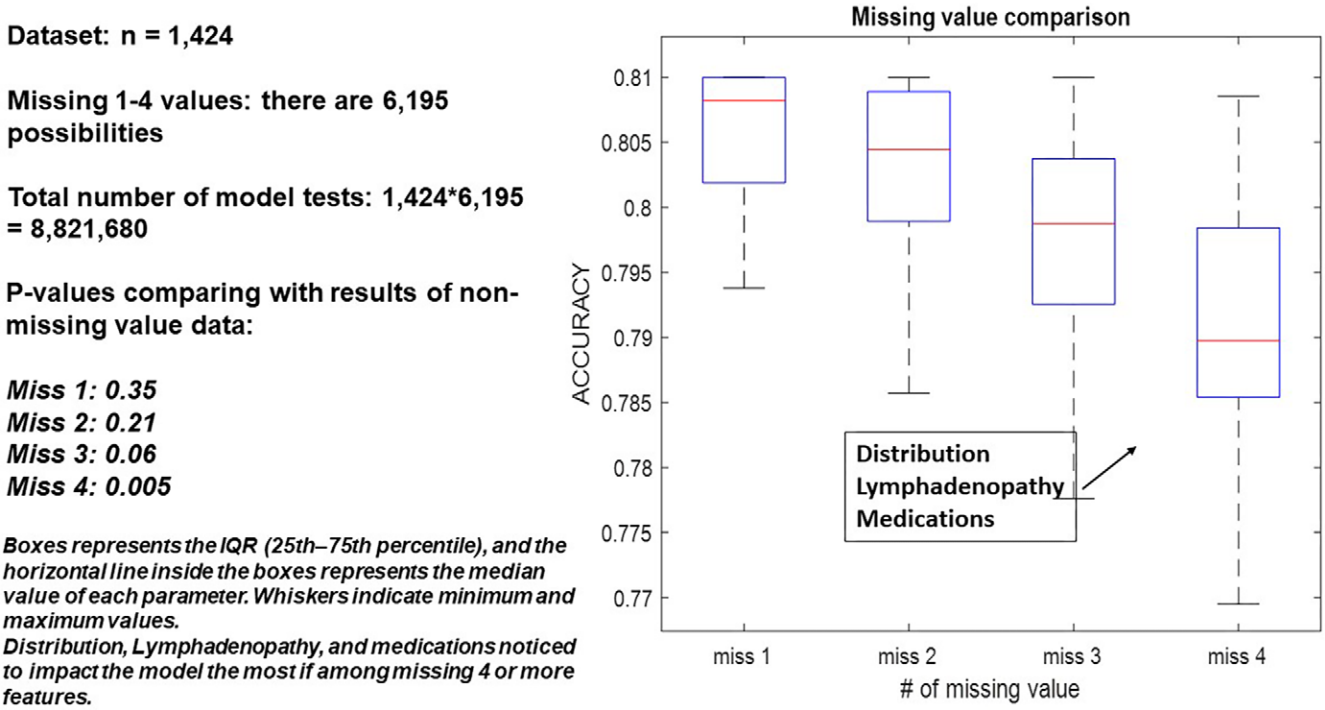

Figure 5: Missing feature analysis of MDACC dataset shows a slight drop in iBRISK accuracy with each additional missing feature. iBRISK = intelligent-augmented breast cancer risk calculator, MDACC = MD Anderson Cancer Center.

and patient risk factors not readily considered in mammography reports, resulting in 53.9%–67.3% of patients from the initial datasets at the three participating sites to be included in the final analysis. However, the model performed well across the sites, suggesting model robustness to external data. Third, 85% of the 20 model variables were needed for robust POM scoring using iBRISK, with the calculator tolerating a maximum of three missing features. The slight accuracy decreases with each additional missing feature using the MDACC dataset emphasize the requirement of precise data curation and annotation for the optimal function of this tool. Fourth, iBRISK substitutes missing data with a default (unknown) or average number value, thus affecting its accuracy. However, demographic and specific information on mammographic calcifications would be available when used in real time. The lack of consistent reporting on mammographic calcification form and calcification distribution underscores the urgent need for consistent structured breast imaging reporting systems that optimize data acquisition, archiving, retrieval, and extraction. The envisaged clinical workflow is an online iBRISK calculator where these 20 features including demographics, history, and mammographic features would be inputted after a mammography examination with a BI-RADS category 4 designation by the radiologist or other providers and the risk score generated. Fifth, findings of previous scans have not been included because of availability constraints, though it can be argued that progression of calcifications and lesions over time is important. Extending our analysis beyond lesion classification as benign versus malignant to clinical outcomes, histology of cancer types, and aggressive versus less aggressive tumors was beyond the scope of this study and should be investigated in future studies. Sixth, most biopsies evaluated in the cost savings analysis were stereotactic biopsies. While other biopsies were performed at this site, the data retrieval and study period occurred during the migration of the electronic medical record system to a new platform; therefore, successful data curation with accurate clinical information of the other biopsy types could not be performed. Of note, stereotactic biopsies are much cheaper. Last, iBRISK POM assessment serves as an adjunct to breast imaging for clinical providers and patients in biopsy decision-making and thus is not a definitive diagnostic tool.

In summary, our study demonstrates that iBRISK can effectively aid in risk stratification of BI-RADS category 4 lesions and reduce overbiopsy of these lesions. Ultimately, the iBRISK calculator will be published as an online interface and made open access, noncommercial, and accessible by health systems and centers worldwide. Future studies aim to improve the model further, particularly by including more granular data and other BI-RADS categories.

Acknowledgments: We are grateful for the life and contributions of Ms Mamta Puppala, who passed away during the preparation of this manuscript, and we dedicate this paper to her memory. She contributed to the provision of study materials and patients. She also worked on data sourcing, collection, validation, assembly, analysis, and interpretation. We also thank our hospital information technology colleagues at the Houston Methodist Hospital, and our Biostatistics and Bioinformatics Shared Resources colleagues at Houston Methodist Neal Cancer Center for their help with this project, as well as Rebecca Danforth, PhD, for proofreading the manuscript.

Author contributions: Guarantors of integrity of entire study, **H.S., B.K., I.W.P., S.T.C.W.**; study concepts/study design or data acquisition or data anal-

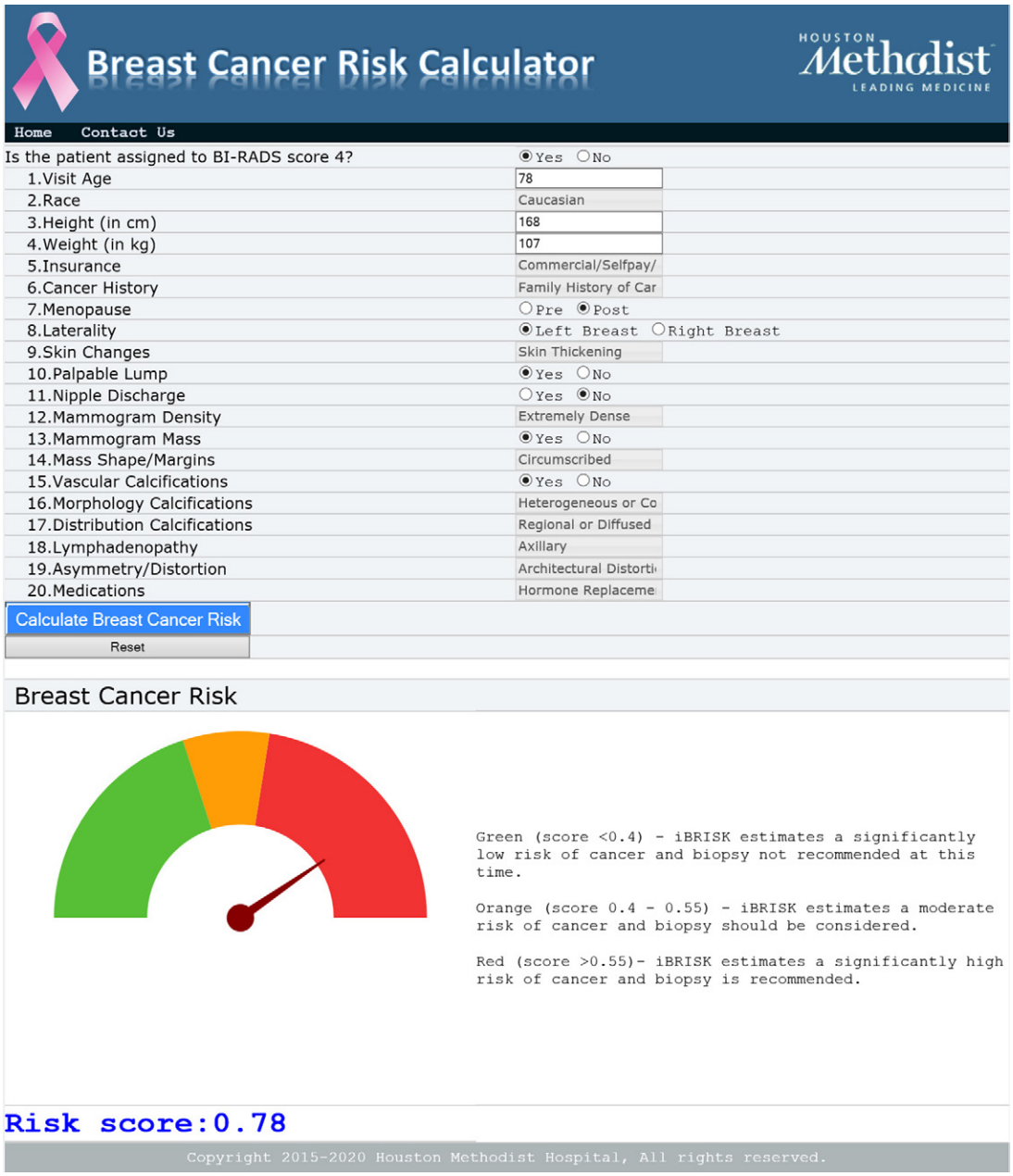

Figure 6: Online iBRISK interface showing 20 fields. iBRISK = intelligent-augmented breast cancer risk calculator.

ysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **C.F.E., T.H., T.A.P., M.E., H.S., W.T.Y.**; clinical studies, **T.H., T.A.P., V.K., M.E., E.B., P.M.O., H.S., I.W.P., K.K., W.T.Y., J.C.C.**; experimental studies, **C.F.E., T.H., T.A.P., V.K., P.M.O., H.S., S.T.C.W.**; statistical analysis, **C.F.E., T.H., H.S., L.W., J.E., Y.C.T.S., B.K., I.W.P., D.S., W.T.Y., S.T.C.W.**; and manuscript editing, **C.F.E., T.H., T.A.P., M.E., P.M.O., H.S., L.W., J.E., Y.C.T.S., B.K., I.W.P., K.K., W.T.Y., J.C.C., S.T.C.W.**

Data sharing: Data generated or analyzed during the study are available from the corresponding author by request.

Disclosures of conflicts of interest: C.F.E. No relevant relationships. **T.H.** No relevant relationships. **T.A.P.** No relevant relationships. **V.K.** Consultant fees from AstraZeneca, Daiichi, Seagen, Gilead, Novartis, Lilly, Pfizer, Genentech, and TerSera; payment or honoraria from AstraZeneca, Daiichi, Seagen, Gilead, Novartis, Lilly, and Genentech. **M.E.** No relevant relationships. **E.B.** No relevant relationships. **P.M.O.** No relevant relationships. **K.A.K.** No relevant relationships. **H.S.** No relevant relationships. **L.W.** No relevant relationships. **J.E.** No relevant relationships. **Y.C.T.S.** Research grants from the National Cancer Institute on topics unrelated to this manuscript. **B.K.** No relevant relationships. **I.W.P.** No relevant relationships. **A.L.C.** Grants or contracts from Novartis and Acrotech. **K.K.** No relevant relationships. **D.S.** No relevant relationships. **W.T.Y.** Research grant with Clarity sponsorship for "Evaluation of individual level breast cancer risk prediction in a cohort of patients with a personal history of breast cancer using a novel software as a medical device;" royalties or licenses from Elsevier. **J.C.C.** Support from the Breast Cancer Research Foundation (BCRF); philanthropic support from M. Neal and R. Neal; National Cancer Institute/National Institutes of Health grant number U01 CA268813; grants or contracts from the Cancer Prevention and Research Institute of Texas (grant no. RP220650); payment or honoraria from Duke and NUS Singapore (August 7, 2022); sole inventor on patent application number 10420838 entitled "Methods for treating cancer using iNOS-inhibitory compositions" held by Houston Methodist Hospital; participation on Merck Triple Negative Breast Cancer Advisory Board (December 13, 2022), Lilly Loxo Advisory Board (December 8, 2022), and BCRF Annual Meeting (October 26, 2022). **S.T.C.W.** No relevant relationships.

References

1. Practice bulletin no. 122: Breast cancer screening. Obstet Gynecol 2011;118(2 Pt 1):372–382.
2. U.S. Department of Health and Human Services. What is breast cancer screening? https://www.cdc.gov/cancer/breast/basic_info/screening.htm. Accessed October 22, 2022.
3. Lauby-Secretan B, Loomis D, Straif K. Breast-Cancer Screening--Viewpoint of the IARC Working Group. N Engl J Med 2015;373(15):1479.
4. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA 2015;314(15):1599–1614. [Published correction appears in JAMA 2016;315(13):1406.]
5. Dahabreh IJ, Wieland LS, Adam GP, Halladay C, Lau J, Trikalinos TA. Core Needle and Open Surgical Biopsy for Diagnosis of Breast Lesions: An Update to the 2009 Report. https://www.ncbi.nlm.nih.gov/books/NBK246878/. Published 2014. Accessed August 20, 2022.
6. Bruening W, Schoelles K, Treadwell J, Launders J, Fontanarosa J, Tipton K. Comparative Effectiveness of Core-Needle and Open Surgical Biopsy for the Diagnosis of Breast Lesions. https://www.ncbi.nlm.nih.gov/books/NBK45220/. Published 2010. Accessed August 20, 2022.
7. D'Orsi CJ. The clinically relevant breast imaging audit. J Breast Imaging 2020;2(1):2–6.
8. Elezaby M, Li G, Bhargavan-Chatfield M, Burnside ES, DeMartini WB. ACR BI-RADS assessment category 4 subdivisions in diagnostic mammography: utilization and outcomes in the national mammography database. Radiology 2018;287(2):416–422.
9. Sickles E, D'Orsi C. ACR BI-RADS Follow-up and Outcome Monitoring. In: ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, Va: American College of Radiology, 2013; 177–189.
10. Masroor I, Rasool M, Saeed SA, Sohail S. To asses inter- and intra-observer variability for breast density and BIRADS assessment categories in mammographic reporting. J Pak Med Assoc 2016;66(2):194–197.
11. Redondo A, Comas M, Macià F, et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. Br J Radiol 2012;85(1019):1465–1470.
12. Gao D, Vahdat LT, Wong S, Chang JC, Mittal V. Microenvironmental regulation of epithelial-mesenchymal transitions in cancer. Cancer Res 2012;72(19):4883–4889.
13. Jin G, Fu C, Zhao H, Cui K, Chang J, Wong ST. A novel method of transcriptional response analysis to facilitate drug repositioning for cancer therapy. Cancer Res 2012;72(1):33–44.
14. Yu KD, Zhu R, Zhan M, et al. Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer. Clin Cancer Res 2013;19(10):2723–2733.
15. Bent CK, Bassett LW, D'Orsi CJ, Sayre JW. The positive predictive value of BI-RADS microcalcification descriptors and final assessment categories. AJR Am J Roentgenol 2010;194(5):1378–1383.
16. Halladay JR, Yankaskas BC, Bowling JM, Alexander C. Positive predictive value of mammography: comparison of interpretations of screening and diagnostic images by the same radiologist and by different radiologists. AJR Am J Roentgenol 2010;195(3):782–785.
17. van Luijt PA, Fracheboud J, Heijnsdijk EA, den Heeten GJ, de Koning HJ; National Evaluation Team for Breast Cancer Screening in Netherlands Study Group (NETB). Nation-wide data on screening performance during the transition to digital mammography: observations in 6 million screens. Eur J Cancer 2013;49(16):3517–3525.
18. Lehman CD, Lee CI, Loving VA, Portillo MS, Peacock S, DeMartini WB. Accuracy and value of breast ultrasound for primary imaging evaluation of symptomatic women 30-39 years of age. AJR Am J Roentgenol 2012;199(5):1169–1177.
19. Ong MS, Mandl KD. National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at $4 billion a year. Health Aff (Millwood) 2015;34(4):576–583.
20. Strigel RM, Burnside ES, Elezaby M, et al. Utility of BI-RADS Assessment Category 4 Subdivisions for Screening Breast MRI. AJR Am J Roentgenol 2017;208(6):1392–1399.
21. McCarthy AM, Keller B, Kontos D, et al. The use of the Gail model, body mass index and SNPs to predict breast cancer among women with abnormal (BI-RADS 4) mammograms. Breast Cancer Res 2015;17(1):1.
22. Clancy K, Aboutalib S, Mohamed A, Sumkin J, Wu S. Deep learning pre-training strategy for mammogram image classification: an evaluation study. J Digit Imaging 2020;33(5):1257–1265.
23. Yala A, Mikhael PG, Strand F, et al. Toward robust mammography-based models for breast cancer risk. Sci Transl Med 2021;13(578):eaba4373.
24. Akselrod-Ballin A, Chorev M, Shoshan Y, et al. Predicting breast cancer by applying deep learning to linked health records and mammograms. Radiology 2019;292(2):331–342.
25. Cui Y, Li Y, Xing D, Bai T, Dong J, Zhu J. Improving the prediction of benign or malignant breast masses using a combination of image biomarkers and clinical parameters. Front Oncol 2021;11:629321. [Published correction appears in Front Oncol 2021;11:694094.]
26. He T, Puppala M, Ezeana CF, et al. A deep learning-based decision support tool for precision risk assessment of breast cancer. JCO Clin Cancer Inform 2019;3(3):1–12.
27. Puppala M, He T, Chen S, et al. METEOR: An Enterprise Health Informatics Environment to Support Evidence-Based Medicine. IEEE Trans Biomed Eng 2015;62(12):2776–2786.
28. Fawcett A. Data Science in 5 Minutes: What is One Hot Encoding? Educative.io. https://www.educative.io/blog/one-hot-encoding. Published February 11, 2021. Accessed August 20, 2022.
29. Hancock JT, Khoshgoftaar TM. Survey on categorical data for neural networks. J Big Data 2020;7:28.
30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–845.
31. SAS Software [computer program]. Version 9.4. Cary, NC: SAS Institute, 2013.
32. Koziełek K, Stranz-Walczak N, Gajdzis P, Karmelita-Katulska K. Evaluation of the positive predictive value (PPV3) of ACR BI-RADS category 4 and 5 based on the outcomes of Invasive Diagnostic Office in an outpatient clinic. Pol J Radiol 2019;84:e185–e189.
33. Ghaemian N, Haji Ghazi Tehrani N, Nabahati M. Accuracy of mammography and ultrasonography and their BI-RADS in detection of breast malignancy. Caspian J Intern Med 2021;12(4):573–579.
34. Kamath J, Cruess DG, Claffey K, Wilson L, Phoenix N, Tannenbaum S. Symptom distress associated with biopsy in women with suspect breast lesions. ISRN Oncol 2012;2012:898327.
35. Lang EV, Berbaum KS, Lutgendorf SK. Large-core breast biopsy: abnormal salivary cortisol profiles associated with uncertainty of diagnosis. Radiology 2009;250(3):631–637.
36. Lotter W, Diab AR, Haslam B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nat Med 2021;27(2):244–249.
37. Marrujo G, Jolly PC, Hall MH. Nonpalpable breast cancer: needle-localized biopsy for diagnosis and considerations for treatment. Am J Surg 1986;151(5):599–602.
38. Meyer JE, Kopans DB, Stomper PC, Lindfors KK. Occult breast abnormalities: percutaneous preoperative needle localization. Radiology 1984;150(2):335–337.
39. Meyer JE, Eberlein TJ, Stomper PC, Sonnenfeld MR. Biopsy of occult breast lesions. Analysis of 1261 abnormalities. JAMA 1990;263(17):2341–2343.
40. Rosenberg AL, Schwartz GF, Feig SA, Patchefsky AS. Clinically occult breast lesions: localization and significance. Radiology 1987;162(1 Pt 1):167–170.
41. Wiratkapun C, Bunyapaiboonsri W, Wibulpolprasert B, Lertsithichai P. Biopsy rate and positive predictive value for breast cancer in BI-RADS category 4 breast lesions. J Med Assoc Thai 2010;93(7):830–837.
42. Ezeana CF, Puppala M, Wang L, Chang JC, Wong STC. A comparative efficacy study of diagnostic digital breast tomosynthesis and digital mammography in BI-RADS 4 breast cancer diagnosis. Eur J Radiol 2022;153:110361.