BMJ Open

Risk prediction models for breast cancer: systematic review and critical appraisal

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055398
Article Type:	Original research
Date Submitted by the Author:	13-Jul-2021
Complete List of Authors:	Zheng, Yadi; Chinese Academy of Medical Sciences & Peking Union Medical College Li, Jiang; Chinese Academy of Medical Sciences & Peking Union Medical College Wu, Zheng; Chinese Academy of Medical Sciences & Peking Union Medical College Li, He; Chinese Academy of Medical Sciences & Peking Union Medical College Cao, Maomao; Chinese Academy of Medical Sciences & Peking Union Medical College Li, Ni; Chinese Academy of Medical Sciences & Peking Union Medical College He, Jie; Chinese Academy of Medical Sciences & Peking Union Medical College
Keywords:	Epidemiology < TROPICAL MEDICINE, Public health < INFECTIOUS DISEASES, Breast tumours < ONCOLOGY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Risk prediction models for breast cancer: systematic review and critical appraisal

Yadi Zheng[#1], Jiang Li[#1], Zheng Wu[1], He Li[1], Maomao Cao[1], Ni Li[*1], Jie He[*2]

[1] Office of Cancer Screening, National Cancer Center/ National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

[2] Department of Thoracic Surgery, National Cancer Center / National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

Corresponding authors: Prof. Ni Li nli@cicams.ac.cn; Prof. Jie He hejie@cicams.ac.cn

[#]Yadi Zheng and Jiang Li contributed equally to this work.

Word count: 2880

Figures/Tables: 5

ABSTRACT

Objectives: To systematically review and critically appraise published studies of risk prediction models for breast cancer in the general population without breast cancer, and provide evidence for future research in the field.

Design: Systematic review using the Prediction model study Risk Of Bias Assessment Tool (PROBAST) framework.

Data sources: PubMed, the Cochrane Library and Embase were searched from inception to August 2020.

Eligibility criteria: We included studies reporting multivariable models to estimate the individualized risk of developing breast cancer among women. Search was limited to English language only.

Data extraction and synthesis: Two reviewers independently screened, reviewed, extracted and assessed studies with discrepancies resolved through discussion or a third reviewer. And risk of bias was assessed according to the PROBAST (Prediction model Risk of Bias Assessment Tool) framework.

Results: 72,353 studies were screened and 36 studies with 43 risk prediction models were included in the review. Most of the studies used logistic regression to develop breast cancer risk prediction models for Caucasian women by case-control data. The most

widely used risk factor was family history and the highest area under the curve was 0.943 (95% confidence internal: 0.919~0.967). And all the models included in the review had high risk of bias.

Conclusions: No risk prediction models were recommended and more key variables should be collected and validated well in the exiting models in the future. And high-quality breast cancer risk prediction models assessed by Prediction model Risk of Bias Assessment Tool should be developed and validated among Asian women.

PROSPERO registration number： CRD42020202570

Strengths and limitations of this study

1. Thoroughly conducted systematic review collecting data from major existing databases.
2. Critically appraised published studies of risk prediction models for breast cancer in the general population and provide evidence for future research in the field.
3. PROBAST was used to assess the quality of prediction models, which was developed through a consensus process involving a group of methodological experts in the area of clinical prediction tools and quality assessment.
4. Studies only about the external validation of the present risk models were not included in the review.
5. Our study highlighted high-quality breast cancer risk prediction models assessed by

PROBAST should be developed and validated among Asian women.

Keywords: breast cancer; risk prediction model; review; quality assessment; Prediction model Risk of Bias Assessment Tool

INTRODUCTION

Breast cancer is a major public health problem, and one of the most severe burdensome cancer among women worldwide [1], accounting for11.7% of new cancer cases and 6.9% of cancer deaths in 2020. And the prevalence of breast cancer is projected to increase over the coming years and ranks first among all cancers in 2020 [2]. Prevention of breast cancer is associated with a reduction in mortality [3], and the methods of predicting women at elevated risk and prevent the disease have been less successful. Numerous breast cancer risk prediction models have been developed to identify the combined effect of risk factors of breast cancer to advise population healthy life, routine screening, genetic testing, and to direct breast cancer research. Risk-stratified screening can improve cost-effectiveness and maximize benefits and minimize harms like overdiagnosis [4]. Individualized prediction model for breast cancer could be used in practice to assist decision making about mass screening or opportunistic screening and treatment strategy.

A recent breast cancer screening guideline [5] suggests that breast cancer screening increase the early detection rate and reduce the incidence if the screening is applied in

appropriate at-risk populations. However, major gaps exist in our knowledge to determine the risk of breast cancer accurately in order to apply these approaches to appropriate populations of women.

A lot of breast cancer risk prediction models have been developed over the past few decades. Many breast cancer risk models have undergone validation including discrimination and calibration in study populations other than those used in initial development, or have been further assessed in comparative studies. Breast cancer related predictors including hormonal factors, environmental factors, family histories, genetic factors and radiographic factors have been based on in these risk models, which would improve the generalizability. For example, the Gail model [6], one of the most famous models, has been widely used and validated worldwide since it was developed in 1989 [7-12].

This study is a systematic and critical review of breast cancer risk prediction models overall by using meta-analysis and the Prediction model Risk of Bias Assessment Tool (PROBAST) [13-14] to assess estimates and the methodological features, in order to find more methods of accurate predicting breast cancer risk, prepare for the development of risk prediction models, and prevent the disease successfully for the future research.

METHODS

Protocol and registration

The current review was designed according to the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [15] and was recorded in the PROSPERO database (registration number: CRD42020202570).

Literature search and eligibility criteria

We systematically searched PubMed, the Cochrane Library and Embase from inception to August 2020. The detailed search strategies were reported in Appendix Table 1. Articles identified from the search were loaded into EndNote X7 and duplicates were removed.

Inclusion criteria: 1) a model used data from cross-sectional studies cohort studies, case-control studies, and randomized controlled trials; 2) a model estimating the individualized risk of breast cancer; 3) a model developed for the general population without breast cancer; 4) reported a multivariable (i.e., at least 2 variables or predictors) model; 5) published in English.

Exclusion criteria: 1) external validation studies that only validated previous models in a different population without adding any additional information such as modifications on the risk factors; 2) models developed by machine learning.

Data extraction

One reviewer screened the search results based on title and abstract, a second reviewer reviewed a random sub-set (10%) of these studies independently. Full text reports were then assessed for eligibility with discrepancies resolved through discussion or a third reviewer.

We extracted information in two categories: 1) For all studies included in the review, we extracted the following information: author, publication year, study design, research method, targeted population, number of risk factors, risk factors, model performance and sample size of development. 2) For studies included validation part, we also extracted the following information: type of validation, study design, targeted population, model performance and sample size of validation. The information was extracted by one reviewer and checked by a second reviewer.

Risk of bias assessment

We used PROBAST to assess the reported prediction models, a new tool designed by a group of experts all over the world to assess the risk of bias and applicability of diagnostic and prognostic prediction models, which can be used in critical appraisal of studies that develop, validate, or update prediction models for individualized predictions [13-14]. In brief, it contains 20 signaling questions in four domains: participants, predictors,

outcome, and statistical analysis. Signaling questions can be answered as yes, probably yes, no, probably no, or no information. A domain where at least one signaling question is answered as no or probably no should be judged as high risk of bias. And only if all domains are judged as low risk of bias, the total bias is judged as low risk as well.

Before putting PROBAST into use, we formed a ten-people study group including prediction model researchers, statisticians, evidence-based medicine specialists etc. to learn and practice the appropriate use of this new tool systematically. Only after everyone understood all these twenty questions totally, we would move to the peer quality assessment part. Risk of bias of every prediction model was assessed by two reviewers independently with discrepancies resolved through discussion or a third reviewer.

If there were more than one models developed in one study, we only assessed the risk of bias once due to their similarity. And we also assessed the risk of external validation of prediction model when it was conducted in the same article that included model development.

Data synthesis and analysis

We calculated and reported descriptive statistics to summarize the characteristics of the models. And we calculated 12 the most frequently used risk factors and classified all risk factors into eight categories: Age, reproductive factors, family history of cancer,

hormone, gene-related factors, lifestyle, medical history and test, and basic information. Classification details can be seen in Appendix Table 2. Then we used network diagram to see the connections of categorized risk factors. And we used forest plot to describe the model performance. The expected observed (E/O) ratio was not included in the forest plot because it was only reported in 6 out of 36 studies. All analyses were performed using Sata 16.0 and NetDraw.

Patient and public involvement

There was no patient or public involvement in this study.

RESILTS

Study selection

A total of 97,964 indexed records (51,193 in PubMed, 3,163 in Cochrane Library and 43,608 in Embase), 25,611 were eliminated as duplicates found in all databases, leaving a total of 72,353 publications. 38 articles were included primarily after screening by title and abstract. 1 brief communication and 1 model that was developed based on the meta-analysis were excluded while full test screening, resulting in 36 studies with 43 models were included in the review eventually. (Figure 1).

Study characteristics

A brief summary of the 36 [6,16-50] included studies is presented in Appendix Table 3.

The included studies were published from 1989 to 2020. And 22 of the studies were conducted over the past ten years with 5 studies published in 2017 especially. Sixteen out of the thirty-six studies used data from case-control studies to develop prediction models [6,17,19,23-26,29-31,39, 40,42,45,48,50], twelve from prospective cohorts [16,18,20-22,27,32-37], seven from nested case-control studies [28,38, 41,43,46,47,49] and one from cross-sectional study [44]. Twenty-eight studies used logistic regression to fit prediction models [6,17-19,22-26,28-32,34,38-50], six used cox proportional hazards regression [20,21,27,33,35,36], one used Poisson regression [16] and one used competing risk regression [37]. Of all forty-three models in thirty-six studies, fourteen models were developed in Caucasian women [6,16,18,23,26,28,29,34,41,44,46,49], thirteen in multiple ethnicities women [20-22,24,27,30,35-38,43,47], eleven in Asian women [17,19,31,32,39,42,48,50], two for African-American women [25,33], two in Hispanic women [40] and one in Nigerian women [45].

The number of risk factors included in the models ranged from three to eighteen. Figure 2 showed the association between different kinds of risk factors after classifying risk factors into eight categories. Figure 2 showed that reproductive factors and family history of cancer were used most frequently and these two kinds of risk factors were used in 37 models together. And reproductive factors together with age, medical history and test together with family history of cancer, reproductive factors together with medical

history and test, family history of cancer together with age, reproductive factors together with basic information and family history of cancer together with basic information were used in more than 20 models, 29, 28, 27, 25, 24, 21, respectively.

Twenty-five studies reported c-statistics [18-22,26-28,30-32,34-39,41,42,44-47,49,50], ranged from 0.59(95% confidence internal: 0.57~0.61) to 0.943(95% confidence internal: 0.919~0.967). Qiu, et al [50] had the highest c-statistics (0.943, 95% confidence internal: 0.919~0.967), and Lee et al [19] and Salih et al [44] reported area under the curve (AUC) over 0.8, 0.867 and 0.864(95% confidence internal: 0.81~0.92), respectively. E/O ratios can be obtained from seven studies [22,27,29,32,35,36,45]. Figure 3 showed that the overall AUC was 0.66(95% confidence internal: 0.66~0.67) for fourteen studies [21, 26, 27, 30, 32, 34, 37, 38, 41, 44, 45, 47, 49, 50] that reported the AUC and 95% confidence internal. And the AUCs of the subgroups in five studies [18, 22, 31, 39, 46] were between 0.6 to 0.7.

In all these thirty-six studies, nine studies assessed prediction models with internal validation [22,26,27,33,39,43-46], eight with external validation [23,25,29,31,37,40,48,50], and one with both [32]. Thirteen studies reported the discriminatory accuracy as the AUC [23,25,27,29,31-33,37,39,40,45,48,50], and ten studies used the expected/observed event ratio (or observed/expected event ratio) to measure the calibration accuracy of the model [23,25,27,29,31,33,37,40,44,48].

Quality assessment

A summary of the quality assessment is shown in Table 1. Overall, all models assessed by PROBAST in the review had high risk of bias, low risk and high risk in outcome domain and analysis domain, respectively. Almost 60% models had low risk in participants domain and more than a half had low risk in predictors domain, 27 models and 30 models respectively. (As shown in Figure 4).

The main reasons for the high risk in analysis domain were model performance measures evaluated inappropriately, categorization of continuous predictors, no reporting of overfitting and optimism in model performance and missing data handled inappropriately (Appendix Table 4).

Table 1. Summary of risk of bias assessment.

Study	Participants	Predictors	Outcome	Analysis	Overall
Gail et al [6]	H	L	L	H	H
Rosner et al [16]	L	L	L	H	H
Ueda et al [17]	H	L	L	H	H
Colditzet al [18]	L	L	L	H	H
Lee et al [19]	H	H	L	H	H

Tice et al [20]	L	L	L	H	H
Tice et al[21]	L	L	L	H	H
Barlow et al [22]	L	L	L	H	H
Decarli et al [23]	H	H	L	H	H
Decarli et al [23*]	L	L	L	H	H
Novotny et al [24]	H	H	L	H	H
Gail et al [25]	H	H	L	H	H
Gail et al [25*]	L	L	L	H	H
Anna et al [26]	H	H	L	H	H
Tice et al [27]	L	L	L	H	H
Tamimi,et al [28]	L	L	L	H	H
Petracci et al [29]	H	H	L	H	H
Petracci et al [29*]	L	L	L	H	H
Dite et al [30]	H	H	L	H	H
Park et al [31]	H	H	L	H	H
Park et al [31*]	L	L	L	H	H
Anothaisintawee et al [32]	H	L	L	H	H

Anothaisintawee et al [32*]	L	L	L	H	H
Boggs et al [33]	L	L	L	H	H
Brentnall et al [34]	L	L	L	H	H
Kerlikowske et al [35]	L	L	L	H	H
Tice et al [36]	L	L	L	H	H
Schonberg et al [37]	L	L	L	H	H
Schonberg et al [37*]	L	L	L	H	H
Shieh et al [38]	L	L	L	H	H
Wang et al [39]	H	H	L	H	H
Banegas, et al [40]	H	L	L	H	H
Banegas et al [40*]	L	L	L	H	H
Eriksson et al [41]	L	L	L	H	H
Hsieh, et al [42]	H	H	L	H	H
Husing et al [43]	L	L	L	H	H
Salih et al [44]	L	L	L	H	H
Wang et al [45]	H	H	L	H	H
Zhang et al [46]	L	L	L	H	H

Clendenen et al [47]	L	H	L	H	H
Wang et al [48]	H	H	L	H	H
Wang et al [48*]	L	L	L	H	H
Abdolell et al [49]	L	L	L	H	H
Qiu et al [50]	H	H	L	H	H
Qiu et al [50*]	H	H	L	H	H

* The external validation was performed in the same study.

L indicates low risk of bias; H indicates high risk of bias.

DISCUSSION

Summary of main results

This systematic review identified 36 studies with 43 risk prediction models developed and/or validated for breast cancer in general population. Most of the studies used logistic regression to develop breast cancer risk prediction models for Caucasian women by case-control data. The most widely used risk factor was family history. And reproductive factors together with family history factor were used in most models. The highest AUC was 0.943 (95% confidence internal: 0.919~0.967) from Qiu, et al [50]. And the overall AUC was 0.66(95% confidence internal: 0.66~0.67) for fourteen studies [21, 26, 27, 30, 32, 34, 37, 38, 41, 44, 45, 47, 49, 50] that reported the AUC and 95% confidence internal. All the

studies presented a high risk of bias due to the high risk in analysis domain, which were mainly because of model performance measures evaluated inappropriately, categorization of continuous predictors, no reporting of overfitting and optimism in model performance and missing data handled inappropriately.

Agreements and disagreements with other reviews

As we can learn from the review, there were more and more risk prediction models of breast cancer over the past thirty years, and most of the models were developed in the Caucasian women, which agreed with the systematic review published by Louro et al in 2019 [51]. Compared with this review, we identified more prediction models and used a newly published tool to assess the quality of included models.

Over the past ten years, some new variables (such as oral contraceptives, diabetes and alcohol consumption) have been included in prediction models. Increased use of the inclusion of common genetic variation in the prediction models was in accord with Louro et al in 2019 [51] and Anothaisintawee et al in 2012 [52]. However, neither of them included models developed with potential biomarkers like tumor-associated antigens. By contrast, we included one model developed by Qiu, et al [50] in 2019 included five tumor-associated antigens. And the model performed well with a high AUC 0.943(95% confidence internal: 0.919,0.967).

Strengths and limitations of the study

We used PROBAST to assess the quality of prediction models, which was developed through a consensus process involving a group of methodological experts in the area of clinical prediction tools and quality assessment and has been used widely in many fields [53-56] since it came out.

Despite the strength, there are three main limitations. Firstly, because of the large number of references retrieved, only one reviewer screened the references by title and abstract. But we checked reference of relevant reviews and primary studies, finding no missed studies. Secondly, quality assessment could be thought to be subjective, which is an inherent bias of systematic review. However, two independent reviewers extracted and assessed the risk prediction models using PROBAST whose authors have indicated essentially objective guidelines and explanations. Moreover, studies only about the external validation of the present risk models were not included in the review. But during the screening of indexed records, we can learn that some models have been validated in many different populations. Take Gail for example, it has been modified and validated in many different ethnicities [57-60].

Implication to research and clinical practice

Ten models [19,30-32,37-39,42,44,49] selected predictors based on univariable analysis,

causing a high risk in analysis domain, which should be avoided. Risk prediction models should include predictors those are well-established and with clinical credibility regardless of any statistical significance [61,62]. Because sometimes predictors only have important relationship with the outcome after adjustment for confounding covariates, and covariates hold no independent predictive power when other covariates are included [13,63].

Some models were high risk in analysis domain because of missing data handled inappropriately, which may lead to biased associations between risk factors and breast cancer as well as biased model performance because of the selectivity of participants [64]. So imputation techniques are supposed to apply when data are missing [65,66].

When developing the risk prediction models, there were only nine studies included internal validation [22,26,27,33,39,43-46], leaving most models without internal validation. Lack of performing internal validation may increase the risk of overfitting [67]. Thus, we suggest that internal validation should be performed before external validation.

PROBAST was created by many international experts, providing a series of guidelines about model development and validation, which can be easily applied and improve clinical practice of prediction models. So, the new and most recommended methodology should be used when a new model is developed or the exiting models are updated.

In the light of the results of our review, it is still hard to recommend any of the models to be applied in the breast cancer screening due to the high risk of bias. And cost-effectiveness should be considered when a model is going to be applied in clinical practice. Because even though the model with some risk factors that cost more to get (e.g., high risk gene) has better model performance, it is still hard to be applied in poor area [68]. What's more, an existing model should be modified or updated before used in another group of people with different characteristics, which may improve the performance of prediction models.

Breast cancer incidence has risen to the first place by 2020 all over the world, which makes it more crucial to develop breast cancer prediction models for different ethnic groups. And in China, we have launched many breast cancer screening programs. For example, Rural Women "two cancers" Check Project Management Solutions have covered 31 provinces and 1437 counties since 2009. And Cancer Screening Program in Urban China conducted by the National Cancer Center has covered 28 provinces and 67 cities with more than 4 million people involved and 2 million people screened by ultrasound and Mammography since 2012, which will provide large data for us to develop a high-quality breast cancer risk prediction model in Chinese and will have great significance for breast cancer prevention of Asian women.

CONCLUSIONS

All 43 models assessed in our review using PROBAST performed the high risk of bias, leaving no model is recommended in the routine screening program. Some new variables, like oral contraceptives, diabetes, and alcohol consumption, have been widely used in prediction models over the past ten years. More key variables should be collected and validated well in the exiting models to improve the model performance. And it is necessary to develop and validate high-quality breast cancer risk predication models among Asian women.

Contributors: YZ and JL conceptualized the study and created the first version of the review protocol. ZW, HL, MC, NL and JH critically reviewed the review protocol and approved it. YZ and HL screened eligible articles. YZ extracted the data, supported by ZW, MC. YZ and JL drafted the first version of the manuscript, supported by NL and JH. All authors contributed to data interpretation and critically assessed it. All authors approved the final version of the manuscript.

Funding: This work was supported by the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences grant number 2019PT320027.

Competing interest: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Patient consent for publication： Not required.

Data availability statement: All data of the current study is present in the main manuscript, figures, tables and online supplemental material.

Ethics statement： This study does not involve human participants.

REFERENCES

1 GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*. 2020;396(10258):1204–1222. doi: 10.1016/S0140-6736(20)30925-9.

2 Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021. doi: 10.3322/caac.21660.

3 Myers ER, Moorman P, Gierisch JM, et al. Benefits and harms of breast cancer screening: A systematic review. *JAMA*. 2015;314(15):1615-1634. doi: 10.1001/jama.2015.13183.

4 Pashayan N, Morris S, Gilbert FJ, et al. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer. *JAMA Oncol*. 2018;4(11): 1504–1510. doi: 10.1001/jamaoncol.2018.1901.

5 Qaseem A, Lin JS, Mustafa RA, Horwitch CA, Wilt TJ; Clinical Guidelines Committee of the American College of Physicians. Screening for breast cancer in average-risk women: A guidance statement from the American college of physicians. *Ann Intern Med*. 2019;170(8):547-560. doi: 10.7326/M18-2147.

6 Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81(24):1879-1886. doi: 10.1093/jnci/81.24.1879.

7 Bondy ML, Newman LA. Assessing breast cancer risk: evolution of the Gail Model. *J Natl Cancer Inst*. 2006;98(17):1172-1173.

8 Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst*. 1999;91(18):1541-1548.

9 Rockhill B, Spiegelman D, Byrne C, et al. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst*. 2001;93(5):358-366.

10 Lin CJ, Block B, Nowalk MP, et al. Breast cancer risk assessment in socioeconomically disadvantaged urban communities. *J Natl Med Assoc*. 2007;99(7):752-756.

11 Bondy ML, Newman LA. Breast cancer risk assessment models: applicability to African-American women. *Cancer*. 2003;97(1 Suppl):230-235.

12 Buron A, Vernet M, Roman M, et al. Can the Gail model increase the predictive value of a positive mammogram in a European population population screening setting? Results from a Spanish cohort. *Breast*. 2013;22(1):83-88.

13 Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1): 51-58. doi: 10.7326/M18-1376.

14 Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Ann Intern Med*. 2019;170(1): W1-W33. doi: 10.7326/M18-1377.

15 Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10): e1001744. doi: 10.1371/journal.pmed.1001744.

16 Rosner B, Colditz GA. Nurses' Health Study: Log-Incidence mathematical model of breast cancer incidence. *J Natl Cancer Inst*. 1996;88(6): 359-64. doi: 10.1093/jnci/88.6.359.

17 Ueda K, Tsukuma H, Tanaka H, et al. Estimation of individualized probabilities of

developing breast cancer for Japanese women. *Breast Cancer*. 2003;10(1):54-62. doi: 10.1007/BF02967626.

18 Colditz GA, Rosner BA, Chen WY, et al. Risk factors for breast cancer according to estrogen and progesterone receptor status. *J Natl Cancer Inst*. 2004;96(3): 218-28. doi: 10.1093/jnci/djh025.

19 Lee EO, Ahn SH, You C, et al. Determining the main risk factors and high-risk groups of breast cancer using a predictive model for breast cancer risk assessment in south Korea. *Cancer Nurs*. 2004;27(5): 400-6. doi: 10.1097/00002820-200409000-00010.

20 Tice JA, Miike R, Adduci K, et al. Nipple aspirate fluid cytology and the Gail Model for breast cancer risk assessment in a screening population. *Cancer epidem biomar*. 2005;14(2): 324-8. doi: 10.1158/1055-9965.EPI-04-0289.

21 Tice JA, Cummings SR, Ziv E, et al. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res*. 2005;94(2): 115-22. doi: 10.1007/s10549-005-5152-4.

22 Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst*. 2006;98:1204-1214. doi:10.1093/jnci/djj331.

23 Decarli A, Calza S, Masala G, et al. Gail Model for prediction of absolute risk of invasive breast cancer: Independent evaluation in the Florence-European prospective investigation into cancer and nutrition cohort. *J Natl Cancer Inst*. 2006;98(23): 1686-93. doi: 10.1093/jnci/djj463.

24 Novotny J, Pecen L, Petruzelka L, et al. Breast cancer risk assessment in the Czech female population-an adjustment of the original Gail model. *Breast Cancer Res Treat*. 2006;95(1): 29-35. doi: 10.1007/s10549-005-9027-5.

25 Gail MH, Costantino JP, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst*. 2007;99(23): 1782-92. doi: 10.1093/jnci/djm223.

26 Crispo A, D'Aiuto G, De Marco M, et al. Gail Model risk factors: Impact of adding an extended family history for breast cancer. *Breast Journal*. 2008;14(3): 221-27. doi:10.1111/j.1524-4741.2008.00566.

27 Tice JA, Cummings SR, Smith-Bindman R, et al. Using clinical factors and mammographic breast density to estimate breast cancer risk: Development and validation of a new predictive model. *Ann Intern Med*. 2008;148(5): 337-47. doi: 10.7326/0003-4819-148-5-200803040-00004.

28 Tamimi RM, Rosner B, Colditz GA. Evaluation of a breast cancer risk prediction

model expanded to include category of prior benign breast disease lesion. *Cancer*. 2010;116(21): 4944-53. doi: 10.1002/cncr.25386.

29 Petracci E, Decarli A, Schairer C, et al. Risk factor modification and projections of absolute breast cancer risk. *J Natl Cancer Inst*. 2011;103(13): 1037-1048. doi:10.1093/jnci/djr172.

30 Dite GS, Mahmoodi M, Bickerstaffe A, et al. Using SNP genotypes to improve the discrimination of a simple breast cancer risk prediction model. *Breast Cancer Res Treat*. 2013;139(3): 887-96. doi: 10.1007/s10549-013-2610-2.

31 Park B, Ma SH, Shin A, et al. Korean risk assessment model for breast cancer risk prediction. *PLOS One*. 2013;8(10): e76736. doi: 10.1371/journal.pone.0076736.

32 Anothaisintawee T, Teerawattananon Y, Wiratkapun C, et al. Development and validation of a breast cancer risk prediction model for Thai women: a cross-sectional study. *Asian Pac J Cancer Prev*. 2014;15(16): 6811-7. doi: 10.7314/apjcp.2014.15.16.6811.

33 Boggs DA, Rosenberg L, Adams-Campbell LL, et al. Prospective approach to breast cancer risk prediction in African American women: The black women's health study model. *J Clin Oncol*. 2015;33(9): 1038-44. doi: 10.1200/JCO.2014.57.2750.

34 Brentnall1 AR, Harkness EF, Astley SM, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res*. 2015;17(1): 147. doi: 10.1186/s13058-015-0653-5.

35 Kerlikowske K, Gard CC, Sprague BL, et al. One versus two breast density measures to predict 5- and 10-year breast cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2015;24(6): 889-897. doi: 10.1158/1055-9965.EPI-15-0035.

36 Tice JA, Miglioretti DL, Li CS, et al. Breast density and benign breast disease: Risk assessment to identify women at high risk of breast cancer. *J Clin Oncol*. 2015;33(28): 3137-3143. doi:10.1200/JCO.2015.60.8869.

37 Schonberg MA, Li VW, Eliassen AH, et al. Accounting for individualized competing mortality risks in estimating postmenopausal breast cancer risk. *Breast Cancer Res Treat.* 2016;160(3): 547-562. doi: 10.1007/s10549-016-4020-8.

38 Shieh Y, Hu D, Ma L, et al. Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res Treat*. 2016;159(3): 513-525. doi: 10.1007/s10549-016-3953-2.

39 Wang F, Dai J, Li M, et al. Risk assessment model for invasive breast cancer in Hong Kong women. *Medicine*. 2016;95(32): e4515.

doi:10.1097/MD.0000000000004515.

40 Banegas MP, John EM, Slattery ML, et al. Projecting individualized absolute invasive breast cancer risk in US Hispanic women. *J Natl Cancer Inst*. 2017;109(2): djw215. doi: 10.1093/jnci/djw215.

41 Eriksson M, Czene K, Pawitan Y, et al. A clinical model for identifying the short-term risk of breast cancer. *Breast Cancer Res*. 2017;19(1):29. doi:10.1186/s13058-017-0820-y.

42 Hsieh YC, Tu SH, Su CT, et al. A polygenic risk score for breast cancer risk in a Taiwanese population. *Support Care Cancer*. 2017;163(1): 131-138. doi:10.1007/s10549-017-4144-5.

43 Hüsing A, Fortner RT, Kühn T, et al. Added value of serum hormone measurements in risk prediction models for breast cancer for women not using exogenous hormones: Results from the EPIC cohort. *Clin Cancer Res*. 2017;23(15): 4181-4189. doi: 10.1158/1078-0432.CCR-16-3011.

44 Salih AM, Alam-Elhuda DM, Alfaki MM, et al. Developing a risk prediction model for breast cancer: A Statistical Utility to Determine Affinity of Neoplasm (SUDAN-CA Breast). *Eur J Med Res*. 2017;22(1): 35. doi:10.1186/s40001-017-0277-6.

45 Wang SF, Ogundiran T, Ademola A, et al. Development of a breast cancer risk

prediction cancer epidemiology, biomarkers & prevention model for women in Nigeria. *Cancer Epidem Biomar*. 2018;27(6): 636-643. doi: 10.1158/1055-9965.EPI-17-1128.

46 Zhang X, Rice M, Tworoger SS, et al. Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. *PLOS Med*. 2018;15(9): e1002644. doi: 10.1371/journal.pmed.1002644.

47 Clendenen TV, Ge W, Koenig KL, et al. Breast cancer risk prediction in women aged 35–50 years: impact of including sex hormone concentrations in the Gail model. *BMC Med Res Methodol*. 2019;19(1):42. doi:10.1186/s13058-019-1126-z.

48 Wang L, Liu L, Zhen Lou,et al.Risk prediction for breast Cancer in Han Chinese women based on a cause-specific Hazard model. *BMC Cancer*. 2019;19(1): 128. doi:10.1186/s12885-019-5321-1.

49 Abdolell M, Payne JI, Caines J, et al. Assessing breast cancer risk within the general screening population: developing a breast cancer risk model to identify higher risk women at mammographic screening. *Eur Radiol*. 2020;30(10):5417-5426. doi:10.1007/s00330-020-06901-x.

50 Qiu C, Wang P, Wang B, et al. Establishment and validation of an

immunodiagnostic model for prediction of breast cancer. *OncoImmunology*. 2020;9(1): e1682382-2. doi:10.1080/2162402X.2019.1682382.

51 Louro J, Posso M, Boon MH, et al. A systematic review and quality assessment of individualised breast cancer risk prediction models. *Br J Cancer*. 2019;121(1): 76–85. doi: 10.1038/s41416-019-0476-8.

52 Anothaisintawee T, Teerawattananon Y, Wiratkapun C, et al. Risk prediction models of breast cancer: A systematic review of model performances. *Breast Cancer Res Treat*. 2012;133(1): 1–10. doi 10.1007/s10549-011-1853-z.

53 Heestermans T, Payne B, Kayode GA, et al. Prognostic models for adverse pregnancy outcomes in low-income and middle-income countries: a systematic review. *BMJ Global Health*. 2019;4(5): e001759. doi:10.1136/bmjgh-2019-001759.

54 Bellou V, Belbasis L, Konstantinidis AK, et al. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ*. 2019;367: l5358. doi: 10.1136/bmj. l5358.

55 de Jong Y, Ramspek CL, van der Endt VHW, et al. A systematic review and external validation of stroke prediction models demonstrates poor performance in dialysis patients. *J Clin Epidemiol*. 2020;123:69-79. doi: 10.1016/j.jclinepi.2020.03.015.

56 Bridge J, Blakey JD, Bonnett LJ. A systematic review of methodology used in the development of prediction models for future asthma exacerbation. *BMC Med Res Methodol*. 2020;20(1): 22. doi: 10.1186/s12874-020-0913-7.

57 Banegas MP, Gail MH, LaCroix A, et al. Evaluating breast cancer risk projections for Hispanic women. *Breast Cancer Res Treat*. 2020;132(1):347–353. doi: 10.1007/s10549-011-1900-9.

58 Min JW, Chang MC, Lee HK, et al. Validation of risk assessment models for predicting the incidence of breast cancer in Korean women. *J Breast Cancer*. 2014;17(3):226–235. doi: 10.4048/jbc.2014.17.3.226.

59 Chay WY, Ong WS, Tan PH, et al. Validation of the Gail model for predicting individual breast cancer risk in a prospective nationwide study of 28,104 Singapore women. *Breast Cancer Res*. 2012;14(1): R19. doi: 10.1186/bcr3104.

60 Adams-Campbell LL, Makambi KH, Palmer JR, et al. Diagnostic accuracy of the Gail model in the Black Women's Health Study. *Breast J*. 2007;13(4):332–336. doi: 10.1111/j.1524-4741.2007. 00439.x.

61 Harrell FE. Regression modeling strategies, with applications to linear models, logistic regression, and survival analysis. New York: Springer, 2001.

62 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in

developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4): 361-87. doi:10.1002/(SICI)1097-0258(19960229)15:4<361: AID-SIM168>3.0.CO;2-4.

63 Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol*. 1996;49(8): 907–16. doi: 10.1016/0895-4356(96)00025-x.

64 Little RJA, Rubin DB. Statistical Analysis With Missing Data. Hoboken, NJ: Wiley; 2002.

65 White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30:377-99. doi:10.1002/sim.4067.

66 Vergouwe Y, Royston P, Moons KG, et al. Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol*. 2010;63:205-14. doi:10.1016/j.jclinepi. 2009.03.017.

67 Royston P, Moons KG, Altman DG, et al. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338(b604). doi: 10.1136/bmj.b604.

68 Arveux P, Wait S, Schaffer P. Building a model to determine the cost-effectiveness of breast cancer screening in France. *Eur J Cancer Care.* 2003;12(2): 143-53. doi: 10.1046/j.1365-2354.2003.00373. x.

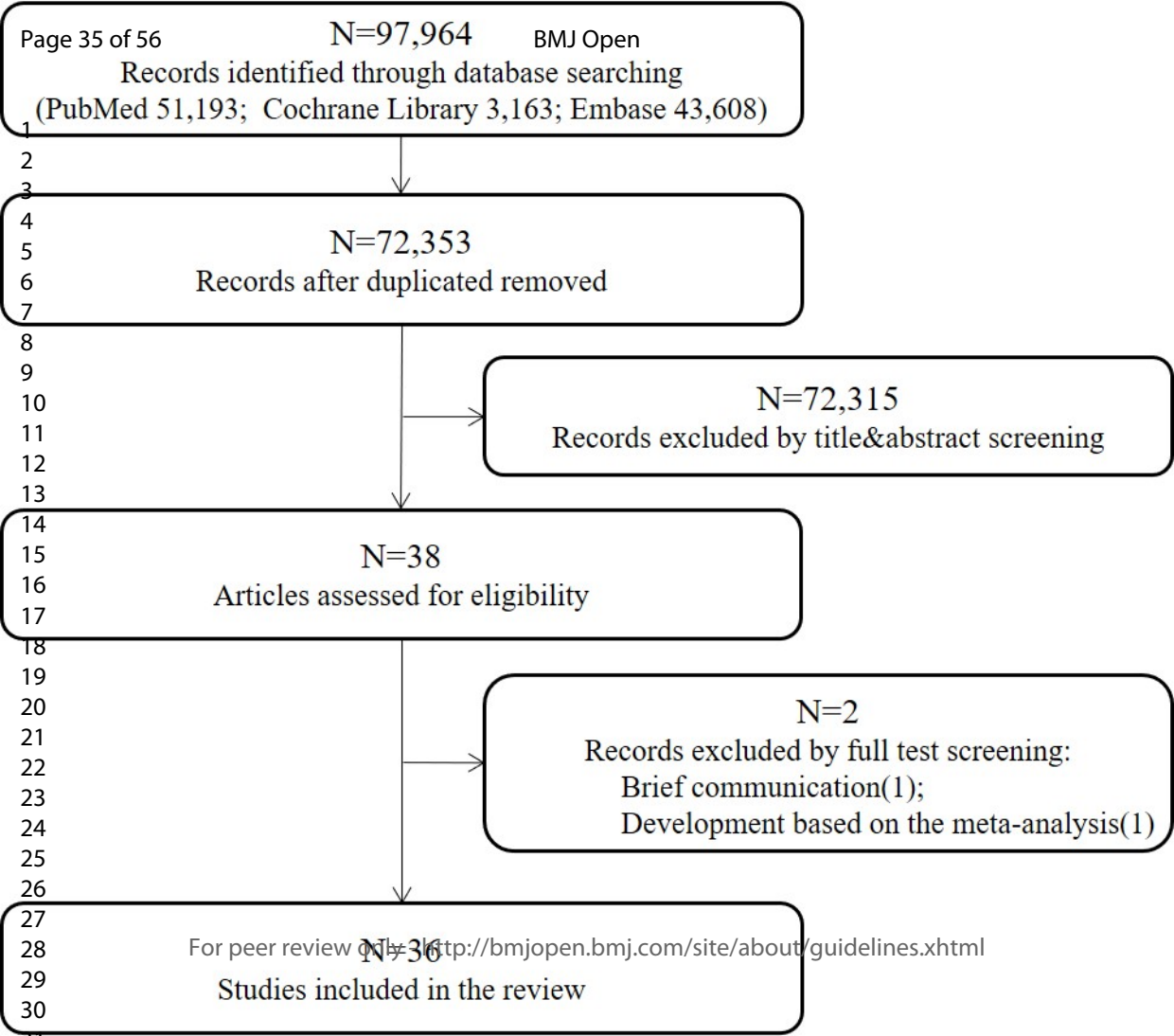
Figure legends:

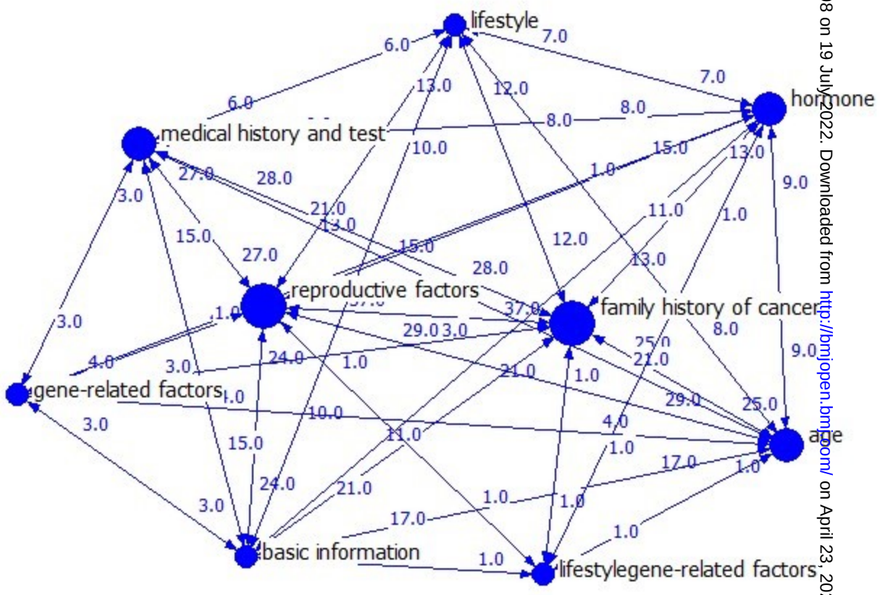
Figure1. PRISMA (preferred reporting items for systematic reviews and meta-analyses) flowchart.

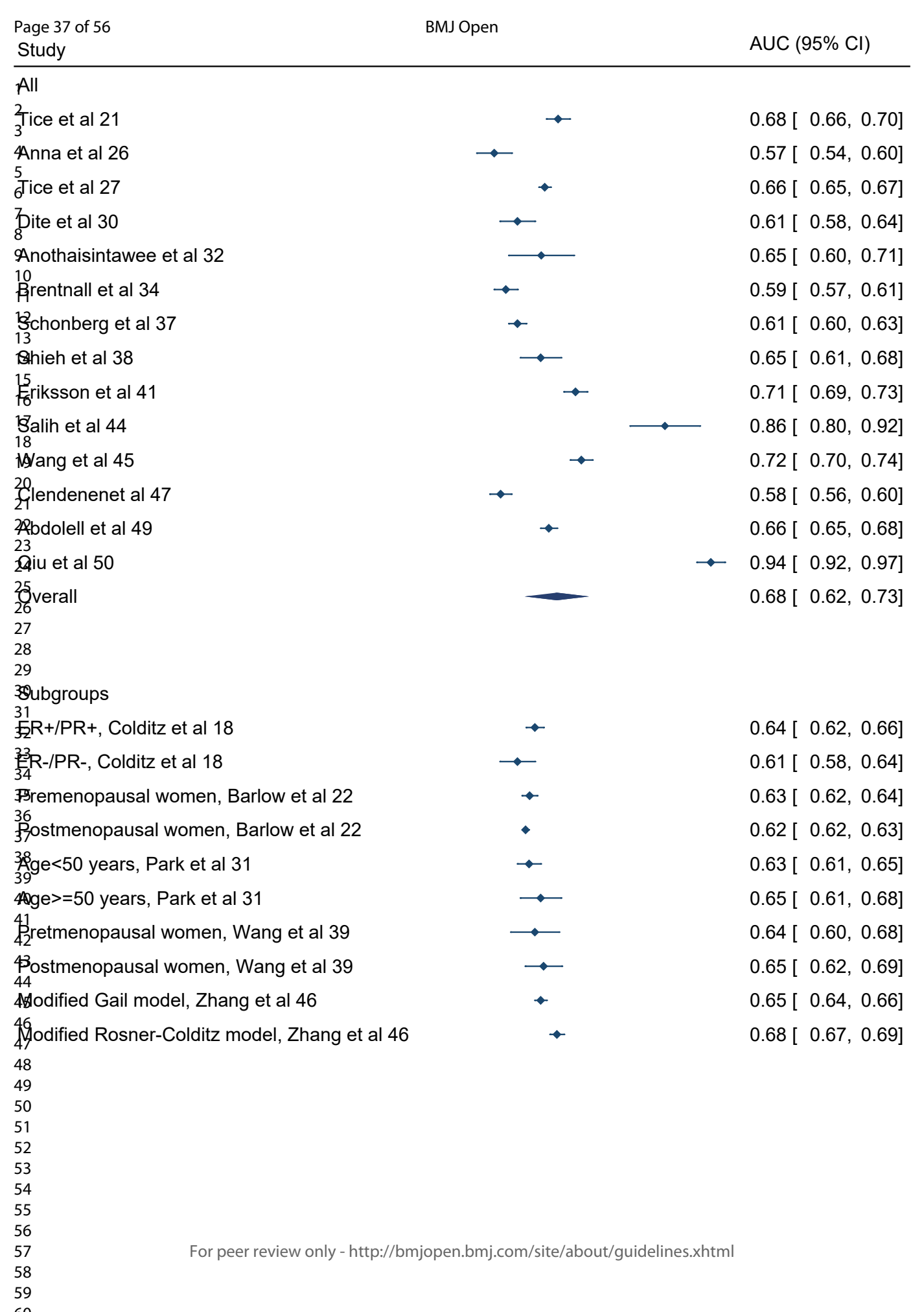
Figure 2. Network diagram of categorized risk factors.

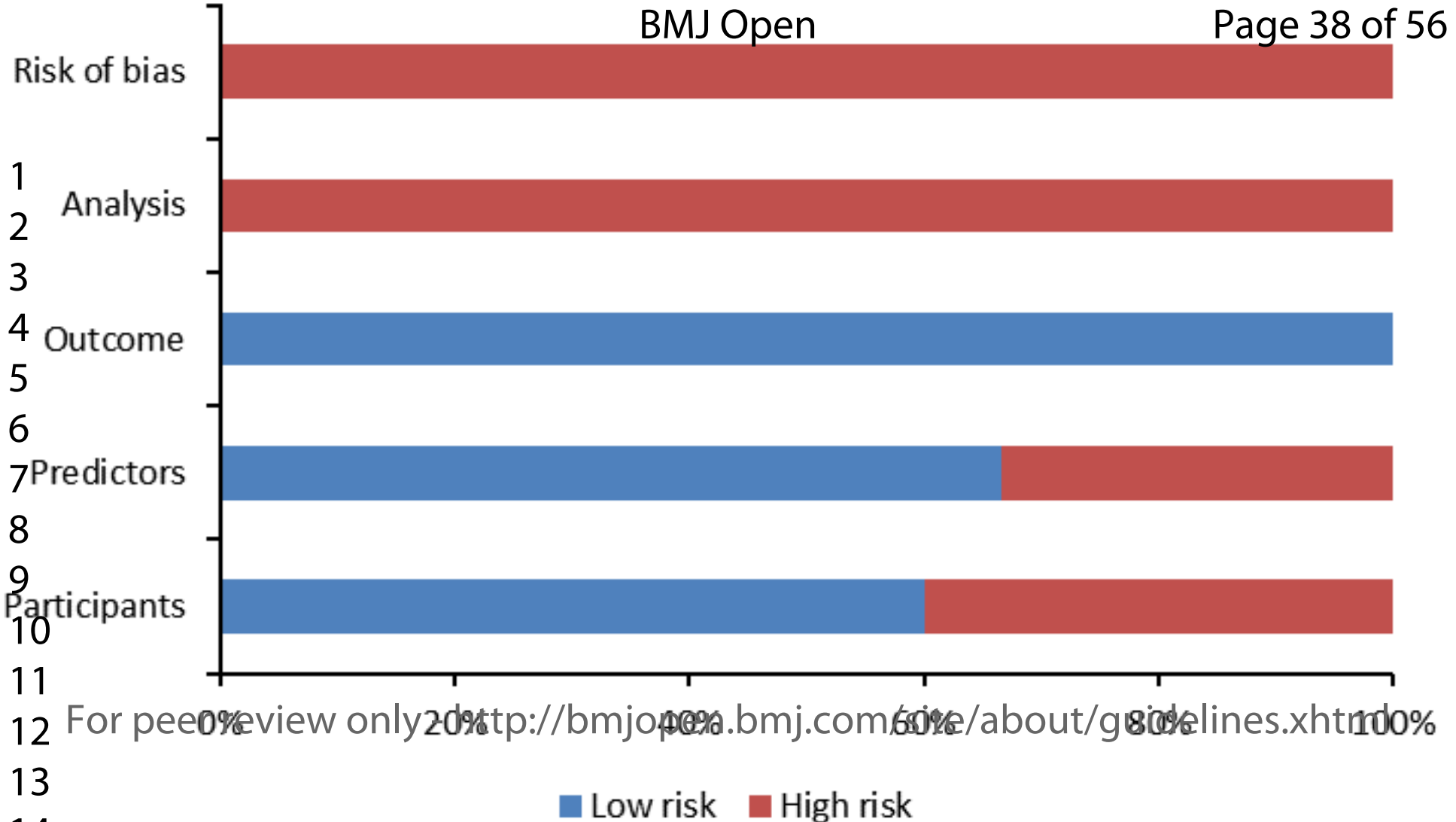
Figure 3. Area under the curve (AUC) and confidence intervals reported by the included studies.

Figure 4. Risk of bias assessment (using PROBAST) of all assessed models based on four domains.






Study
AUC (95% CI)
All
Tice et al 21
0.68 [0.66, 0.70]
Anna et al 26
0.57 [0.54, 0.60]
Tice et al 27
0.66 [0.65, 0.67]
Dite et al 30
0.61 [0.58, 0.64]
Anothaisintawee et al 32
0.65 [0.60, 0.71]
Brentnall et al 34
0.59 [0.57, 0.61]
Schonberg et al 37
0.61 [0.60, 0.63]
Shieh et al 38
0.65 [0.61, 0.68]
Eriksson et al 41
0.71 [0.69, 0.73]
Salih et al 44
0.86 [0.80, 0.92]
Wang et al 45
0.72 [0.70, 0.74]
Clendenenet al 47
0.58 [0.56, 0.60]
Abdolell et al 49
0.66 [0.65, 0.68]
Qiu et al 50
0.94 [0.92, 0.97]
Overall
0.68 [0.62, 0.73]
Subgroups
ER+/PR+, Colditz et al 18
0.64 [0.62, 0.66]
ER-/PR-, Colditz et al 18
0.61 [0.58, 0.64]
Premenopausal women, Barlow et al 22
0.63 [0.62, 0.64]
Postmenopausal women, Barlow et al 22
0.62 [0.62, 0.63]
Age<50 years, Park et al 31
0.63 [0.61, 0.65]
Age>=50 years, Park et al 31
0.65 [0.61, 0.68]
Pretmenopausal women, Wang et al 39
0.64 [0.60, 0.68]
Postmenopausal women, Wang et al 39
0.65 [0.62, 0.69]
Modified Gail model, Zhang et al 46
0.65 [0.64, 0.66]
Modified Rosner-Colditz model, Zhang et al 46
0.68 [0.67, 0.69]



Appendix

Appendix Table 1. Searching strategy.

Searching strategy
Take PubMed for example:
#1 "Breast Neoplasms"[Mesh] OR "Breast Carcinoma In Situ"[Mesh] OR "Breast Neoplasms, Male"[Mesh] OR "Carcinoma, Ductal, Breast"[Mesh] OR "Carcinoma, Lobular"[Mesh] OR "Hereditary Breast and Ovarian Cancer Syndrome"[Mesh] OR "Inflammatory Breast Neoplasms"[Mesh] OR "Triple Negative Breast Neoplasms"[Mesh] OR "Unilateral Breast Neoplasms"[Mesh] OR phyllodes tumor[Title/Abstract] OR breast sarcoma[Title/Abstract] OR mamma cancer*[Title/Abstract] OR mammary cancer*[Title/Abstract] OR mammary gland cancer*[Title/Abstract] OR Mammary Ductal Carcinoma*[Title/Abstract] OR breast gland cancer*[Title/Abstract] OR breast gland neoplasm*[Title/Abstract] OR Breast Neoplasm*[Title/Abstract] OR Breast Tumor*[Title/Abstract] OR Breast Cancer*[Title/Abstract] OR Mammary Cancer*[Title/Abstract] OR Breast Malignant Neoplasm*[Title/Abstract] OR Breast Malignant Tumor*[Title/Abstract] OR Human Mammary Carcinoma*[Title/Abstract] OR Human Mammary Neoplasm*[Title/Abstract] OR Breast Carcinoma*[Title/Abstract] OR Lobular Carcinoma*[Title/Abstract] 383,395
#2 ("Machine Learning"[Mesh] OR "Regression Analysis"[Mesh] OR "Multivariate Analysis"[Mesh] OR "Models, Biological"[Mesh] OR "Models, Statistical"[Mesh] OR "Neural Networks, Computer"[Mesh] OR "Algorithms"[Mesh] OR "Artificial Intelligence"[Mesh]) AND "Risk Assessment" [Mesh] 49,055
#3 predict*[Title/Abstract] AND (outcome*[Title/Abstract] OR mortality[Title/Abstract] OR index[Title/Abstract] OR rule*[Title/Abstract] OR decision*[Title/Abstract] OR scor*[Title/Abstract]) 576,113
#4 risk*[Title/Abstract] AND (predict*[Title/Abstract] OR calculate*[Title/Abstract] OR

assess*[Title/Abstract] OR scor*[Title/Abstract] OR algorithm[Title/Abstract]) 945,703 #5 model*[Title/Abstract] AND (logistic[Title/Abstract] OR statistic*[Title/Abstract] OR risk*[Title/Abstract] OR predict*[Title/Abstract]) 877,551 #6 “area under the curve”[Title/Abstract] OR “area under the receiver operator characteristic curve”[Title/Abstract] OR AUC[Title/Abstract] OR scor* system[Title/Abstract] OR “summary receiver operating characteristic”[Title/Abstract] OR SROC[Title/Abstract] 197,599 #7 OR/1-5 2,031,685 #8 #6 AND #7 51,193

Appendix Table 2. Classification of risk factors.

age	
reproductive factors	age at menarche, age at first birth, menopause, age at subsequent births, menstrual regularity, total menstrual duration, breastfeeding, breast density, parity, reproductive characteristics, microcalcifications and masses, abortions, breast volume
family history of cancer	family history of breast cancer, family history of any cancer
hormone	hormone therapy, oral contraceptives, estrogen plus progestin use, testosterone, estradiol, sex hormone-binding globulin, Insulin-like growth factor-I, estrone sulphate, prolactin, anti-Müllerian hormone
gene-related factors	polygenic risk score, rs2981582 (FGFR2), rs3803662(TOX3), rs889312(MAP3K1), rs13387042(2q35), rs13281615(8q24), rs4415084 (FGF10), rs3817198 (LSP1), rs981782(HCN1), rs10822013(ZNF365), rs3784099(RAD51B)
lifestyle	alcohol consumption, smoking status, exercise, light at night, sleep quality, vegetables and fruits, cereals, life satisfaction score

medical history and test	previous biopsies, benign breast disease, nipple aspirate fluid cytology, prior breast procedure, prior false-positive mammogram, breast inflammatory, benign breast category, benign breast disease, atypical hyperplasia, mammogram in past 2 years, diabetes, myocardial infarction, stroke, emphysema, congestive heart failure, p53, CyclinB1, p16, p62,14-3-3ξ
basic information	body mass index, weight, education, ethnicity, occupational activity, height

Appendix Table 3. Summary of the 36 included studies.

Author	Year	develop							validate				
		Study design	Research method	Targeted population	No of risk factors	Risk factors	Model development (AUC (95%CI); E/O ratio (95%)	Sample size of development	Type of validation	study design	Targeted population	Model validation (AUC (95%CI); E/O ratio (95%)	Sample size of validation
Gail et al [6]	1989	Case-control study	Logistic regression	Caucasian; 20–79 years	5	Age, age at menarche, age at first birth, number of previous biopsies, number of first degree relatives with breast cancer	AUC: none; E/O ratio: none	2,852cases/ 3,146 controls	None	None	None	None	None
Rosner et al [16]	1996	Prospective cohort study	Poisson regression	Caucasian; 30–64 years	5	Age, age at menarche, age at first birth, menopause, age at subsequent births	AUC: none; E/O ratio: none	2,249 cases/ 89,132 total	None	None	None	None	None
Ueda et al [17]	2003	Case-control study	Logistic regression	Asian women; age was not specified.	4	Age at menarche, age at first birth, family history of breast cancer, body mass index	AUC: none; E/O ratio: none	376 cases/ 430 controls	None	None	None	None	None
Colditz et al [18]	2004	Prospective cohort study	Logistic regression	Caucasian; 30-64 years	11	Age, age at menarche, age at first birth, menopause, age at subsequent births, benign breast disease, postmenopausal hormone use, family history of breast cancer in a first-degree relative, weight, body mass index, alcohol consumption	AUC: ER+/PR+: 0.64 (0.63,0.66); ER-/PR-: 0.61 (0.58, 0.64); E/O ratio: none	2,846 cases/ 66,145 total	None	None	None	None	None
Lee et al [19]	2004	Case-control study	Logistic regression	Asian women; age was not specified.	1) Hospitalized controls: 5 2) Nur	1) Hospitalized controls: family history, menstrual regularity, total menstrual duration, age at first full-term pregnancy, duration of breastfeeding 2) Nurse/teacher controls: age, education level, menstrual regularity, drinking status, smoking status	AUC: 1) Hospitalized controls: 0.714; 2) Nurse/teacher controls: 0.867; E/O ratio: none	1) Hospitalized controls: 384 cases/ 166 controls; 2) Nurse/teacher controls: 384 cases/	None	None	None	None	None

					se/teacher controls: 5			104 controls					
Tice et al [20]	2005	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 18 years and older	6	Age, age at menarche, previous biopsy, age at first birth, first degree breast cancer, nipple aspirate fluid cytology	AUC: 0.64; E/O ratio: none	400 cases/ 6,904 total	None	None	None	None	None
Tice et al [21]	2005	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35 years and older	6	Age, age at menarche, previous biopsy, age at first birth, first degree breast cancer, breast density	AUC: 0.68 (0.66,0.70); E/O ratio: none	955 cases/ 81,777 total	None	None	None	None	None
Barlow et al [22]	2006	Prospective cohort study	Logistic regression	Multiple ethnicities, 35-84 years	1) Premenopausal women: 4 2) Postmenopausal women: 10	1) Premenopausal women: age, breast density, family history of breast cancer, a prior breast procedure 2) Postmenopausal women: age, breast density, race, ethnicity, family history of breast cancer, a prior breast procedure, body mass index, natural menopause, hormone therapy, a prior false-positive mammogram	AUC: Premenopausal women: 0.631 (0.618, 0.644); postmenopausal women: 0.624 (0.619, 0.630) E/O ratio [a]: Premenopausal women: 1.000 postmenopausal women: 1.001	1) Premenopausal women: 1,726 cases/ 568,215 total; 2) postmenopausal women: 9,300 cases/ 1,642,824 total	Internal validation	None	None	None	None
Decarli et al [23]	2006	Case-control study	Logistic regression	Caucasian; 20–74 years	5	Age , age of menarche, number of breast biopsies, age at first live birth, first degree breast cancer	AUC: none; E/O ratio: none	2569 cases/ 2588 controls	External validation	Prospective cohort study	Caucasian; 35-64 years	AUC: 0.59; E/O ratio: 0.96(0.84, 1.11)	194 cases /10,031 total

Novotny et al [24]	2006	Case-control study	Logistic regression	Multiple ethnicities; 23-84 years	8	Age of menarche, number of biopsies, age at first childbirth, number of breast cancer cases in first-degree relatives, number of any cancer cases in first-degree relatives, breast inflammatory, body mass index, number of conceptions	AUC: none; E/O ratio: none	2299 cases/ controls	None	None	None	None	None
Gail et al [25]	2007	Case-control study	Logistic regression	African-American Women; 35-64 years	5	Age, age at menarche, number of affected mother or sisters, age at first live birth, number of previous benign biopsy examinations	AUC: none; E/O ratio: none	1607 cases/ 1647 controls	External validation	Prospective cohort study	African American women; 50-79 years	AUC: 0.555 (0.535,0.575); E/O ratio: 0.93b	350 cases /14,059 total
Anna et al [26]	2008	Case-control study	Logistic regression	Caucasian; age was not specified	5	Age, age at menarche, number of biopsies, age at first live birth, family history	AUC: 0.57 (0.54, 0.60); E/O ratio: none	558 cases/ 1207 controls	Internal validation	None	None	None	None
Tice et al [27]	2008	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35 years or older	5	Age, ethnicity, first degree breast cancer, previous biopsies, breast density	AUC: 0.657 (0.65,0.67); E/O ratio: 1.00 (0.98,1.03)	14,766 cases/ 1095484 total	Internal validation	Prospective cohort study	Multiple ethnicities; 35 years or older	AUC: 0.660(0.65,0.66); E/O ratio: 1.03(0.99,1.06)	3,465 cases/ 251,789 total
Tamimi et al [28]	2010	Nested case-control study	Logistic regression	Caucasian; 40-79 years	11	The type of benign breast disease, age, age at menarche, age at first birth and at each subsequent birth, age at menopause and type of menopause, history of benign breast diseases, family history of breast cancer in mother or sister, height, weight at age 18 years, current use of postmenopausal hormones (including type and duration of use), alcohol intake	AUC: 0.635; E/O ratio: none	240 cases/ 1036 controls	None	None	None	None	None

Petracci et al [29]	2011	Case-control study	Logistic regression	Caucasian; 20-74 years	8	Reproductive characteristics, education, occupational activity, family history, biopsy history, alcohol consumption, leisure physical activity, body mass index.	AUC: none; E/O ratio: 1.10 (0.96,1.26)	2569 cases/ 2588 controls	External validation	Prospective cohort study	Caucasian; 35-64 years	AUC: Age<50: 0.62(0.555,0.689) ; age>=50: 0.57 (0.519,0.614); E/O ratio: 1.10(0.96,1.26)	206 cases/ 8,426 total
Dite et al [30]	2013	Case-control study	Logistic regression	Multiple ethnicities; 35-59 years	13	Age, ethnicity, age at menarche, age at birth of first child, number of first-degree relatives with breast cancer, number of biopsies, presence of atypical hyperplasia, rs2981582(FGFR2), rs3803662(TOX3), rs889312(MAP3K1), rs13387042(2q35), rs13281615(8q24), rs4415084 (FGF10), rs3817198 (LSP1)	AUC: 0.61 (0.58,0.64); E/O ratio: none	962 cases/ 463 controls	None	None	None	None	None
Park et al [31]	2013	Case-control study	Logistic regression	Asian women; age was not specified.	1) Age<50 years: 7 2) Age>=50 years: 7	1)Age<50 years: a family history of breast cancer in first-degree relatives, age at menarche, menopausal status, age at first full-term pregnancy, duration of breast feeding, oral contraceptive usage, exercise. 2)Age>=50 years: a family history of breast cancer in first degree relatives, age at menarche, age at menopause, experience of pregnancy, body mass index, oral contraceptive usage, exercise	AUC: Age<50 years: 0.63 (0.61-0.65); Age>=50 years: 0.65 (0.61- 0.68); E/O ratio: none	3,789 cases/ 3,789 controls	External validation	Prospective cohort study	None	1)Korean Multi-Center Cohort (KMCC): AUC: 0.61(0.49,0.72); E/O ratio: 0.97(0.67,1.40) 2)National Cancer Center (NCC) cohort: AUC: 0.89(0.85,0.93) E/O ratio: 0.96(0.70,1.37)	1) KMCC: 29cases/ 6148 total; 2)NCC: 36 cases/ 7546 total
Anothaisintawee et al [32]	2014	Cross-sectional study	Logistic regression	Asian women; age was not specified	4	Age, menopausal status, body mass index, use of oral contraceptives	AUC: 0.651 (0.595, 0.707); O/E ratio: 1.00 (0.82, 1.21) [b]	107cases/ 15,718total	Internal and external validation	Cross-sectional study	Asian women; 18 years or older	Internal validation: AUC: 0.646(0.642,0.650); E/O ratio: none; External validation:	35 cases/ 4,978 total

												AUC: 0.609(0.511,0.706); O/E ratio: 0.97 (0.68, 1.35)[b]	
Boggs et al [33]	2015	Prospective cohort study	Cox proportional hazards regression	African-American Women; 30-69 years	9	Family history, previous biopsy, body mass index at age 18 years, age at menarche, age at first birth, oral contraceptive use, bilateral oophorectomy, estrogen plus progestin use, height	AUC: none; E/O ratio: none	896 cases/ 55,093 total	Internal validation	Prospective cohort study	African American Women; 30-69 years	AUC: 0.59 (0.56, 0.61); E/O ratio: 0.96(0.88,1.05)	506 cases/ 48,193 total
Brentnall et al [34]	2015	Prospective cohort study	Logistic regression	Caucasian; 47-73 years	1) Gail model +Density residual: :8 2) Tyrer-Cuzick+density residual: 11	1) Gail model+ Density residual: Age, Ethnicity, age at menarche, age at first birth, number of previous biopsies, benign disease, number of first degree relatives with breast cancer, density residual 2) Tyrer-Cuzick+ density residual: Age, gen phenotype, family history, age at menarche, age at first birth, menopause, atypical Hyperplasia, lobular carcinoma in situ, height, body mass index, density residual	(1) Primary (invasive+ DCIS): 1)Gail model+ Density residual: AUC: 0.59(0.57,0.61); E/O ratio: none; 2)Tyrer- Cuzick+ density residual: AUC: 0.61(0.59,0.63); E/O ratio: none; (2) Secondary(invasive): 1)Gail model+ Density residual: AUC: 0.59(0.57,0.61); E/O ratio: none; 2)Tyrer-Cuzick+ density residual: AUC: 0.61(0.58–0.63); E/O ratio: none	697 cases/ 50,628 total	None	None	None	None	None

Kerlikowske et al [35]	2015	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35-74 years	5	Age, ethnicity, first degree breast cancer, previous biopsies, changes in breast density	AUC: 5-year risk model: 0.640; 10-year risk model: 0.628; E/O ratio: 5-year risk model: 0.98(0.96,1.00); 10-year risk model: 0.95(0.94,0.96)	13,715 cases/ 722,654 total	None	None	None	None	None
Tice et al [36]	2015	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35-74 years	6	Age, race/ethnicity, family history of breast cancer, history of breast biopsy, benign breast disease diagnoses, breast density	AUC: 0.665; E/O ratio: 5 Years: 1.04(1.02 ,1.06); 10 years: 1.05 (1.03,1.06)	17908 cases/ 1,135,977 total	None	None	None	None	None
Schonberg et al [37]	2016	Prospective cohort study	Competing risk regression	Multiple ethnicities; 57–85 years	16	Age at study entry, postmenopausal hormone use, number of first-degree relatives with history of breast cancer and age at diagnosis, history of breast biopsy, highest body mass index in past 10 years, age at menopause, age at first birth and parity, average alcohol use per day (highest average use in past 10 years), cigarette use, mammogram in past 2 years, limited in moderate daily activity, diabetes, myocardial infarction, stroke, emphysema, congestive heart failure	AUC： 0.61 (0.60,0.63); E/O ratio: none	73,066 total	External validation	Prospective cohort study	Multiple ethnicities; 55-91 years	AUC: 0.57 (0.55,0.58); E/O ratio: 0.92(0.88,0.97)	74,887 total
Shieh et al [38]	2016	Nested case-control study	Logistic regression	Multiple ethnicities; 36-86 years	7	Age, ethnicity, first degree breast cancer, previous biopsies, breast density, polygenic risk score, body mass index	AUC:0.65(0.61,0.68); E/O ratio: none	486 cases/ 495 controls	None	None	None	None	None

Wang et al [39]	2016	Case-control study	Logistic regression	Asian women; 20-84 years	1)Premenopausal: 5; 2)Postmenopausal: 11	1) Premenopausal: age, number of parity, case number of breast cancer in first-degree relatives, light at night, sleep quality; 2) Postmenopausal: age, number of parity, case number of breast cancer in first-degree relatives, light at night, body mass index, age at menarche, age at first give birth, ever breast feeding, ever using of oral contraceptive, hormone replacement treatment, history of benign breast diseases.	1) Pretmenopausal women: AUC: 0.640(0.598,0.681); E/O ratio: none; 2) Postmenopausal women: 0.655(0.621,0.653); E/O ratio: none	923 cases / 918 controls	Internal validation	Case-control study	Asian women; 20-84 years	1) Premenopausal: average AUC: 0.621; 3) Postmenopausal: Average AUC: 0.632	None

Banegas et al [40]	2017	Case-control study	Logistic regression	Hispanic Women; 35-79 years	1) The US-born Hispanic risk model :3; 2) the foreign-born Hispanic risk model :4	1) The US-born the Hispanic risk model: age at first full-term pregnancy, biopsy for benign breast disease, family history of breast cancer; 3) The foreign-born the Hispanic risk model: age at first full-term pregnancy, biopsy for benign breast disease, family history of breast cancer, age at menarche	None	1086 cases/ 411 controls	External validation	Prospective cohort study	Hispanic Women; 50-79 years	1)US-born Hispanics: AUC: 0.564 (0.485, 0.644); O/E:1.07 (0.81 ,1.40) [b]; 2)Foreign-born Hispanics: AUC: 0.625 (0.487 ,0.764); O/E: 0.66 (0.41,1.07) [b] 4) Hispanics of unknown nativity: AUC: 0.582(0.509,0.656); O/E: 0.89(0.69,1.14) [b]	130 cases/ 6,220 total
Eriksson et al [41]	2017	Nested case-control study	Logistic regression	Caucasian; 40-74 years	7	MD, computer-aided detection of microcalcifications and masses, use of hormone replacement therapy, family history of breast cancer, menopausal status, age, body mass index	AUC: 0.71(0.69,0.73); E/O ratio: none	433cases / 1732 controls	None	None	None	None	None
Hsieh et al [42]	2017	Case-control study	Logistic regression	Asian women; 20-90 years	11	FGFR2 (rs2981582), HCN1 (rs981782), MAP3K1 (rs889312), TOX3(rs3803662), ZNF365(rs10822013), RAD51B(rs3784099), age, body mass index, age at menarche, parity, menopausal status	AUC: 0.6652; E/O ratio: none	446 cases/ 514 controls	None	None	None	None	None

Husing et al [43]	2017	Nested case-control study	Logistic regression	Multiple ethnicities; 26-77 years	13	Menopausal status, age at menarche, age at menopause, duration of postmenopausal hormones use, parity, number of children and age at first full term pregnancy, family history of breast cancer, alcohol consumption at recruitment, body mass index, measurements of testosterone, estradiol, sex hormone binding globulin, Insulin-like growth factor-I	AUC: none; E/O ratio: none	1,217 cases/ 1,976 controls	Internal validation	None	None	None	None
Salih et al [44]	2017	Cross-sectional study	Logistic regression	Caucasian; 32–74 years	5	Age, age at menarche, family history, vegetables and fruits weekly servings, type of cereals used	AUC: 0.864(0.81,0.92)	63 cases/ 90 controls	Internal validation	None	None	O/E ratio: 0.78[b]	None
Wang et al [45]	2018	Case–control study	Logistic regression	Nigerian women; age was not specified	9	Age, age at menarche, parity, duration of breastfeeding, family history of breast cancer, height, body mass index, benign breast diseases, alcohol consumption	AUC: 0.720(0.701,0.739); E/O ratio: 1.01 (0.93,1.09)	1,208 cases/ 1,484 controls	Internal validation	None	Nigerian women; 20-79 years	AUC: 0.694 (0.666,0.721); E/O ratio: none	603 cases/ 741 controls

Zhang et al [46]	2018	Nested case-control study	Logistic regression	Caucasian; 34-70 years	1) Gail model + PRS + MD + T + E1S +PRL: 10; 2) Rosner-Colditz model + PRS + MD + T + E1S + PRL: 16	1) Gail model+ PRS + MD + T + E1S +PRL: Age, age at menarche, previous biopsies, age at first birth, first degree breast cancer, PRS, MD, E1S, T, PRL 2) Rosner-Colditz model+ PRS + MD + T + E1S + PRL: age, age at menarche, age at first birth, menopause, age at subsequent births, benign breast disease, hormone replacement therapy, first degree breast cancer, weight, body mass index, alcohol, PRS, MD, E1S, T, PRL	AUC: Gail model+ PRS + MD + T + E1S +PRL: 0.65(0.64,0.66); Rosner-Colditz model+ PRS + MD + T + E1S + PRL: 0.678 (0.666,0.690); E/O ratio: none	4,006 cases / 7,874 controls	Internal validation	None	None	None	None
Clendenen et al [47]	2019	Nested case-control study	Logistic regression	Multiple ethnicities; 35–50 years	6	Age at menarche, age at first live birth, number of benign breast biopsies, number of first-degree family members with breast cancer, AMH, tT	AUC: 0.581(0.562,0.599); E/O ratio: none	1,762 cases/ 1,890 controls	None	None	None	None	None
Wang et al [48]	2019	Case-control study	Logistic regression	Asian women; 25-70 years	6	Number of abortions, age at first live birth, benign breast disease history, body mass index,	None	328 cases / 656 controls	External validation	Prospective cohort study	Asian women	AUC: 0.64 (0.55,0.72); E/O ratio: 1.03 (0.74,1.49)	34 cases/ 13,176 total

						breast cancer family history, life satisfaction score						
Abdolell et al [49]	2020	Nested case-control study	Logistic regression	Caucasian; 40-75 years	5	Age at screen, percent mammographic density, breast volume, core biopsy history, family history	AUC: 0.664(0.650,0.678); E/O ratio: none	1,882 cases/ 5,888 controls	None	None	None	None
Qiu et al [50]	2020	Case–control study	Logistic regression	Asian women; 29-81 years	5	p53, CyclinB1, p16, p62,14-3-3ξ	AUC:0.943(0.919,0.967); E/O ratio: none	184 cases/ 184 controls	External validation	Case-control study	Asian women; 24-78 years	AUC: 0.916(0.886,0.947); E/O ratio: none	197 cases/ 109 controls

[a]E/O ratios were calculated based on the original information. [b]The original publication reported the Observed/Expected ratio.
ER: estrogen receptor; PR: progesterone receptor; PRS: polygenic risk score; MD: mammographic density; E1S: estrone sulphate; T: testosterone; PRL: prolactin; AMH: anti-Müllerian hormone.

Appendix Table 4. Risk of bias assessment of included based on PROBAST.

Study	Participants			Predictors				Outcome							Analysis									Overall
	1.1	1.2		2.1	2.2	2.3		3.1	3.2	3.3	3.4	3.5	3.6		4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	
Gail et al [6]	N	Y	H	Y	PY	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	N	N	Y	PY	N	N	Y	H	H
Rosner et al [16]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	NI	N	N	Y	H	H
Ueda et al [17]	N	NI	H	Y	PY	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	Y	Y	PY	N	N	Y	H	H
Colditzet al [18]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	L	Y	N	Y	N	Y	N	N	Y	Y	H	H
Lee et al [19]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	Y	N	PY	N	N	Y	H	H
Tice et al [20]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	PY	Y	N	N	N	Y	H	H
Tice et al [21]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	PY	Y	N	N	N	Y	H	H

Barlow et al [22]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	PY	NI	Y	N	N	Y	Y	H	H
Decarli et al [23]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	N	N	Y	N	Y	N	Y	H	H
Decarli et al [23*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	N	NI	-	NI	Y	-	-	H	H
Novotny et al [24]	N	PY	H	Y	PN	Y	H	PY	Y	Y	Y	Y	Y	L	Y	N	N	N	Y	PY	N	N	Y	H	H
Gail et al [25]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	PY	N	N	Y	H	H
Gail et al [25*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	-	Y	Y	-	-	H	H
Anna et al [26]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	PY	N	N	Y	H	H
Tice et al [27]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	N	N	Y	Y	Y	Y	Y	H	H
Tamimi,et al [28]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	NI	N	N	Y	H	H
Petracci et al [29]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	Y	PY	N	N	Y	H	H
Petracci et al [29*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	-	Y	PN	-	-	H	H
Dite et al [30]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	N	PY	N	N	Y	H	H
Park et al [31]	N	Y	H	Y	PY	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	N	PY	N	N	Y	H	H
Park et al [31*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	NI	-	PY	Y	-	-	H	H
Anothaisintawee et al [32]	Y	Y	H	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	Y	N	PY	PN	N	Y	H	H
Anothaisintawee et al [32*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	Y	-	PY	PN	-	-	H	H
Boggs et al [33]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	NI	Y	N	Y	Y	H	H
Brentnall et al [34]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	Y	N	N	N	Y	H	H
Kerlikowske et al [35]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	N	PN	Y	Y	H	H
Tice et al [36]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	Y	PN	Y	Y	H	H
Schonberg et al [37]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	PY	N	Y	N	N	Y	H	H
Schonberg et al [37*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	-	Y	PN	-	-	H	H
Shieh et al [38]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	N	N	N	Y	Y	H	H
Wang et al [39]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	N	PY	N	Y	Y	H	H
Banegas, et al [40]	N	Y	H	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	PY	N	NI	Y	H	H
Banegas et al [40*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	-	PY	Y	-	-	H	H

Eriksson et al [41]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	NI	PY	N	Y	Y	H	H
Hsieh, et al [42]	N	NI	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	N	PY	N	Y	Y	H	H
Husing et al [43]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	PY	N	Y	Y	H	H
Salih et al [44]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	N	PY	N	Y	Y	H	H
Wang et al [45]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	Y	PY	PN	Y	Y	H	H
Zhang et al [46]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	N	Y	Y	H	H
Clendenen et al [47]	Y	Y	L	PN	Y	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	N	N	Y	H	H
Wang et al [48]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	Y	N	N	Y	H	H
Wang et al [48*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	N	-	Y	PN	-	-	H	H
Abdolell et al [49]	Y	PY	L	Y	Y	Y	L	PY	Y	Y	Y	Y	Y	L	Y	Y	Y	N	N	PY	N	N	Y	H	H
Qiu et al [50]	N	NI	H	Y	N	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	PY	N	N	Y	H	H
Qiu et al [50*]	N	NI	H	Y	N	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	-	PY	N	-	-	H	H

* The external validation was performed in the same study.

L: low risk of bias; H: high risk of bias; Y: yes; N: no; PY: probably yes; PN: probably no; NI: no information; -: not applicable.

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review.	1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	2,3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	4,5
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	5
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	6
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	6
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	6 and Appendix Table 1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	7
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	7
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	8,9 and Appendix Table 2
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	8,9 and Appendix Table 2
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	7,8
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	8,9
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	8,9
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	8,9 and Appendix Table 3
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	8,9
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	8,9
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Not performed
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Not performed

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Not performed
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Not performed
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	9 and figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	9
Study characteristics	17	Cite each included study and present its characteristics.	9,10
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	12,13,14,15
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	9,10,11 and Appendix Table 3
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Appendix Table 3, table 1, figure 2, figure 3 and figure 4
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	9, 10,11,12, 13,14,15 and Appendix Table 4
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Not performed
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Not performed
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not performed
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Not performed
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	15, 16
	23b	Discuss any limitations of the evidence included in the review.	17
	23c	Discuss any limitations of the review processes used.	17
	23d	Discuss implications of the results for practice, policy, and future research.	17, 18, 19
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	3,6
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Not performed

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not performed
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	20
Competing interests	26	Declare any competing interests of review authors.	20
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Appendix Table 1,2,3

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

BMJ Open

Risk prediction models for breast cancer: a systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055398.R1
Article Type:	Original research
Date Submitted by the Author:	21-Jan-2022
Complete List of Authors:	Zheng, Yadi; Chinese Academy of Medical Sciences & Peking Union Medical College Li, Jiang; Chinese Academy of Medical Sciences & Peking Union Medical College Wu, Zheng; Chinese Academy of Medical Sciences & Peking Union Medical College Li, He; Chinese Academy of Medical Sciences & Peking Union Medical College Cao, Maomao; Chinese Academy of Medical Sciences & Peking Union Medical College Li, Ni; Chinese Academy of Medical Sciences & Peking Union Medical College He, Jie; Chinese Academy of Medical Sciences & Peking Union Medical College
Primary Subject Heading:	Public health
Secondary Subject Heading:	Epidemiology, Oncology, Public health
Keywords:	Epidemiology < TROPICAL MEDICINE, Public health < INFECTIOUS DISEASES, Breast tumours < ONCOLOGY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Risk prediction models for breast cancer: a systematic review

Yadi Zheng[1], Jiang Li[*1], Zheng Wu[1], He Li[1], Maomao Cao[1], Ni Li[*1], Jie He[2]

[1] Office of Cancer Screening, National Cancer Center/ National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

[2] Department of Thoracic Surgery, National Cancer Center / National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

Corresponding authors: Prof. Jiang Li lij@cicams.ac.cn; Prof. Ni Li nli@cicams.ac.cn

Word count: 2896

Figures/Tables: 5

ABSTRACT

Objectives: To systematically review and critically appraise published studies of risk prediction models for breast cancer in the general population without breast cancer, and provide evidence for future research in the field.

Design: Systematic review using the Prediction model study Risk Of Bias Assessment Tool (PROBAST) framework.

Data sources: PubMed, the Cochrane Library and Embase were searched from inception to 16 December, 2021.

Eligibility criteria: We included studies reporting multivariable models to estimate the individualized risk of developing female breast cancer among different ethnic groups. Search was limited to English language only.

Data extraction and synthesis: Two reviewers independently screened, reviewed, extracted and assessed studies with discrepancies resolved through discussion or a third reviewer. Risk of bias was assessed according to the PROBAST (Prediction model Risk of Bias Assessment Tool) framework.

Results: 63,894 studies were screened and 40 studies with 47 risk prediction models were included in the review. Most of the studies used logistic regression to develop breast cancer risk prediction models for Caucasian women by case-control data. The most

widely used risk factor was reproductive factors and the highest area under the curve was 0.943 (95% confidence interval: 0.919~0.967). All the models included in the review had high risk of bias.

Conclusions: No breast cancer risk prediction models were recommended for different ethnic groups and more key variables like breast density and single-nucleotide polymorphisms (SNPs) should be collected and well validated in the existing models in the future. High-quality breast cancer risk prediction models assessed by PROBAST should be developed and validated, especially among Asian women.

PROSPERO registration number： CRD42020202570

Strengths and limitations of this study

1. Thoroughly conducted systematic review collecting data from major existing databases.
2. Critically appraised published studies of risk prediction models for breast cancer in the general population and provide evidence for future research in the field.
3. PROBAST was used to assess the quality of prediction models, which was developed through a consensus process involving a group of methodological experts in the area of clinical prediction tools and quality assessment.
4. Studies only about the external validation of the present risk models were not included in the review.

5. Our study highlighted high-quality breast cancer risk prediction models assessed by PROBAST should be developed and validated among different ethnic groups, especially among Asian women.

Keywords: breast cancer; risk prediction model; review; quality assessment; Prediction model Risk of Bias Assessment Tool

INTRODUCTION

Breast cancer is a major public health problem, and one of the most severe burdensome cancer among women worldwide [1], accounting for11.7% of new cancer cases and 6.9% of cancer deaths in 2020. The prevalence of breast cancer is projected to increase over the coming years and is the most common cancer in women in 2020 [2]. Breast cancer prevention is associated with a reduction in mortality [3], and more researches are needed to improve the methods of identifying women at elevated risk and preventing the disease. Numerous breast cancer risk prediction models have been developed to identify the combined effect of risk factors for breast cancer, guide routine screening and genetic testing, and reduce the burden of breast cancer. Risk-stratified screening can improve cost-effectiveness and maximize benefits and minimize harms like overdiagnosis [4]. Individualized prediction model for breast cancer could be used in practice to assist decision making about mass screening or opportunistic screening and treatment strategy.

A recent breast cancer screening guideline [5] suggests that breast cancer screening increase the early detection rate and reduce the incidence if the screening is applied in appropriate at-risk populations. However, major gaps exist in our knowledge to determine the risk of breast cancer accurately in order to apply these approaches to appropriate populations of women.

A lot of breast cancer risk prediction models have been developed over the past few decades. Many breast cancer risk models have undergone validation including discrimination and calibration in study populations other than those used in initial development, or have been further assessed in comparative studies. Breast cancer related predictors including hormonal factors, environmental factors, family histories, genetic factors and radiographic factors have been based on in these risk models, which would improve the generalizability. For example, the Gail model [6], one of the most famous models, has been widely used and validated worldwide since it was developed in 1989 [7-12].

This study is a systematic review of breast cancer risk prediction models by using meta-analysis and the Prediction model Risk of Bias Assessment Tool (PROBAST) [13-14]. The aim of our study is to systematically review published studies of risk prediction models for breast cancer in the general population, find more methods of predicting

female breast cancer risk among one or more ethnic groups, prepare for the development of risk prediction models, and provide evidence for future research in the field.

METHODS

Protocol and registration

The current review was designed according to the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [15] and was recorded in the PROSPERO database (registration number: CRD42020202570).

Literature search and eligibility criteria

We systematically searched PubMed, the Cochrane Library and Embase from inception to 16 December, 2021. The detailed search strategies were reported in Appendix Table 1. Articles identified from the search were loaded into EndNote X7 and duplicates were removed.

Inclusion criteria: 1) a model used data from cross-sectional studies, cohort studies, case-control studies, and randomized controlled trials; 2) a model estimating the individualized risk of female breast cancer among one or more ethnic groups; 3) a model developed for the general population without breast cancer; 4) reported a multivariable (i.e., at least 2 variables or predictors) model; 5) published in English.

Exclusion criteria: 1) external validation studies that only validated previous models in a different population without adding any additional information such as modifications on the risk factors; 2) models developed by machine learning.

Data extraction

Two reviewers screened the search results independently. Full text reports were then assessed for eligibility with discrepancies resolved through discussion or a third reviewer.

We extracted information in two categories: 1) For all studies included in the review, we extracted the following information: author, publication year, study design, research method, targeted population, number of risk factors, risk factors, model performance and sample size of development. 2) For studies included validation part, we also extracted the following information: type of validation, study design, targeted population, model performance and sample size of validation. The information was extracted by one reviewer and checked by a second reviewer.

Risk of bias assessment

We used PROBAST to assess the reported prediction models, which is a new tool designed by a group of experts all over the world to assess the risk of bias and applicability of diagnostic and prognostic prediction models. It can be used in critical appraisal of studies that develop, validate, or update prediction models for individualized predictions

[13-14]. In brief, it contains 20 signaling questions in four domains: participants, predictors, outcome, and statistical analysis. Signaling questions can be answered as yes, probably yes, no, probably no, or no information. A domain where at least one signaling question is answered as no or probably no should be judged as high risk of bias. Only if all domains are judged as low risk of bias, the total bias is judged as low risk as well.

Before putting PROBAST into use, we formed a ten-people study group including prediction model researchers, statisticians, evidence-based medicine specialists etc. to learn and practice the appropriate use of this new tool systematically. Only after everyone understood all these twenty questions totally, we would move to the peer quality assessment part. Risk of bias of every prediction model was assessed by two reviewers independently with discrepancies resolved through discussion or a third reviewer.

If there were more than one models developed in one study, we only assessed the risk of bias once due to their similarity. We also assessed the risk of external validation of prediction model when it was conducted in the same article that included model development.

Data synthesis and analysis

We calculated and reported descriptive statistics to summarize the characteristics of the models. We calculated the most frequently used risk factors and classified all risk

factors into eight categories: Age, reproductive factors, family history of cancer, hormone, gene-related factors, lifestyle, medical history and test, and basic information. Classification details can be seen in Appendix Table 2. Then we used network diagram to see the connections of categorized risk factors. We used forest plot to describe the model performance. The expected observed (E/O) ratio was not included in the forest plot because it was only reported in 7 out of 40 studies. All analyses were performed using Stata 16.0 and NetDraw.

Patient and public involvement

There was no patient or public involvement in this study.

RESULTS

Study selection

A total of 92,519 indexed records (54,653 in PubMed, 30,374 in Cochrane Library and 7,492 in Embase), 28,625 were eliminated as duplicates found in all databases, leaving a total of 63,894 publications. 43 articles were included primarily after screening by title and abstract. 3 studies which were only about the external validation of previous models were excluded while full test screening, resulting in 40 studies with 47 models were included in the review eventually. (Figure 1).

Study characteristics

A brief summary of the 40 [6,16-54] included studies is presented in Appendix Table 3. The included studies were published from 1989 to 2021. 25 of the studies were conducted over the past ten years with 5 studies published in 2017 especially. Seventeen out of the forty studies used data from case-control studies to develop prediction models [6,17,19,23-26,29-31,39, 41,43,46,49,51,54], thirteen from prospective cohorts [16,18,20-22,27,33-37,40,52], eight from nested case-control studies [28,38, 42,44,47,48,50,53] and two from cross-sectional study [32,45]. Thirty-one studies used logistic regression to fit prediction models [6,17-19,22-26,28-32,34,38-51,53,54], seven used cox proportional hazards regression [20,21,27,33,35,36,52], one used Poisson regression [16] and one used competing risk regression [37]. Of all forty-seven models in forty studies, sixteen models were developed in Caucasian women [6,16,18,23,26,28,29,34,40,42,45,47,50,53], thirteen in multiple ethnicities women [20-22,24,27,30,35-38,44,48], twelve in Asian women [17,19,31,32,39,43,49,51,52], two in African-American women [25,33], two in Hispanic women [41], one in Nigerian women [46] and one in Cypriot Women [54].

The association between eight categories of risk factors was shown in Figure 2. Reproductive factors had the biggest node size, which meant that this factor was most frequently connected with other factors among prediction models. The number between two factors meant the times these two factors were included in the same models, some of which were over thirty. For instance, reproductive factors and family history of cancer

were included in the same models for forty times, and reproductive factors and age were included in the same models for thirty-one times.

Twenty-nine studies reported c-statistics [18-22,26-28,30-32,34-40,42,43,45-48,50-54], ranged from 0.59(95% confidence interval: 0.57~0.61) to 0.943(95% confidence interval: 0.919~0.967). Qiu, et al [51] had the highest c-statistics (0.943, 95% confidence interval: 0.919~0.967), and Lee et al [19] and Salih et al [45] reported area under the curve (AUC) over 0.8, 0.867 and 0.864(95% confidence interval: 0.81~0.92), respectively. E/O ratios can be obtained from eight studies [22,27,29,32,35,36,46,52]. Figure 3 showed that the overall AUC was 0.68(95% confidence interval: 0.63~0.73) for sixteen studies [21,26,27,30,32,34,37,38,42,45,46,48,50,51,52,54] that reported the AUC and 95% confidence interval. The AUCs of the subgroups in five studies [18,22,31,39,47] were between 0.6 to 0.7.

In all these forty studies, nine studies assessed prediction models with internal validation [22,26,27,33,39,44-47], ten with external validation [23,25,29,31,37,41,49,51-53], and one with both [32]. Fifteen studies reported the discriminatory accuracy as the AUC [23,25,27,29,31-33,37,39,41,46,49,51-53], and eleven studies used the expected/observed event ratio (or observed/expected event ratio) to measure the calibration accuracy of the model [23,25,27,29,31,33,37,41,45,49,52].

Quality assessment

A summary of the quality assessment is shown in Table 1. Overall, all models assessed by PROBAST in the review had high risk of bias. There was a low and high risk of bias in the outcome and analysis domains respectively. Over 60% models had low risk in participants domain and about 70% models had low risk in predictors domain, 32 models and 36 models respectively. (As shown in Figure 4).

The main reasons for the high risk in analysis domain were model performance measures evaluated inappropriately, categorization of continuous predictors, no reporting of overfitting and optimism in model performance and missing data handled inappropriately (Appendix Table 4).

Table 1. Summary of risk of bias assessment.

Study	Participants	Predictors	Outcome	Analysis	Overall
Gail et al [6]	H	L	L	H	H
Rosner et al [16]	L	L	L	H	H
Ueda et al [17]	H	L	L	H	H
Colditzet al [18]	L	L	L	H	H
Lee et al [19]	H	H	L	H	H
Tice et al [20]	L	L	L	H	H
Tice et al[21]	L	L	L	H	H
Barlow et al [22]	L	L	L	H	H
Decarli et al [23]	H	H	L	H	H
Decarli et al [23*]	L	L	L	H	H
Novotny et al [24]	H	H	L	H	H
Gail et al [25]	H	H	L	H	H
Gail et al [25*]	L	L	L	H	H
Anna et al [26]	H	H	L	H	H

Tice et al [27]	L	L	L	H	H
Tamimi,et al [28]	L	L	L	H	H
Petracci et al [29]	H	H	L	H	H
Petracci et al [29*]	L	L	L	H	H
Dite et al [30]	H	H	L	H	H
Park et al [31]	H	H	L	H	H
Park et al [31*]	L	L	L	H	H
Anothaisintawee et al [32]	H	L	L	H	H
Anothaisintawee et al [32*]	L	L	L	H	H
Boggs et al [33]	L	L	L	H	H
Brentnall et al [34]	L	L	L	H	H
Kerlikowske et al [35]	L	L	L	H	H
Tice et al [36]	L	L	L	H	H
Schonberg et al [37]	L	L	L	H	H
Schonberg et al [37*]	L	L	L	H	H
Shieh et al [38]	L	L	L	H	H
Wang et al [39]	H	H	L	H	H
Mass et al [40]	L	L	L	H	H
Banegas, et al [41]	H	L	L	H	H
Banegas et al [41*]	L	L	L	H	H
Eriksson et al [42]	L	L	L	H	H
Hsieh, et al [43]	H	H	L	H	H
Husing et al [44]	L	L	L	H	H
Salih et al [45]	L	L	L	H	H
Wang et al [46]	H	H	L	H	H
Zhang et al [47]	L	L	L	H	H
Clendenen et al [48]	L	H	L	H	H
Wang et al [49]	H	H	L	H	H
Wang et al [49*]	L	L	L	H	H
Abdolell et al [50]	L	L	L	H	H
Qiu et al [51]	H	H	L	H	H
Qiu et al [51*]	H	H	L	H	H
Han et al [52]	L	L	L	H	H
Han et al[52]*	L	L	L	H	H

Rosner et al [53]	L	L	L	H	H
Rosner et al [53]*	L	L	L	H	H
Yiangou et al [54]	H	L	L	H	H

* The external validation was performed in the same study.
L indicates low risk of bias; H indicates high risk of bias.

DISCUSSION

Summary of main results

This systematic review identified 40 studies with 47 risk prediction models developed and/or validated for breast cancer among different ethnic groups. Most of the studies used logistic regression to develop breast cancer risk prediction models for Caucasian women by case-control data. The most widely used risk factor was reproductive factors. Reproductive factors together with family history factor were used in most models. The highest AUC was 0.943 (95% confidence interval: 0.919~0.967) from Qiu, et al [51]. The overall AUC was 0.68(95% confidence interval: 0.63~0.73) for sixteen studies [21,26,27,30,32,34,37,38,42,45,46,48,50,51,52,54] that reported the AUC and 95% confidence interval. All the studies presented a high risk of bias due to the high risk in analysis domain, which were mainly because of model performance measures evaluated inappropriately, categorization of continuous predictors, no reporting of overfitting and optimism in model performance and missing data handled inappropriately.

Agreements and disagreements with other reviews

As we can learn from the review, there were more and more risk prediction models of breast cancer over the past thirty years. Most of the models were developed in the Caucasian women, which agreed with the systematic review published by Louro et al in 2019 [55]. Compared with this review, we identified more prediction models and used a newly published tool to assess the quality of included models.

Over the past ten years, some new variables (such as oral contraceptives, diabetes, and alcohol consumption) have been included in prediction models. Increased use of the inclusion of common genetic variation in the prediction models was in accord with Louro et al in 2019 [55] and Anothaisintawee et al in 2012 [56]. However, neither of them included models developed with potential biomarkers like tumor-associated antigens. By contrast, we included one model developed by Qiu, et al [51] in 2019 included five tumor-associated antigens. The model performed well with a high AUC 0.943(95% confidence interval: 0.919,0.967).

Strengths and limitations of the study

PROBAST was developed through a consensus process involving a group of methodological experts in the area of clinical prediction tools and quality assessment. We used it to assess the quality of prediction models, which has been used widely in many fields [57-60] since it came out.

Despite the strength, there are three main limitations. Firstly, we didn't systematically search gray literature. Therefore, some models may not be identified. Secondly, quality assessment could be thought to be subjective, which is an inherent bias of systematic review. However, two independent reviewers extracted and assessed the risk prediction models using PROBAST whose authors have indicated essentially objective guidelines and explanations. Moreover, studies only about the external validation of the present risk models were not included in the review, but the original developments of these risk models were covered. For instance, the study describes the original developments of Gail model [6] was included in our research, while the studies only about the external validation of Gail model [61-64] were not included.

Implication to research and clinical practice

Eleven models [19,30-32,37-39,43,45,50,54] selected predictors based on univariable analysis, causing a high risk in analysis domain, which should be avoided. Risk prediction models should include predictors those are well-established and with clinical credibility regardless of any statistical significance [65,66]. Because sometimes predictors only have important relationship with the outcome after adjustment for confounding covariates, and covariates hold no independent predictive power when other covariates are included [13,67].

Some models were high risk in analysis domain because of missing data handled

inappropriately, which may lead to biased associations between risk factors and breast cancer as well as biased model performance because of the selectivity of participants [68]. So imputation techniques are supposed to apply when data are missing [69,70].

When developing the risk prediction models, there were only nine studies included internal validation [22,26,27,33,39,44-47], leaving most models without internal validation. Lack of performing internal validation may increase the risk of overfitting [71]. Thus, we suggest that internal validation should be performed before external validation.

PROBAST was created by many international experts, providing a series of guidelines about model development and validation, which can be easily applied and improve clinical practice of prediction models. So, the new and most recommended methodology should be used when a new model is developed or the existing models are updated.

In the light of the results of our review, it is still hard to recommend any of the models to be applied in the breast cancer screening due to the high risk of bias. More key variables like mammographic density and single-nucleotide polymorphisms (SNPs) should be well collected and validated in the existing models to improve the model performance. High mammographic density is a strong risk factor for breast cancer [72,73], and several studies have found that mammographic density improves the accuracy of risk-

prediction models [74,75]. Studies have shown that adding SNPs into risk-prediction models can improve model performance with promising results [76-78]. Cost-effectiveness should be considered when a model is going to be applied in clinical practice. Because even though the model with some risk factors that cost more to get (e.g., high risk gene) has better model performance, it is still hard to be applied in poor area [79]. What's more, an existing model should be modified or updated before used in another group of people with different characteristics, which may improve the performance of prediction models.

Breast cancer incidence has risen to the first place by 2020 all over the world, which makes it more crucial to develop breast cancer prediction models for different ethnic groups. In China, we have launched many breast cancer screening programs. For example, Rural Women "two cancers" Check Project Management Solutions have covered 31 provinces and 1437 counties since 2009. Cancer Screening Program in Urban China conducted by the National Cancer Center has covered 28 provinces and 67 cities with more than 4 million people involved and 2 million people screened by ultrasound and Mammography since 2012. The program will provide large data for us to develop a high-quality breast cancer risk prediction model in Chinese and will have great significance for breast cancer prevention of Asian women.

CONCLUSIONS

All 47 models assessed in our review using PROBAST performed the high risk of bias, leaving no model is recommended in the routine screening program. Some new variables, like oral contraceptives, diabetes, and alcohol consumption, have been widely used in prediction models over the past ten years. More key variables like breast density and SNPs should be collected and well validated in the existing models to improve the model performance. It is necessary to develop and validate high-quality breast cancer risk predication models among different ethnic groups, especially among Asian women.

Contributors: YZ and JL conceptualized the study and created the first version of the review protocol. ZW, HL, MC, NL, and JH critically reviewed the review protocol and approved it. YZ and HL screened eligible articles. YZ extracted the data, supported by ZW, MC. YZ drafted the first version of the manuscript, supported by JL, NL, and JH. All authors contributed to data interpretation and critically assessed it. All authors approved the final version of the manuscript.

Funding: This work was supported by the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (grant number: 2019PT320027).

Competing interest: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Patient consent for publication： Not required.

Data availability statement: All data of the current study is present in the main manuscript, figures, tables, and online supplemental material.

Ethics statement: This study does not involve human participants.

REFERENCES

1 GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*. 2020;396(10258):1204–1222. doi: 10.1016/S0140-6736(20)30925-9.

2 Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021. doi: 10.3322/caac.21660.

3 Myers ER, Moorman P, Gierisch JM, et al. Benefits and harms of breast cancer screening: A systematic review. *JAMA*. 2015;314(15):1615-1634. doi: 10.1001/jama.2015.13183.

4 Pashayan N, Morris S, Gilbert FJ, et al. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer. *JAMA Oncol*. 2018;4(11): 1504–1510. doi: 10.1001/jamaoncol.2018.1901.

5 Qaseem A, Lin JS, Mustafa RA, Horwitch CA, Wilt TJ; Clinical Guidelines

Committee of the American College of Physicians. Screening for breast cancer in average-risk women: A guidance statement from the American college of physicians. *Ann Intern Med*. 2019;170(8):547-560. doi: 10.7326/M18-2147.

6 Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81(24):1879-1886. doi: 10.1093/jnci/81.24.1879.

7 Bondy ML, Newman LA. Assessing breast cancer risk: evolution of the Gail Model. *J Natl Cancer Inst*. 2006;98(17):1172-1173.

8 Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst*. 1999;91(18):1541-1548.

9 Rockhill B, Spiegelman D, Byrne C, et al. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst*. 2001;93(5):358-366.

10 Lin CJ, Block B, Nowalk MP, et al. Breast cancer risk assessment in socioeconomically disadvantaged urban communities. *J Natl Med Assoc*. 2007;99(7):752-756.

11 Bondy ML, Newman LA. Breast cancer risk assessment models: applicability to

African-American women. *Cancer*. 2003;97(1 Suppl):230-235.

12 Buron A, Vernet M, Roman M, et al. Can the Gail model increase the predictive value of a positive mammogram in a European population population screening setting? Results from a Spanish cohort. *Breast*. 2013;22(1):83-88.

13 Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1): 51-58. doi: 10.7326/M18-1376.

14 Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Ann Intern Med*. 2019;170(1): W1-W33. doi: 10.7326/M18-1377.

15 Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10): e1001744. doi: 10.1371/journal.pmed.1001744.

16 Rosner B, Colditz GA. Nurses' Health Study: Log-Incidence mathematical model of breast cancer incidence. *J Natl Cancer Inst*. 1996;88(6): 359-64. doi: 10.1093/jnci/88.6.359.

17 Ueda K, Tsukuma H, Tanaka H, et al. Estimation of individualized probabilities of developing breast cancer for Japanese women. *Breast Cancer*. 2003;10(1):54-62.

doi: 10.1007/BF02967626.

18 Colditz GA, Rosner BA, Chen WY, et al. Risk factors for breast cancer according to estrogen and progesterone receptor status. *J Natl Cancer Inst*. 2004;96(3): 218-28. doi: 10.1093/jnci/djh025.

19 Lee EO, Ahn SH, You C, et al. Determining the main risk factors and high-risk groups of breast cancer using a predictive model for breast cancer risk assessment in south Korea. *Cancer Nurs*. 2004;27(5): 400-6. doi: 10.1097/00002820-200409000-00010.

20 Tice JA, Miike R, Adduci K, et al. Nipple aspirate fluid cytology and the Gail Model for breast cancer risk assessment in a screening population. *Cancer epidem biomar*. 2005;14(2): 324-8. doi: 10.1158/1055-9965.EPI-04-0289.

21 Tice JA, Cummings SR, Ziv E, et al. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res*. 2005;94(2): 115-22. doi: 10.1007/s10549-005-5152-4.

22 Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst*. 2006;98:1204-1214. doi:10.1093/jnci/djj331.

23 Decarli A, Calza S, Masala G, et al. Gail Model for prediction of absolute risk of

invasive breast cancer: Independent evaluation in the Florence-European prospective investigation into cancer and nutrition cohort. *J Natl Cancer Inst*. 2006;98(23): 1686-93. doi: 10.1093/jnci/djj463.

24 Novotny J, Pecen L, Petruzelka L, et al. Breast cancer risk assessment in the Czech female population-an adjustment of the original Gail model. *Breast Cancer Res Treat*. 2006;95(1): 29-35. doi: 10.1007/s10549-005-9027-5.

25 Gail MH, Costantino JP, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst*. 2007;99(23): 1782-92. doi: 10.1093/jnci/djm223.

26 Crispo A, D'Aiuto G, De Marco M, et al. Gail Model risk factors: Impact of adding an extended family history for breast cancer. *Breast Journal*. 2008;14(3): 221-27. doi:10.1111/j.1524-4741.2008.00566.

27 Tice JA, Cummings SR, Smith-Bindman R, et al. Using clinical factors and mammographic breast density to estimate breast cancer risk: Development and validation of a new predictive model. *Ann Intern Med*. 2008;148(5): 337-47. doi: 10.7326/0003-4819-148-5-200803040-00004.

28 Tamimi RM, Rosner B, Colditz GA. Evaluation of a breast cancer risk prediction model expanded to include category of prior benign breast disease lesion. *Cancer*.

2010;116(21): 4944-53. doi: 10.1002/cncr.25386.

29 Petracci E, Decarli A, Schairer C, et al. Risk factor modification and projections of absolute breast cancer risk. *J Natl Cancer Inst*. 2011;103(13): 1037-1048. doi:10.1093/jnci/djr172.

30 Dite GS, Mahmoodi M, Bickerstaffe A, et al. Using SNP genotypes to improve the discrimination of a simple breast cancer risk prediction model. *Breast Cancer Res Treat*. 2013;139(3): 887-96. doi: 10.1007/s10549-013-2610-2.

31 Park B, Ma SH, Shin A, et al. Korean risk assessment model for breast cancer risk prediction. *PLOS One*. 2013;8(10): e76736. doi: 10.1371/journal.pone.0076736.

32 Anothaisintawee T, Teerawattananon Y, Wiratkapun C, et al. Development and validation of a breast cancer risk prediction model for Thai women: a cross-sectional study. *Asian Pac J Cancer Prev*. 2014;15(16): 6811-7. doi: 10.7314/apjcp.2014.15.16.6811.

33 Boggs DA, Rosenberg L, Adams-Campbell LL, et al. Prospective approach to breast cancer risk prediction in African American women: The black women's health study model. *J Clin Oncol*. 2015;33(9): 1038-44. doi: 10.1200/JCO.2014.57.2750.

34 Brentnall1 AR, Harkness EF, Astley SM, et al. Mammographic density adds

accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res*. 2015;17(1): 147. doi: 10.1186/s13058-015-0653-5.

35 Kerlikowske K, Gard CC, Sprague BL, et al. One versus two breast density measures to predict 5- and 10-year breast cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2015;24(6): 889-897. doi: 10.1158/1055-9965.EPI-15-0035.

36 Tice JA, Miglioretti DL, Li CS, et al. Breast density and benign breast disease: Risk assessment to identify women at high risk of breast cancer. *J Clin Oncol*. 2015;33(28): 3137-3143. doi:10.1200/JCO.2015.60.8869.

37 Schonberg MA, Li VW, Eliassen AH, et al. Accounting for individualized competing mortality risks in estimating postmenopausal breast cancer risk. *Breast Cancer Res Treat.* 2016;160(3): 547-562. doi: 10.1007/s10549-016-4020-8.

38 Shieh Y, Hu D, Ma L, et al. Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res Treat*. 2016;159(3): 513-525. doi: 10.1007/s10549-016-3953-2.

39 Wang F, Dai J, Li M, et al. Risk assessment model for invasive breast cancer in Hong Kong women. *Medicine*. 2016;95(32): e4515. doi:10.1097/MD.0000000000004515.

40 Maas P, Barrdahl M, Joshi AD, et al. Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol.* 2016;2(10): 1295-1302. doi: 10.1001/jamaoncol.2016.1025.

41 Banegas MP, John EM, Slattery ML, et al. Projecting individualized absolute invasive breast cancer risk in US Hispanic women. *J Natl Cancer Inst*. 2017;109(2): djw215. doi: 10.1093/jnci/djw215.

42 Eriksson M, Czene K, Pawitan Y, et al. A clinical model for identifying the short-term risk of breast cancer. *Breast Cancer Res*. 2017;19(1):29. doi:10.1186/s13058-017-0820-y.

43 Hsieh YC, Tu SH, Su CT, et al. A polygenic risk score for breast cancer risk in a Taiwanese population. *Support Care Cancer*. 2017;163(1): 131-138. doi:10.1007/s10549-017-4144-5.

44 Hüsing A, Fortner RT, Kühn T, et al. Added value of serum hormone measurements in risk prediction models for breast cancer for women not using exogenous hormones: Results from the EPIC cohort. *Clin Cancer Res*. 2017;23(15): 4181-4189. doi: 10.1158/1078-0432.CCR-16-3011.

45 Salih AM, Alam-Elhuda DM, Alfaki MM, et al. Developing a risk prediction model for breast cancer: A Statistical Utility to Determine Affinity of Neoplasm (SUDAN-

CA Breast). *Eur J Med Res*. 2017;22(1): 35. doi:10.1186/s40001-017-0277-6.

46 Wang SF, Ogundiran T, Ademola A, et al. Development of a breast cancer risk prediction cancer epidemiology, biomarkers & prevention model for women in Nigeria. *Cancer Epidem Biomar*. 2018;27(6): 636-643. doi: 10.1158/1055-9965.EPI-17-1128.

47 Zhang X, Rice M, Tworoger SS, et al. Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. *PLOS Med*. 2018;15(9): e1002644. doi: 10.1371/journal.pmed.1002644.

48 Clendenen TV, Ge W, Koenig KL, et al. Breast cancer risk prediction in women aged 35–50 years: impact of including sex hormone concentrations in the Gail model. *BMC Med Res Methodol*. 2019;19(1):42. doi:10.1186/s13058-019-1126-z.

49 Wang L, Liu L, Zhen Lou,et al.Risk prediction for breast Cancer in Han Chinese women based on a cause-specific Hazard model. *BMC Cancer*. 2019;19(1): 128. doi:10.1186/s12885-019-5321-1.

50 Abdolell M, Payne JI, Caines J, et al. Assessing breast cancer risk within the general screening population: developing a breast cancer risk model to identify higher risk women at mammographic screening. *Eur Radiol*. 2020;30(10):5417-

5426. doi:10.1007/s00330-020-06901-x.

51 Qiu C, Wang P, Wang B, et al. Establishment and validation of an immunodiagnostic model for prediction of breast cancer. *OncoImmunology*. 2020;9(1): e1682382-2. doi:10.1080/2162402X.2019.1682382.

52 Han Y, Lv J, Yu C, et al. Development and external validation of a breast cancer absolute risk prediction model in Chinese population. *Breast Cancer Res.* 2021;23(1): 62. doi: 10.1186/s13058-021-01439-2.

53 Rosner B, Tamimi RM, Kraft P, et al. Simplified Breast Risk Tool Integrating Questionnaire Risk Factors, Mammographic Density, and Polygenic Risk Score: Development and Validation. *Cancer Epidemiol Biomarkers Prev.* 2021;30(4): 600-607. doi: 10.1158/1055-9965.EPI-20-0900.

54 Yiangou K, Kyriacou K, Kakouri E, et al. Combination of a 15-SNP Polygenic Risk Score and Classical Risk Factors for the Prediction of Breast Cancer Risk in Cypriot Women. *Cancers (Basel).* 2021;13(18): 4568. doi: 10.3390/cancers13184568.

55 Louro J, Posso M, Boon MH, et al. A systematic review and quality assessment of individualised breast cancer risk prediction models. *Br J Cancer*. 2019;121(1): 76–85. doi: 10.1038/s41416-019-0476-8.

56 Anothaisintawee T, Teerawattananon Y, Wiratkapun C, et al. Risk prediction models of breast cancer: A systematic review of model performances. *Breast Cancer Res Treat*. 2012;133(1): 1–10. doi 10.1007/s10549-011-1853-z.

57 Heestermans T, Payne B, Kayode GA, et al. Prognostic models for adverse pregnancy outcomes in low-income and middle-income countries: a systematic review. *BMJ Global Health*. 2019;4(5): e001759. doi:10.1136/bmjgh-2019-001759.

58 Bellou V, Belbasis L, Konstantinidis AK, et al. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ*. 2019;367: l5358. doi: 10.1136/bmj. l5358.

59 de Jong Y, Ramspek CL, van der Endt VHW, et al. A systematic review and external validation of stroke prediction models demonstrates poor performance in dialysis patients. *J Clin Epidemiol*. 2020;123:69-79. doi: 10.1016/j.jclinepi.2020.03.015.

60 Bridge J, Blakey JD, Bonnett LJ. A systematic review of methodology used in the development of prediction models for future asthma exacerbation. *BMC Med Res Methodol*. 2020;20(1): 22. doi: 10.1186/s12874-020-0913-7.

61 Banegas MP, Gail MH, LaCroix A, et al. Evaluating breast cancer risk projections for Hispanic women. *Breast Cancer Res Treat*. 2020;132(1):347–353. doi:

10.1007/s10549-011-1900-9.

62 Min JW, Chang MC, Lee HK, et al. Validation of risk assessment models for predicting the incidence of breast cancer in Korean women. *J Breast Cancer*. 2014;17(3):226–235. doi: 10.4048/jbc.2014.17.3.226.

63 Chay WY, Ong WS, Tan PH, et al. Validation of the Gail model for predicting individual breast cancer risk in a prospective nationwide study of 28,104 Singapore women. *Breast Cancer Res*. 2012;14(1): R19. doi: 10.1186/bcr3104.

64 Adams-Campbell LL, Makambi KH, Palmer JR, et al. Diagnostic accuracy of the Gail model in the Black Women's Health Study. *Breast J*. 2007;13(4):332–336. doi: 10.1111/j.1524-4741.2007. 00439.x.

65 Harrell FE. Regression modeling strategies, with applications to linear models, logistic regression, and survival analysis. New York: Springer, 2001.

66 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4): 361-87. doi:10.1002/(SICI)1097-0258(19960229)15:4<361: AID-SIM168>3.0.CO;2-4.

67 Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol*. 1996;49(8): 907–16.

doi: 10.1016/0895-4356(96)00025-x.

68 Little RJA, Rubin DB. Statistical Analysis With Missing Data. Hoboken, NJ: Wiley; 2002.

69 White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30:377-99. doi:10.1002/sim.4067.

70 Vergouwe Y, Royston P, Moons KG, et al. Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol*. 2010; 63:205-14. doi: 10.1016/j.jclinepi. 2009.03.017.

71 Royston P, Moons KG, Altman DG, et al. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338(b604). doi: 10.1136/bmj.b604.

72 Huo CW, Chew GL, Britt KL, et al. Mammographic density—a review on the current understanding of its association with breast cancer. *Breast Cancer Res Treat.* 2014; 144:479–502. doi: 10.1007/s10549-014-2901-2.

73 Vachon CM, van Gils CH, Sellers TA, et al. Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Res.* 2007; 9:217. doi: 10.1186/bcr1829. 74.

74 Warwick J, Birke H, Stone J, et al. Mammographic breast density refines Tyrer-Cuzick estimates of breast cancer risk in high-risk women: findings from the placebo arm of the International Breast Cancer Intervention Study I. *Breast Cancer Res.*

2014;16(5):451. doi: 10.1186/s13058-014-0451-5.

75 Brentnall AR, Harkness EF, Astley SM, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res.* 2015;17(1):147. doi: 10.1186/s13058-015-0653-5.

76 Brentnall AR, Evans DG, Cuzick J. Distribution of breast cancer risk from SNPs and classical risk factors in women of routine screening age in the UK. *Br J Cancer.* 2014;110(3):827-828. doi: 10.1038/bjc.2013.747.

77 Evans DG, Warwick J, Astley SM, et al. Assessing individual breast cancer risk within the U.K. National Health Service Breast Screening Program: a new paradigm for cancer prevention. *Cancer Prev Res (Phila).* 2012;5(7):943-951. doi: 10.1158/1940-6207.CAPR-11-0458.

78 Dite GS, MacInnis RJ, Bickerstaffe A, et al. Breast cancer risk prediction using clinical models and 77 independent risk-associated SNPs for women aged under 50 years: Australian Breast Cancer Family Registry. *Cancer Epidemiol Biomarkers Prev.* 2016; 25(2):359-365. doi: 10.1158/1055-9965.EPI-15-0838.

79 Arveux P, Wait S, Schaffer P. Building a model to determine the cost-effectiveness of breast cancer screening in France. *Eur J Cancer Care.* 2003;12(2): 143-53. doi:

10.1046/j.1365-2354.2003.00373. x.

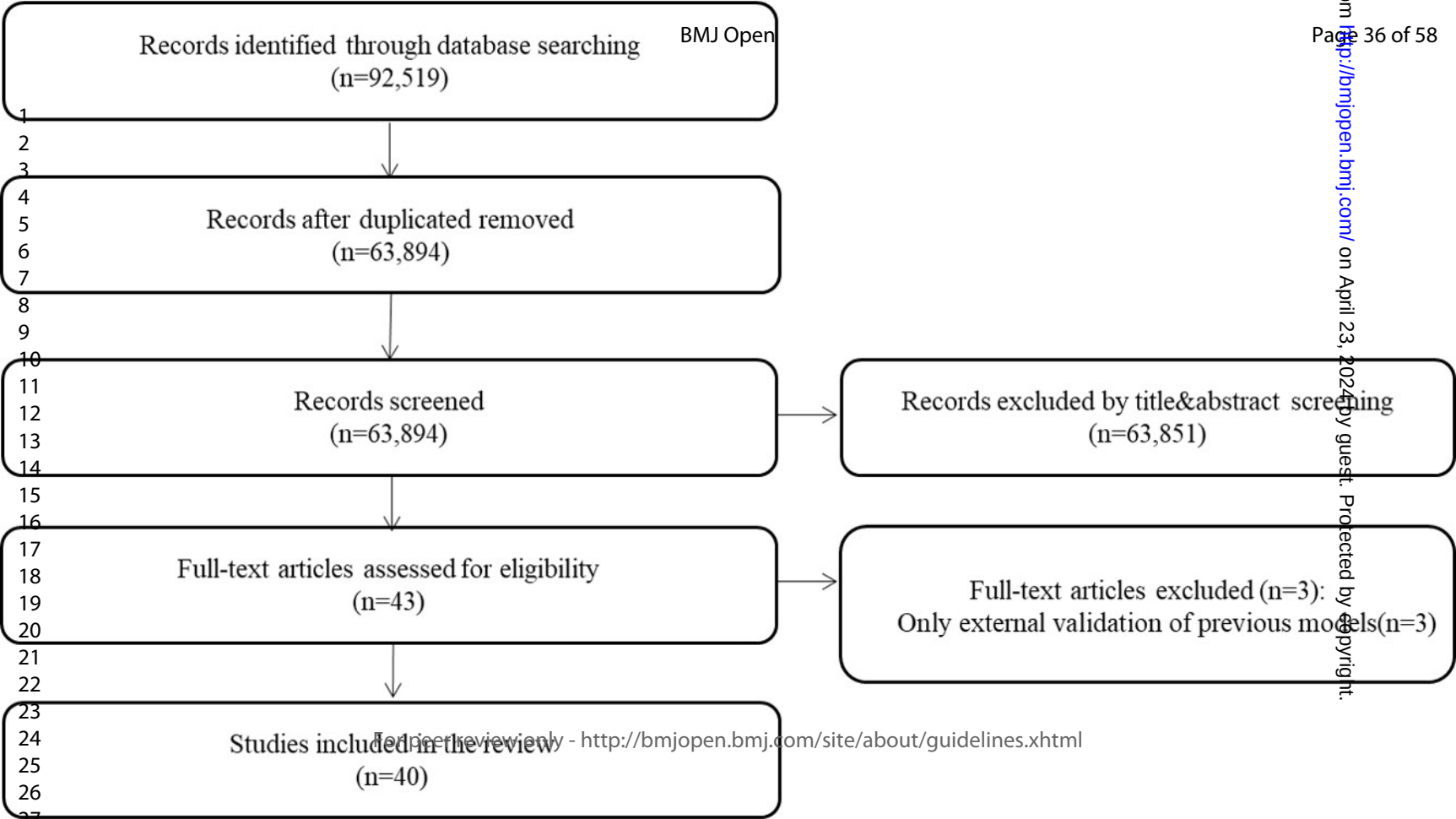
Figure legends:

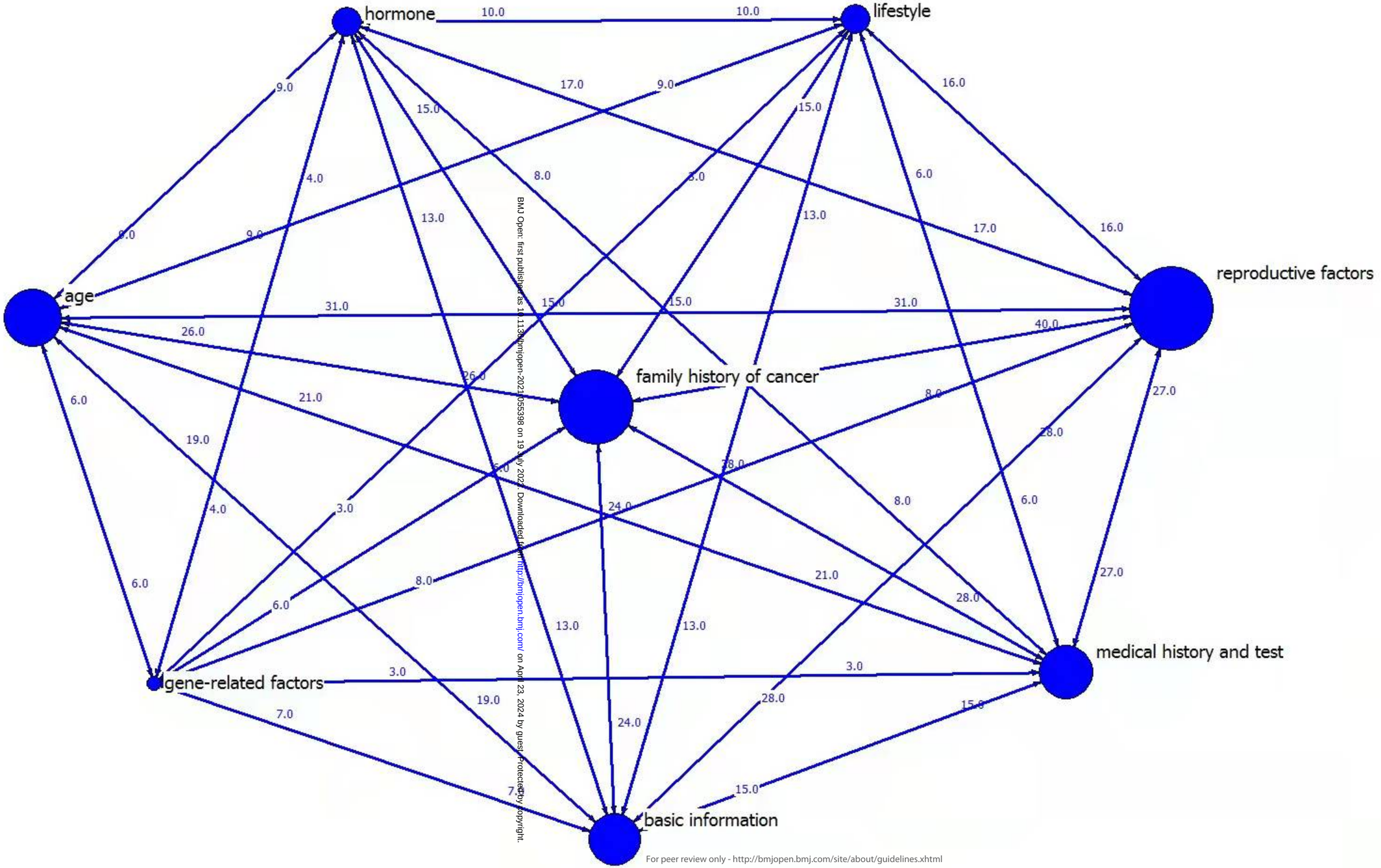
Figure1. PRISMA (preferred reporting items for systematic reviews and meta-analyses) flowchart.

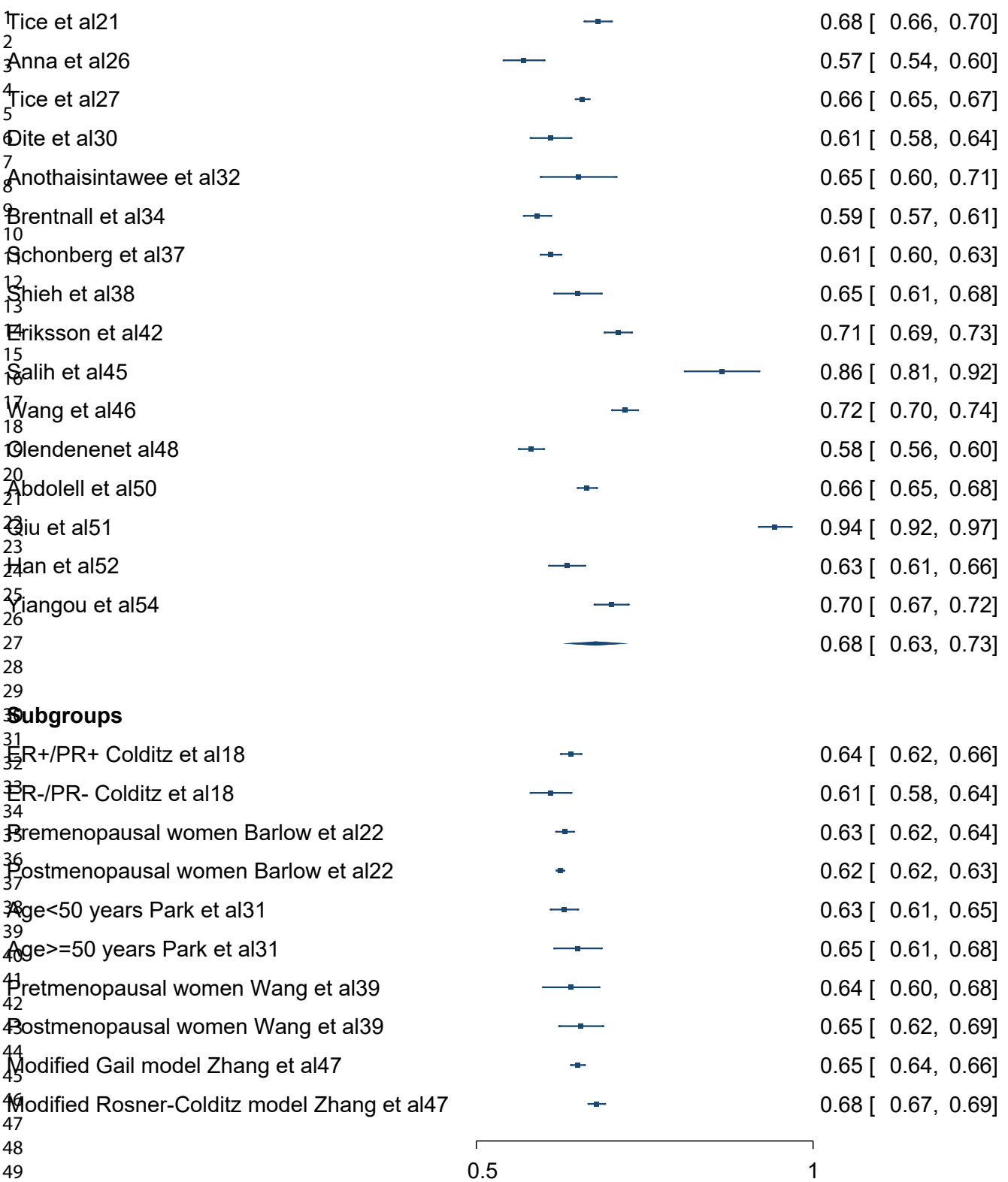
Figure 2. Network diagram of eight categorized risk factors (age, basic information, family history of cancer, gene-related factors, hormone, lifestyle, medical history and test, and reproductive factors).

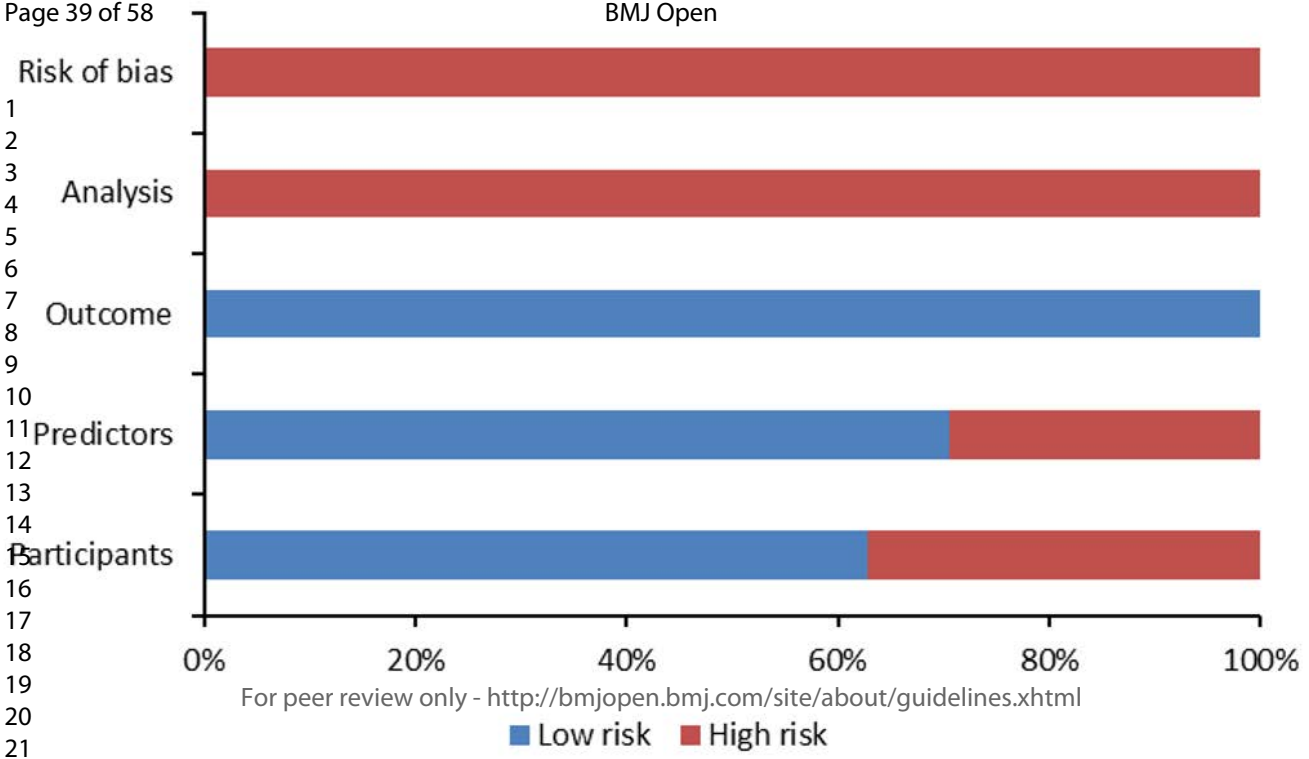
Figure 3. Area under the curve (AUC) and confidence intervals reported by the included studies.

Figure 4. Risk of bias assessment (using PROBAST) of all assessed models based on four domains.






All
Tice et al21 0.68 [0.66, 0.70]
Anna et al26 0.57 [0.54, 0.60]
Tice et al27 0.66 [0.65, 0.67]
Dite et al30 0.61 [0.58, 0.64]
Anothaisintawee et al32 0.65 [0.60, 0.71]
Brentnall et al34 0.59 [0.57, 0.61]
Schonberg et al37 0.61 [0.60, 0.63]
Shieh et al38 0.65 [0.61, 0.68]
Eriksson et al42 0.71 [0.69, 0.73]
Salih et al45 0.86 [0.81, 0.92]
Wang et al46 0.72 [0.70, 0.74]
Olendenenet al48 0.58 [0.56, 0.60]
Abdolell et al50 0.66 [0.65, 0.68]
Qiu et al51 0.94 [0.92, 0.97]
Han et al52 0.63 [0.61, 0.66]
Yiangou et al54 0.70 [0.67, 0.72]
0.68 [0.63, 0.73]
Subgroups
ER+/PR+ Colditz et al18 0.64 [0.62, 0.66]
ER-/PR- Colditz et al18 0.61 [0.58, 0.64]
Premenopausal women Barlow et al22 0.63 [0.62, 0.64]
Postmenopausal women Barlow et al22 0.62 [0.62, 0.63]
Age<50 years Park et al31 0.63 [0.61, 0.65]
Age>=50 years Park et al31 0.65 [0.61, 0.68]
Pretmenopausal women Wang et al39 0.64 [0.60, 0.68]
Postmenopausal women Wang et al39 0.65 [0.62, 0.69]
Modified Gail model Zhang et al47 0.65 [0.64, 0.66]
Modified Rosner-Colditz model Zhang et al47 0.68 [0.67, 0.69]
0.5
1



Appendix

Appendix Table 1. Searching strategy.

Searching strategy
Take PubMed for example:
#1 "Breast Neoplasms"[Mesh] OR "Breast Carcinoma In Situ"[Mesh] OR "Breast Neoplasms, Male"[Mesh] OR "Carcinoma, Ductal, Breast"[Mesh] OR "Carcinoma, Lobular"[Mesh] OR "Hereditary Breast and Ovarian Cancer Syndrome"[Mesh] OR "Inflammatory Breast Neoplasms"[Mesh] OR "Triple Negative Breast Neoplasms"[Mesh] OR "Unilateral Breast Neoplasms"[Mesh] OR phyllodes tumor[Title/Abstract] OR breast sarcoma[Title/Abstract] OR mamma cancer*[Title/Abstract] OR mammary cancer*[Title/Abstract] OR mammary gland cancer*[Title/Abstract] OR Mammary Ductal Carcinoma*[Title/Abstract] OR breast gland cancer*[Title/Abstract] OR breast gland neoplasm*[Title/Abstract] OR Breast Neoplasm*[Title/Abstract] OR Breast Tumor*[Title/Abstract] OR Breast Cancer*[Title/Abstract] OR Mammary Cancer*[Title/Abstract] OR Breast Malignant Neoplasm*[Title/Abstract] OR Breast Malignant Tumor*[Title/Abstract] OR Human Mammary Carcinoma*[Title/Abstract] OR Human Mammary Neoplasm*[Title/Abstract] OR Breast Carcinoma*[Title/Abstract] OR Lobular Carcinoma*[Title/Abstract] 418,670
#2 ("Regression Analysis"[Mesh] OR "Multivariate Analysis"[Mesh] OR "Models, Biological"[Mesh] OR "Models, Statistical"[Mesh] OR "Algorithms"[Mesh]) AND "Risk Assessment" [Mesh] 52,269
#3 predict*[Title/Abstract] AND (outcome*[Title/Abstract] OR index[Title/Abstract] OR rule*[Title/Abstract] OR decision*[Title/Abstract] OR scor*[Title/Abstract]) 624,639
#4 risk*[Title/Abstract] AND (predict*[Title/Abstract] OR calculate*[Title/Abstract] OR assess*[Title/Abstract] OR scor*[Title/Abstract] OR algorithm[Title/Abstract]) 1,109,068
#5 model*[Title/Abstract] AND (logistic[Title/Abstract] OR statistic*[Title/Abstract] OR risk*[Title/Abstract] OR predict*[Title/Abstract]) 1,1035,123
#6 OR/2-5 2,195,108
#7 #1 AND #6 54,653

Appendix Table 2. Classification of risk factors.

age	
reproductive factors	age at menarche, age at first birth, menopause, age at subsequent births, menstrual regularity, total menstrual duration, breastfeeding, breast density, parity, reproductive characteristics, microcalcifications and masses, abortions, breast volume
family history of cancer	family history of breast cancer, family history of any cancer
hormone	hormone therapy, oral contraceptives, estrogen plus progestin use, testosterone, estradiol, sex hormone binding globulin, Insulin-like growth factor-I, estrone sulphate, prolactin, anti-Müllerian hormone
gene-related factors	polygenic risk score, rs2981582 (FGFR2), rs3803662(TOX3), rs889312(MAP3K1), rs13387042(2q35), rs13281615(8q24), rs4415084 (FGF10), rs3817198 (LSP1), rs981782(HCN1), rs10822013(ZNF365), rs3784099(RAD51B)
lifestyle	alcohol consumption, smoking status, exercise, light at night, sleep quality, vegetables and fruits, cereals, life satisfaction score
medical history and test	previous biopsies, benign breast disease, nipple aspirate fluid cytology, prior breast procedure, prior false-positive mammogram, breast inflammatory, benign breast category, benign breast disease, atypical hyperplasia, mammogram in past 2 years, diabetes, myocardial infarction, stroke, emphysema, congestive heart failure, p53, CyclinB1, p16, p62,14-3-3ξ
basic information	body mass index, weight, education, ethnicity, occupational activity, height, residence area

Appendix Table 3. Summary of the 40 included studies.

Author	Year	develop							validate				
		Study design	Research method	Targeted population	No of risk factors	Risk factors	Model development (AUC (95%CI); E/O ratio (95%)	Sample size of development	Type of validation	study design	Targeted population	Model validation (AUC (95%CI); E/O ratio (95%)	Sample size of validation
Gail et al [6]	1989	Case-control study	Logistic regression	Caucasian; 20–79 years	5	Age, age at menarche, age at first birth, number of previous biopsies, number of first degree relatives with breast cancer	AUC: none; E/O ratio: none	2,852cases/ 3,146 controls	None	None	None	None	None
Rosner et al [16]	1996	Prospective cohort study	Poisson regression	Caucasian; 30–64 years	5	Age, age at menarche, age at first birth, menopause, age at subsequent births	AUC: none; E/O ratio: none	2,249 cases/ 89,132 total	None	None	None	None	None
Ueda et al [17]	2003	Case-control study	Logistic regression	Asian women; age was not specified.	4	Age at menarche, age at first birth, family history of breast cancer, body mass index	AUC: none; E/O ratio: none	376 cases/ 430 controls	None	None	None	None	None
Colditz et al [18]	2004	Prospective cohort study	Logistic regression	Caucasian; 30-64 years	11	Age, age at menarche, age at first birth, menopause, age at subsequent births, benign breast disease, postmenopausal hormone use, family history of breast cancer in a first-degree relative, weight, body mass index, alcohol consumption	AUC: ER+/PR+: 0.64 (0.63,0.66); ER-/PR-: 0.61 (0.58, 0.64); E/O ratio: none	2,846 cases/ 66,145 total	None	None	None	None	None
Lee et al [19]	2004	Case-control study	Logistic regression	Asian women; age was not specified.	1) Hospitalized controls: 5 2) Nur	1) Hospitalized controls: family history, menstrual regularity, total menstrual duration, age at first full-term pregnancy, duration of breastfeeding 2) Nurse/teacher controls: age, education level, menstrual regularity, drinking status, smoking status	AUC: 1) Hospitalized controls: 0.714; 2) Nurse/teacher controls: 0.867; E/O ratio: none	1) Hospitalized controls: 384 cases/ 166 controls; 2) Nurse/teacher controls: 384 cases/	None	None	None	None	None

					se/teacher controls: 5			104 controls					
Tice et al [20]	2005	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 18 years and older	6	Age, age at menarche, previous biopsy, age at first birth, first degree breast cancer, nipple aspirate fluid cytology	AUC: 0.64; E/O ratio: none	400 cases/ 6,904 total	None	None	None	None	None
Tice et al [21]	2005	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35 years and older	6	Age, age at menarche, previous biopsy, age at first birth, first degree breast cancer, breast density	AUC: 0.68 (0.66,0.70); E/O ratio: none	955 cases/ 81,777 total	None	None	None	None	None
Barlow et al [22]	2006	Prospective cohort study	Logistic regression	Multiple ethnicities, 35-84 years	1) Premenopausal women: 4 2) Postmenopausal women: 10	1) Premenopausal women: age, breast density, family history of breast cancer, a prior breast procedure 2) Postmenopausal women: age, breast density, race, ethnicity, family history of breast cancer, a prior breast procedure, body mass index, natural menopause, hormone therapy, a prior false-positive mammogram	AUC: Premenopausal women: 0.631 (0.618, 0.644); postmenopausal women: 0.624 (0.619, 0.630) E/O ratio [a]: Premenopausal women: 1.000 postmenopausal women: 1.001	1) Premenopausal women: 1,726 cases/ 568,215 total; 2) postmenopausal women: 9,300 cases/ 1,642,824 total	Internal validation	None	None	None	None
Decarli et al [23]	2006	Case-control study	Logistic regression	Caucasian; 20–74 years	5	Age , age of menarche, number of breast biopsies, age at first live birth, first degree breast cancer	AUC: none; E/O ratio: none	2569 cases/ 2588 controls	External validation	Prospective cohort study	Caucasian; 35-64 years	AUC: 0.59; E/O ratio: 0.96(0.84, 1.11)	194 cases /10,031 total

Novotny et al [24]	2006	Case-control study	Logistic regression	Multiple ethnicities; 23-84 years	8	Age of menarche, number of biopsies, age at first childbirth, number of breast cancer cases in first-degree relatives, number of any cancer cases in first-degree relatives, breast inflammatory, body mass index, number of conceptions	AUC: none; E/O ratio: none	2299 cases/ controls	None	None	None	None	None
Gail et al [25]	2007	Case-control study	Logistic regression	African-American Women; 35-64 years	5	Age, age at menarche, number of affected mother or sisters, age at first live birth, number of previous benign biopsy examinations	AUC: none; E/O ratio: none	1607 cases/ 1647 controls	External validation	Prospective cohort study	African American women; 50-79 years	AUC: 0.555 (0.535,0.575); E/O ratio: 0.93b	350 cases /14,059 total
Anna et al [26]	2008	Case-control study	Logistic regression	Caucasian; age was not specified	5	Age, age at menarche, number of biopsies, age at first live birth, family history	AUC: 0.57 (0.54, 0.60); E/O ratio: none	558 cases/ 1207 controls	Internal validation	None	None	None	None
Tice et al [27]	2008	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35 years or older	5	Age, ethnicity, first degree breast cancer, previous biopsies, breast density	AUC: 0.657 (0.65,0.67); E/O ratio: 1.00 (0.98,1.03)	14,766 cases/ 1095484 total	Internal validation	Prospective cohort study	Multiple ethnicities; 35 years or older	AUC: 0.660(0.65,0.66); E/O ratio: 1.03(0.99,1.06)	3,465 cases/ 251,789 total
Tammi et al [28]	2010	Nested case-control study	Logistic regression	Caucasian; 40-79 years	11	The type of benign breast disease, age, age at menarche, age at first birth and at each subsequent birth, age at menopause and type of menopause, history of benign breast diseases, family history of breast cancer in mother or sister, height, weight at age 18 years, current use of postmenopausal hormones (including type and duration of use), alcohol intake	AUC: 0.635; E/O ratio: none	240 cases/ 1036 controls	None	None	None	None	None

Petracci et al [29]	2011	Case-control study	Logistic regression	Caucasian; 20-74 years	8	Reproductive characteristics, education, occupational activity, family history, biopsy history, alcohol consumption, leisure physical activity, body mass index.	AUC: none; E/O ratio: 1.10 (0.96,1.26)	2569 cases/ 2588 controls	External validation	Prospective cohort study	Caucasian; 35-64 years	AUC: Age<50: 0.62(0.555,0.689) ; age>=50: 0.57 (0.519,0.614); E/O ratio: 1.10(0.96,1.26)	206 cases/ 8,426 total
Dite et al [30]	2013	Case-control study	Logistic regression	Multiple ethnicities; 35-59 years	13	Age, ethnicity, age at menarche, age at birth of first child, number of first-degree relatives with breast cancer, number of biopsies, presence of atypical hyperplasia, rs2981582(FGFR2), rs3803662(TOX3), rs889312(MAP3K1), rs13387042(2q35), rs13281615(8q24), rs4415084 (FGF10), rs3817198 (LSP1)	AUC: 0.61 (0.58,0.64); E/O ratio: none	962 cases/ 463 controls	None	None	None	None	None
Park et al [31]	2013	Case-control study	Logistic regression	Asian women; age was not specified.	1) Age<50 years: 7 2) Age>=50 years: 7	1)Age<50 years: a family history of breast cancer in first-degree relatives, age at menarche, menopausal status, age at first full-term pregnancy, duration of breast feeding, oral contraceptive usage, exercise. 2)Age>=50 years: a family history of breast cancer in first degree relatives, age at menarche, age at menopause, experience of pregnancy, body mass index, oral contraceptive usage, exercise	AUC: Age<50 years: 0.63 (0.61-0.65); Age>=50 years: 0.65 (0.61- 0.68); E/O ratio: none	3,789 cases/ 3,789 controls	External validation	Prospective cohort study	None	1)Korean Multi-Center Cohort (KMCC): AUC: 0.61(0.49,0.72); E/O ratio: 0.97(0.67,1.40) 2)National Cancer Center (NCC) cohort: AUC: 0.89(0.85,0.93) E/O ratio: 0.96(0.70,1.37)	1) KMCC: 29cases/ 6148 total; 2)NCC: 36 cases/ 7546 total
Apothaisintawee et al [32]	2014	Cross-sectional study	Logistic regression	Asian women; age was not specified	4	Age, menopausal status, body mass index, use of oral contraceptives	AUC: 0.651 (0.595, 0.707); O/E ratio: 1.00 (0.82, 1.21) [b]	107cases/ 15,718total	Internal and external validation	Cross-sectional study	Asian women; 18 years or older	Internal validation: AUC: 0.646(0.642,0.650); E/O ratio: none; External validation:	35 cases/ 4,978 total

												AUC: 0.609(0.511,0.706); O/E ratio: 0.97 (0.68, 1.35)[b]	
Boggs et al [33]	2015	Prospective cohort study	Cox proportional hazards regression	African-American Women; 30-69 years	9	Family history, previous biopsy, body mass index at age 18 years, age at menarche, age at first birth, oral contraceptive use, bilateral oophorectomy, estrogen plus progestin use, height	AUC: none; E/O ratio: none	896 cases/ 55,093 total	Internal validation	Prospective cohort study	African American Women; 30-69 years	AUC: 0.59 (0.56, 0.61); E/O ratio: 0.96(0.88,1.05)	506 cases/ 48,193 total
Brentnall et al [34]	2015	Prospective cohort study	Logistic regression	Caucasian; 47-73 years	1) Gail model +Density residual: :8 2) Tyrer-Cuzick+density residual: 11	1) Gail model+ Density residual: Age, Ethnicity, age at menarche, age at first birth, number of previous biopsies, benign disease, number of first degree relatives with breast cancer, density residual 2) Tyrer-Cuzick+ density residual: Age, gen phenotype, family history, age at menarche, age at first birth, menopause, atypical Hyperplasia, lobular carcinoma in situ, height, body mass index, density residual	(1) Primary (invasive+ DCIS): 1)Gail model+ Density residual: AUC: 0.59(0.57,0.61); E/O ratio: none; 2)Tyrer- Cuzick+ density residual: AUC: 0.61(0.59,0.63); E/O ratio: none; (2) Secondary(invasive): 1)Gail model+ Density residual: AUC: 0.59(0.57,0.61); E/O ratio: none; 2)Tyrer-Cuzick+ density residual: AUC: 0.61(0.58–0.63); E/O ratio: none	697 cases/ 50,628 total	None	None	None	None	None

Kerlikowske et al [35]	2015	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35-74 years	5	Age, ethnicity, first degree breast cancer, previous biopsies, changes in breast density	AUC: 5-year risk model: 0.640; 10-year risk model: 0.628; E/O ratio: 5-year risk model: 0.98(0.96,1.00); 10-year risk model: 0.95(0.94,0.96)	13,715 cases/ 722,654 total	None	None	None	None	None
Tice et al [36]	2015	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35-74 years	6	Age, race/ethnicity, family history of breast cancer, history of breast biopsy, benign breast disease diagnoses, breast density	AUC: 0.665; E/O ratio: 5 Years: 1.04(1.02 ,1.06); 10 years: 1.05 (1.03,1.06)	17908 cases/ 1,135,977 total	None	None	None	None	None
Tyrer et al [37]	2016	Prospective cohort study	Competing risk regression	Multiple ethnicities; 57–85 years	16	Age at study entry, postmenopausal hormone use, number of first-degree relatives with history of breast cancer and age at diagnosis, history of breast biopsy, highest body mass index in past 10 years, age at menopause, age at first birth and parity, average alcohol use per day (highest average use in past 10 years), cigarette use, mammogram in past 2 years, limited in moderate daily activity, diabetes, myocardial infarction, stroke, emphysema, congestive heart failure	AUC： 0.61 (0.60,0.63); E/O ratio: none	73,066 total	External validation	Prospective cohort study	Multiple ethnicities; 55-91 years	AUC: 0.57 (0.55,0.58); E/O ratio: 0.92(0.88,0.97)	74,887 total
Gail et al [38]	2016	Nested case-control study	Logistic regression	Multiple ethnicities; 36-86 years	7	Age, ethnicity, first degree breast cancer, previous biopsies, breast density, polygenic risk score, body mass index	AUC:0.65(0.61,0.68); E/O ratio: none	486 cases/ 495 controls	None	None	None	None	None

Wang et al [39]	2016	Case-control study	Logistic regression	Asian women; 20-84 years	1)Premenopausal: 5; 2)Postmenopausal: 11	1) Premenopausal: age, number of parity, case number of breast cancer in first-degree relatives, light at night, sleep quality; 2) Postmenopausal: age, number of parity, case number of breast cancer in first-degree relatives, light at night, body mass index, age at menarche, age at first give birth, ever breast feeding, ever using of oral contraceptive, hormone replacement treatment, history of benign breast diseases.	1) Pretmenopausal women: AUC: 0.640(0.598,0.681); E/O ratio: none; 2) Postmenopausal women: 0.655(0.621,0.686); E/O ratio: none	923 cases / 918 controls	Internal validation	Case-control study	Asian women; 20-84 years	1) Premenopausal: average AUC: 0.621; 3) Postmenopausal: Average AUC: 0.632	None
Maas et al [40]	2016	Prospective cohort study	Logistic regression	Caucasian	11	Age at menarche, menopause, age at first birth, parity, alcohol consumption, height, smoking status, BMI, family history, hormone therapy, PRS	AUC: 0.640; E/O ratio: none	17,171 cases / 19,862 controls	None	None	None	None	None

Banegas et al [41]	2017	Case-control study	Logistic regression	Hispanic Women; 35-79 years	1) The US-born Hispanic risk model :3; 2) the foreign-born Hispanic risk model :4	1) The US-born the Hispanic risk model: age at first full-term pregnancy, biopsy for benign breast disease, family history of breast cancer; 2) The foreign-born the Hispanic risk model: age at first full-term pregnancy, biopsy for benign breast disease, family history of breast cancer, age at menarche	None	1086 cases/ 411 controls	External validation	Prospective cohort study	Hispanic Women; 50-79 years	1)US-born Hispanics: AUC: 0.564 (0.485, 0.644); O/E:1.07 (0.81 ,1.40) [b]; 2)Foreign-born Hispanics: AUC: 0.625 (0.487 ,0.764); O/E: 0.66 (0.41,1.07) [b] 3) Hispanics of unknown nativity: AUC: 0.582(0.509,0.656); O/E: 0.89(0.69,1.14) [b]	130 cases/ 6,220 total
Eriksson et al [42]	2017	Nested case-control study	Logistic regression	Caucasian; 40-74 years	7	MD, computer-aided detection of microcalcifications and masses, use of hormone replacement therapy, family history of breast cancer, menopausal status, age, body mass index	AUC: 0.71(0.69,0.73); E/O ratio: none	433cases / 1732 controls	None	None	None	None	None
Hsieh et al [43]	2017	Case-control study	Logistic regression	Asian women; 20-90 years	11	FGFR2 (rs2981582), HCN1 (rs981782), MAP3K1 (rs889312), TOX3(rs3803662), ZNF365(rs10822013), RAD51B(rs3784099), age, body mass index, age at menarche, parity, menopausal status	AUC: 0.6652; E/O ratio: none	446 cases/ 514 controls	None	None	None	None	None

Husing et al [44]	2017	Nested case-control study	Logistic regression	Multiple ethnicities; 26-77 years	13	Menopausal status, age at menarche, age at menopause, duration of postmenopausal hormones use, parity, number of children and age at first full term pregnancy, family history of breast cancer, alcohol consumption at recruitment, body mass index, measurements of testosterone, estradiol, sex hormone binding globulin, Insulin-like growth factor-I	AUC: none; E/O ratio: none	1,217 cases/ 1,976 controls	Internal validation	None	None	None	None
Salih et al [45]	2017	Cross-sectional study	Logistic regression	Caucasian; 32–74 years	5	Age, age at menarche, family history, vegetables and fruits weekly servings, type of cereals used	AUC: 0.864(0.81,0.92)	63 cases/ 90 controls	Internal validation	None	None	O/E ratio: 0.78[b]	None
Wang et al [46]	2018	Case–control study	Logistic regression	Nigerian women; age was not specified	9	Age, age at menarche, parity, duration of breastfeeding, family history of breast cancer, height, body mass index, benign breast diseases, alcohol consumption	AUC: 0.720(0.701,0.739); E/O ratio: 1.01 (0.93,1.09)	1,208 cases/ 1,484 controls	Internal validation	None	Nigerian women; 20-79 years	AUC: 0.694 (0.666,0.721); E/O ratio: none	603 cases/ 741 controls

Zhang et al [47]	2018	Nested case-control study	Logistic regression	Caucasian; 34-70 years	1) Gail model + PRS + MD + T + E1S +PRL: 10; 2) Rosner-Colditz model + PRS + MD + T + E1S + PRL: 16	1) Gail model+ PRS + MD + T + E1S +PRL: Age, age at menarche, previous biopsies, age at first birth, first degree breast cancer, PRS, MD, E1S, T, PRL 2) Rosner-Colditz model+ PRS + MD + T + E1S + PRL: age, age at menarche, age at first birth, menopause, age at subsequent births, benign breast disease, hormone replacement therapy, first degree breast cancer, weight, body mass index, alcohol, PRS, MD, E1S, T, PRL	AUC: Gail model+ PRS + MD + T + E1S +PRL: 0.65(0.64,0.66); Rosner-Colditz model+ PRS + MD + T + E1S + PRL: 0.678 (0.666,0.690); E/O ratio: none	4,006 cases / 7,874 controls	Internal validation	None	None	None	None
Tendenenet al [48]	2019	Nested case-control study	Logistic regression	Multiple ethnicities; 35–50 years	6	Age at menarche, age at first live birth, number of benign breast biopsies, number of first-degree family members with breast cancer, AMH, tT	AUC: 0.581(0.562,0.599); E/O ratio: none	1,762 cases/ 1,890 controls	None	None	None	None	None
Wang et al [49]	2019	Case-control study	Logistic regression	Asian women; 25-70 years	6	Number of abortions, age at first live birth, benign breast disease history, body mass index,	None	328 cases / 656 controls	External validation	Prospective cohort study	Asian women	AUC: 0.64 (0.55,0.72); E/O ratio: 1.03 (0.74,1.49)	34 cases/ 13,176 total

						breast cancer family history, life satisfaction score							
Abdolell et al [50]	2020	Nested case-control study	Logistic regression	Caucasian; 40-75 years	5	Age at screen, percent mammographic density, breast volume, core biopsy history, family history	AUC: 0.664(0.650,0.678); E/O ratio: none	1,882 cases/ 5,888 controls	None	None	None	None	None
Qiu et al [51]	2020	Case–control study	Logistic regression	Asian women; 29-81 years	5	p53, CyclinB1, p16, p62,14-3-3ξ	AUC:0.943(0.919,0.967); E/O ratio: none	184 cases/ 184 controls	External validation	Case-control study	Asian women; 24-78 years	AUC: 0.916(0.886,0.947); E/O ratio: none	197 cases/ 109 controls
Han et al [52]	2021	Prospective cohort study	Cox regression	Asian women; 30-79 years	8	age, residence area, education, BMI, height, family history of cancer, parity, age at menarche	AUC: 0.634(0.608,0.661); E/O ratio: 1.01(0.94,1.09)	2,287 cases/ 300,824 total	External validation	Prospective cohort study	Asian women;	AUC: 0.585(0.564,0.605) E/O ratio: 0.94(0.89,0.99)	73,203 total
Kraft et al [53]	2021	Nested case-control study	Logistic regression	Caucasian; 40-75 years	4	Age, breast density, questionnaire score, PRS	AUC: 0.658 E/O ratio: none	2,799 cases/ 75,557 controls	External validation	Nested case-control study	Caucasian; 40-75 years	AUC: 0.687	438 cases/ 898 controls
Yiangou et al [54]	2021	Case–control study	Logistic regression	Cypriot Women	11	menopause, age at menarche, parity, age at first birth, breastfeeding, height, BMI, hormone therapy, smoking status, family history, PRS	AUC: 0.70 (0.67,0.72) E/O ratio: none	1,109 cases/ 1,177 controls	None	None	None	None	None

[a]E/O ratios were calculated based on the original information. [b]The original publication reported the Observed/Expected ratio.
ER: estrogen receptor; PR: progesterone receptor; PRS: polygenic risk score; MD: mammographic density; E1S: estrone sulphate; T: testosterone; PRL: prolactin; AMH: anti-Müllerian hormone; NI: no information.

Appendix Table 4. Risk of bias assessment of included models based on PROBAST.

Study	Participants			Predictors				Outcome							Analysis									Overall
	1.1	1.2		2.1	2.2	2.3		3.1	3.2	3.3	3.4	3.5	3.6		4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9		
Gail et al [6]	N	Y	H	Y	PY	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	N	N	Y	PY	N	N	Y	H	H
Rosner et al [16]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	NI	N	N	Y	H	H
Ueda et al [17]	N	NI	H	Y	PY	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	Y	Y	PY	N	N	Y	H	H
Colditzet al [18]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	L	Y	N	Y	N	Y	N	N	Y	Y	H	H
Lee et al [19]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	Y	N	PY	N	N	Y	H	H
Tice et al [20]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	PY	Y	N	N	N	Y	H	H
Tice et al [21]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	PY	Y	N	N	N	Y	H	H
Barlow et al [22]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	PY	NI	Y	N	N	Y	Y	H	H
Decarli et al [23]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	N	N	Y	N	Y	N	Y	H	H
Decarli et al [23*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	N	NI	-	NI	Y	-	-	H	H
Novotny et al [24]	N	PY	H	Y	PN	Y	H	PY	Y	Y	Y	Y	Y	L	Y	N	N	N	Y	PY	N	N	Y	H	H
Gail et al [25]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	PY	N	N	Y	H	H
Gail et al [25*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	-	Y	Y	-	-	H	H
Anna et al [26]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	PY	N	N	Y	H	H
Tice et al [27]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	N	N	Y	Y	Y	Y	Y	H	H
Tamimi,et al [28]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	NI	N	N	Y	H	H
Petracci et al [29]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	Y	PY	N	N	Y	H	H
Petracci et al [29*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	-	Y	PN	-	-	H	H
Dite et al [30]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	N	PY	N	N	Y	H	H
Park et al [31]	N	Y	H	Y	PY	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	N	PY	N	N	Y	H	H
Park et al [31*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	NI	-	PY	Y	-	-	H	H
Anothaisintawee et al [32]	Y	Y	H	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	Y	N	PY	PN	N	Y	H	H

Anothaisintawee et al [32*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	Y	-	PY	PN	-	-	H	H
Boggs et al [33]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	NI	Y	N	Y	Y	H	H
Brentnall et al [34]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	Y	N	N	N	Y	H	H
Kerlikowske et al [35]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	N	PN	Y	Y	H	H
Tice et al [36]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	Y	PN	Y	Y	H	H
Schonberg et al [37]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	PY	N	Y	N	N	Y	H	H
Schonberg et al [37*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	-	Y	PN	-	-	H	H
Shieh et al [38]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	N	N	N	Y	Y	H	H
Wang et al [39]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	N	PY	N	Y	Y	H	H
Maas et al [40]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	PN	PN	Y	H	H
Banegas, et al [41]	N	Y	H	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	PY	N	NI	Y	H	H
Banegas et al [41*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	-	PY	Y	-	-	H	H
Eriksson et al [42]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	NI	PY	N	Y	Y	H	H
Hsieh, et al [43]	N	NI	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	N	PY	N	Y	Y	H	H
Husing et al [44]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	PY	N	Y	Y	H	H
Salih et al [45]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	N	PY	N	Y	Y	H	H
Wang et al [46]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	Y	PY	PN	Y	Y	H	H
Zhang et al [47]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	N	Y	Y	H	H
Clendenen et al [48]	Y	Y	L	PN	Y	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	N	N	Y	H	H
Wang et al [49]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	Y	N	N	Y	H	H
Wang et al [49*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	N	-	Y	PN	-	-	H	H
Abdolell et al [50]	Y	PY	L	Y	Y	Y	L	PY	Y	Y	Y	Y	Y	L	Y	Y	Y	N	N	PY	N	N	Y	H	H
Qiu et al [51]	N	NI	H	Y	N	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	PY	N	N	Y	H	H
Qiu et al [51*]	N	NI	H	Y	N	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	-	PY	N	-	-	H	H
Han et al [52]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	Y	PY	Y	Y	Y	H	H
Han et al [52*]	Y	Y	L	Y	Y	Y	L	PY	Y	Y	Y	Y	Y	L	Y	Y	Y	N	-	NI	Y	-	-	H	H

Rosner et al [53]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	Y	PY	N	N	Y	H	H
Rosner et al [53*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	-	PY	N	-	-	H	H
Yiangou et al [54]	N	Y	H	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	PN	N	PY	N	PN	Y	H	H

* The external validation was performed in the same study.

L: low risk of bias; H: high risk of bias; Y: yes; N: no; PY: probably yes; PN: probably no; NI: no information; -: not applicable.

1.1. Were appropriate data sources used, e.g., cohort, RCT, or nested case–control study data?

1.2. Were all inclusions and exclusions of participants appropriate?

2.1. Were predictors defined and assessed in a similar way for all participants?

2.2. Were predictor assessments made without knowledge of outcome data?

2.3. Are all predictors available at the time the model is intended to be used?

3.1. Was the outcome determined appropriately?

3.2. Was a prespecified or standard outcome definition used?

3.3. Were predictors excluded from the outcome definition?

3.4. Was the outcome defined and determined in a similar way for all participants?

3.5. Was the outcome determined without knowledge of predictor information?

3.6. Was the time interval between predictor assessment and outcome determination appropriate?

4.1. Were there a reasonable number of participants with the outcome?

4.2. Were continuous and categorical predictors handled appropriately?

4.3. Were all enrolled participants included in the analysis?

4.4. Were participants with missing data handled appropriately?

4.5. Was selection of predictors based on univariable analysis avoided? (Development studies only)

4.6. Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted for appropriately?

4.7. Were relevant model performance measures evaluated appropriately?

4.8. Were model overfitting, underfitting, and optimism in model performance accounted for? (Development studies only)

4.9. Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis? (Development studies only)

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review.	1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	2,3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	4,5
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	5,6
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	6,7
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	6
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	6 and Appendix Table 1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	7
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	7
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	8,9 and Appendix Table 2
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	8,9 and Appendix Table 2
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	7,8
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	8,9
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	8,9
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	8,9 and Appendix Table 3
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	8,9

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	8,9
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Not performed
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Not performed
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Not performed
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Not performed
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	9 and figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	9
Study characteristics	17	Cite each included study and present its characteristics.	9,10
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	12,13,14,15
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	9,10,11 and Appendix Table 3
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Appendix Table 3, table 1, figure 2, figure 3 and figure 4
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	9, 10,11,12, 13,14 and Appendix Table 4
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Not performed
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Not performed
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not performed
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Not performed
DISCUSSION			

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	14,15
	23b	Discuss any limitations of the evidence included in the review.	15,16
	23c	Discuss any limitations of the review processes used.	15,16
	23d	Discuss implications of the results for practice, policy, and future research.	16,17,18
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	3,6
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Not performed
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not performed
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	19
Competing interests	26	Declare any competing interests of review authors.	19
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Appendix Table 1,2,3

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

BMJ Open

Risk prediction models for breast cancer: a systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055398.R2
Article Type:	Original research
Date Submitted by the Author:	14-Apr-2022
Complete List of Authors:	Zheng, Yadi; National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Office of Cancer Screening Li, Jiang; National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Office of Cancer Screening Wu, Zheng; National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Office of Cancer Screening Li, He; National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Office of Cancer Screening Cao, Maomao; National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Office of Cancer Screening Li, Ni; National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Office of Cancer Screening He, Jie; National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Department of Thoracic Surgery
Primary Subject Heading:	Public health
Secondary Subject Heading:	Epidemiology, Oncology, Public health
Keywords:	Epidemiology < TROPICAL MEDICINE, Public health < INFECTIOUS DISEASES, Breast tumours < ONCOLOGY, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Risk prediction models for breast cancer: a systematic review

Yadi Zheng[1], Jiang Li[*1], Zheng Wu[1], He Li[1], Maomao Cao[1], Ni Li[*1], Jie He[2]

[1] Office of Cancer Screening, National Cancer Center/ National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

[2] Department of Thoracic Surgery, National Cancer Center / National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

Corresponding authors: Prof. Ni Li nli@cicams.ac.cn; Prof. Jiang Li lij@cicams.ac.cn

Word count: 2772

Figures/Tables: 5

ABSTRACT

Objectives: To systematically review and critically appraise published studies of risk prediction models for breast cancer in the general population without breast cancer, and provide evidence for future research in the field.

Design: Systematic review using the Prediction model study Risk Of Bias Assessment Tool (PROBAST) framework.

Data sources: PubMed, the Cochrane Library and Embase were searched from inception to 16 December, 2021.

Eligibility criteria: We included studies reporting multivariable models to estimate the individualized risk of developing female breast cancer among different ethnic groups. Search was limited to English language only.

Data extraction and synthesis: Two reviewers independently screened, reviewed, extracted, and assessed studies with discrepancies resolved through discussion or a third reviewer. Risk of bias was assessed according to the PROBAST (Prediction model Risk of Bias Assessment Tool) framework.

Results: 63,894 studies were screened and 40 studies with 47 risk prediction models were included in the review. Most of the studies used logistic regression to develop breast cancer risk prediction models for Caucasian women by case-control data. The most

widely used risk factor was reproductive factors and the highest area under the curve was 0.943 (95% confidence interval: 0.919~0.967). All the models included in the review had high risk of bias.

Conclusions: No risk prediction models for breast cancer were recommended for different ethnic groups and models incorporating mammographic density or single-nucleotide polymorphisms (SNPs) among Asian women are few and poorly needed. High-quality breast cancer risk prediction models assessed by PROBAST should be developed and validated, especially among Asian women.

PROSPERO registration number： CRD42020202570

Strengths and limitations of this study

1. Thoroughly conducted systematic review collecting data from major existing databases.
2. Critically appraised published studies of risk prediction models for breast cancer in the general population and provide evidence for future research in the field.
3. PROBAST was used to assess the quality of prediction models, which was developed through a consensus process involving a group of methodological experts in the area of clinical prediction tools and quality assessment.
4. Studies only about the external validation of the present risk models were not included in the review.

5. Our study highlighted high-quality breast cancer risk prediction models assessed by PROBAST should be developed and validated among different ethnic groups, especially among Asian women.

Keywords: breast cancer; risk prediction model; review; quality assessment; Prediction model Risk of Bias Assessment Tool

INTRODUCTION

Breast cancer is a major public health problem, and one of the most severe burdensome cancer among women worldwide [1], accounting for11.7% of new cancer cases and 6.9% of cancer deaths in 2020. The prevalence of breast cancer is projected to increase over the coming years and is the most common cancer in women in 2020 [2]. Breast cancer prevention is associated with a reduction in mortality [3], and more researches are needed to improve the methods of identifying women at elevated risk and preventing the disease. Numerous breast cancer risk prediction models have been developed to identify the combined effect of risk factors for breast cancer, guide routine screening and genetic testing, and reduce the burden of breast cancer. Risk-stratified screening can improve cost-effectiveness and maximize benefits and minimize harms like overdiagnosis [4]. Individualized prediction model for breast cancer could be used in practice to assist decision making about mass screening or opportunistic screening and treatment strategy.

A recent breast cancer screening guideline [5] suggests that breast cancer screening increase the early detection rate and reduce the incidence if the screening is applied in appropriate at-risk populations. However, major gaps exist in our knowledge to determine the risk of breast cancer accurately in order to apply these approaches to appropriate populations of women.

A lot of breast cancer risk prediction models have been developed over the past few decades. Many breast cancer risk models have undergone validation including discrimination and calibration in study populations other than those used in initial development, or have been further assessed in comparative studies. Breast cancer related predictors including hormonal factors, environmental factors, family histories, genetic factors and radiographic factors have been based on in these risk models, which would improve the generalizability. For example, the Gail model [6], one of the most famous models, has been widely used and validated worldwide since it was developed in 1989 [7-12].

This study is a systematic review of breast cancer risk prediction models by using meta-analysis and the Prediction model Risk of Bias Assessment Tool (PROBAST) [13-14]. The aim of our study is to systematically review published studies of risk prediction models for breast cancer in the general population, find more methods of predicting

female breast cancer risk among one or more ethnic groups, prepare for the development of risk prediction models, and provide evidence for future research in the field.

METHODS

Protocol and registration

The current review was designed according to the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [15] and was recorded in the PROSPERO database (registration number: CRD42020202570).

Literature search and eligibility criteria

We systematically searched PubMed, the Cochrane Library and Embase from inception to 16 December, 2021. The detailed search strategies were reported in Appendix Table 1. Articles identified from the search were loaded into EndNote X7 and duplicates were removed.

Inclusion criteria: 1) a model used data from cross-sectional studies, cohort studies, case-control studies, and randomized controlled trials; 2) a model estimating the individualized risk of female breast cancer among one or more ethnic groups; 3) a model developed for the general population without breast cancer; 4) reported a multivariable (i.e., at least 2 variables or predictors) model; 5) published in English.

Exclusion criteria: 1) external validation studies that only validated previous models in a different population without adding any additional information such as modifications on the risk factors; 2) models developed by machine learning.

Data extraction

Two reviewers screened the search results independently. Full text reports were then assessed for eligibility with discrepancies resolved through discussion or a third reviewer.

We extracted information in two categories: 1) For all studies included in the review, we extracted the following information: author, publication year, study design, research method, targeted population, number of risk factors, risk factors, model performance and sample size of development. 2) For studies included validation part, we also extracted the following information: type of validation, study design, targeted population, model performance and sample size of validation. The information was extracted by one reviewer and checked by a second reviewer.

Risk of bias assessment

We used PROBAST to assess the reported prediction models, which is a new tool designed by a group of experts all over the world to assess the risk of bias and applicability of diagnostic and prognostic prediction models. It can be used in critical appraisal of studies that develop, validate, or update prediction models for individualized predictions

[13-14]. In brief, it contains 20 signaling questions in four domains: participants, predictors, outcome, and statistical analysis. Signaling questions can be answered as yes, probably yes, no, probably no, or no information. A domain where at least one signaling question is answered as no or probably no should be judged as high risk of bias. Only if all domains are judged as low risk of bias, the total bias is judged as low risk as well.

Before putting PROBAST into use, we formed a ten-people study group including prediction model researchers, statisticians, evidence-based medicine specialists etc. to learn and practice the appropriate use of this new tool systematically. Only after everyone understood all these twenty questions totally, we would move to the peer quality assessment part. Risk of bias of every prediction model was assessed by two reviewers independently with discrepancies resolved through discussion or a third reviewer.

If there were more than one models developed in one study, we only assessed the risk of bias once due to their similarity. We also assessed the risk of external validation of prediction model when it was conducted in the same article that included model development.

Data synthesis and analysis

We calculated and reported descriptive statistics to summarize the characteristics of the models. We calculated the most frequently used risk factors and classified all risk

factors into eight categories: Age, reproductive factors, family history of cancer, hormone, gene-related factors, lifestyle, medical history and test, and basic information. Classification details can be seen in Appendix Table 2. Then we used network diagram to see the connections of categorized risk factors. We used forest plot to describe the model performance. The expected observed (E/O) ratio was not included in the forest plot because it was only reported in 7 out of 40 studies. All analyses were performed using Stata 16.0 and NetDraw.

Patient and public involvement

There was no patient or public involvement in this study.

RESULTS

Study selection

A total of 92,519 indexed records (54,653 in PubMed, 30,374 in Cochrane Library and 7,492 in Embase), 28,625 were eliminated as duplicates found in all databases, leaving a total of 63,894 publications. 43 articles were included primarily after screening by title and abstract. 3 studies which were only about the external validation of previous models were excluded while full test screening, resulting in 40 studies with 47 models were included in the review eventually. (Figure 1).

Study characteristics

A brief summary of the 40 [6,16-54] included studies is presented in Appendix Table 3. The included studies were published from 1989 to 2021. 25 of the studies were conducted over the past ten years with 5 studies published in 2017 especially. Seventeen out of the forty studies used data from case-control studies to develop prediction models [6,17,19,23-26,29-31,39, 41,43,46,49,51,54], thirteen from prospective cohorts [16,18,20-22,27,33-37,40,52], eight from nested case-control studies [28,38, 42,44,47,48,50,53] and two from cross-sectional study [32,45]. Thirty-one studies used logistic regression to fit prediction models [6,17-19,22-26,28-32,34,38-51,53,54], seven used cox proportional hazards regression [20,21,27,33,35,36,52], one used Poisson regression [16] and one used competing risk regression [37]. Of all forty-seven models in forty studies, sixteen models were developed in Caucasian women [6,16,18,23,26,28,29,34,40,42,45,47,50,53], thirteen in multiple ethnicities women [20-22,24,27,30,35-38,44,48], twelve in Asian women [17,19,31,32,39,43,49,51,52], two in African-American women [25,33], two in Hispanic women [41], one in Nigerian women [46] and one in Cypriot Women [54].

The association between eight categories of risk factors was shown in Figure 2. Reproductive factors had the biggest node size, which meant that this factor was most frequently connected with other factors among prediction models. The number between two factors meant the times these two factors were included in the same models, some of which were over thirty. For instance, reproductive factors and family history of cancer

were included in the same models for forty times, and reproductive factors and age were included in the same models for thirty-one times.

Twenty-nine studies reported c-statistics [18-22,26-28,30-32,34-40,42,43,45-48,50-54], ranged from 0.59(95% confidence interval: 0.57~0.61) to 0.943(95% confidence interval: 0.919~0.967). Qiu, et al [51] had the highest c-statistics (0.943, 95% confidence interval: 0.919~0.967), and Lee et al [19] and Salih et al [45] reported area under the curve (AUC) over 0.8, 0.867 and 0.864(95% confidence interval: 0.81~0.92), respectively. E/O ratios can be obtained from eight studies [22,27,29,32,35,36,46,52]. Figure 3 showed that the overall AUC was 0.68(95% confidence interval: 0.63~0.73) for sixteen studies [21,26,27,30,32,34,37,38,42,45,46,48,50,51,52,54] that reported the AUC and 95% confidence interval. The AUCs of the subgroups in five studies [18,22,31,39,47] were between 0.6 to 0.7.

In all these forty studies, nine studies assessed prediction models with internal validation [22,26,27,33,39,44-47], ten with external validation [23,25,29,31,37,41,49,51-53], and one with both [32]. Fifteen studies reported the discriminatory accuracy as the AUC [23,25,27,29,31-33,37,39,41,46,49,51-53], and eleven studies used the expected/observed event ratio (or observed/expected event ratio) to measure the calibration accuracy of the model [23,25,27,29,31,33,37,41,45,49,52].

Quality assessment

A summary of the quality assessment is shown in Table 1. Overall, all models assessed by PROBAST in the review had high risk of bias. There was a low and high risk of bias in the outcome and analysis domains respectively. Over 60% models had low risk in participants domain and about 70% models had low risk in predictors domain, 32 models and 36 models respectively. (As shown in Figure 4).

The main reasons for the high risk in analysis domain were model performance measures evaluated inappropriately, categorization of continuous predictors, no reporting of overfitting and optimism in model performance and missing data handled inappropriately (Appendix Table 4).

Table 1. Summary of risk of bias assessment.

Study	Participants	Predictors	Outcome	Analysis	Overall
Gail et al [6]	H	L	L	H	H
Rosner et al [16]	L	L	L	H	H
Ueda et al [17]	H	L	L	H	H
Colditzet al [18]	L	L	L	H	H
Lee et al [19]	H	H	L	H	H
Tice et al [20]	L	L	L	H	H
Tice et al[21]	L	L	L	H	H
Barlow et al [22]	L	L	L	H	H
Decarli et al [23]	H	H	L	H	H
Decarli et al [23*]	L	L	L	H	H
Novotny et al [24]	H	H	L	H	H
Gail et al [25]	H	H	L	H	H
Gail et al [25*]	L	L	L	H	H
Anna et al [26]	H	H	L	H	H

Tice et al [27]	L	L	L	H	H
Tamimi,et al [28]	L	L	L	H	H
Petracci et al [29]	H	H	L	H	H
Petracci et al [29*]	L	L	L	H	H
Dite et al [30]	H	H	L	H	H
Park et al [31]	H	H	L	H	H
Park et al [31*]	L	L	L	H	H
Anothaisintawee et al [32]	H	L	L	H	H
Anothaisintawee et al [32*]	L	L	L	H	H
Boggs et al [33]	L	L	L	H	H
Brentnall et al [34]	L	L	L	H	H
Kerlikowske et al [35]	L	L	L	H	H
Tice et al [36]	L	L	L	H	H
Schonberg et al [37]	L	L	L	H	H
Schonberg et al [37*]	L	L	L	H	H
Shieh et al [38]	L	L	L	H	H
Wang et al [39]	H	H	L	H	H
Mass et al [40]	L	L	L	H	H
Banegas, et al [41]	H	L	L	H	H
Banegas et al [41*]	L	L	L	H	H
Eriksson et al [42]	L	L	L	H	H
Hsieh, et al [43]	H	H	L	H	H
Husing et al [44]	L	L	L	H	H
Salih et al [45]	L	L	L	H	H
Wang et al [46]	H	H	L	H	H
Zhang et al [47]	L	L	L	H	H
Clendenen et al [48]	L	H	L	H	H
Wang et al [49]	H	H	L	H	H
Wang et al [49*]	L	L	L	H	H
Abdolell et al [50]	L	L	L	H	H
Qiu et al [51]	H	H	L	H	H
Qiu et al [51*]	H	H	L	H	H
Han et al [52]	L	L	L	H	H
Han et al[52]*	L	L	L	H	H

Rosner et al [53]	L	L	L	H	H
Rosner et al [53]*	L	L	L	H	H
Yiangou et al [54]	H	L	L	H	H

* The external validation was performed in the same study.
L indicates low risk of bias; H indicates high risk of bias.

DISCUSSION

Summary of main results

This systematic review identified 40 studies with 47 risk prediction models developed and/or validated for breast cancer among different ethnic groups. Most of the studies used logistic regression to develop breast cancer risk prediction models for Caucasian women by case-control data. The most widely used risk factor was reproductive factors. Reproductive factors together with family history factor were used in most models. The highest AUC was 0.943 (95% confidence interval: 0.919~0.967) from Qiu, et al [51]. The overall AUC was 0.68(95% confidence interval: 0.63~0.73) for sixteen studies [21,26,27,30,32,34,37,38,42,45,46,48,50,51,52,54] that reported the AUC and 95% confidence interval. All the studies presented a high risk of bias due to the high risk in analysis domain, which were mainly because of model performance measures evaluated inappropriately, categorization of continuous predictors, no reporting of overfitting and optimism in model performance and missing data handled inappropriately.

Agreements and disagreements with other reviews

As we can learn from the review, there were more and more risk prediction models of breast cancer over the past thirty years. Most of the models were developed in the Caucasian women, which agreed with the systematic review published by Louro et al in 2019 [55]. Compared with this review, we identified more prediction models and used a newly published tool to assess the quality of included models.

Over the past ten years, some new variables (such as oral contraceptives, diabetes, and alcohol consumption) have been included in prediction models. Increased use of the inclusion of common genetic variation in the prediction models was in accord with Louro et al in 2019 [55] and Anothaisintawee et al in 2012 [56]. However, neither of them included models developed with potential biomarkers like tumor-associated antigens. By contrast, we included one model developed by Qiu, et al [51] in 2019 included five tumor-associated antigens. The model performed well with a high AUC 0.943(95% confidence interval: 0.919,0.967).

Strengths and limitations of the study

PROBAST was developed through a consensus process involving a group of methodological experts in the field of clinical prediction tools and quality assessment. We used it to assess the quality of prediction models, which has been used widely in many fields [57-60] since it came out.

Despite the strength, there are four main limitations. Firstly, we didn't systematically search gray literature. Therefore, some models may not be identified. Secondly, quality assessment could be thought to be subjective, which is an inherent bias of systematic review. However, two independent reviewers extracted and assessed the risk prediction models using PROBAST whose authors have indicated essentially objective guidelines and explanations. Moreover, studies only about the external validation of the present risk models were not included in the review, but the original developments of these risk models were covered. For instance, the study describes the original developments of Gail model [6] was included in our research, while the studies only about the external validation of Gail model [61-64] were not included. What's more, papers about genetically oriented models like BOADICEA[65,66] and BRACAPRO[67] were not included in our study because some rare truncating/pathogenic variants like BRCA1 and BRCA2 are needed to be tested, which might be too expensive to use for general population in the mass screening[55].

Implication to research and clinical practice

Eleven models [19,30-32,37-39,43,45,50,54] selected predictors based on univariable analysis, causing a high risk in analysis domain, which should be avoided. Risk prediction models should include predictors those are well-established and with clinical credibility regardless of any statistical significance [68,69]. Because sometimes predictors only have

important relationship with the outcome after adjustment for confounding covariates, and covariates hold no independent predictive power when other covariates are included [13,70].

Some models were high risk in analysis domain because of missing data handled inappropriately, which may lead to biased associations between risk factors and breast cancer as well as biased model performance because of the selectivity of participants [71]. So imputation techniques are supposed to apply when data are missing [72,73].

When developing the risk prediction models, there were only nine studies included internal validation [22,26,27,33,39,44-47], leaving most models without internal validation. Lack of performing internal validation may increase the risk of overfitting [74]. Thus, we suggest that internal validation should be performed before external validation.

PROBAST was created by many international experts, providing a series of guidelines about model development and validation, which can be easily applied and improve clinical practice of prediction models. So, the new and most recommended methodology should be used when a new model is developed or the existing models are updated.

In the light of the results of our review, it is still hard to recommend any of the models to be applied in the breast cancer screening due to the high risk of bias. Adding variables like mammographic density or single-nucleotide polymorphisms (SNPs) to risk-

prediction models can improve the model performance and has been well validated in the general population of European-ancestry women[40,75-80]. But the model incorporating breast density or SNPs among Asian women is few and poorly needed. Cost-effectiveness should be considered when a model is going to be applied in clinical practice. Because even though the model with some risk factors that cost more to get (e.g., high risk gene) has better model performance, it is still hard to be applied in poor area [81]. What's more, an existing model should be modified or updated before used in another group of people with different characteristics, which may improve the performance of prediction models.

Breast cancer incidence has risen to the first place by 2020 all over the world, which makes it more crucial to develop breast cancer prediction models for different ethnic groups. In China, we have launched many breast cancer screening programs. For example, Rural Women "two cancers" Check Project Management Solutions have covered 31 provinces and 1437 counties since 2009. Cancer Screening Program in Urban China conducted by the National Cancer Center has covered 28 provinces and 67 cities with more than 4 million people involved and 2 million people screened by ultrasound and Mammography since 2012. The program will provide large data for us to develop a high-quality breast cancer risk prediction model in Chinese and will have great significance for breast cancer prevention of Asian women.

CONCLUSIONS

All 47 models assessed in our review using PROBAST performed the high risk of bias, leaving no model is recommended in the routine screening program. Some new variables, like oral contraceptives, diabetes, and alcohol consumption, have been widely used in prediction models over the past ten years. Models incorporating mammographic density or SNPs among Asian women are few and poorly needed. It is necessary to develop and validate high-quality breast cancer risk predication models among different ethnic groups, especially among Asian women.

Contributors: YZ and JL conceptualized the study and created the first version of the review protocol. ZW, HL, MC, NL, and JH critically reviewed the review protocol and approved it. YZ and HL screened eligible articles. YZ extracted the data, supported by ZW, MC. YZ drafted the first version of the manuscript, supported by JL, NL, and JH. All authors contributed to data interpretation and critically assessed it. All authors approved the final version of the manuscript.

Funding: This work was supported by the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (grant number: 2019PT320027).

Competing interest: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Patient consent for publication： Not required.

Data availability statement: All data of the current study is present in the main manuscript, figures, tables, and online supplemental material.

Ethics statement： This study does not involve human participants.

REFERENCES

1 GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*. 2020;396(10258):1204–1222. doi: 10.1016/S0140-6736(20)30925-9.

2 Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021. doi: 10.3322/caac.21660.

3 Myers ER, Moorman P, Gierisch JM, et al. Benefits and harms of breast cancer screening: A systematic review. *JAMA*. 2015;314(15):1615-1634. doi: 10.1001/jama.2015.13183.

4 Pashayan N, Morris S, Gilbert FJ, et al. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer. *JAMA Oncol*. 2018;4(11): 1504–1510. doi: 10.1001/jamaoncol.2018.1901.

5 Qaseem A, Lin JS, Mustafa RA, Horwitch CA, Wilt TJ; Clinical Guidelines Committee of the American College of Physicians. Screening for breast cancer in average-risk women: A guidance statement from the American college of physicians. *Ann Intern Med*. 2019;170(8):547-560. doi: 10.7326/M18-2147.

6 Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81(24):1879-1886. doi: 10.1093/jnci/81.24.1879.

7 Bondy ML, Newman LA. Assessing breast cancer risk: evolution of the Gail Model. *J Natl Cancer Inst*. 2006;98(17):1172-1173.

8 Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst*. 1999;91(18):1541-1548.

9 Rockhill B, Spiegelman D, Byrne C, et al. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst*. 2001;93(5):358-366.

10 Lin CJ, Block B, Nowalk MP, et al. Breast cancer risk assessment in socioeconomically disadvantaged urban communities. *J Natl Med Assoc*. 2007;99(7):752-756.

11 Bondy ML, Newman LA. Breast cancer risk assessment models: applicability to African-American women. *Cancer*. 2003;97(1 Suppl):230-235.

12 Buron A, Vernet M, Roman M, et al. Can the Gail model increase the predictive value of a positive mammogram in a European population population screening setting? Results from a Spanish cohort. *Breast*. 2013;22(1):83-88.

13 Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1): 51-58. doi: 10.7326/M18-1376.

14 Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Ann Intern Med*. 2019;170(1): W1-W33. doi: 10.7326/M18-1377.

15 Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10): e1001744. doi: 10.1371/journal.pmed.1001744.

16 Rosner B, Colditz GA. Nurses' Health Study: Log-Incidence mathematical model of breast cancer incidence. *J Natl Cancer Inst*. 1996;88(6): 359-64. doi: 10.1093/jnci/88.6.359.

17 Ueda K, Tsukuma H, Tanaka H, et al. Estimation of individualized probabilities of

developing breast cancer for Japanese women. *Breast Cancer*. 2003;10(1):54-62. doi: 10.1007/BF02967626.

18 Colditz GA, Rosner BA, Chen WY, et al. Risk factors for breast cancer according to estrogen and progesterone receptor status. *J Natl Cancer Inst*. 2004;96(3): 218-28. doi: 10.1093/jnci/djh025.

19 Lee EO, Ahn SH, You C, et al. Determining the main risk factors and high-risk groups of breast cancer using a predictive model for breast cancer risk assessment in south Korea. *Cancer Nurs*. 2004;27(5): 400-6. doi: 10.1097/00002820-200409000-00010.

20 Tice JA, Miike R, Adduci K, et al. Nipple aspirate fluid cytology and the Gail Model for breast cancer risk assessment in a screening population. *Cancer epidem biomar*. 2005;14(2): 324-8. doi: 10.1158/1055-9965.EPI-04-0289.

21 Tice JA, Cummings SR, Ziv E, et al. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res*. 2005;94(2): 115-22. doi: 10.1007/s10549-005-5152-4.

22 Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst*. 2006;98:1204-1214. doi:10.1093/jnci/djj331.

23 Decarli A, Calza S, Masala G, et al. Gail Model for prediction of absolute risk of invasive breast cancer: Independent evaluation in the Florence-European prospective investigation into cancer and nutrition cohort. *J Natl Cancer Inst*. 2006;98(23): 1686-93. doi: 10.1093/jnci/djj463.

24 Novotny J, Pecen L, Petruzelka L, et al. Breast cancer risk assessment in the Czech female population-an adjustment of the original Gail model. *Breast Cancer Res Treat*. 2006;95(1): 29-35. doi: 10.1007/s10549-005-9027-5.

25 Gail MH, Costantino JP, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst*. 2007;99(23): 1782-92. doi: 10.1093/jnci/djm223.

26 Crispo A, D'Aiuto G, De Marco M, et al. Gail Model risk factors: Impact of adding an extended family history for breast cancer. *Breast Journal*. 2008;14(3): 221-27. doi:10.1111/j.1524-4741.2008.00566.

27 Tice JA, Cummings SR, Smith-Bindman R, et al. Using clinical factors and mammographic breast density to estimate breast cancer risk: Development and validation of a new predictive model. *Ann Intern Med*. 2008;148(5): 337-47. doi: 10.7326/0003-4819-148-5-200803040-00004.

28 Tamimi RM, Rosner B, Colditz GA. Evaluation of a breast cancer risk prediction

model expanded to include category of prior benign breast disease lesion. *Cancer*. 2010;116(21): 4944-53. doi: 10.1002/cncr.25386.

29 Petracci E, Decarli A, Schairer C, et al. Risk factor modification and projections of absolute breast cancer risk. *J Natl Cancer Inst*. 2011;103(13): 1037-1048. doi:10.1093/jnci/djr172.

30 Dite GS, Mahmoodi M, Bickerstaffe A, et al. Using SNP genotypes to improve the discrimination of a simple breast cancer risk prediction model. *Breast Cancer Res Treat*. 2013;139(3): 887-96. doi: 10.1007/s10549-013-2610-2.

31 Park B, Ma SH, Shin A, et al. Korean risk assessment model for breast cancer risk prediction. *PLOS One*. 2013;8(10): e76736. doi: 10.1371/journal.pone.0076736.

32 Anothaisintawee T, Teerawattananon Y, Wiratkapun C, et al. Development and validation of a breast cancer risk prediction model for Thai women: a cross-sectional study. *Asian Pac J Cancer Prev*. 2014;15(16): 6811-7. doi: 10.7314/apjcp.2014.15.16.6811.

33 Boggs DA, Rosenberg L, Adams-Campbell LL, et al. Prospective approach to breast cancer risk prediction in African American women: The black women's health study model. *J Clin Oncol*. 2015;33(9): 1038-44. doi: 10.1200/JCO.2014.57.2750.

34 Brentnall1 AR, Harkness EF, Astley SM, et al. Mammographic density adds

accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res*. 2015;17(1): 147. doi: 10.1186/s13058-015-0653-5.

35 Kerlikowske K, Gard CC, Sprague BL, et al. One versus two breast density measures to predict 5- and 10-year breast cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2015;24(6): 889-897. doi: 10.1158/1055-9965.EPI-15-0035.

36 Tice JA, Miglioretti DL, Li CS, et al. Breast density and benign breast disease: Risk assessment to identify women at high risk of breast cancer. *J Clin Oncol*. 2015;33(28): 3137-3143. doi:10.1200/JCO.2015.60.8869.

37 Schonberg MA, Li VW, Eliassen AH, et al. Accounting for individualized competing mortality risks in estimating postmenopausal breast cancer risk. *Breast Cancer Res Treat.* 2016;160(3): 547-562. doi: 10.1007/s10549-016-4020-8.

38 Shieh Y, Hu D, Ma L, et al. Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Res Treat*. 2016;159(3): 513-525. doi: 10.1007/s10549-016-3953-2.

39 Wang F, Dai J, Li M, et al. Risk assessment model for invasive breast cancer in Hong Kong women. *Medicine*. 2016;95(32): e4515. doi:10.1097/MD.0000000000004515.

40 Maas P, Barrdahl M, Joshi AD, et al. Breast cancer risk from modifiable and

nonmodifiable risk factors among white women in the United States. *JAMA Oncol.* 2016;2(10): 1295-1302. doi: 10.1001/jamaoncol.2016.1025.

41 Banegas MP, John EM, Slattery ML, et al. Projecting individualized absolute invasive breast cancer risk in US Hispanic women. *J Natl Cancer Inst*. 2017;109(2): djw215. doi: 10.1093/jnci/djw215.

42 Eriksson M, Czene K, Pawitan Y, et al. A clinical model for identifying the short-term risk of breast cancer. *Breast Cancer Res*. 2017;19(1):29. doi:10.1186/s13058-017-0820-y.

43 Hsieh YC, Tu SH, Su CT, et al. A polygenic risk score for breast cancer risk in a Taiwanese population. *Support Care Cancer*. 2017;163(1): 131-138. doi:10.1007/s10549-017-4144-5.

44 Hüsing A, Fortner RT, Kühn T, et al. Added value of serum hormone measurements in risk prediction models for breast cancer for women not using exogenous hormones: Results from the EPIC cohort. *Clin Cancer Res*. 2017;23(15): 4181-4189. doi: 10.1158/1078-0432.CCR-16-3011.

45 Salih AM, Alam-Elhuda DM, Alfaki MM, et al. Developing a risk prediction model for breast cancer: A Statistical Utility to Determine Affinity of Neoplasm (SUDAN-CA Breast). *Eur J Med Res*. 2017;22(1): 35. doi:10.1186/s40001-017-0277-6.

46 Wang SF, Ogundiran T, Ademola A, et al. Development of a breast cancer risk prediction cancer epidemiology, biomarkers & prevention model for women in Nigeria. *Cancer Epidem Biomar*. 2018;27(6): 636-643. doi: 10.1158/1055-9965.EPI-17-1128.

47 Zhang X, Rice M, Tworoger SS, et al. Addition of a polygenic risk score, mammographic density, and endogenous hormones to existing breast cancer risk prediction models: A nested case-control study. *PLOS Med*. 2018;15(9): e1002644. doi: 10.1371/journal.pmed.1002644.

48 Clendenen TV, Ge W, Koenig KL, et al. Breast cancer risk prediction in women aged 35–50 years: impact of including sex hormone concentrations in the Gail model. *BMC Med Res Methodol*. 2019;19(1):42. doi:10.1186/s13058-019-1126-z.

49 Wang L, Liu L, Zhen Lou,et al.Risk prediction for breast Cancer in Han Chinese women based on a cause-specific Hazard model. *BMC Cancer*. 2019;19(1): 128. doi:10.1186/s12885-019-5321-1.

50 Abdolell M, Payne JI, Caines J, et al. Assessing breast cancer risk within the general screening population: developing a breast cancer risk model to identify higher risk women at mammographic screening. *Eur Radiol*. 2020;30(10):5417-5426. doi:10.1007/s00330-020-06901-x.

51 Qiu C, Wang P, Wang B, et al. Establishment and validation of an immunodiagnostic model for prediction of breast cancer. *OncoImmunology*. 2020;9(1): e1682382-2. doi:10.1080/2162402X.2019.1682382.

52 Han Y, Lv J, Yu C, et al. Development and external validation of a breast cancer absolute risk prediction model in Chinese population. *Breast Cancer Res.* 2021;23(1): 62. doi: 10.1186/s13058-021-01439-2.

53 Rosner B, Tamimi RM, Kraft P, et al. Simplified Breast Risk Tool Integrating Questionnaire Risk Factors, Mammographic Density, and Polygenic Risk Score: Development and Validation. *Cancer Epidemiol Biomarkers Prev.* 2021;30(4): 600-607. doi: 10.1158/1055-9965.EPI-20-0900.

54 Yiangou K, Kyriacou K, Kakouri E, et al. Combination of a 15-SNP Polygenic Risk Score and Classical Risk Factors for the Prediction of Breast Cancer Risk in Cypriot Women. *Cancers (Basel)*. 2021;13(18): 4568. doi: 10.3390/cancers13184568.

55 Louro J, Posso M, Boon MH, et al. A systematic review and quality assessment of individualised breast cancer risk prediction models. *Br J Cancer*. 2019;121(1): 76–85. doi: 10.1038/s41416-019-0476-8.

56 Anothaisintawee T, Teerawattananon Y, Wiratkapun C, et al. Risk prediction models of breast cancer: A systematic review of model performances. *Breast Cancer Res*

Treat. 2012;133(1): 1–10. doi 10.1007/s10549-011-1853-z.

57 Heestermans T, Payne B, Kayode GA, et al. Prognostic models for adverse pregnancy outcomes in low-income and middle-income countries: a systematic review. *BMJ Global Health*. 2019;4(5): e001759. doi:10.1136/bmjgh-2019-001759.

58 Bellou V, Belbasis L, Konstantinidis AK, et al. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ*. 2019;367: l5358. doi: 10.1136/bmj. l5358.

59 de Jong Y, Ramspek CL, van der Endt VHW, et al. A systematic review and external validation of stroke prediction models demonstrates poor performance in dialysis patients. *J Clin Epidemiol*. 2020;123: 69-79. doi: 10.1016/j.jclinepi.2020.03.015.

60 Bridge J, Blakey JD, Bonnett LJ. A systematic review of methodology used in the development of prediction models for future asthma exacerbation. *BMC Med Res Methodol*. 2020;20(1): 22. doi: 10.1186/s12874-020-0913-7.

61 Banegas MP, Gail MH, LaCroix A, et al. Evaluating breast cancer risk projections for Hispanic women. *Breast Cancer Res Treat*. 2020;132(1):347–353. doi: 10.1007/s10549-011-1900-9.

62 Min JW, Chang MC, Lee HK, et al. Validation of risk assessment models for predicting the incidence of breast cancer in Korean women. *J Breast Cancer*.

2014;17(3):226–235. doi: 10.4048/jbc.2014.17.3.226.

63 Chay WY, Ong WS, Tan PH, et al. Validation of the Gail model for predicting individual breast cancer risk in a prospective nationwide study of 28,104 Singapore women. *Breast Cancer Res*. 2012;14(1): R19. doi: 10.1186/bcr3104.

64 Adams-Campbell LL, Makambi KH, Palmer JR, et al. Diagnostic accuracy of the Gail model in the Black Women's Health Study. *Breast J*. 2007;13(4):332–336. doi: 10.1111/j.1524-4741.2007. 00439.x.

65 Antoniou AC, Pharoah PP, Smith P, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. Br J Cancer. 2004;91(8):1580-90. doi: 10.1038/sj.bjc.6602175.

66 Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med. 2019;21(8):1708-1718. doi: 10.1038/s41436-018-0406-9. 67

67 Berry DA, Parmigiani G, Sanchez J, et al. Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. J Natl Cancer Inst. 1997;89(3):227-38. doi: 10.1093/jnci/89.3.227.

68 Harrell FE. Regression modeling strategies, with applications to linear models, logistic regression, and survival analysis. New York: Springer, 2001.

69 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4): 361-87. doi:10.1002/(SICI)1097-0258(19960229)15:4<361: AID-SIM168>3.0.CO;2-4.

70 Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol*. 1996;49(8): 907–16. doi: 10.1016/0895-4356(96)00025-x.

71 Little RJA, Rubin DB. Statistical Analysis With Missing Data. Hoboken, NJ: Wiley; 2002.

72 White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30: 377-99. doi:10.1002/sim.4067.

73 Vergouwe Y, Royston P, Moons KG, et al. Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol*. 2010; 63:205-14. doi: 10.1016/j.jclinepi. 2009.03.017.

74 Royston P, Moons KG, Altman DG, et al. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338(b604). doi: 10.1136/bmj.b604.

75 Warwick J, Birke H, Stone J, et al. Mammographic breast density refines Tyrer-Cuzick estimates of breast cancer risk in high-risk women: findings from the placebo

arm of the International Breast Cancer Intervention Study I. *Breast Cancer Res.* 2014;16(5):451. doi: 10.1186/s13058-014-0451-5.

76 Brentnall AR, Harkness EF, Astley SM, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. *Breast Cancer Res.* 2015;17(1):147. doi: 10.1186/s13058-015-0653-5.

77 Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet*. 2019;104(1):21-34. doi: 10.1016/j.ajhg.2018.11.002.

78 Pal Choudhury P, Wilcox AN, Brook MN, et al. Comparative Validation of Breast Cancer Risk Prediction Models and Projections for Future Risk Stratification. *J Natl Cancer Inst*. 2020;112(3):278-285. doi: 10.1093/jnci/djz113.

79 Hurson AN, Pal Choudhury P, Gao C, et al. Prospective evaluation of a breast-cancer risk model integrating classical risk factors and polygenic risk in 15 cohorts from six countries. *Int J Epidemiol*. 2022;50(6):1897-1911. doi: 10.1093/ije/dyab036.

80 Pal Choudhury P, Brook MN, Hurson AN, et al. Comparative validation of the BOADICEA and Tyrer-Cuzick breast cancer risk models incorporating classical risk factors and polygenic risk in a population-based prospective cohort of women of

European ancestry. *Breast Cancer Res*. 2021;23(1):22. doi: 10.1186/s13058-021-01399-7.

81 Arveux P, Wait S, Schaffer P. Building a model to determine the cost-effectiveness of breast cancer screening in France. *Eur J Cancer Care.* 2003;12(2): 143-53. doi: 10.1046/j.1365-2354.2003.00373. x.

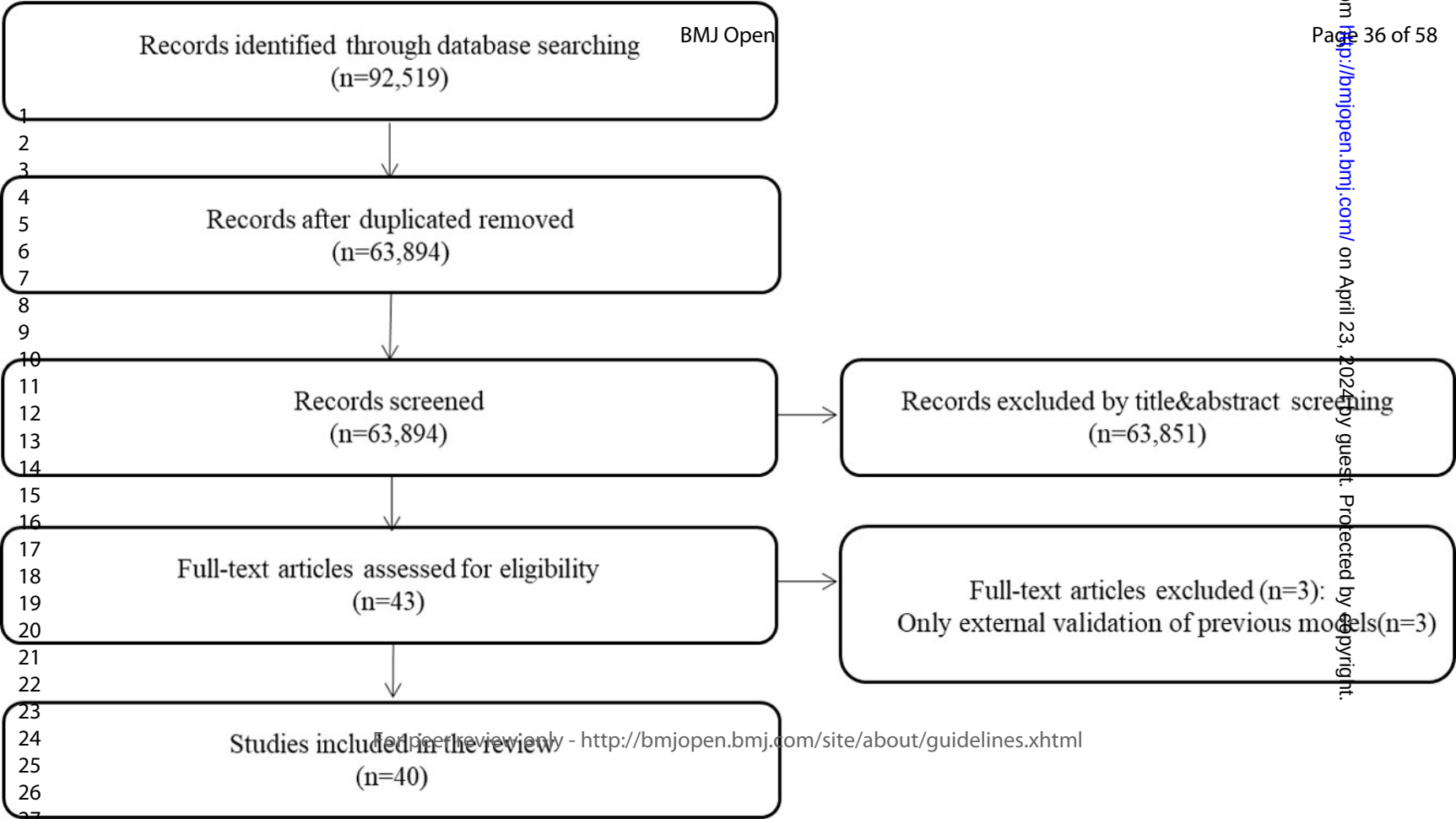
Figure legends:

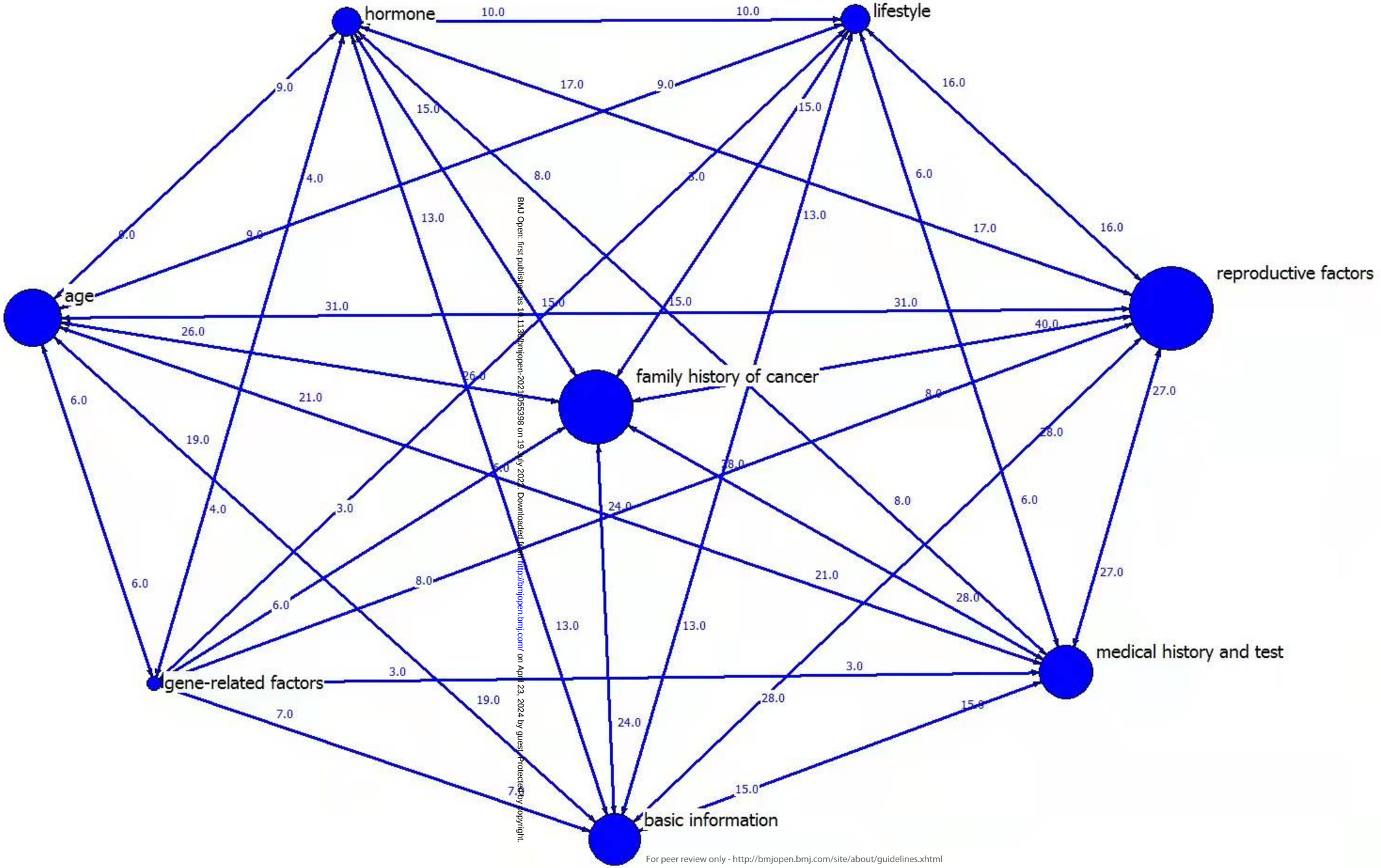
Figure1. PRISMA (preferred reporting items for systematic reviews and meta-analyses) flowchart.

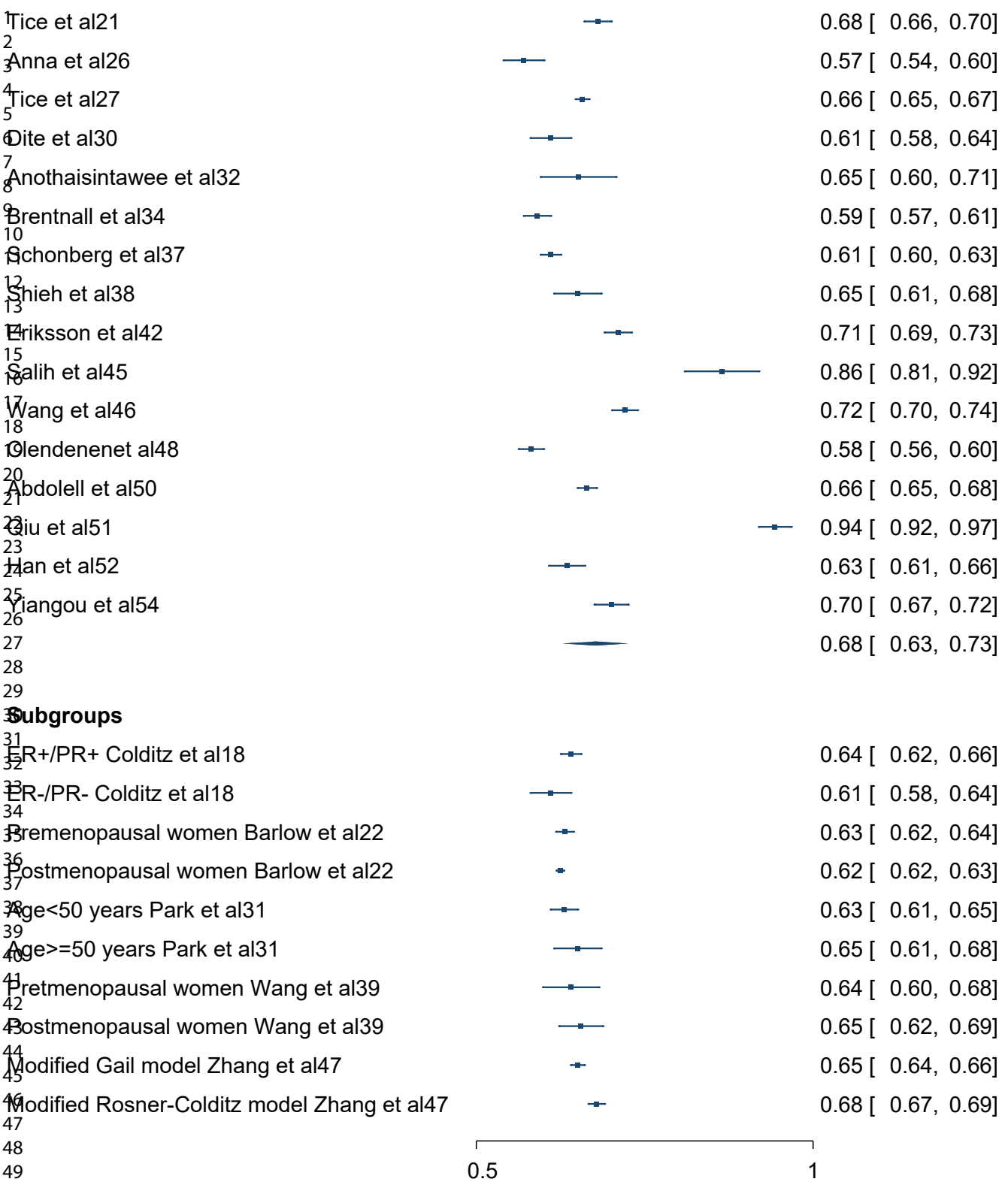
Figure 2. Network diagram of eight categorized risk factors (age, basic information, family history of cancer, gene-related factors, hormone, lifestyle, medical history and test, and reproductive factors).

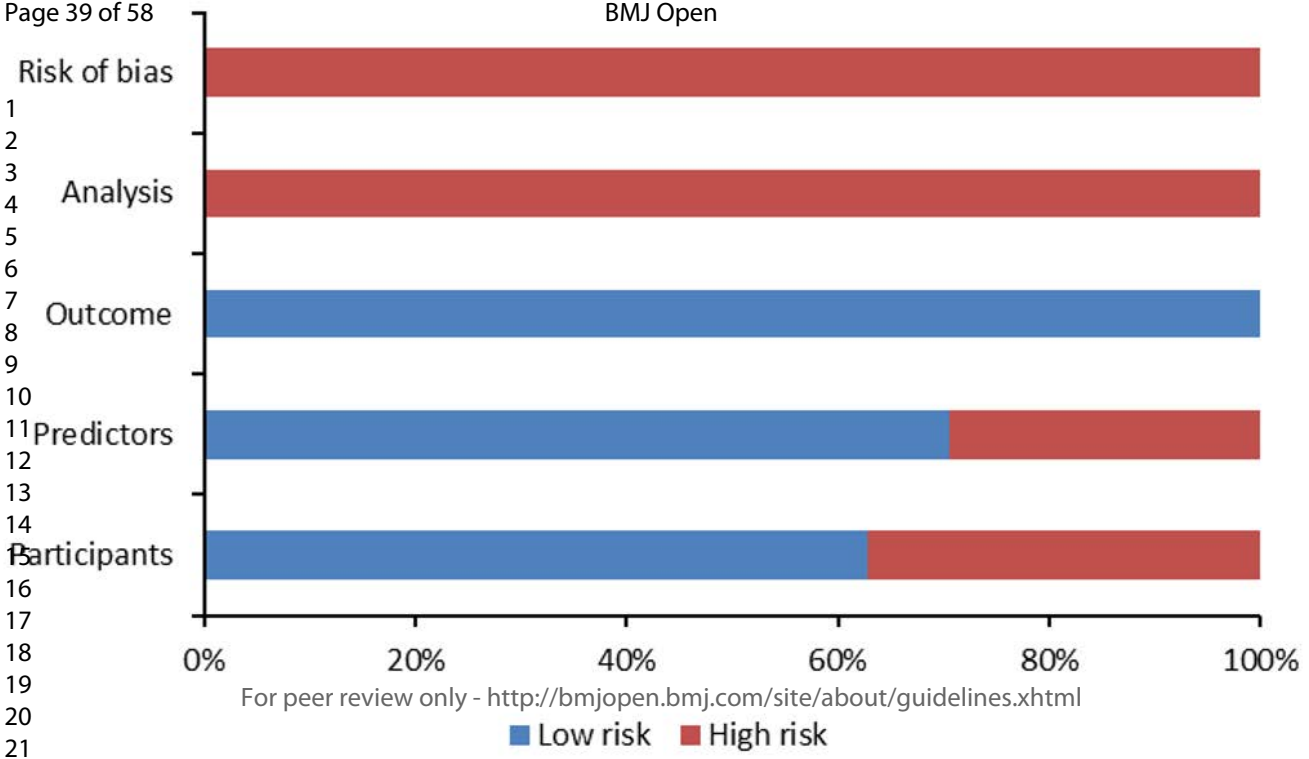
Figure 3. Area under the curve (AUC) and confidence intervals reported by the included studies.

Figure 4. Risk of bias assessment (using PROBAST) of all assessed models based on four domains.






All
Tice et al21 0.68 [0.66, 0.70]
Anna et al26 0.57 [0.54, 0.60]
Tice et al27 0.66 [0.65, 0.67]
Dite et al30 0.61 [0.58, 0.64]
Anothaisintawee et al32 0.65 [0.60, 0.71]
Brentnall et al34 0.59 [0.57, 0.61]
Schonberg et al37 0.61 [0.60, 0.63]
Shieh et al38 0.65 [0.61, 0.68]
Eriksson et al42 0.71 [0.69, 0.73]
Salih et al45 0.86 [0.81, 0.92]
Wang et al46 0.72 [0.70, 0.74]
Olendenenet al48 0.58 [0.56, 0.60]
Abdolell et al50 0.66 [0.65, 0.68]
Qiu et al51 0.94 [0.92, 0.97]
Han et al52 0.63 [0.61, 0.66]
Yiangou et al54 0.70 [0.67, 0.72]
0.68 [0.63, 0.73]
Subgroups
ER+/PR+ Colditz et al18 0.64 [0.62, 0.66]
ER-/PR- Colditz et al18 0.61 [0.58, 0.64]
Premenopausal women Barlow et al22 0.63 [0.62, 0.64]
Postmenopausal women Barlow et al22 0.62 [0.62, 0.63]
Age<50 years Park et al31 0.63 [0.61, 0.65]
Age>=50 years Park et al31 0.65 [0.61, 0.68]
Pretmenopausal women Wang et al39 0.64 [0.60, 0.68]
Postmenopausal women Wang et al39 0.65 [0.62, 0.69]
Modified Gail model Zhang et al47 0.65 [0.64, 0.66]
Modified Rosner-Colditz model Zhang et al47 0.68 [0.67, 0.69]
0.5
1



Appendix

Appendix Table 1. Searching strategy.

Searching strategy
Take PubMed for example:
#1 "Breast Neoplasms"[Mesh] OR "Breast Carcinoma In Situ"[Mesh] OR "Breast Neoplasms, Male"[Mesh] OR "Carcinoma, Ductal, Breast"[Mesh] OR "Carcinoma, Lobular"[Mesh] OR "Hereditary Breast and Ovarian Cancer Syndrome"[Mesh] OR "Inflammatory Breast Neoplasms"[Mesh] OR "Triple Negative Breast Neoplasms"[Mesh] OR "Unilateral Breast Neoplasms"[Mesh] OR phyllodes tumor[Title/Abstract] OR breast sarcoma[Title/Abstract] OR mamma cancer*[Title/Abstract] OR mammary cancer*[Title/Abstract] OR mammary gland cancer*[Title/Abstract] OR Mammary Ductal Carcinoma*[Title/Abstract] OR breast gland cancer*[Title/Abstract] OR breast gland neoplasm*[Title/Abstract] OR Breast Neoplasm*[Title/Abstract] OR Breast Tumor*[Title/Abstract] OR Breast Cancer*[Title/Abstract] OR Mammary Cancer*[Title/Abstract] OR Breast Malignant Neoplasm*[Title/Abstract] OR Breast Malignant Tumor*[Title/Abstract] OR Human Mammary Carcinoma*[Title/Abstract] OR Human Mammary Neoplasm*[Title/Abstract] OR Breast Carcinoma*[Title/Abstract] OR Lobular Carcinoma*[Title/Abstract] 418,670
#2 ("Regression Analysis"[Mesh] OR "Multivariate Analysis"[Mesh] OR "Models, Biological"[Mesh] OR "Models, Statistical"[Mesh] OR "Algorithms"[Mesh]) AND "Risk Assessment" [Mesh] 52,269
#3 predict*[Title/Abstract] AND (outcome*[Title/Abstract] OR index[Title/Abstract] OR rule*[Title/Abstract] OR decision*[Title/Abstract] OR scor*[Title/Abstract]) 624,639
#4 risk*[Title/Abstract] AND (predict*[Title/Abstract] OR calculate*[Title/Abstract] OR assess*[Title/Abstract] OR scor*[Title/Abstract] OR algorithm[Title/Abstract]) 1,109,068
#5 model*[Title/Abstract] AND (logistic[Title/Abstract] OR statistic*[Title/Abstract] OR risk*[Title/Abstract] OR predict*[Title/Abstract]) 1,1035,123
#6 OR/2-5 2,195,108
#7 #1 AND #6 54,653

Appendix Table 2. Classification of risk factors.

age	
reproductive factors	age at menarche, age at first birth, menopause, age at subsequent births, menstrual regularity, total menstrual duration, breastfeeding, breast density, parity, reproductive characteristics, microcalcifications and masses, abortions, breast volume
family history of cancer	family history of breast cancer, family history of any cancer
hormone	hormone therapy, oral contraceptives, estrogen plus progestin use, testosterone, estradiol, sex hormone binding globulin, Insulin-like growth factor-I, estrone sulphate, prolactin, anti-Müllerian hormone
gene-related factors	polygenic risk score, rs2981582 (FGFR2), rs3803662(TOX3), rs889312(MAP3K1), rs13387042(2q35), rs13281615(8q24), rs4415084 (FGF10), rs3817198 (LSP1), rs981782(HCN1), rs10822013(ZNF365), rs3784099(RAD51B)
lifestyle	alcohol consumption, smoking status, exercise, light at night, sleep quality, vegetables and fruits, cereals, life satisfaction score
medical history and test	previous biopsies, benign breast disease, nipple aspirate fluid cytology, prior breast procedure, prior false-positive mammogram, breast inflammatory, benign breast category, benign breast disease, atypical hyperplasia, mammogram in past 2 years, diabetes, myocardial infarction, stroke, emphysema, congestive heart failure, p53, CyclinB1, p16, p62,14-3-3ξ
basic information	body mass index, weight, education, ethnicity, occupational activity, height, residence area

Appendix Table 3. Summary of the 40 included studies.

Author	Year	develop							validate				
		Study design	Research method	Targeted population	No of risk factors	Risk factors	Model development (AUC (95%CI); E/O ratio (95%)	Sample size of development	Type of validation	study design	Targeted population	Model validation (AUC (95%CI); E/O ratio (95%)	Sample size of validation
Gail et al [6]	1989	Case-control study	Logistic regression	Caucasian; 20–79 years	5	Age, age at menarche, age at first birth, number of previous biopsies, number of first degree relatives with breast cancer	AUC: none; E/O ratio: none	2,852cases/ 3,146 controls	None	None	None	None	None
Rosner et al [16]	1996	Prospective cohort study	Poisson regression	Caucasian; 30–64 years	5	Age, age at menarche, age at first birth, menopause, age at subsequent births	AUC: none; E/O ratio: none	2,249 cases/ 89,132 total	None	None	None	None	None
Ueda et al [17]	2003	Case-control study	Logistic regression	Asian women; age was not specified.	4	Age at menarche, age at first birth, family history of breast cancer, body mass index	AUC: none; E/O ratio: none	376 cases/ 430 controls	None	None	None	None	None
Colditz et al [18]	2004	Prospective cohort study	Logistic regression	Caucasian; 30-64 years	11	Age, age at menarche, age at first birth, menopause, age at subsequent births, benign breast disease, postmenopausal hormone use, family history of breast cancer in a first-degree relative, weight, body mass index, alcohol consumption	AUC: ER+/PR+: 0.64 (0.63,0.66); ER-/PR-: 0.61 (0.58, 0.64); E/O ratio: none	2,846 cases/ 66,145 total	None	None	None	None	None
Lee et al [19]	2004	Case-control study	Logistic regression	Asian women; age was not specified.	1) Hospitalized controls: 5 2) Nur	1) Hospitalized controls: family history, menstrual regularity, total menstrual duration, age at first full-term pregnancy, duration of breastfeeding 2) Nurse/teacher controls: age, education level, menstrual regularity, drinking status, smoking status	AUC: 1) Hospitalized controls: 0.714; 2) Nurse/teacher controls: 0.867; E/O ratio: none	1) Hospitalized controls: 384 cases/ 166 controls; 2) Nurse/teacher controls: 384 cases/	None	None	None	None	None

					se/teacher controls: 5			104 controls					
Tice et al [20]	2005	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 18 years and older	6	Age, age at menarche, previous biopsy, age at first birth, first degree breast cancer, nipple aspirate fluid cytology	AUC: 0.64; E/O ratio: none	400 cases/ 6,904 total	None	None	None	None	None
Tice et al [21]	2005	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35 years and older	6	Age, age at menarche, previous biopsy, age at first birth, first degree breast cancer, breast density	AUC: 0.68 (0.66,0.70); E/O ratio: none	955 cases/ 81,777 total	None	None	None	None	None
Barlow et al [22]	2006	Prospective cohort study	Logistic regression	Multiple ethnicities, 35-84 years	1) Premenopausal women: 4 2) Postmenopausal women: 10	1) Premenopausal women: age, breast density, family history of breast cancer, a prior breast procedure 2) Postmenopausal women: age, breast density, race, ethnicity, family history of breast cancer, a prior breast procedure, body mass index, natural menopause, hormone therapy, a prior false-positive mammogram	AUC: Premenopausal women: 0.631 (0.618, 0.644); postmenopausal women: 0.624 (0.619, 0.630) E/O ratio [a]: Premenopausal women: 1.000 postmenopausal women: 1.001	1) Premenopausal women: 1,726 cases/ 568,215 total; 2) postmenopausal women: 9,300 cases/ 1,642,824 total	Internal validation	None	None	None	None
Decarli et al [23]	2006	Case-control study	Logistic regression	Caucasian; 20–74 years	5	Age , age of menarche, number of breast biopsies, age at first live birth, first degree breast cancer	AUC: none; E/O ratio: none	2569 cases/ 2588 controls	External validation	Prospective cohort study	Caucasian; 35-64 years	AUC: 0.59; E/O ratio: 0.96(0.84, 1.11)	194 cases /10,031 total

Novotny et al [24]	2006	Case-control study	Logistic regression	Multiple ethnicities; 23-84 years	8	Age of menarche, number of biopsies, age at first childbirth, number of breast cancer cases in first-degree relatives, number of any cancer cases in first-degree relatives, breast inflammatory, body mass index, number of conceptions	AUC: none; E/O ratio: none	2299 cases/ controls	None	None	None	None	None
Gail et al [25]	2007	Case-control study	Logistic regression	African-American Women; 35-64 years	5	Age, age at menarche, number of affected mother or sisters, age at first live birth, number of previous benign biopsy examinations	AUC: none; E/O ratio: none	1607 cases/ 1647 controls	External validation	Prospective cohort study	African American women; 50-79 years	AUC: 0.555 (0.535,0.575); E/O ratio: 0.93b	350 cases /14,059 total
Anna et al [26]	2008	Case-control study	Logistic regression	Caucasian; age was not specified	5	Age, age at menarche, number of biopsies, age at first live birth, family history	AUC: 0.57 (0.54, 0.60); E/O ratio: none	558 cases/ 1207 controls	Internal validation	None	None	None	None
Tice et al [27]	2008	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35 years or older	5	Age, ethnicity, first degree breast cancer, previous biopsies, breast density	AUC: 0.657 (0.65,0.67); E/O ratio: 1.00 (0.98,1.03)	14,766 cases/ 1095484 total	Internal validation	Prospective cohort study	Multiple ethnicities; 35 years or older	AUC: 0.660(0.65,0.66); E/O ratio: 1.03(0.99,1.06)	3,465 cases/ 251,789 total
Tammi et al [28]	2010	Nested case-control study	Logistic regression	Caucasian; 40-79 years	11	The type of benign breast disease, age, age at menarche, age at first birth and at each subsequent birth, age at menopause and type of menopause, history of benign breast diseases, family history of breast cancer in mother or sister, height, weight at age 18 years, current use of postmenopausal hormones (including type and duration of use), alcohol intake	AUC: 0.635; E/O ratio: none	240 cases/ 1036 controls	None	None	None	None	None

Petracci et al [29]	2011	Case-control study	Logistic regression	Caucasian; 20-74 years	8	Reproductive characteristics, education, occupational activity, family history, biopsy history, alcohol consumption, leisure physical activity, body mass index.	AUC: none; E/O ratio: 1.10 (0.96,1.26)	2569 cases/ 2588 controls	External validation	Prospective cohort study	Caucasian; 35-64 years	AUC: Age<50: 0.62(0.555,0.689) ; age>=50: 0.57 (0.519,0.614); E/O ratio: 1.10(0.96,1.26)	206 cases/ 8,426 total
Dite et al [30]	2013	Case-control study	Logistic regression	Multiple ethnicities; 35-59 years	13	Age, ethnicity, age at menarche, age at birth of first child, number of first-degree relatives with breast cancer, number of biopsies, presence of atypical hyperplasia, rs2981582(FGFR2), rs3803662(TOX3), rs889312(MAP3K1), rs13387042(2q35), rs13281615(8q24), rs4415084 (FGF10), rs3817198 (LSP1)	AUC: 0.61 (0.58,0.64); E/O ratio: none	962 cases/ 463 controls	None	None	None	None	None
Park et al [31]	2013	Case-control study	Logistic regression	Asian women; age was not specified.	1) Age<50 years: 7 2) Age>=50 years: 7	1)Age<50 years: a family history of breast cancer in first-degree relatives, age at menarche, menopausal status, age at first full-term pregnancy, duration of breast feeding, oral contraceptive usage, exercise. 2)Age>=50 years: a family history of breast cancer in first degree relatives, age at menarche, age at menopause, experience of pregnancy, body mass index, oral contraceptive usage, exercise	AUC: Age<50 years: 0.63 (0.61-0.65); Age>=50 years: 0.65 (0.61- 0.68); E/O ratio: none	3,789 cases/ 3,789 controls	External validation	Prospective cohort study	None	1)Korean Multi-Center Cohort (KMCC): AUC: 0.61(0.49,0.72); E/O ratio: 0.97(0.67,1.40) 2)National Cancer Center (NCC) cohort: AUC: 0.89(0.85,0.93) E/O ratio: 0.96(0.70,1.37)	1) KMCC: 29cases/ 6148 total; 2)NCC: 36 cases/ 7546 total
Apothaisintawee et al [32]	2014	Cross-sectional study	Logistic regression	Asian women; age was not specified	4	Age, menopausal status, body mass index, use of oral contraceptives	AUC: 0.651 (0.595, 0.707); O/E ratio: 1.00 (0.82, 1.21) [b]	107cases/ 15,718total	Internal and external validation	Cross-sectional study	Asian women; 18 years or older	Internal validation: AUC: 0.646(0.642,0.650); E/O ratio: none; External validation:	35 cases/ 4,978 total

												AUC: 0.609(0.511,0.706); O/E ratio: 0.97 (0.68, 1.35)[b]	
Boggs et al [33]	2015	Prospective cohort study	Cox proportional hazards regression	African-American Women; 30-69 years	9	Family history, previous biopsy, body mass index at age 18 years, age at menarche, age at first birth, oral contraceptive use, bilateral oophorectomy, estrogen plus progestin use, height	AUC: none; E/O ratio: none	896 cases/ 55,093 total	Internal validation	Prospective cohort study	African American Women; 30-69 years	AUC: 0.59 (0.56, 0.61); E/O ratio: 0.96(0.88,1.05)	506 cases/ 48,193 total
Brentnall et al [34]	2015	Prospective cohort study	Logistic regression	Caucasian; 47-73 years	1) Gail model +Density residual: :8 2) Tyrer-Cuzick+density residual: 11	1) Gail model+ Density residual: Age, Ethnicity, age at menarche, age at first birth, number of previous biopsies, benign disease, number of first degree relatives with breast cancer, density residual 2) Tyrer-Cuzick+ density residual: Age, gen phenotype, family history, age at menarche, age at first birth, menopause, atypical Hyperplasia, lobular carcinoma in situ, height, body mass index, density residual	(1) Primary (invasive+ DCIS): 1)Gail model+ Density residual: AUC: 0.59(0.57,0.61); E/O ratio: none; 2)Tyrer- Cuzick+ density residual: AUC: 0.61(0.59,0.63); E/O ratio: none; (2) Secondary(invasive): 1)Gail model+ Density residual: AUC: 0.59(0.57,0.61); E/O ratio: none; 2)Tyrer-Cuzick+ density residual: AUC: 0.61(0.58–0.63); E/O ratio: none	697 cases/ 50,628 total	None	None	None	None	None

Kerlikowske et al [35]	2015	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35-74 years	5	Age, ethnicity, first degree breast cancer, previous biopsies, changes in breast density	AUC: 5-year risk model: 0.640; 10-year risk model: 0.628; E/O ratio: 5-year risk model: 0.98(0.96,1.00); 10-year risk model: 0.95(0.94,0.96)	13,715 cases/ 722,654 total	None	None	None	None	None
Tice et al [36]	2015	Prospective cohort study	Cox proportional hazards regression	Multiple ethnicities; 35-74 years	6	Age, race/ethnicity, family history of breast cancer, history of breast biopsy, benign breast disease diagnoses, breast density	AUC: 0.665; E/O ratio: 5 Years: 1.04(1.02 ,1.06); 10 years: 1.05 (1.03,1.06)	17908 cases/ 1,135,977 total	None	None	None	None	None
Tyrer et al [37]	2016	Prospective cohort study	Competing risk regression	Multiple ethnicities; 57–85 years	16	Age at study entry, postmenopausal hormone use, number of first-degree relatives with history of breast cancer and age at diagnosis, history of breast biopsy, highest body mass index in past 10 years, age at menopause, age at first birth and parity, average alcohol use per day (highest average use in past 10 years), cigarette use, mammogram in past 2 years, limited in moderate daily activity, diabetes, myocardial infarction, stroke, emphysema, congestive heart failure	AUC： 0.61 (0.60,0.63); E/O ratio: none	73,066 total	External validation	Prospective cohort study	Multiple ethnicities; 55-91 years	AUC: 0.57 (0.55,0.58); E/O ratio: 0.92(0.88,0.97)	74,887 total
Gail et al [38]	2016	Nested case-control study	Logistic regression	Multiple ethnicities; 36-86 years	7	Age, ethnicity, first degree breast cancer, previous biopsies, breast density, polygenic risk score, body mass index	AUC:0.65(0.61,0.68); E/O ratio: none	486 cases/ 495 controls	None	None	None	None	None

Wang et al [39]	2016	Case-control study	Logistic regression	Asian women; 20-84 years	1)Premenopausal: 5; 2)Postmenopausal: 11	1) Premenopausal: age, number of parity, case number of breast cancer in first-degree relatives, light at night, sleep quality; 2) Postmenopausal: age, number of parity, case number of breast cancer in first-degree relatives, light at night, body mass index, age at menarche, age at first give birth, ever breast feeding, ever using of oral contraceptive, hormone replacement treatment, history of benign breast diseases.	1) Pretmenopausal women: AUC: 0.640(0.598,0.681); E/O ratio: none; 2) Postmenopausal women: 0.655(0.621,0.686); E/O ratio: none	923 cases / 918 controls	Internal validation	Case-control study	Asian women; 20-84 years	1) Premenopausal: average AUC: 0.621; 3) Postmenopausal: Average AUC: 0.632	None
Maas et al [40]	2016	Prospective cohort study	Logistic regression	Caucasian	11	Age at menarche, menopause, age at first birth, parity, alcohol consumption, height, smoking status, BMI, family history, hormone therapy, PRS	AUC: 0.640; E/O ratio: none	17,171 cases / 19,862 controls	None	None	None	None	None

Banegas et al [41]	2017	Case-control study	Logistic regression	Hispanic Women; 35-79 years	1) The US-born Hispanic risk model :3; 2) the foreign-born Hispanic risk model :4	1) The US-born the Hispanic risk model: age at first full-term pregnancy, biopsy for benign breast disease, family history of breast cancer; 2) The foreign-born the Hispanic risk model: age at first full-term pregnancy, biopsy for benign breast disease, family history of breast cancer, age at menarche	None	1086 cases/ 411 controls	External validation	Prospective cohort study	Hispanic Women; 50-79 years	1)US-born Hispanics: AUC: 0.564 (0.485, 0.644); O/E:1.07 (0.81 ,1.40) [b]; 2)Foreign-born Hispanics: AUC: 0.625 (0.487 ,0.764); O/E: 0.66 (0.41,1.07) [b] 3) Hispanics of unknown nativity: AUC: 0.582(0.509,0.656); O/E: 0.89(0.69,1.14) [b]	130 cases/ 6,220 total
Eriksson et al [42]	2017	Nested case-control study	Logistic regression	Caucasian; 40-74 years	7	MD, computer-aided detection of microcalcifications and masses, use of hormone replacement therapy, family history of breast cancer, menopausal status, age, body mass index	AUC: 0.71(0.69,0.73); E/O ratio: none	433cases / 1732 controls	None	None	None	None	None
Hsieh et al [43]	2017	Case-control study	Logistic regression	Asian women; 20-90 years	11	FGFR2 (rs2981582), HCN1 (rs981782), MAP3K1 (rs889312), TOX3(rs3803662), ZNF365(rs10822013), RAD51B(rs3784099), age, body mass index, age at menarche, parity, menopausal status	AUC: 0.6652; E/O ratio: none	446 cases/ 514 controls	None	None	None	None	None

Husing et al [44]	2017	Nested case-control study	Logistic regression	Multiple ethnicities; 26-77 years	13	Menopausal status, age at menarche, age at menopause, duration of postmenopausal hormones use, parity, number of children and age at first full term pregnancy, family history of breast cancer, alcohol consumption at recruitment, body mass index, measurements of testosterone, estradiol, sex hormone binding globulin, Insulin-like growth factor-I	AUC: none; E/O ratio: none	1,217 cases/ 1,976 controls	Internal validation	None	None	None	None
Salih et al [45]	2017	Cross-sectional study	Logistic regression	Caucasian; 32–74 years	5	Age, age at menarche, family history, vegetables and fruits weekly servings, type of cereals used	AUC: 0.864(0.81,0.92)	63 cases/ 90 controls	Internal validation	None	None	O/E ratio: 0.78[b]	None
Wang et al [46]	2018	Case–control study	Logistic regression	Nigerian women; age was not specified	9	Age, age at menarche, parity, duration of breastfeeding, family history of breast cancer, height, body mass index, benign breast diseases, alcohol consumption	AUC: 0.720(0.701,0.739); E/O ratio: 1.01 (0.93,1.09)	1,208 cases/ 1,484 controls	Internal validation	None	Nigerian women; 20-79 years	AUC: 0.694 (0.666,0.721); E/O ratio: none	603 cases/ 741 controls

Zhang et al [47]	2018	Nested case-control study	Logistic regression	Caucasian; 34-70 years	1) Gail model + PRS + MD + T + E1S +PRL: 10; 2) Rosner-Colditz model + PRS + MD + T + E1S + PRL: 16	1) Gail model+ PRS + MD + T + E1S +PRL: Age, age at menarche, previous biopsies, age at first birth, first degree breast cancer, PRS, MD, E1S, T, PRL 2) Rosner-Colditz model+ PRS + MD + T + E1S + PRL: age, age at menarche, age at first birth, menopause, age at subsequent births, benign breast disease, hormone replacement therapy, first degree breast cancer, weight, body mass index, alcohol, PRS, MD, E1S, T, PRL	AUC: Gail model+ PRS + MD + T + E1S +PRL: 0.65(0.64,0.66); Rosner-Colditz model+ PRS + MD + T + E1S + PRL: 0.678 (0.666,0.690); E/O ratio: none	4,006 cases / 7,874 controls	Internal validation	None	None	None	None
Tendenenet al [48]	2019	Nested case-control study	Logistic regression	Multiple ethnicities; 35–50 years	6	Age at menarche, age at first live birth, number of benign breast biopsies, number of first-degree family members with breast cancer, AMH, tT	AUC: 0.581(0.562,0.599); E/O ratio: none	1,762 cases/ 1,890 controls	None	None	None	None	None
Wang et al [49]	2019	Case-control study	Logistic regression	Asian women; 25-70 years	6	Number of abortions, age at first live birth, benign breast disease history, body mass index,	None	328 cases / 656 controls	External validation	Prospective cohort study	Asian women	AUC: 0.64 (0.55,0.72); E/O ratio: 1.03 (0.74,1.49)	34 cases/ 13,176 total

						breast cancer family history, life satisfaction score							
Abdolell et al [50]	2020	Nested case-control study	Logistic regression	Caucasian; 40-75 years	5	Age at screen, percent mammographic density, breast volume, core biopsy history, family history	AUC: 0.664(0.650,0.678); E/O ratio: none	1,882 cases/ 5,888 controls	None	None	None	None	None
Qiu et al [51]	2020	Case–control study	Logistic regression	Asian women; 29-81 years	5	p53, CyclinB1, p16, p62,14-3-3ξ	AUC:0.943(0.919,0.967); E/O ratio: none	184 cases/ 184 controls	External validation	Case-control study	Asian women; 24-78 years	AUC: 0.916(0.886,0.947); E/O ratio: none	197 cases/ 109 controls
Han et al [52]	2021	Prospective cohort study	Cox regression	Asian women; 30-79 years	8	age, residence area, education, BMI, height, family history of cancer, parity, age at menarche	AUC: 0.634(0.608,0.661); E/O ratio: 1.01(0.94,1.09)	2,287 cases/ 300,824 total	External validation	Prospective cohort study	Asian women;	AUC: 0.585(0.564,0.605) E/O ratio: 0.94(0.89,0.99)	73,203 total
Kraft et al [53]	2021	Nested case-control study	Logistic regression	Caucasian; 40-75 years	4	Age, breast density, questionnaire score, PRS	AUC: 0.658 E/O ratio: none	2,799 cases/ 75,557 controls	External validation	Nested case-control study	Caucasian; 40-75 years	AUC: 0.687	438 cases/ 898 controls
Yiangou et al [54]	2021	Case–control study	Logistic regression	Cypriot Women	11	menopause, age at menarche, parity, age at first birth, breastfeeding, height, BMI, hormone therapy, smoking status, family history, PRS	AUC: 0.70 (0.67,0.72) E/O ratio: none	1,109 cases/ 1,177 controls	None	None	None	None	None

[a]E/O ratios were calculated based on the original information. [b]The original publication reported the Observed/Expected ratio.
ER: estrogen receptor; PR: progesterone receptor; PRS: polygenic risk score; MD: mammographic density; E1S: estrone sulphate; T: testosterone; PRL: prolactin; AMH: anti-Müllerian hormone; NI: no information.

Appendix Table 4. Risk of bias assessment of included models based on PROBAST.

Study	Participants			Predictors				Outcome							Analysis									Overall
	1.1	1.2		2.1	2.2	2.3		3.1	3.2	3.3	3.4	3.5	3.6		4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9		
Gail et al [6]	N	Y	H	Y	PY	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	N	N	Y	PY	N	N	Y	H	H
Rosner et al [16]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	NI	N	N	Y	H	H
Ueda et al [17]	N	NI	H	Y	PY	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	Y	Y	PY	N	N	Y	H	H
Colditzet al [18]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	L	Y	N	Y	N	Y	N	N	Y	Y	H	H
Lee et al [19]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	Y	N	PY	N	N	Y	H	H
Tice et al [20]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	PY	Y	N	N	N	Y	H	H
Tice et al [21]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	PY	Y	N	N	N	Y	H	H
Barlow et al [22]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	PY	NI	Y	N	N	Y	Y	H	H
Decarli et al [23]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	N	N	Y	N	Y	N	Y	H	H
Decarli et al [23*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	N	NI	-	NI	Y	-	-	H	H
Novotny et al [24]	N	PY	H	Y	PN	Y	H	PY	Y	Y	Y	Y	Y	L	Y	N	N	N	Y	PY	N	N	Y	H	H
Gail et al [25]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	PY	N	N	Y	H	H
Gail et al [25*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	-	Y	Y	-	-	H	H
Anna et al [26]	NY	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	PY	N	N	Y	H	H
Tice et al [27]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	N	N	Y	Y	Y	Y	Y	H	H
Tamimi,et al [28]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	NI	N	N	Y	H	H
Petracci et al [29]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	Y	PY	N	N	Y	H	H
Petracci et al [29*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	-	Y	PN	-	-	H	H
Dite et al [30]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	N	PY	N	N	Y	H	H
Park et al [31]	N	Y	H	Y	PY	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	N	PY	N	N	Y	H	H
Park et al [31*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	NI	-	PY	Y	-	-	H	H
Anothaisintawee et al [32]	Y	Y	H	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	Y	N	PY	PN	N	Y	H	H

Anothaisintawee et al [32*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	Y	-	PY	PN	-	-	H	H
Boggs et al [33]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	NI	Y	N	Y	Y	H	H
Brentnall et al [34]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	Y	N	N	N	Y	H	H
Kerlikowske et al [35]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	N	PN	Y	Y	H	H
Tice et al [36]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	Y	PN	Y	Y	H	H
Schonberg et al [37]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	PY	N	Y	N	N	Y	H	H
Schonberg et al [37*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	-	Y	PN	-	-	H	H
Shieh et al [38]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	N	N	N	Y	Y	H	H
Wang et al [39]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	N	PY	N	Y	Y	H	H
Maas et al [40]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	PN	PN	Y	H	H
Banegas, et al [41]	N	Y	H	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	PY	N	NI	Y	H	H
Banegas et al [41*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	-	PY	Y	-	-	H	H
Eriksson et al [42]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	NI	PY	N	Y	Y	H	H
Hsieh, et al [43]	N	NI	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	NI	N	PY	N	Y	Y	H	H
Husing et al [44]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	PY	N	Y	Y	H	H
Salih et al [45]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	N	PY	N	Y	Y	H	H
Wang et al [46]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	Y	PY	PN	Y	Y	H	H
Zhang et al [47]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	N	Y	Y	H	H
Clendenen et al [48]	Y	Y	L	PN	Y	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	Y	Y	PY	N	N	Y	H	H
Wang et al [49]	N	Y	H	Y	PN	Y	H	Y	Y	Y	Y	Y	Y	L	Y	N	Y	N	Y	Y	N	N	Y	H	H
Wang et al [49*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	N	N	Y	N	-	Y	PN	-	-	H	H
Abdolell et al [50]	Y	PY	L	Y	Y	Y	L	PY	Y	Y	Y	Y	Y	L	Y	Y	Y	N	N	PY	N	N	Y	H	H
Qiu et al [51]	N	NI	H	Y	N	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	Y	PY	N	N	Y	H	H
Qiu et al [51*]	N	NI	H	Y	N	Y	H	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	NI	-	PY	N	-	-	H	H
Han et al [52]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	Y	PY	Y	Y	Y	H	H
Han et al [52*]	Y	Y	L	Y	Y	Y	L	PY	Y	Y	Y	Y	Y	L	Y	Y	Y	N	-	NI	Y	-	-	H	H

Rosner et al [53]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	Y	PY	N	N	Y	H	H
Rosner et al [53*]	Y	Y	L	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	N	-	PY	N	-	-	H	H
Yiangou et al [54]	N	Y	H	Y	Y	Y	L	Y	Y	Y	Y	Y	Y	L	Y	Y	Y	PN	N	PY	N	PN	Y	H	H

* The external validation was performed in the same study.

L: low risk of bias; H: high risk of bias; Y: yes; N: no; PY: probably yes; PN: probably no; NI: no information; -: not applicable.

1.1. Were appropriate data sources used, e.g., cohort, RCT, or nested case–control study data?

1.2. Were all inclusions and exclusions of participants appropriate?

2.1. Were predictors defined and assessed in a similar way for all participants?

2.2. Were predictor assessments made without knowledge of outcome data?

2.3. Are all predictors available at the time the model is intended to be used?

3.1. Was the outcome determined appropriately?

3.2. Was a prespecified or standard outcome definition used?

3.3. Were predictors excluded from the outcome definition?

3.4. Was the outcome defined and determined in a similar way for all participants?

3.5. Was the outcome determined without knowledge of predictor information?

3.6. Was the time interval between predictor assessment and outcome determination appropriate?

4.1. Were there a reasonable number of participants with the outcome?

4.2. Were continuous and categorical predictors handled appropriately?

4.3. Were all enrolled participants included in the analysis?

4.4. Were participants with missing data handled appropriately?

4.5. Was selection of predictors based on univariable analysis avoided? (Development studies only)

4.6. Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted for appropriately?

4.7. Were relevant model performance measures evaluated appropriately?

4.8. Were model overfitting, underfitting, and optimism in model performance accounted for? (Development studies only)

4.9. Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis? (Development studies only)

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review.	1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	2,3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	4,5
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	5,6
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	6,7
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	6
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	6 and Appendix Table 1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	7
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	7
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	8,9 and Appendix Table 2
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	8,9 and Appendix Table 2
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	7,8
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	8,9
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	8,9
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	8,9 and Appendix Table 3
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	8,9

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	8,9
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Not performed
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Not performed
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Not performed
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Not performed
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	9 and figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	9
Study characteristics	17	Cite each included study and present its characteristics.	9,10
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	12,13,14,15
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	9,10,11 and Appendix Table 3
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Appendix Table 3, table 1, figure 2, figure 3 and figure 4
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	9, 10,11,12, 13,14 and Appendix Table 4
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Not performed
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Not performed
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not performed
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Not performed
DISCUSSION			

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	14,15
	23b	Discuss any limitations of the evidence included in the review.	15,16
	23c	Discuss any limitations of the review processes used.	15,16
	23d	Discuss implications of the results for practice, policy, and future research.	16,17,18
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	3,6
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Not performed
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not performed
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	19
Competing interests	26	Declare any competing interests of review authors.	19
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Appendix Table 1,2,3

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/