

TITLE: Predicting the likelihood of carrying a *BRCA1* or *BRCA2* mutation in Asian breast cancer patients

Boon Hong Ang¹, Weang Kee Ho^{1, 2}, Eldarina Wijaya¹, Pui Yoke Kwan¹, Pei Sze Ng¹, Sook Yee Yoon¹, Siti Norhidayu Hasan¹, Joanna Mei Ch'wan Lim¹, Tiara Hasan¹, Mei-Chee Tai¹, Jamie Allen³, Andrew Lee³, Nur Aishah Mohd Taib^{4,5}, Cheng Har Yip⁶, Mikael Hartman⁷, Swee Ho Lim⁸, Ern Yu Tan⁹, Benita Kiat Tee Tan^{10,11,12}, Su-Ming Tan¹³, Veronique Kiak Mien Tan^{10,11}, Peh Joo Ho¹⁴, Alexis Jiaying Khng¹⁴, Alison M. Dunning¹⁵, Jingmei Li^{14,16}, Douglas F. Easton^{3,16}, Antonis C. Antoniou^{3,16}, Soo Hwang Teo^{1,4,16*}

AFFILIATIONS

¹Cancer Research Malaysia, 1, Jalan SS12/1A, 47500 Subang Jaya, Malaysia

²School of Mathematical Sciences, Faculty of Science and Engineering, University of Nottingham Malaysia, Jalan Broga, Semenyih 43500 Selangor, Malaysia

³Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

⁴University Malaya Cancer Research Institute, Faculty of Medicine, University of Malaya, Jalan Universiti, 50630 Kuala Lumpur, Malaysia

⁵Department of Surgery, Faculty of Medicine, University of Malaya, Jalan Universiti, 50630 Kuala Lumpur, Malaysia

⁶Subang Jaya Medical Centre, No. 1, Jalan SS12/1A, 47500 Subang Jaya, Malaysia

⁷Department of Surgery, National University Hospital and NUHS, 1E Kent Ridge Road, 119228 Singapore, Singapore

⁸Breast Department, KK Women's and Children's Hospital, Singapore 100 Bukit Timah Road, 229899, Singapore

⁹Department of General Surgery, Tan Tock Seng Hospital, Singapore 308433, Singapore

¹⁰Division of Surgical Oncology, National Cancer Centre, Singapore, Singapore

¹¹Department of General Surgery, Singapore General Hospital, Singapore, Singapore

¹²Department of General Surgery, Sengkang General Hospital, Singapore

¹³Division of Breast Surgery, Department of General Surgery, Changi General Hospital, Singapore, Singapore

¹⁴Human Genetics, Genome Institute of Singapore, 60 Biopolis St, 138672 Singapore, Singapore

¹⁵Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, United Kingdom

¹⁶Joint senior authors

#Malaysian Breast Cancer Genetic Study (MyBrCa) and Singapore Breast Cancer Cohort (SGBCC) contributors are listed in the Acknowledgments section.

FINANCIAL SUPPORT

This study was supported in whole, or in part, by Wellcome Trust [Grant number: v203477/Z/16/Z]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

The Malaysian Breast Cancer Genetic Study was established using funds from the Malaysian Ministry of Science, and the Malaysian Ministry of Higher Education High Impact Research Grant [Grant number: UM.C/HIR/MOHE/06], and additional funding was received from Yayasan Sime Darby, Yayasan PETRONAS, Estee Lauder Group of Companies, Khind Starfish Foundation, and other donors of Cancer Research Malaysia.

SGBCC was supported by the National Research Foundation Singapore [Grant number: (NRF-NRFF2017-02, awarded to JL Li)], National University of Singapore (NUS) start-up Grant, National University Cancer Institute Singapore (NCIS) Centre Grant, [Grant number: NMRC/CG/NCIS/2010, NMRC/CG/012/2013, CGAug16M005], Breast Cancer Prevention Programme (BCPP), Asian Breast Cancer Research Fund, and the National Medical Research Council (NMRC) Clinician Scientist Award (SI Category) [Grant

number: NMRC/CSA-SI/0015/2017]. ACA is supported by Cancer Research UK [Grant number: (C12292/ A20861, and PPRPGM-Nov20\100002)].

***CORRESPONDING AUTHOR**

Soo Hwang Teo: Cancer Research Malaysia, 1 Jalan SS12/1A, 47500 Subang Jaya, Selangor, Malaysia, Email: soohwang.teo@cancerresearch.my; Phone + 603 5650 9797

RUNNING HEAD: BRCA germline carrier prediction in Asians

ABSTRACT WORD COUNT: 257

CONTEXT SUMMARY WORD COUNT: 150

TEXT WORD COUNT: 2999

ABSTRACT

PURPOSE With the development of PARP inhibitors for treatment of cancer patients with an altered *BRCA1* or *BRCA2* gene, there is an urgent need to ensure that there are appropriate strategies for identifying mutation carriers whilst balancing the increased demand for and cost of cancer genetics services. To date, the majority of mutation prediction tools have been developed in women of European descent where the age and cancer-subtype distributions are different from that in Asian women.

METHODS In this study, we built a new model (ARiCa: Asian Risk Calculator) for estimating the likelihood of carrying a pathogenic variant in *BRCA1* or *BRCA2* gene, using germline *BRCA* genetic testing results in a cross-sectional population-based study of 8,162 Asian breast cancer patients. We compared the model performance to existing mutation prediction models. The models were evaluated for discrimination and calibration.

RESULTS ARiCa included age of diagnosis, ethnicity, bilateral breast cancer, tumour biomarkers, and family history of breast cancer or ovarian cancer as predictors. The inclusion of tumour grade improved significantly the model performance. The full model was calibrated (Hosmer-Lemeshow p-value=0.614) and discriminated well between *BRCA* and non-*BRCA* pathogenic variant carriers (Area Under Receiver Operating Curve 0.80, 95% Confidence Interval=0.75-0.84). Addition of grade to the existing clinical genetic testing criteria targeting breast cancer patients below 45 years reduced the proportion of patients referred for genetic counselling and testing from 37% to 33% (p-value=0.003), thereby improving the overall efficacy.

CONCLUSION Population-specific customisation of mutation prediction models and clinical genetic testing criteria improved the accuracy of BRCA mutation prediction in Asian patients.

CONTEXT SUMMARY

Key objective: Increasing breast cancer incidence and limited resources pose a significant challenge to genetic counselling and testing in many low- and middle-income countries in Asia. Whilst existing mutation prediction models underestimate proportion of carriers in Asian women, a logistic regression model was developed and validated based on South East Asian breast cancer patients unselected for age of diagnosis and family history of cancer to estimate the likelihood of carrying a pathogenic variant in *BRCA1* or *BRCA2* gene, called ARiCa (**A**siAn **R**isk **C**alculator).

Knowledge generated: ARiCa outperformed existing mutation prediction models. Discrimination of mutation prediction model and efficacy of clinical genetic testing criteria were significantly enhanced by the inclusion of tumour grade.

Relevance: Population-specific customisation of mutation prediction tools is important to enable more accurate BRCA mutation prediction in diverse populations for referral of breast cancer patients for genetic counselling and testing who may benefit from the selection of therapy.

INTRODUCTION

Germline genetic testing for *BRCA1* and *BRCA2* (*BRCA*) has enabled risk management for individuals at elevated cancer risk, and with the advent of PARP inhibitor treatment, enabled treatment selection with improved outcomes.¹ In high resource countries, clinical genetics services are well established and patients are referred for germline *BRCA* genetic testing using criteria based on age of onset of cancer, breast cancer histology, and cancer family history.^{2,3}

Similar genetic testing criteria have been incorporated into clinical practice guidelines in Asian countries, but these pose significant resource challenges, particularly in low- and middle-income Asian countries with limited clinical genetics services.⁴ Notably, because of the shifting reproductive and behavioural patterns, the incidence in many Asian countries have doubled or tripled in the past 40 years.⁵ This dramatic increase in incidence in younger generations means that the mean age of diagnosis for breast cancer in Asian women is approximately 10 years younger than that in European women,^{6,7} thus, a higher proportion of breast cancer patients fulfil clinical genetic testing criteria for referral, exacerbating the challenges in access to genetic services in Asian countries.⁴

In European or North American populations, models have been developed for predicting the likelihood of carrying germline *BRCA* pathogenic variants (PVs), mostly using data of breast cancer patients ascertained through genetic clinics, and these models are well-calibrated for these populations.^{8,9,10,11,12,13,14,15} Evaluation of such models in high-risk breast cancer patients of Asian descent living in North America¹⁶ or in Asia^{17,18,19,20,21}

showed that these models underestimated the proportion of *BRCA* PVs carriers, especially for *BRCA2* PVs carriers^{16,17,20,21} and for breast cancer patients with no family history of breast cancer.^{17,18} Recently, a *BRCA* carrier prediction algorithm (KOHCal) was developed in South Korea, based on high-risk breast cancer patients and was found to have better discrimination and calibration in South Koreans than models built on women of European descent.²²

With the approval of PARP inhibitors for treatment of breast and ovarian cancer patients with *BRCA* PVs, there is an urgent need to determine the performance of these models in diverse populations. To date, no studies have evaluated the performance of clinical genetic testing criteria or developed a *BRCA* carrier prediction model in a population-based study of Asian breast cancer patients. In this study, we evaluated the performance of existing *BRCA* carrier prediction models, developed a new prediction model, and customised clinical genetic testing criteria in a population-based study of 8,162 Asian breast cancer patients from Malaysia and Singapore unselected for age of diagnosis and family history of cancer.

METHODS

Study population

The study participants were women diagnosed clinically with breast cancer (invasive and non-invasive) who were recruited in the Malaysian Breast Cancer Genetic (MyBrCa) study²³ and the Singapore Breast Cancer Cohort (SGBCC) study. Cases were recruited from two hospitals in Malaysia and six hospitals in Singapore. Germline DNA were

sequenced in two batches, using targeted sequencing panels described previously.²⁴ Carriers of pathogenic variants in non-BRCA genes were treated as non-carriers.

Statistical analyses

Existing BRCA carrier prediction models

Three existing BRCA carrier prediction models were evaluated in this study; two empirical models (PENNI, KOHCal) and a genetic risk model BOADICEA 5.0.²⁵ Model performance was determined based on model calibration, assessed using Hosmer-Lemeshow (HL) test, and discrimination, assessed using area under receiver operating curve (AUC).

Development and validation of population-specific BRCA carrier prediction model

The study sample was randomly split into training and validation sets, comprising 70% and 30% of the samples, respectively. Candidate predictors of BRCA PV included age of breast cancer diagnosis, ethnicity, bilateral breast cancer, pathological features, and family history of breast or ovarian cancer (**Supplementary Table 1**). Missing data in the training set were imputed using multiple imputation by chained equations, whilst missing data in the validation set were imputed using single or multiple imputation by chained equations, under the missing at random (MAR) assumption.^{26,27} Given that multiple imputation generates more than one imputed dataset, results for single imputation in the validation dataset is presented in the main Figures and Tables, with results of multiple imputation included in Supplementary data where relevant. Additional sensitivity test was performed to ensure that the validation test results are comparable after single and

multiple imputations. BRCA carrier prediction models were built based on logistic regression method using the training set. Model calibration and discrimination were evaluated in the validation set using HL test and AUC, respectively. The optimal carrier probability threshold for genetic testing was chosen based on the intersection of sensitivity and specificity curves.²⁸

Customisation and evaluation of existing clinical criteria for germline BRCA genetic testing

Modified Clinical Criteria (MCC) were developed starting with the MCGplus Criteria² by considering combinations of age of diagnosis of proband in 5 year intervals, with and without considering grade, resulting in a total of 96 different categories. The efficacy of MCC was evaluated in the validation set based on detection ratio (number of patients to be screened to detect one carrier).

All the data were analysed using Stata version 13.0 (Stata Corp., College Station., Texas, USA) and a p-value < 0.05 (two-tailed) was deemed to be statistically significant. See Supplemental Methods for more details.

RESULTS

Study population characteristics

In this cross-sectional population-based study of 8,162 breast cancer patients, 323 (4.0%) had germline *BRCA1* or *BRCA2* PVs (**Supplementary Table 1**). The majority of patients were Chinese (75.4%), with a mean age of diagnosis of 52.3 years (SD=10.77).

Compared to Chinese women, Indian women had a higher proportion of HR- and TNBC breast cancer cases, whereas Malay women had higher proportions of HER2+ and Luminal B breast cancer cases. There was a higher proportion of carriers amongst Indian and Malay women. Whilst the tumour characteristics tested were significantly associated with *BRCA1* status, these were also associated with *BRCA2* status, with the exception of ER, PR, HR, and TNBC (**Supplementary Table 2**).

Development and validation of population-specific BRCA carrier prediction model

Prediction models were developed using 5,714 breast cancer cases (228 *BRCA* carriers) and validated using 2,448 cases (95 *BRCA* carriers) (**Supplementary Fig 1**). Collinearity tests showed that ER, PR, HER2, TNBC, HR, and immune-histochemical subtypes were correlated (correlation coefficients, $r > 0.40$). Hence, six combinations of tumour biomarkers along with the remaining predictors were considered in the analyses: (a) TNBC, (b) ER, (c) ER and HER2, (d) HR and HER2, (e) HER2, and (f) immune-histochemical subtypes. The best-performing model was selected based on the highest AUC and the lowest non-significant HL score in the validation set (**Supplementary Table 3**). Model (a) (AUC=0.86, HL=2.63) and Model (e) (AUC=0.75, HL=10.89) were the best-performing models for *BRCA1* and *BRCA2* PVs carrier status, whereas Model (c) was the best-performing model for overall *BRCA* (AUC=0.80, HL=5.43). The predictive performance of Model (c) by mutation-type were similar to the respective best-performing models (*BRCA1*: Model (c) versus Model (a) – AUC (HL): 0.86 (4.33) versus 0.86 (2.63); *BRCA2*: Model (c) versus Model (e) – AUC (HL): 0.75 (12.15) versus 0.75 (10.89)). Analyses after multiple imputation showed similar results. Hence, Model (c), subsequently

referred to as ARiCa (**A**si**R**i**C**a), was selected as the final model for predicting overall *BRCA* PVs carrier status. We evaluated the performance of ARiCa by ethnicity and found that the model had high discriminatory power and well-calibrated across ethnic groups (**Supplementary Table 4**).

In ARiCa, younger age of diagnosis, Indian ethnicity, bilateral breast cancer, ER-negativity, HER2-negativity, higher grade, and presence of first degree family history of breast or ovarian cancer were associated with overall *BRCA* PVs carrier status (**Table 1**). These variables were also associated with *BRCA1* PVs carrier status except grade, whereas *BRCA2* was only associated with younger age, HER2-negativity, higher grade and first degree family history of breast cancer. Notably, both Malay and Indian ethnicities were associated with higher odds of being *BRCA1* PVs carriers compared to Chinese ethnicity.

We determined the optimal carrier probability threshold for ARiCa in the validation set as the intercept of sensitivity and specificity (**Supplementary Fig 2**). At the optimal threshold, corresponding to a mutation prevalence of 4%, 31% (95%CI=29-33) of breast cancer patients would require germline *BRCA* genetic testing and 71% (95%CI=61-80) of *BRCA* PVs carriers would be identified (**Supplementary Table 5**). Performance of ARiCa were consistent across imputed validation sets after multiple imputation.

Comparison of BRCA carrier prediction models

We compared the performance of ARiCa with models which have been developed in other populations using data for 2,426 patients from the validation set for whom data are available for variables required in all considered models. For overall *BRCA*, ARiCa had the highest AUC (0.80), followed by PENNII (0.74), BOADICEA (0.73), and KOHCal (0.71) (**Fig 1**). The AUCs for *BRCA1* were similar across models, but ARiCa had significantly better discriminatory ability than PENNII, BOADICEA, and KOHCal for *BRCA2* (0.75, 0.69, 0.65, and 0.63 respectively).

All models were well-calibrated; ARiCa had the lowest HL for overall *BRCA* (**Fig 2**). There was no significant difference between the observed proportion and expected probability for *BRCA1* and *BRCA2* PVs carriers and a majority were distributed close to the bisector (**Supplementary Fig 3**).

We compared the efficacy measures (sensitivity, specificity) at the optimal and the conventional 10% and 20% thresholds.^{17,21,29} All models had poor sensitivity at the 10% and 20%, so we focused on the lower optimal thresholds (**Table 2**).¹⁷ At the respective optimal thresholds for each model, the sensitivities of all models for overall *BRCA* were 63-71% and the specificities were 67-71%. Whilst all models achieved a sensitivity of 83% for *BRCA1*, KOHCal (56%), BOADICEA (56%), and PENNII (51%) had lower sensitivity for *BRCA2* than ARiCa (66%). ARiCa achieved relatively high sensitivity and specificity for overall *BRCA* (71%, 71%), *BRCA1* (83%, 70%), and *BRCA2* (66%, 70%) at the optimal threshold.

Customisation and evaluation of existing clinical criteria for germline *BRCA* genetic testing

We evaluated the NCCN and Mainstreaming Cancer Genetics (UK) clinical genetic testing criteria in the validation set. Applying the NCCN and MCG Criteria would lead to 37% and 39% being referred with 72% and 69% of *BRCA* PVs carriers identified, respectively. Addition of family history variables to the MCG Criteria (MCGPlus) increased the screening rate from 39% to 49% and improved the detection rate from 69% to 81% (data not shown). Whilst the expanded NCCN Criteria detected the highest detection rate (96%), more than three-quarters of breast cancer patients (88%) would need to be screened.

Given that patients in this study had a younger age of diagnosis for breast cancer than those in the Western populations and grade was a significant predictor of *BRCA* PVs carrier status, we customised MCGplus Criteria by considering several combinations of age of diagnosis for breast cancer and higher-grade breast cancer (grade 2 or 3) to improve the overall efficacy. Applying 96 Modified Clinical Criteria (48 MCC with grade and 48 MCC without grade) (**Fig 3**), we found that the detection rates for *BRCA1* (78-92%) were higher than for *BRCA2* (46-93%). Notably, at similar detection rate, the addition of grade resulted in reduction in screening rate of 1% to 10% (average=4%) for overall *BRCA*. Similarly, the addition of grade resulted in reduction in screening rate of 3% to 12% (average=5%) for existing clinical genetic testing criteria. There was no difference in reduction rates between *BRCA1* and *BRCA2*.

We identified 3 clinical criteria categories (MCC 17, 29, 33) from **Fig 3** with similar screening rates to ARiCa for overall *BRCA* (screening rate=31%) (**Supplementary Table 5**). These criteria had identical criteria for grade 2 or 3 breast cancer (≤ 40) and bilateral breast cancer (≤ 60), but they had different thresholds for age of diagnosis of proband with TNBC and family history of breast or ovarian cancer (**Table 3**).

We also identified 3 clinical criteria categories (NCCN with grade, MCC 10, 45) from **Fig 3** with similar detection rates to ARiCa for overall *BRCA* (detection rate=71%) (**Supplementary Table 5**). These criteria had different combinations of age of diagnosis of proband with grade 2 or 3 breast cancer, TNBC, bilateral breast cancer, and family history of breast or ovarian cancer (**Table 3**).

All 6 modified criteria resulted in lower detection ratios, when compared to existing clinical genetic testing criteria (Expanded NCCN=24:1; MCGplus=15:1; MCG=14:1; NCCN=13:1) (**Table 3**). NCCN with grade outperformed the 5 modified criteria by achieving a higher detection rate (69%) at the lowest detection ratio (12:1). Nonetheless, all 6 modified criteria still underperformed ARiCa. For instance, MCC (17, 29, 33) had lower detection rates of 63-66% compared to 71% for ARiCa. Similarly, NCCN with grade and MCC (10, 45) had higher screening rate of 33-38% compared to 31% for ARiCa (**Table 3**).

DISCUSSION

Whilst germline *BRCA1* or *BRCA2* PVs testing has an established role in risk management, this is increasingly relevant in the selection of therapy.¹ We showed that logistic regression models built based on a large Asian population-based study of breast cancer patients, unselected for age of diagnosis and family history, outperformed the genetic risk model (BOADICEA) developed using data on European-ancestry populations and the empirical models (PENNII, KOHCal) developed using breast cancer patients with early onset or familial breast cancer. The Modified Clinical Criteria (MCC) customised to the Asian breast cancer patients in combination with presence of grade were more efficient than existing clinical genetic testing criteria.

In multivariable regression analyses, we found that the risk factors significantly associated with *BRCA* PVs carrier status in this study were consistent with previously published findings from Asian countries, including younger age of diagnosis, bilateral breast cancer, ER-negative status, HER2-negative status, higher grade, and presence of first degree family history of breast cancer or ovarian cancer.^{20,22,30,31,32,33,34,35,36,37,38}

We found that all the *BRCA* carrier prediction models and the Modified Clinical Criteria (MCC) were more sensitive (sensitivity/detection rate) and accurate (discrimination) for *BRCA1* than *BRCA2*, which is likely to be driven by the stronger association between *BRCA1* and the ER-negative status, TNBC subtype, and ovarian cancer family history.^{39,40,41,42,43} Indeed, several studies have previously demonstrated that the use of pathologic characteristics, namely ER and TNBC improved the sensitivity and

discrimination for *BRCA1* when selecting individuals for germline genetic testing in high-risk breast cancer patients.^{44,45,46,47,48}

All three existing models tested performed similarly in our study population (AUC=0.71-0.74) as previously reported in other Asian populations (AUC=0.69-0.76), but the AUCs were lower than those reported in women of European descent, especially for BOADICEA (AUC=0.77).^{13,17,22,29} BOADICEA had lower discriminatory ability for overall *BRCA* and *BRCA2*, consistent with the observation that BOADICEA performed better at lower thresholds because it underestimated carrier probability of Asian breast cancer patients with the lowest sensitivity at conventional thresholds, particularly those with germline *BRCA2* PVs.¹⁷ Nevertheless, BOADICEA outperformed PENNII for *BRCA1*.¹⁷ These observations are not surprising. Given that BOADICEA is a genetic risk model, it relies on population-specific parameters for breast cancer incidences, PV frequencies, and tumour-pathology distributions as input parameters. Customisation of BOADICEA using population-specific parameters and addition of tumour grade (a clear predictor of carrier status in our analysis) could substantially improve its discrimination.

In terms of calibration, models built on Asian populations had better calibration for *BRCA2*, whilst models built on women of European descent had better calibration for *BRCA1*. Notably, this was evident in KOHCal and BOADICEA that appeared to be calibrated with the lowest HL for *BRCA2* and *BRCA1*, respectively. A possible explanation could be due to the variation in mutation prevalence. Whilst *BRCA2* mutations are more

common than *BRCA1* mutations in Asian, it is the opposite in many European populations.¹⁷

Notwithstanding BRCA carrier prediction models have good discrimination, there are challenges in their implementation in resource constrained settings. Clinical genetic testing criteria are likely to continue as a mainstay for referral of patients for genetic counselling and testing. In our evaluation, NCCN with grade had a sensitivity of 69% at a screening rate of 33%. This is marginally better than the existing NCCN and MCG criteria where at similar detection rates of 69-72%, about 37-39% patients would be referred for genetic counselling and testing. It is possible that this improvement is because of the age threshold for TNBC (≤ 60 vs no age restriction), bilateral breast cancer (46-50 vs ≤ 60), and first degree family history of breast cancer (46-50 vs no age restriction), but the inclusion of grade is also an important consideration. Indeed, higher grade was identified as a strong predictor not only for *BRCA1* PVs carriers but also for *BRCA2* PVs carriers.⁴² Previous studies have shown that inclusion of grade can improve the sensitivity and discrimination of germline *BRCA* PVs prediction in high-risk breast cancer patients.^{44,45} Given that *BRCA1* (91%) and *BRCA2* (89%) PVs carriers were of higher grade than in non-carriers (76%), future improvements in BRCA carrier prediction tools could include grade.^{45,49}

Limitations and strengths

The validation sample was relatively small with only 95 *BRCA* PVs carriers. Future independent studies should aim to assess the models developed here. The analysis was

also restricted to *BRCA1* or *BRCA2*, but gene-panels that include additional susceptibility genes are now widely used, which include additional genes (e.g., *PALB2*) that may be relevant in informing treatment.³ However, the present sample size is too small to allow the prediction of carrying PVs in other genes. Although grade was identified as a potential variable to include in risk prediction models, it is noteworthy that quality assurance may be required for this and other variables in order to ensure model accuracy. Finally, whilst ARiCa was shown to perform equally well across different ethnic groups in Malaysia and Singapore, studies in other Asian populations are needed to evaluate its utility in these populations.

Despite the limitations, this is the first study to develop a logistic regression BRCA carrier prediction model and customise clinical genetic testing criteria for use in mainstream germline *BRCA* genetic testing based on unselected sample of breast cancer patients in South East Asia.

CONCLUSION

With the advent of germline genetic testing for treatment selection, more women may consider genetic testing as part of their treatment plans. Given that Asian women have a younger age of diagnosis for breast cancer and different distribution of breast cancer subtypes compared to women of European descent, population-specific customisation of BRCA carrier prediction tools is important to enable more accurate BRCA mutation prediction in diverse populations.

AUTHORS CONTRIBUTIONS AND COMPETING INTEREST STATEMENT

Conception and design: BHA, WKH, PYK, ACA, DFE, SHT

Administrative support (Data and sample management): EW, MCT, JA, SNH, TH

Provision of study materials or patients (Provided DNA samples and/or phenotypic data): NAMT, CHY, MH, SHL, EYT, BKTT, SMT, VKMT, JL

Collection and assembly of data (Panel gene sequencing and Sanger validation):
AMD, PSN, SYY, SNH, JMCL, TH, PJH, AJK

Data analysis and interpretation (Bioinformatics and statistical analyses): BHA, WKH, EW, MCT, JA, AL

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

The authors confirm that they have no competing financial interests.

ACKNOWLEDGEMENTS

MyBrCa thanks study participants and all research staff at Cancer Research Malaysia, University Malaya, and Sime Darby Medical Centre who assisted in recruitment and interviews (including, Meow Keong Thong, Daphne SC Lee, Sheau-Yee Lee, Sze-Yee Phuah, Kah-Nyin Lai, Shao Yan Lau, Shivaani Mariapun, Siu Wan Wong, Heamanthaa Padmanabhan, Hanani Che Halim, Leelavathy Krishnan) for their contributions and commitment to this study. For SGBCC, we want to thank the program manager Jenny Liu, clinical research coordinators/research assistants Siew-Li Tan, Siok-Hoon Yeo, Ting-Ting Koh, Amanda Ong, Jin-Yee Lee, Michelle Mok, Ying-Jia Chew, Jing-Jing Hong, and

Hui-Min Lau for their contributions in recruitment, Yen-Shing Yeoh for data preparation, and Alexis Khng for processing the DNA samples. We also want to thank all the participants' support to SGBCC.

REFERENCES

1. Tung NM, Garber JE: BRCA 1/2 testing: therapeutic implications for breast cancer management. *Br J Cancer* 119:141, 2018
2. Kemp Z, Turnbull A, Yost S, et al: Evaluation of cancer-based criteria for use in mainstream BRCA1 and BRCA2 genetic testing in patients with breast cancer. *JAMA Netw Open* 2:e194428-e194428, 2019
3. Yadav S, Hu C, Hart SN, et al: Evaluation of Germline Genetic Testing Criteria in a Hospital-Based Series of Women with Breast Cancer. *J Clin Oncol* 38:1409, 2020
4. Nakamura S, Kwong A, Kim SW, et al: Current status of the management of hereditary breast and ovarian cancer in Asia: first report by the Asian BRCA consortium. *Public Health Genomics* 19:53-60, 2016
5. Boyle P, Levin B: World cancer report 2008. IARC Press, International Agency for Research on Cancer, 2008
6. Lin CH, Yap YS, Lee KH, et al: Contrasting epidemiology and clinicopathology of female breast cancer in Asians vs the US population. *J Natl Cancer Inst* 111:1298-1306, 2019
7. Song QK, Li J, Huang R, et al: Age of diagnosis of breast cancer in china: almost 10 years earlier than in the United States and the European union. *Asian Pac J Cancer Prev* 15:10021-10025, 2014
8. Apicella C, Andrews L, Hodgson SV, et al: Log odds of carrying an Ancestral Mutation in BRCA1 or BRCA2 for a Defined personal and family history in an Ashkenazi Jewish woman (LAMBDA). *Breast Cancer Res* 5:R206, 2003

9. Berry DA, Iversen Jr ES, Gudbjartsson DF, et al: BRCAPRO validation, sensitivity of genetic testing of BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes. *J Clin Oncol* 20:2701-2712, 2002
10. Couch FJ, DeShano ML, Blackwood MA, et al: BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer. *N Engl J Med* 336:1409-1415, 1997
11. de la Hoya M, Osorio A, Godino J, et al: Association between BRCA1 and BRCA2 mutations and cancer phenotype in Spanish breast/ovarian cancer families: implications for genetic testing. *Int J Cancer* 97:466-471, 2002
12. Frank TS, Deffenbaugh AM, Reid JE, et al: Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10,000 individuals. *J Clin Oncol* 20:1480-1490, 2002
13. Lindor NM, Johnson KJ, Harve H, et al: Predicting BRCA1 and BRCA2 gene mutation carriers: comparison of PENN II model to previous study. *Fam Cancer* 9:495-502, 2010
14. Shattuck-Eidens D, Oliphant A, McClure M, et al: BRCA1 sequence analysis in women at high risk for susceptibility mutations: risk factor analysis and implications for genetic testing. *JAMA* 278:1242-1250, 1997
15. Vahteristo P, Eerola H, Tamminen A, et al: A probability model for predicting BRCA1 and BRCA2 mutations in breast and breast-ovarian cancer families. *Br J Cancer* 84:704-708, 2001
16. Kurian AW, Gong, GD, Chun NM, et al: Performance of BRCA1/2 mutation prediction models in Asian Americans. *J Clin Oncol* 26:4752, 2008

17. Hung FH, Wang YA, Jian JW, et al: Evaluating BRCA mutation risk predictive models in a Chinese cohort in Taiwan. *Sci Rep* 9:1-10, 2019
18. Kang E, Park SK, Yang JJ, et al: Accuracy of BRCA1/2 mutation prediction models in Korean breast cancer patients. *Breast Cancer Res Treat* 134:1189-1197, 2012
19. Kim H, Choi DH: Distribution of BRCA1 and BRCA2 mutations in Asian patients with breast cancer. *J Breast Cancer* 16:357-365, 2013
20. Thirthagiri E, Lee SY, Kang P, et al: Evaluation of BRCA1 and BRCA2 mutations and risk-prediction models in a typical Asian country (Malaysia) with a relatively low incidence of breast cancer. *Breast Cancer Res* 10:R59, 2008
21. Kwong A, Wong CH, Suen DT, et al: Accuracy of BRCA1/2 mutation prediction models for different ethnicities and genders: experience in a southern Chinese cohort. *World J Surg* 36:702-713, 2012
22. Kang E, Park SK, Lee, JW, et al: KOHBRA BRCA risk calculator (KOHCal): a model for predicting BRCA1 and BRCA2 mutations in Korean breast cancer patients. *J Hum Genet*, 61:365-371, 2016
23. Tan MM, Ho WK, Yoon SY, et al: A case-control study of breast cancer risk factors in 7,663 women in Malaysia. *PloS One* 13:e0203469, 2018
24. Dorling et al: Breast cancer risk genes: association analysis of rare coding variants in 34 genes in 60,466 cases and 53,461 controls. *N Engl J Med*, 2021
25. Lee A, Mavaddat N, Wilcox AN, et al: BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med* 21:1708-1718, 2019

26. Howlader N, Cronin KA, Kurian AW, et al: Differences in breast cancer survival by molecular subtypes in the United States. *Cancer Epidemiol Biomarkers Prev* 27:619-626, 2018
27. Stuart EA, Azur M, Frangakis C, et al: Multiple imputation with large data sets: a case study of the Children's Mental Health Initiative. *Am JEpidemiol* 169:1133-1139, 2009
28. Hosmer DW, Lemeshow SL: *Applied logistic regression* (ed 2). New York, A Wiley-Interscience, 2000
29. Antoniou AC, Hardy R, Walker L, et al: Predicting the likelihood of carrying a BRCA1 or BRCA2 mutation: validation of BOADICEA, BRCAPRO, IBIS, Myriad and the Manchester scoring system using data from UK genetics clinics. *J Med Genet* 45:425-431, 2008
30. Kwong A, Wong LP, Wong HN, et al: Clinical and pathological characteristics of Chinese patients with BRCA related breast cancer. *Hugo J* 3:63-76, 2009
31. Liu Y, Ide Y, Inuzuka M, et al: BRCA1/BRCA2 mutations in Japanese women with ductal carcinoma in situ. *Mol Genet Genomic Med* 7:e493, 2019
32. Noh JM, Han BK, Choi DH, et al: Association between BRCA mutation status, pathological findings, and magnetic resonance imaging features in patients with breast cancer at risk for the mutation. *J Breast Cancer* 16:308-314, 2013
33. Wong ES, Shekar S, Chan CH, et al: Predictive factors for BRCA1 and BRCA2 genetic testing in an Asian clinic-based population. *PloS One* 10:e0134408, 2015

34. Yip CH, Taib NA, Choo WY, et al: Clinical and pathologic differences between BRCA1-, BRCA2-, and non-BRCA-associated breast cancers in a multiracial developing country. *World J Surg* 33:2077-2081, 2009
35. Yu JH, Lee JW, Son BH, et al: Characteristics of BRCA1/2 mutation-positive breast cancers in Korea: a comparison study based on multicenter data and the Korean Breast Cancer Registry. *J Breast Cancer* 17:129-135, 2014
36. Yang XR, Devi BC, Sung H, et al: Prevalence and spectrum of germline rare variants in BRCA1/2 and PALB2 among breast cancer cases in Sarawak, Malaysia. *Breast Cancer Res Treat* 165:687-97, 2017
37. Kwong A, Shin VY, Ho JC, et al: Comprehensive spectrum of BRCA1 and BRCA2 deleterious mutations in breast cancer in Asian countries. *J Med Genet* 5:15-23, 2016
38. Wen WX, Allen J, Lai KN, et al: Inherited mutations in BRCA1 and BRCA2 in an unselected multiethnic cohort of Asian patients with breast cancer and healthy controls from Malaysia. *J Med Genet* 55:97-103, 2018
39. Chen H, Wu J, Zhang Z, et al: Association between BRCA status and triple-negative breast cancer: a meta-analysis. *Front Pharmacol* 9:909, 2018
40. Lakhani SR, Van De Vijver MJ, Jacquemier J, et al: The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2. *J Clin Oncol* 20:2310-2318, 2002

41. Tung N, Wang Y, Collins LC, et al: Estrogen receptor positive breast cancers in BRCA1 mutation carriers: clinical risk factors and pathologic features. *Breast Cancer Res* 12:R12, 2010
42. Spurdle AB, Couch FJ, Parsons MT, et al: Refined histopathological predictors of BRCA1 and BRCA2 mutation status: a large-scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia. *Breast Cancer Res* 16:1-6, 2014
43. Ford D, Easton DF, Stratton M, et al: Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *Am J Hum Genet* 62:676-89, 1998
44. Farshid G, Balleine RL, Cummings M, et al: Morphology of breast cancer as a means of triage of patients for BRCA1 genetic testing. *Am J Surg Pathol* 30:1357-1366, 2006
45. James PA, Doherty R, Harris M, et al: Optimal selection of individuals for BRCA mutation testing: a comparison of available methods. *J Clin Oncol* 24:707-715, 2006
46. Lidereau R, Eisinger F, Champègne MH, et al: Major improvement in the efficacy of BRCA1 mutation screening using morphoclinical features of breast cancer. *Cancer Res* 60:1206-1210, 2000
47. Phuah SY, Looi LM, Hassan N, et al: Triple-negative breast cancer and PTEN (phosphatase and tensin homologue) loss are predictors of BRCA1 germline mutations in women with early-onset and familial breast cancer, but not in women with isolated late-onset breast cancer. *Breast Cancer Res* 14:R142, 2012

48. Robertson L, Hanson H, Seal S, et al: BRCA1 testing should be offered to individuals with triple-negative breast cancer diagnosed below 50 years. *Br J Cancer* 106:1234-1238, 2012
49. Stratton MR: Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. *The Lancet* 349:1505-1510, 1997

FIGURE LEGENDS

Fig 1. ROC curves of BRCA carrier prediction models

Sample: 2426 MyBrCa and SGBCC breast cancer patients with pedigree available in single imputed validation set.

Fig 2. Observed proportion and expected probability of BRCA carrier prediction models

Sample: 2426 MyBrCa and SGBCC breast cancer patients with pedigree available in single imputed validation set.

Abbreviation: HL, Hosmer-Lemeshow.

Fig 3. Eligible patients and BRCA PVs carriers detected for clinical criteria with and without grade

Sample: 2448 MyBrCa and SGBCC breast cancer patients in single imputed validation set.

Abbreviation: MCC, Modified Clinical Criteria.

Supplementary Fig 1. Study design and sample selection

Supplementary Fig 2. Optimal carrier probability threshold of ARiCa

Sample: 2448 MyBrCa and SGBCC breast cancer patients in single imputed validation set.

Supplementary Fig 3. Observed proportion and expected probability of BRCA carrier prediction models by type of germline BRCA PVs

Sample: 2426 MyBrCa and SGBCC breast cancer patients with pedigree available in single imputed validation set.

Abbreviation: HL, Hosmer-Lemeshow.

FIGURES AND TABLES

Table 1. Multivariable regression of ARiCa

Variable	Category	<i>BRCA</i> vs Non- <i>BRCA</i> (n=5,714)				<i>BRCA1</i> vs Non- <i>BRCA1</i> (n=5,714)				<i>BRCA2</i> vs Non- <i>BRCA2</i> (n=5,714)			
		Odds Ratio	95% CI		P-value	Odds Ratio	95% CI		P-value	Odds Ratio	95% CI		P-value
Age*		0.94	0.93	0.96	<0.001	0.93	0.91	0.95	<0.001	0.95	0.94	0.97	<0.001
Ethnicity	Chinese	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-
	Malay	1.26	0.87	1.82	0.220	1.86	1.05	3.29	0.034	0.98	0.61	1.57	0.920
	Indian	2.06	1.37	3.09	<0.001	3.20	1.75	5.82	<0.001	1.35	0.77	2.36	0.295
	Other	1.50	0.45	4.96	0.511	1.79	0.23	13.91	0.58	1.37	0.33	5.77	0.667
Bilateral	Unilateral	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-
	Contralateral	2.56	1.57	4.17	<0.001	4.31	2.15	8.61	<0.001	1.56	0.80	3.06	0.195
	Ipsilateral	1.21	0.46	3.21	0.689	1.70	0.45	6.49	0.437	0.86	0.20	3.59	0.833
ER	ER+	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-
	ER-	1.62	1.17	2.24	0.004	5.59	3.15	9.92	<0.001	0.71	0.45	1.10	0.126
HER2	HER2+	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-
	HER2-	2.35	1.59	3.48	<0.001	3.11	1.61	6.01	0.001	1.84	1.13	3.00	0.015
Grade	One	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-
	Two	3.18	1.38	7.32	0.006	2.27	0.41	12.49	0.346	3.62	1.39	9.39	0.008
	Three	4.02	1.72	9.42	0.001	2.91	0.53	16.02	0.219	4.34	1.62	11.66	0.004
FHBC	No	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-
	Yes	3.01	2.23	4.07	<0.001	3.48	2.13	5.69	<0.001	2.56	1.77	3.71	<0.001
FHOC	No	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-
	Yes	4.57	2.51	8.30	<0.001	7.95	3.63	17.41	<0.001	1.93	0.75	4.92	0.170

Sample: 5714 MyBrCa and SGBCC breast cancer patients in multiply imputed training set.

Abbreviations: Bilateral, Bilateral Breast Cancer; FHBC, First Degree Family History for Breast Cancer; FHOC, First Degree Family History for Ovarian Cancer; 95% CI, 95% Confidence Interval.

*Age of diagnosis for breast cancer of proband.

Fig 1. ROC curves of BRCA carrier prediction models

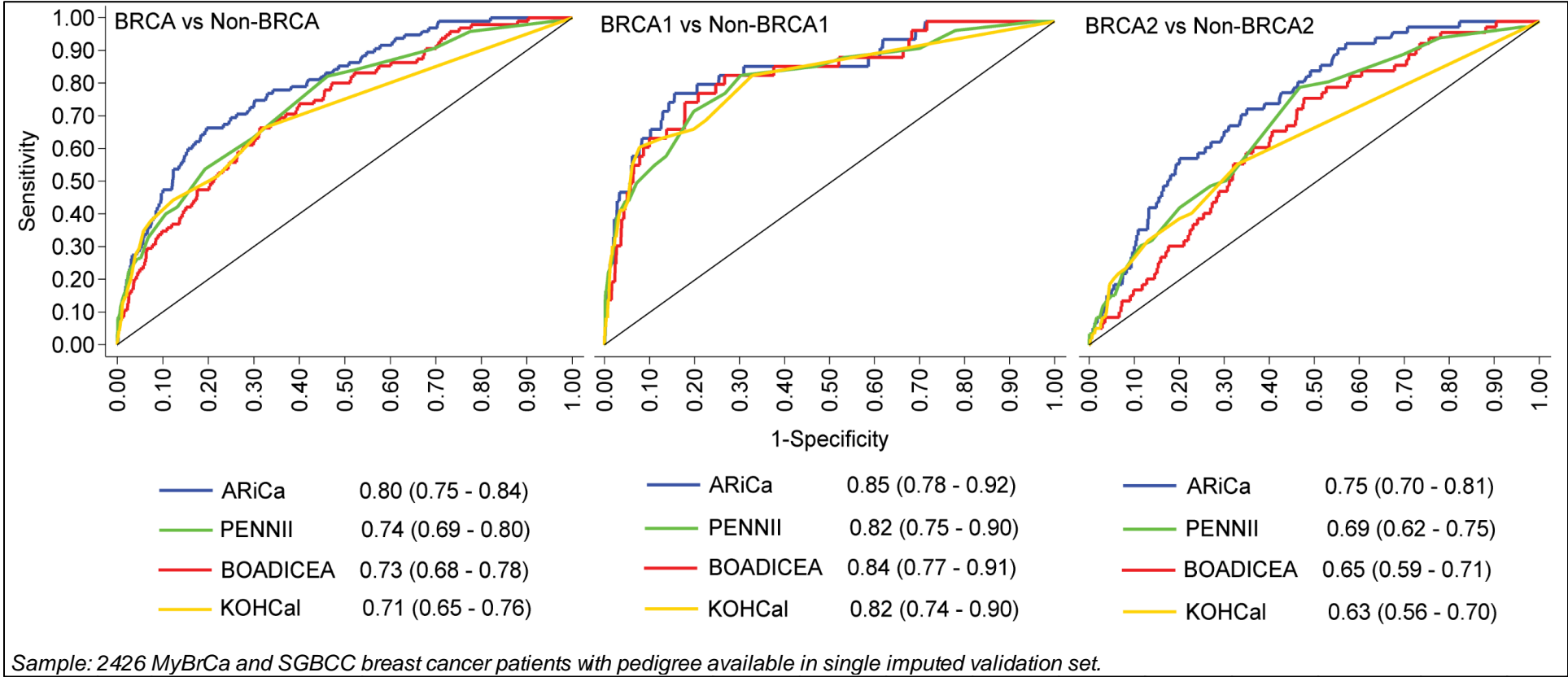


Fig 2. Observed proportion and expected probability of BRCA carrier prediction models

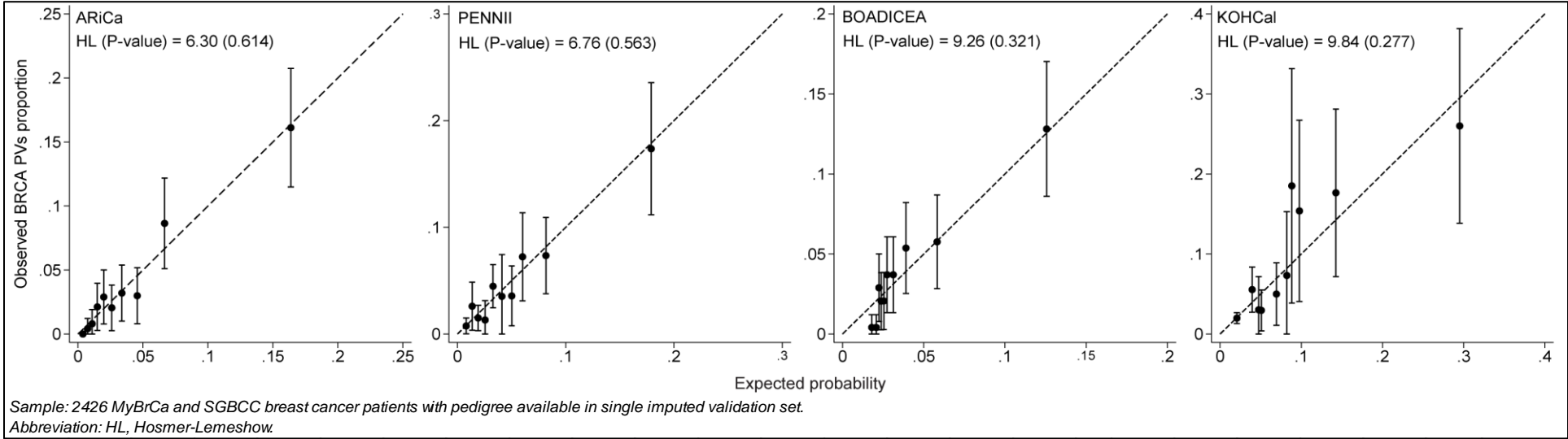


Table 2. Performance of BRCA carrier prediction models at different thresholds

Threshold (%)	Model	<i>BRCA</i> vs Non- <i>BRCA</i> (n=2,426)		<i>BRCA1</i> vs Non- <i>BRCA1</i> (n=2,426)		<i>BRCA2</i> vs Non- <i>BRCA2</i> (n=2,426)	
		Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
4.0*	ARiCa	71	71	83	70	66	70
8.0*	PENNII	63	70	83	70	51	69
2.2*	BOADICEA	66	67	83	67	56	67
4.0*	KOHCaI	66	68	83	67	56	67
10.0**	ARiCa	34	94	58	93	19	93
	PENNII	60	74	78	73	49	73
	BOADICEA	22	95	44	95	8	95
	KOHCaI	44	88	64	87	32	87
20.0**	ARiCa	15	99	25	98	8	98
	PENNII	15	99	25	98	8	98
	BOADICEA	7	99	14	99	3	99
	KOHCaI	22	97	42	97	10	96

Sample: 2426 MyBrCa and SGBCC breast cancer patients with pedigree available in single imputed validation set.

**Optimal threshold.*

***Conventional threshold.*

Fig 3. Eligible patients and *BRCA* PVs carriers detected for clinical criteria with and without grade

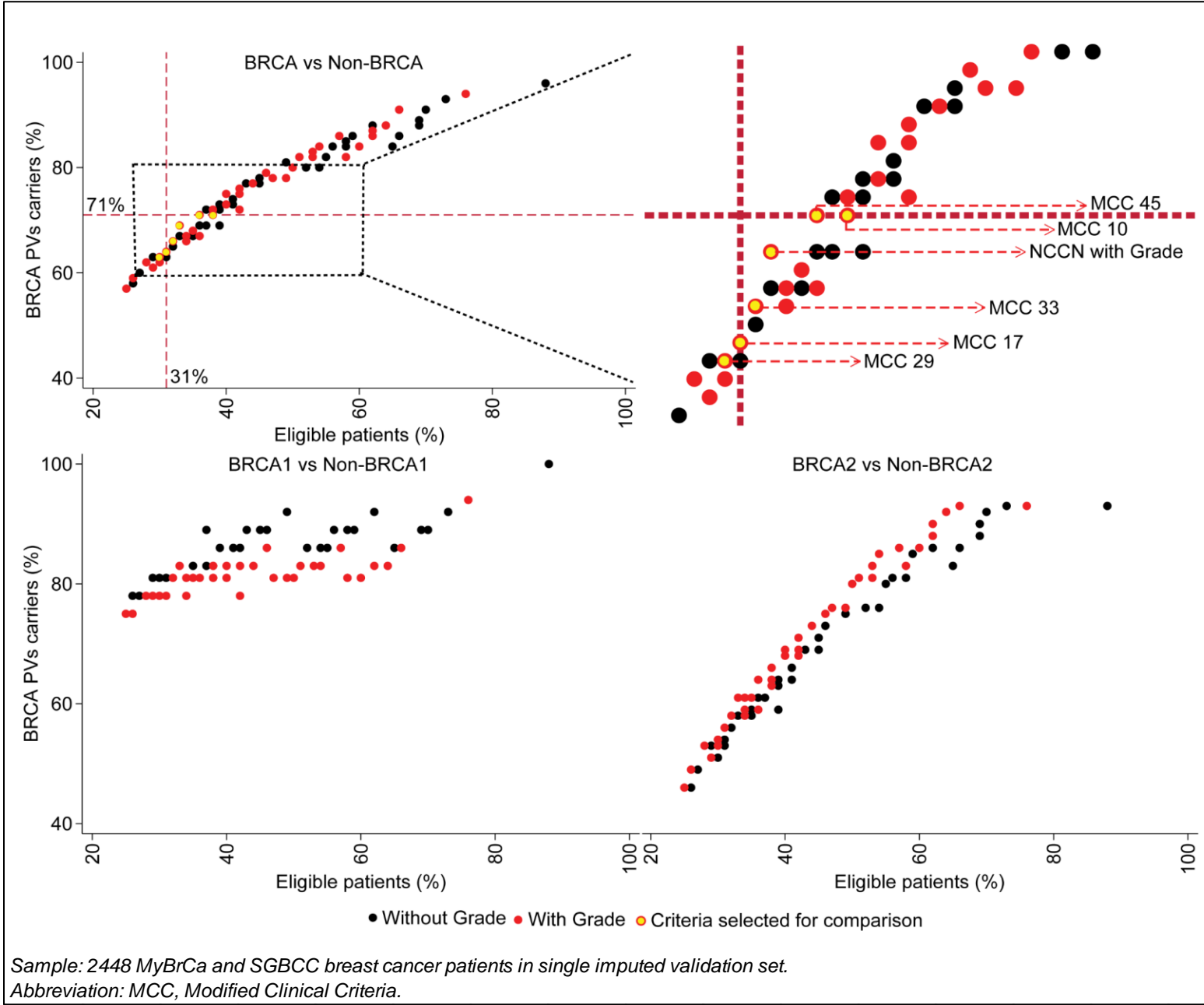


Table 3. Evaluation of clinical criteria with grade

Criteria	BC + Grade*	TNBC*	Bilateral*	OC	FHBC*	FHOC*	Eligible patients (%)**	BRCA PVs carriers (%)**	Detection ratio
MCC 29	≤40	≤45	≤60				30.0	63.0	12 : 1
MCC 17	≤40	≤60	≤60		≤60	≤60	31.0	64.0	13 : 1
MCC 33	≤40	≤50	≤60				32.0	66.0	12 : 1
NCCN with grade	≤45	≤60	46-50		46-50		33.0	69.0	12 : 1
MCC 45	≤40		≤60				38.0	71.0	14 : 1
MCC 10	≤45	≤50	≤60		≤60	≤60	36.0	71.0	13 : 1

Sample: 2448 MyBrCa and SGBCC breast cancer patients in single imputed validation set.

Abbreviations: MCC, Modified Clinical Criteria; BC, Breast Cancer of proband; TNBC, Triple Negative Breast Cancer; Bilateral, Bilateral Breast Cancer; OC, Ovarian Cancer; FHBC, one or more first degree relatives with Breast Cancer; FHOC, one or more first degree relatives with Ovarian Cancer.

**Age of diagnosis for breast cancer of proband.*

***Fulfilled at least one criterion.*

SUPPLEMENTAL METHODS

Study population

Participants donated a blood or saliva sample that was processed and stored, completed a questionnaire that included information on lifestyle risk factors for breast cancer, and provided written informed consent. Germline DNA were sequenced in two batches, using targeted sequencing panels that target the coding regions and exon-intron boundaries of known and suspected breast cancer susceptibility gene, which included *BRCA1* and *BRCA2* (*BRCA*) genes.¹ Target enrichment was performed using the Fluidigm Access Array system (n=2,441) or the Fluidigm Juno system (n=5,721) and sequenced on Illumina HiSeq 2500 or HiSeq 4000. All identified protein truncating variants and Class 4 and Class 5 missense variants were confirmed by Sanger sequencing and were considered as PVs for the purpose of this study. Recruitment and genetic studies were approved by the Ethics Committees of University Malaya Medical Centre [UM 842.9], Subang Jaya Medical Centre [reference no:201208.1], NHG Domain Specific Review Board [NHG DSRB Ref:2009/00501], and SingHealth Centralised Institutional Review Board [CIRB Ref:2010/632/B].

Statistical analyses

Existing BRCA carrier prediction models

The predicted likelihood of carrying a *BRCA* pathogenic variant was generated for each patient by using batch version for BOADICEA 5.0 which is implemented in the CanRisk tool,² including personal (cancer history, demographic, pathology, lifestyle, anthropometric, menstrual and reproductive) and family history information (cancer history, demographic) based on UK incidences,³ using web application for PENNII,⁴ or based on beta-coefficients from multivariable logistic regression for KOHCal.⁵

Development and validation of population-specific BRCA carrier prediction model

Independent variables that were considered for the ARiCa (**A**siatic **R**isk **C**alculator) model development, include age of diagnosis for breast cancer, ethnicity (Chinese/Malay/Indian/Other), bilateral breast cancer, pathological features, grade, immune-histochemical subtypes, and presence of first degree family history of breast cancer or ovarian cancer. Pathological features include estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) evaluated separately or in combination. Grade index of carcinomas assigned based on the morphology of cancer cells from excised tumours viewed under microscope were obtained from histopathological reports: grade 1 for well-differentiated carcinomas, grade 2 for moderately differentiated carcinomas, and grade 3 for poorly differentiated carcinomas. The training and validation sets were selected randomly but such that the prevalence of *BRCA* PVs was the same (4%) in both sets. Equality tests were performed to ensure that the distributions of independent and outcome variables were similar between training and validation sets.

Tumour biomarkers and grade had missing rates greater than 10%: 11% for ER, 14% for PR, 26% for HER2, and 15% for grade (**Supplementary Table 1**). First degree family history of ovarian cancer, parity, menopausal status, and BMI were also included in the imputation model as they were shown to be important predictors of the independent variables in this study. We evaluated the MAR assumption by assessing the association of missingness status of each variable by remaining variables included in the imputation model (data not shown). We generated 100 imputed datasets for analyses. Hormone receptor (HR+:ER+, PR+), TNBC, and immune-histochemical subtypes were not directly imputed, but instead derived from the imputed ER, PR, and HER2 status data. Classification of immune-histochemical subtypes are as follow: TNBC (ER-, PR-, HER2-), HER2-enriched (ER-, PR-, HER2+), Luminal A (ER+/-, PR+/-, HER2-), and Luminal B (ER+/-, PR+/-, HER2+). Each imputed dataset was

analysed separately and combined according to Rubin's rules.⁶ The proportion of missing data in the validation set was similar to that in the training set (**Supplementary Table 1**).

The area under receiver operating curve (AUC) was used to assess the ability to discriminate *BRCA* PVs carriers from non-carriers. AUC values range from 0 to 1, with AUC of 1 indicating perfect discrimination.⁷ The Hosmer-Lemeshow (HL) test was performed to measure calibration that provides an indication of the overall fit of models to the data by comparing the observed and expected number of *BRCA* PVs carriers in deciles of predicted carrier probabilities.⁸ Our analyses considered that pathogenic variants are protein truncating variants and known pathogenic missense variants. Since the *BRCA* testing did not involve screening for large re-arrangements the mutation testing sensitivity will be somewhat lower than 100%, however large re-arrangements account for a small proportion of *BRCA* pathogenic variants. Our modelling approach assumed that the mutation testing sensitivity is 100%. Performance measures reported, included sensitivity/detection rate (*BRCA* PVs carriers detected, %), specificity (non-carriers detected, %), screening rate (eligible patients, %), and detection ratio (number of patients to be screened to detect one carrier).

Customisation and evaluation of Modified Clinical Criteria for germline BRCA genetic testing

Combinations of age of diagnosis of proband in years are as follow: a) breast cancer diagnosed at age (≤ 40 , ≤ 45 , ≤ 50 , ≤ 55) or grade 2 or 3 breast cancer diagnosed at age (≤ 40 , ≤ 45 , ≤ 50 , ≤ 55), b) TNBC (≤ 40 , ≤ 45 , ≤ 50 , ≤ 55 , ≤ 60 , any age), c) bilateral breast cancer (≤ 60 , any age), d) one or more first degree relatives with breast cancer (≤ 60 , any age), e) one or more first degree relatives with ovarian cancer (≤ 60 , any age). Screening rate (eligible patients, %) and sensitivity/detection rate (*BRCA* PVs carriers detected, %) were calculated for each criterion as well as for the overall clinical criteria whereby at least one criterion was fulfilled. To enable direct comparison with mutation prediction models, the efficacy of

MCC was evaluated in the validation set by comparing detection ratio (number of patients to be screened to detect one carrier).

SUPPLEMENTAL MATERIALS

Supplementary Table 1. Characteristics of study population

Variable	Total n (%) (n=8,162)	Missing** n (%)	Chinese n (%) (n=6,140)	Malay n (%) (n=1,207)	Indian n (%) (n=718)	P-value
Age*	52.26 (10.77)	56 (0.7)	52.79 (10.80)	49.28 (10.19)	52.53 (10.53)	0.031
Ethnicity	-	14 (0.2)	-	-	-	-
Chinese	6140 (75.4)	-	-	-	-	-
Malay	1207 (14.8)	-	-	-	-	-
Indian	718 (8.8)	-	-	-	-	-
Other	83 (1.0)	-	-	-	-	-
Bilateral	-	55 (0.7)	-	-	-	0.051
Unilateral	7621 (94.0)	-	5727 (94.0)	1138 (94.5)	671 (93.6)	-
Contralateral	352 (4.3)	-	252 (4.1)	55 (4.6)	38 (5.3)	-
Ipsilateral	134 (1.7)	-	115 (1.9)	11 (0.9)	8 (1.1)	-
ER	-	917 (11.2)	-	-	-	<0.001
ER+	5171 (71.4)	-	3945 (72.8)	743 (67.8)	426 (66.6)	-
ER-	2074 (28.6)	-	1477 (27.2)	355 (32.2)	214 (33.4)	-
PR	-	1182 (14.5)	-	-	-	0.029
PR+	4394 (62.9)	-	3344 (63.9)	635 (59.8)	375 (61.6)	-
PR-	2586 (37.1)	-	1885 (26.1)	426 (40.2)	234 (38.4)	-
HR	-	1186 (14.5)	-	-	-	0.003
HR+	5215 (74.8)	-	3962 (75.8)	762 (72.0)	434 (71.3)	-
HR-	1761 (25.2)	-	1264 (24.2)	298 (28.0)	175 (28.7)	-
HER2	-	2141 (26.2)	-	-	-	0.001
HER2+	1821 (30.3)	-	1330 (29.7)	327 (35.0)	147 (27.3)	-
HER2-	4200 (69.7)	-	3149 (70.3)	605 (65.0)	391 (72.7)	-
Grade	-	1221 (15.0)	-	-	-	0.002
One	986 (14.2)	-	779 (14.8)	116 (11.4)	85 (14.4)	-
Two	3082 (44.4)	-	2363 (44.9)	429 (42.1)	256 (43.2)	-
Three	2874 (41.4)	-	2116 (40.3)	474 (46.5)	251 (42.4)	-
Subtypes	-	2340 (28.7)	-	-	-	<0.001

Variable	Total n (%) (n=8,162)	Missing** n (%)	Chinese n (%) (n=6,140)	Malay n (%) (n=1,207)	Indian n (%) (n=718)	P-value
Luminal A	3316 (57.0)	-	2527 (58.2)	460 (51.1)	289 (56.3)	-
Luminal B	975 (16.7)	-	716 (16.5)	181 (20.1)	70 (13.6)	-
TNBC	798 (13.7)	-	560 (12.9)	135 (15.0)	88 (17.2)	-
HER2-enriched	733 (12.6)	-	535 (12.4)	124 (13.8)	66 (12.9)	-
TNBC	-	1148 (14.1)	-	-	-	0.007
No	6216 (88.6)	-	4695 (89.3)	925 (87.3)	527 (85.7)	-
Yes	798 (11.4)	-	560 (10.7)	135 (12.7)	88 (14.3)	-
FHBC	-	117 (1.4)	-	-	-	0.008
No	6823 (84.8)	-	5085 (84.2)	1046 (92.9)	611 (85.8)	-
Yes	1222 (15.2)	-	957 (15.8)	148 (7.1)	101 (14.2)	-
FHOC	-	-	-	-	-	0.203
No	8031 (98.4)	-	6042 (98.4)	1191 (98.7)	701 (97.6)	-
Yes	131 (1.6)	-	98 (1.6)	16 (1.3)	17 (2.4)	-
Carrier	-	-	-	-	-	<0.001
Non-carrier	7839 (96.0)	-	5,942 (96.8)	1136 (94.1)	668 (93.0)	-
<i>BRCA1</i>	122 (1.5)	-	66 (1.1)	29 (2.4)	25 (3.5)	-
<i>BRCA2</i>	201 (2.5)	-	132 (2.1)	42 (3.5)	25 (3.5)	-

Sample: 8162 MyBrCa and SGBCC breast cancer patients before imputation.

Abbreviations: Bilateral, Bilateral Breast Cancer; Subtypes, Immune-histochemical Subtypes; FHBC, First Degree Family History for Breast Cancer; FHOC, First Degree Family History for Ovarian Cancer; Carrier, BRCA PVs Carrier Status.

*Age of diagnosis for breast cancer of proband in Mean (SD).

***Training set: Age of diagnosis (0.8%), Ethnicity (0.1%), Bilateral Breast Cancer (0.6%), ER (11%), PR (14%), HER2 (26%), Grade (15%), and First Degree Family History of Breast Cancer (1%).

***Validation set: Age of diagnosis (0.5%), Ethnicity (0.3%), Bilateral Breast Cancer (0.7%), ER (11%), PR (15%), HER2 (26%), Grade (16%), and First Degree Family History of Breast Cancer (2%).

Supplementary Table 2. Comparison of *BRCA* PVs carrier status

Variable	<i>BRCA1</i> n (%) (n=122)	Non-<i>BRCA1</i> n (%) (n=8,040)	P-value	<i>BRCA2</i> n (%) (n=201)	Non-<i>BRCA2</i> n (%) (n=7,961)	P-value
Age*	43.93 (10.78)	52.39 (10.72)	<0.001	47.59 (10.16)	52.38 (10.75)	<0.001
Ethnicity	-	-	<0.001	-	-	0.013
Chinese	66 (54.1)	6074 (75.7)	-	132 (65.7)	6008 (75.6)	-
Malay	29 (23.8)	1178 (14.7)	-	42 (20.9)	1165 (14.7)	-
Indian	25 (20.5)	693 (8.6)	-	25 (12.4)	693 (8.7)	-
Other	2 (1.6)	81 (1.0)	-	2 (1.0)	81 (1.0)	-
Bilateral	-	-	<0.001	-	-	0.035
Unilateral	100 (82.0)	7521 (94.2)	-	180 (90.5)	7441 (94.1)	-
Contralateral	18 (14.8)	334 (4.2)	-	16 (8.0)	336 (4.2)	-
Ipsilateral	4 (3.2)	130 (1.6)	-	3 (1.5)	131 (1.7)	-
ER	-	-	<0.001	-	-	0.340
ER+	25 (23.4)	5146 (72.1)	-	132 (74.6)	5039 (71.3)	-
ER-	82 (76.6)	1992 (27.9)	-	45 (25.4)	2029 (28.7)	-
PR	-	-	<0.001	-	-	0.902
PR+	23 (22.1)	4371 (63.6)	-	105 (62.5)	4289 (63.0)	-
PR-	81 (77.9)	2505 (36.4)	-	63 (37.5)	2523 (37.0)	-
HR	-	-	<0.001	-	-	0.249
HR+	30 (28.8)	5185 (75.5)	-	132 (78.6)	5083 (74.7)	-
HR-	74 (71.2)	1687 (24.5)	-	35 (21.4)	1725 (25.3)	-
HER2	-	-	<0.001	-	-	<0.001
HER2+	13 (13.3)	1808 (30.5)	-	25 (17.1)	1796 (30.6)	-
HER2-	85 (86.7)	4115 (69.5)	-	121 (82.9)	4079 (69.4)	-
Grade	-	-	<0.001	-	-	<0.001
One	2 (2.0)	984 (14.4)	-	4 (2.4)	982 (14.5)	-
Two	27 (27.6)	3054 (44.6)	-	83 (49.7)	2998 (44.3)	-
Three	69 (70.4)	2805 (39.0)	-	80 (47.9)	2794 (41.2)	-
Subtypes	-	-	<0.001	-	-	0.001

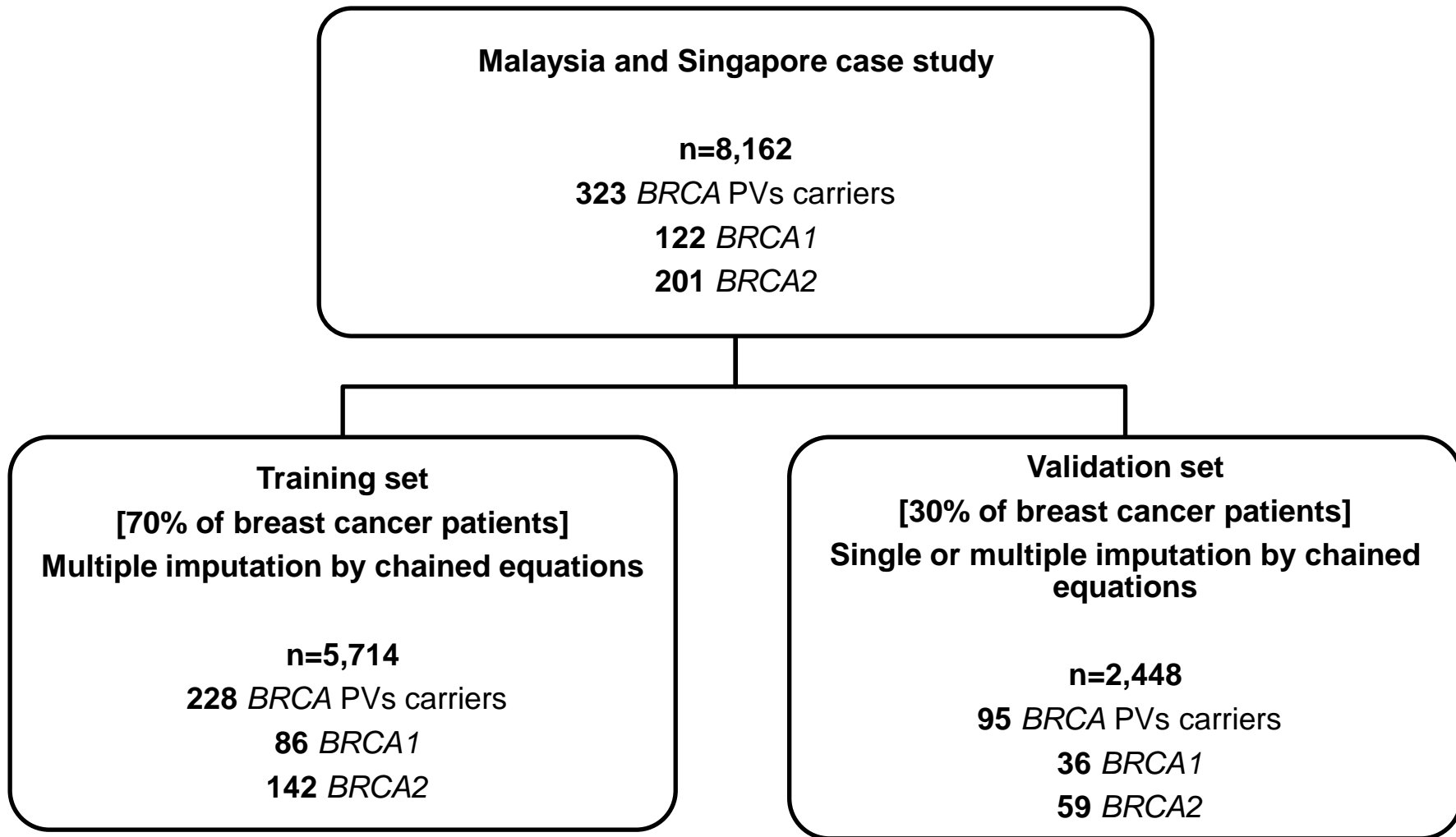
Variable	BRCA1 n (%) (n=122)	Non-BRCA1 n (%) (n=8,040)	P-value	BRCA2 n (%) (n=201)	Non-BRCA2 n (%) (n=7,961)	P-value
Luminal A	20 (21.5)	3296 (57.5)	-	92 (65.6)	3224 (56.7)	-
Luminal B	6 (6.5)	969 (17.0)	-	19 (13.6)	958 (16.9)	-
TNBC	63 (67.7)	735 (12.8)	-	26 (18.6)	772 (13.6)	-
HER2-enriched	4 (4.3)	729 (12.7)	-	5 (3.6)	728 (12.8)	-
TNBC	-	-	<0.001	-	-	0.118
No	38 (37.6)	6178 (89.4)	-	146 (84.9)	6070 (88.7)	-
Yes	63 (62.4)	735 (10.6)	-	26 (15.1)	772 (11.3)	-
FHBC	-	-	<0.001	-	-	<0.001
No	75 (62.0)	6748 (85.2)	-	141 (70.5)	6682 (85.2)	-
Yes	46 (38.0)	1176 (14.8)	-	59 (29.5)	1163 (14.8)	-
FHOC	-	-	<0.001	-	-	0.001
No	106 (86.9)	7925 (98.6)	-	192 (95.5)	7839 (98.5)	-
Yes	16 (13.1)	115 (1.4)	-	9 (4.5)	122 (1.5)	-

Sample: 8162 MyBrCa and SGBCC breast cancer patients before imputation.

Abbreviations: Bilateral, Bilateral Breast Cancer; Subtypes, Immune-histochemical Subtypes; FHBC, First Degree Family History for Breast Cancer; FHOC, First Degree Family History for Ovarian Cancer; Carrier, BRCA PVs Carrier Status.

*Age of diagnosis for breast cancer of proband in Mean (SD).

Supplementary Fig 1. Study design and sample selection



Supplementary Table 3. Discrimination and calibration of multivariable regression models for selection of ARiCa

Model	<i>BRCA</i> vs Non- <i>BRCA</i> (n=2,448)		<i>BRCA1</i> vs Non- <i>BRCA1</i> (n=2,448)		<i>BRCA2</i> vs Non- <i>BRCA2</i> (n=2,448)	
	AUC (95% CI)	HL (P-value)	AUC (95% CI)	HL (P-value)	AUC (95% CI)	HL (P-value)
Single imputation						
(a) TNBC	0.78 (0.74 - 0.83)	7.79 (0.454)	0.86 (0.79 - 0.93)	2.63 (0.955)	0.72 (0.66 - 0.77)	6.19 (0.626)
(b) ER	0.77 (0.72 - 0.82)	2.56 (0.959)	0.84 (0.77 - 0.91)	5.28 (0.728)	0.72 (0.66 - 0.78)	12.57 (0.128)
(c) ER + HER2	0.80 (0.75 - 0.84)	5.43 (0.711)	0.86 (0.79 - 0.93)	4.33 (0.826)	0.75 (0.69 - 0.80)	12.15 (0.145)
(d) HR + HER2	0.79 (0.75 - 0.84)	4.22 (0.836)	0.86 (0.79 - 0.93)	6.12 (0.634)	0.74 (0.69 - 0.80)	12.26 (0.140)
(e) HER2	0.79 (0.75 - 0.84)	6.85 (0.552)	0.85 (0.79 - 0.90)	12.45 (0.132)	0.75 (0.69 - 0.80)	10.89 (0.208)
(f) Subtypes	0.80 (0.75 - 0.84)	7.81 (0.452)	0.86 (0.79 - 0.92)	3.82 (0.873)	0.75 (0.69 - 0.80)	15.14 (0.056)
Multiple imputation						
(a) TNBC	0.78 (0.77 - 0.78)	6.36 (0.701)	0.85 (0.84 - 0.85)	5.39 (0.878)	0.71 (0.70 - 0.71)	5.25 (0.786)
(b) ER	0.77 (0.76 - 0.77)	3.73 (0.926)	0.84 (0.83 - 0.84)	5.86 (0.805)	0.71 (0.70 - 0.71)	13.75 (0.185)
(c) ER + HER2	0.80 (0.79 - 0.80)	5.08 (0.937)	0.86 (0.85 - 0.86)	5.56 (0.824)	0.73 (0.72 - 0.73)	8.79 (0.689)
(d) HR + HER2	0.78 (0.77 - 0.78)	5.12 (0.946)	0.85 (0.84 - 0.85)	5.25 (0.887)	0.72 (0.71 - 0.72)	7.02 (0.831)
(e) HER2	0.78 (0.77 - 0.78)	6.78 (0.842)	0.84 (0.83 - 0.84)	7.20 (0.938)	0.72 (0.71 - 0.72)	9.01 (0.485)
(f) Subtypes	0.79 (0.78 - 0.79)	5.90 (0.756)	0.85 (0.84 - 0.85)	4.75 (0.816)	0.72 (0.71 - 0.72)	10.42 (0.563)

Sample: 2448 MyBrCa and SGBCC breast cancer patients in imputed validation set.

Abbreviations: AUC, Area Under Curve; 95% CI, 95% Confidence Interval; HL, Hosmer-Lemeshow; TNBC, Triple Negative Breast Cancer; Subtypes, Immune-histochemical subtypes.

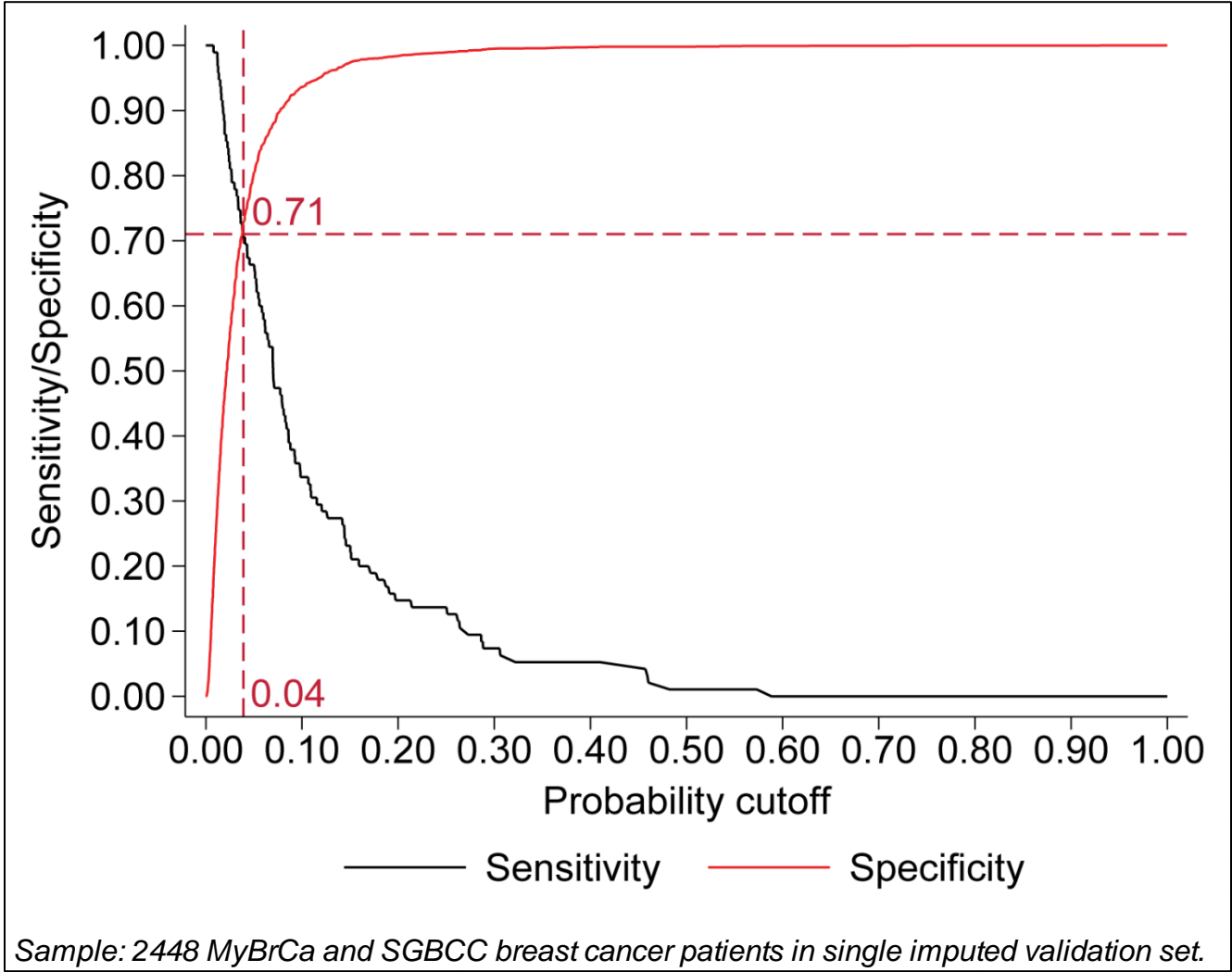
Supplementary Table 4. Discrimination and calibration of ARiCa by ethnicity

Mutation	Chinese (n=1,816)		Malay (n=361)		Indian (n=248)	
	AUC (95% CI)	HL (P-value)	AUC (95% CI)	HL (P-value)	AUC (95% CI)	HL (P-value)
Single imputation						
BRCA vs Non-BRCA	0.80 (0.74 - 0.86)	9.57 (0.297)	0.73 (0.63 - 0.83)	6.09 (0.637)	0.77 (0.65 - 0.88)	6.01 (0.647)
BRCA1 vs Non-BRCA1	0.88 (0.80 - 0.96)	6.10 (0.636)	0.73 (0.51 - 0.95)	9.37 (0.312)	0.91 (0.81 - 1.00)	1.78 (0.987)
BRCA2 vs Non-BRCA2	0.76 (0.69 - 0.84)	6.14 (0.631)	0.72 (0.60 - 0.83)	5.86 (0.663)	0.69 (0.57 - 0.80)	9.85 (0.276)
Multiple imputation						
BRCA vs Non-BRCA	0.79 (0.77 - 0.81)	7.97 (0.635)	0.75 (0.71 - 0.79)	5.92 (0.945)	0.77 (0.75 - 0.79)	5.25 (0.821)
BRCA1 vs Non-BRCA1	0.86 (0.84 - 0.88)	5.41 (0.826)	0.79 (0.73 - 0.85)	8.60 (0.712)	0.91 (0.90 - 0.92)	1.82 (0.987)
BRCA2 vs Non-BRCA2	0.76 (0.74 - 0.78)	7.17 (0.785)	0.68 (0.64 - 0.72)	6.28 (0.909)	0.70 (0.66 - 0.74)	8.64 (0.655)

Sample: 2448 MyBrCa and SGBCC breast cancer patients in imputed validation set.

Abbreviations: AUC, Area Under Curve; 95% CI, 95% Confidence Interval; HL, Hosmer Lemeshow.

Supplementary Fig 2. Optimal carrier probability threshold of ARiCa



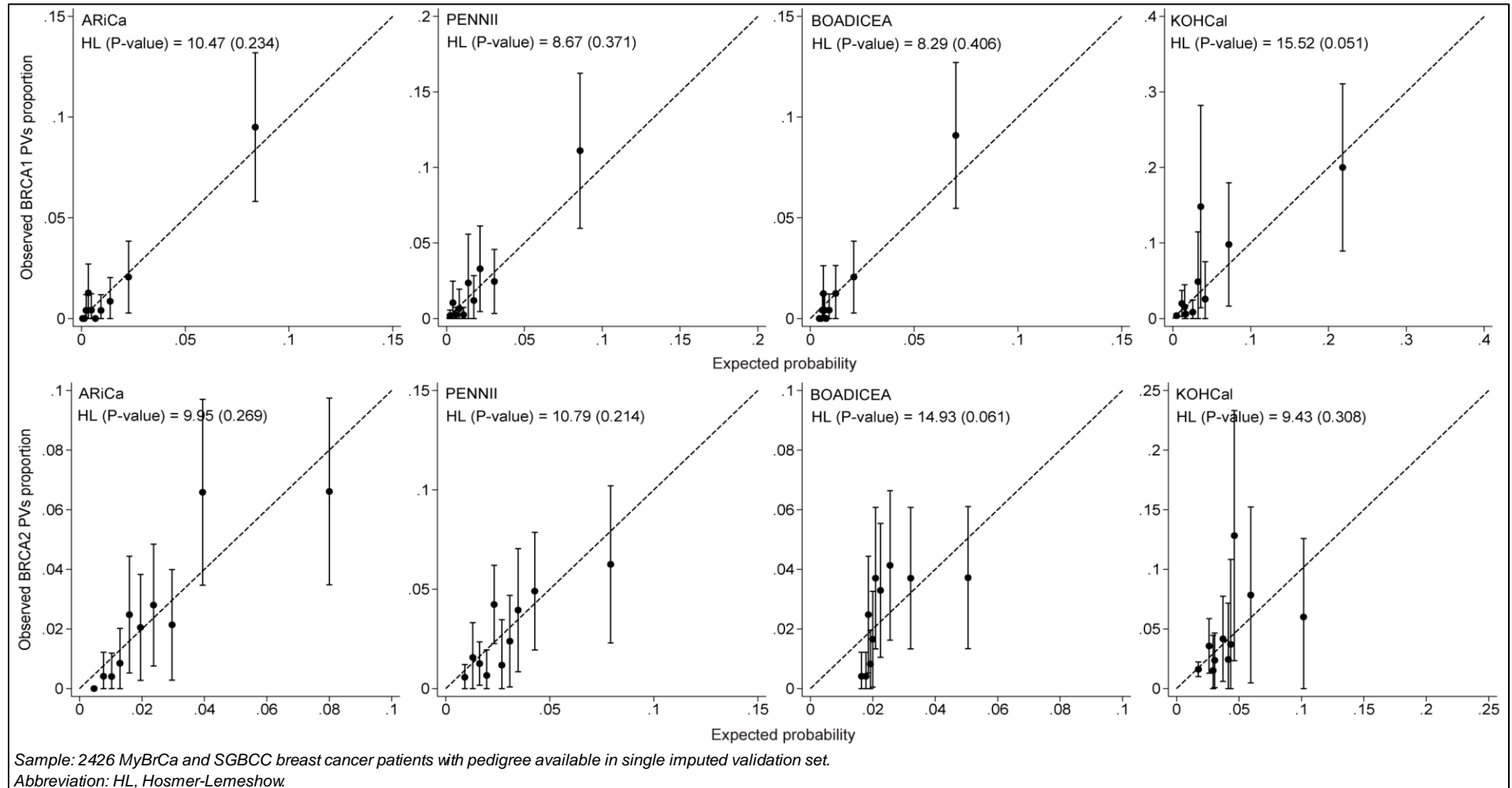
Supplementary Table 5. Performance of ARiCa at optimal carrier probability threshold

Model		Single imputation (n=2,448)	Multiple imputation (n=2,448)
Threshold (%)		4.0	4.0
Sensitivity (%)			
	Overall <i>BRCA</i>	71 (61 - 80)*	70 (69 - 70)
	<i>BRCA1</i>	83 (67 - 94)*	84 (83 - 85)
	<i>BRCA2</i>	66 (53 - 78)*	63 (62 - 64)
Specificity (%)			
	Overall <i>BRCA</i>	71 (69 - 73)*	71 (70 - 71)
	<i>BRCA1</i>	70 (68 - 72)*	69 (68 - 69)
	<i>BRCA2</i>	70 (68 - 72)*	70 (69 - 70)
Eligible patients (%)		31 (29 - 33)*	31 (30 - 31)
Detection ratio		11 : 1	11 : 1

Sample: 2448 MyBrCa and SGBCC breast cancer patients in imputed validation set.

**95% Confidence Intervals generated using normal approximation to binomial distribution with continuity correction.*

Supplementary Fig 3. Observed proportion and expected probability of BRCA carrier prediction models by type of germline BRCA PVs



REFERENCES

1. Dorling et al: Breast cancer risk genes: association analysis of rare coding variants in 34 genes in 60,466 cases and 53,461 controls. *N Engl J Med*, 2021
2. Carver T, Hartley S, Lee A, et al: CanRisk Tool—A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. *Cancer Epidemiol Biomarkers Prev* 30:469-73, 2021
3. Lee A, Mavaddat N, Wilcox AN, et al: BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med* 21:1708-1718, 2019
4. Lindor NM, Johnson KJ, Harve H, et al: Predicting BRCA1 and BRCA2 gene mutation carriers: comparison of PENN II model to previous study. *Fam Cancer* 9:495-502, 2010
5. Kang E, Park SK, Lee, JW, et al: KOHBRA BRCA risk calculator (KOHCal): a model for predicting BRCA1 and BRCA2 mutations in Korean breast cancer patients. *J Hum Genet*, 61:365-371, 2016
6. Little RJ, Rubin DB: *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons, 2019
7. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29-36, 1982
8. Archer KJ, Lemeshow S: Goodness-of-fit test for a logistic regression model fitted using survey sample data. *Stata J* 6:97-105, 2006