A systematic review and quality assessment of individualised breast cancer risk prediction models

Javier Louro1,2,3, Margarita Posso1,2, Michele Hilton Boon4, Marta Román1,2, Laia Domingo1,2, Xavier Castells©1,2 and María Sala1,2

BACKGROUND: Individualised breast cancer risk prediction models may be key for planning risk-based screening approaches. Our aim was to conduct a systematic review and quality assessment of these models addressed to women in the general population.

METHODS: We followed the Cochrane Collaboration methods searching in Medline, EMBASE and The Cochrane Library databases up to February 2018. We included studies reporting a model to estimate the individualised risk of breast cancer in women in the general population. Study quality was assessed by two independent reviewers. Results are narratively summarised.

RESULTS: We included 24 studies out of the 2976 citations initially retrieved. Twenty studies were based on four models, the Breast Cancer Risk Assessment Tool (BCRAT), the Breast Cancer Surveillance Consortium (BCSC), the Rosner & Colditz model, and the International Breast Cancer Intervention Study (IBIS), whereas four studies addressed other original models. Four of the studies included genetic information. The quality of the studies was moderate with some limitations in the discriminative power and data inputs. A maximum AUROC value of 0.71 was reported in the study conducted in a screening context.

CONCLUSION: Individualised risk prediction models are promising tools for implementing risk-based screening policies. However, it is a challenge to recommend any of them since they need further improvement in their quality and discriminatory capacity.

British Journal of Cancer (2019) 121:76–85; https://doi.org/10.1038/s41416-019-0476-8
Intervention, Comparison, Outcomes (PICO) question of this systematic review is the following: Should individualised breast cancer risk prediction models vs. no risk prediction models be used to develop risk-based screening approaches for women in the general population?

We retrieved relevant literature by using a combination of controlled vocabulary and keyword search terms in the following databases: (i) Medline (accessed through PubMed); (ii) The Cochrane Library; and (iii) EMBASE (accessed through Ovid). Terms related to breast cancer recurrence were excluded in order to avoid retrieving citations out of the scope of this systematic review. We adapted the search algorithms to the requirements of each database and used validated filters to retrieve systematic reviews and primary studies as needed. We reviewed references of included studies that could potentially fulfil our eligibility criteria. The detailed search strategy is reported in Supplementary table 1.

We searched primary studies of individualised breast cancer risk models searching each database from its inception up to February 2018.

Study selection
Eligible studies were those published in English that reported a model to estimate the individualised risk of breast cancer in women in the general population. We included models that assessed more than one risk factor and reported the quantitative characteristics of the risk prediction model. If multiple publications were based on the same individualised risk model, the most extensive report of the model in terms of risk factors reported was chosen. We excluded external validation studies that replicated previous models without adding any additional information such as a new design for collecting the inputs data, modifications on the risk factors or the risk model method.

Articles identified from the search were loaded into EndNote X7.7.1 for Windows (2008, Version 12.0.4) and duplicates were removed.

Data extraction and quality assessment
One reviewer screened the search results based on title and abstract, and a second reviewer performed a quality check of the study screening by reviewing 20% of the references. Two reviewers independently confirmed eligibility based on the full text of the relevant articles. In case of disagreement between researchers, the inclusion of studies was determined by consensus. We reported the result of this process with a PRISMA flowchart (Fig. 1).

We used a predefined form to extract the following information from included studies: author, publication date, country, study design, the name of the model if available, sample characteristics, sample size, type of breast cancer, the method of analysis, and validation of the model. Data abstraction was conducted by one reviewer and checked by another.

Two reviewers carried out the assessment of the risk of bias independently and final quality assessment was based on consensus. We used the ISPOR-AMCP-NPC Questionnaire to assess the relevance and credibility of each risk prediction study and the following sources of limitations: (i) internal and external validation; (ii) bias due to the study design for risk estimates; (iii) limitations in data inputs; (iv) appropriateness of the model analysis; (v) reporting bias; (vi) interpretation bias; and (vii) conflict of interest. The risk of bias for each domain was rated as low, high or unclear. For systematic reviews we used the AMSTAR 2 critical appraisal tool.

Data synthesis and analysis
We evaluated the model validation by assessing both the discriminative power and the calibration accuracy estimated for the women in the general population. When available in the included publication, we extracted the area under the receiver operating characteristic curve (AUROC), the net reclassification index (NRI) and the expected observed (E/O) ratio. The NRI was not included in the tables because it was only reported in 2 out of 24 articles. The characteristics of the included models and the risk prediction outcomes reported preclude the possibility to pool data across studies. Therefore, a narrative synthesis has been conducted. Key study characteristics, validation and accuracy of individual risk models, and methodological quality are described in tables and summarised in a narrative manner. Results are presented according to the original model that they reported.

RESULTS
Study inclusion
The database searches for primary studies retrieved 2974 citations, of which 79 were considered potentially relevant. These 79 studies were screened in full text. We found a systematic review of Anothaisintawee et al.7 which we used as a source of primary studies. In addition, two studies were included after a manual inspection of papers’ references.12,13 After the full text was checked, 24 studies12–35 met the inclusion criteria and were considered in the evidence synthesis. Details about study inclusion with reasons for exclusion are described in the flowchart (Fig. 1), and a list of references to excluded studies is provided in Supplementary table 2.

Fig. 1 PRISMA flowchart
Characteristics of the included studies
The included studies can be grouped according to the risk model that they reported, the Breast Cancer Risk Assessment Tool (BCRAT), the Breast Cancer Surveillance Consortium (BCSC), the Rosner & Colditz model, the International Breast Cancer Intervention Study (IBIS), and other original models. The study by Zhang et al.\textsuperscript{13} is included in two of the groups (BCRAT and Rosner & Colditz models) because it provides information of both models and presents its results separately. A brief summary of the 24 included studies is presented in Table 1 and the extended characteristics in Supplementary table 3.

a. Breast Cancer Risk Assessment Tool ‘BCRAT’ model. This model was first published in the United States in 1989 assessing age, family history of breast cancer, age at first birth, menarche, and previous biopsies as risk factors for predicting individualised breast cancer risk.\textsuperscript{22} After this first publication, eight studies were identified that were based on BCRAT model but modified the data collection design, assessed additional risk factors or changed the statistical method. In addition to the five risk factors proposed in 1989, other variables such as body mass index (BMI), weight, hormone replacement therapy (HRT), alcohol consumption, physical activity, diet, breast density, atypical hyperplasia, breast inflammatory disease, parity, a polygenic risk score or hormones have been included in updated versions (Table 1).\textsuperscript{13,14,16,17,20,23,25,26,30}

b. Breast Cancer Surveillance Consortium ‘BCSC’ model. One relevant variation of the BCRAT model opens the path to the emergence of the BCSC model first published by Tice et al. in 2008 in the United States.\textsuperscript{31} In this study, Tice et al. used data from a cohort to create an individualised risk prediction model that combines age, family history, previous biopsies, breast density, and ethnicity. The BCSC model has been further evaluated by other authors\textsuperscript{12,24,29,32} and it currently includes previous benign breast diseases and polygenic risk score using SNPs as risk factors (Table 1).

c. Rosner & Colditz model. Parallel to the BCSC model, another model based on the ‘Nurses’ Health Study’ cohort developed by Rosner & Colditz in 1996 was also developed in the United States. This model currently includes 11 risk factors: age, menarche, menopause, age at first birth, age at subsequent births, previous benign breast disease, HRT, family history, weight, BMI, alcohol consumption, and oestriadiol levels.\textsuperscript{16,19,20,25,36} In the same way as in the BCRAT, Zhang et al.\textsuperscript{13} analysed this model adding breast density, a polygenic risk score and endogenous hormones as risk factors.

d. International Breast Cancer Intervention Study ‘IBIS’ model. The IBIS model\textsuperscript{33} includes genetic information adding the BRCA genes as a hypothetical susceptibility gene.

e. Other models. Four studies reporting different models were also identified.\textsuperscript{15,21,34,35} Apart from the above-mentioned risk factors, the models also assessed other variables such as abortion, breastfeeding, height, and previous mammography results. Particularly relevant is the Eriksson model\textsuperscript{21} since it was the only one targeted to the screening population. In this study, the authors included risk factors that were available at mammography screening examination: age, BMI, HRT, family history, menopause, breast density, and presence of microcalcifications and/or masses in the screen-mammogram.

Discriminatory accuracy
Fifteen out of the 24 studies reported the discriminatory accuracy as the AUROC (Table 1 and Fig. 2).

Calibration accuracy
Nine out of the 24 studies reported the calibration accuracy as the E/O ratio (Table 1).
| Study ID    | Targeted population | Risk factors (number of categories)                                                                 | Discriminatory accuracy (AUROC)* | Calibration (E/O ratio)* |
|------------|---------------------|----------------------------------------------------------------------------------------------------|---------------------------------|-------------------------|
| BCRAT model |                     |                                                                                                   |                                 |                         |
| Banegas 2017 | Hispanic, 25–79 years | Age (2), Menarche (3), Previous biopsies (2), Age at first birth (3), First degree breast cancer (2) | US-born: 0.56                   | US-born: 0.93b          |
| Boyle 2004  | Caucasian, 20–74 years | Age (n), Menarche (3), Age at first birth (3), First degree breast cancer (2), BMI (3), Alcohol (3), Physical activity (3), HRT (2), Diet beta-carotene/vitE (5), Diet fruits/vegetables (5) | 0.6                            | 1.03b                   |
| Chen 2006   | Caucasian, 35–74 years | Age (2), Weight (6), Breast density (5), Menarche (3), Previous biopsies (3), Age at first birth (4), First degree breast cancer (3), Atypical hyperplasia (2) | None                           | None                    |
| Decarli 2006 | Caucasian, 20–74 years | Age (2), Menarche (3), Previous biopsies (3), Age at first birth (4), First degree breast cancer (3) | 0.59                           | 0.96                    |
| Gail 1989   | African-American, 35–64 years | Age (2), Menarche (3), Previous biopsies (3), Age at first birth (4), First degree breast cancer (3) | None                           | None                    |
| Gail 2007   | Asian, 20–55 years | Age (2), Menarche (3), Previous biopsies (3), Age at first birth (4), First degree breast cancer (n), Ethnicity (6) | 0.61                           | 0.85b                   |
| Matsuno 2011 | Multiple ethnicities, 23–84 years | Age (2), Menarche (3), Previous biopsies (3), Age at first birth (4), First degree breast cancer (3), First degree family history of cancer (5), Parity (n), Breast inflammatory disease (2) | None                           | None                    |
| Novotny 2006 | Multiple ethnicities, older than 35 years | Age (2), Menarche (3), Previous biopsies (3), Age at first birth (4), First degree breast cancer (3), Breast density (4) | None                           | None                    |
| Tice 2005   | Caucasian 30–64 years | Age (2), Menarche (3), Previous biopsies (3), Age at first birth (4), First degree breast cancer (2), Polygenic Risk Score(n), Mammographic density (n), Estrone Sulphate (n), Testosterone (n), Prolactin (n) | 0.65                           | None                    |
| Breast Cancer Surveillance Consortium ‘BCSC’ model |                 |                                                                                                   |                                 |                         |
| Kerlikowske 2015 | Multiple ethnicities, 35–74 years | Age (n), Ethnicity (6), First degree breast cancer (2), Previous biopsies (2), Changes in breast density (16) | 0.64                           | 5-years: 0.98 10-years: 0.95 |
| Shieh 2016  | Multiple ethnicities. Age was not specified | Age (n), Ethnicity (6), First degree breast cancer (2), Previous biopsies (2), Breast density (4), Polygenic risk score (n), BMI (n) | 0.65                           | None                    |
| Tice 2008  | Multiple ethnicities, 35–84 years | Age (n), Ethnicity (4), First degree breast cancer (2), Previous biopsies (2), Breast density (BI-RADS) | 0.66                           | 1.03                    |
| Tice 2015  | Multiple ethnicities, 35–74 years | Age (n), Ethnicity (4), First degree breast cancer (2), Breast density (4), Benign breast disease (6) | 0.67                           | 5-years: 1.04 10-years: 1.05 |
| Vachon 2015 | Multiple ethnicities. Age was not specified | Age (n), Ethnicity (6), First degree breast cancer (2), Previous biopsies (2), Breast density (4), Polygenic risk score (n) | 0.69                           | None                    |
| Rosner & Colditz model based on the ‘Nurses’ Health Study’ |             |                                                                                                   |                                 |                         |
| Colditz 2000 | Caucasian, 30–64 years | Age (n), Menarche (n), Age at first birth (n), Menopause (n), Age at subsequent births (n), Benign breast disease (2), HRT (n), First degree breast cancer (2), Weight (n), BMI (n), Alcohol (n) | None                           | None                    |
| Colditz 2004 | Caucasian, 30–64 years | Age (n), Menarche (n), Age at first birth (n), Menopause (n), Age at subsequent births (n), Benign breast disease (2), HRT (n), First degree breast cancer (2), Weight (n), BMI (n), Alcohol (n) | ER+/PR−: 0.64                  | None                    |
| Rosner 1996 | Caucasian, 30–64 years | Age (n), Menarche (n), Age at first birth (n), Menopause (n), Age at subsequent births (n) | None                           | None                    |
| Rosner 2008 | Caucasian, 30–64 years | Age (n), Menarche (n), Age at first birth (n), Menopause (n), Age at subsequent births (n), Benign breast disease (2), HRT (n), First degree breast cancer (2), Weight (n), BMI (n), Alcohol (n), Estradiol levels (n) | 0.65                           | None                    |
| Zhang 2018  | Caucasian 30–64 years | Age (n), Menarche (n), Age at first birth (n), Menopause (n), Age at subsequent births (n), Benign breast disease (2), HRT (n), First degree breast cancer (2), Weight (n), BMI (n), Alcohol (n), Polygenic Risk Score(n), Mammographic density (n), Estrone Sulphate (n), Testosterone (n), Prolactin (n) | 0.68                           | None                    |
| Study ID     | Targeted population                  | Risk factors (number of categories)                                                                 | Discriminatory accuracy (AUROC) | Calibration (E/O ratio) |
|-------------|--------------------------------------|---------------------------------------------------------------------------------------------------|--------------------------------|------------------------|
| BCRAT model | None                                 | Age (2), Menarche (2), Family history (2), Height (2), BMI (2), Breastfeeding (2), Abortion (2)   | None                           | None                   |
| Tyrer et al. | Post menopause: 1.05, Pre menopause: 0.64 | Age (5), Menarche (2), Menopause (2), Breast density (3), Previous breast biopsy (2), Breast cancer family history (2) | 0.71                           | None                   |
| Ueda et al. | Post menopause: 1.01, Pre menopause: 0.63 | Age (12), Menarche (2), Menopause (2), First degree breast cancer (2), Breastfeeding (2), Abortion (2) | None                           | None                   |
| Tyrer et al. | Post menopause: 1.05, Pre menopause: 0.64 | Age (5), Menarche (2), Menopause (2), Previous breast biopsy (2), Breast cancer family history (2) | 0.71                           | None                   |
|           |                                      | Age (5), Menarche (2), Menopause (2), Previous breast biopsy (2), Breast cancer family history (2) | 0.71                           | None                   |

The quality of the included studies was moderate due to some limitations in the discriminative power, study design, and data inputs. The studies did not show important limitations with regards to the validation, appropriateness of the model analysis, reporting or interpretation of the results (Fig. 3). A summary of the risk of bias assessment per each source of limitation is presented here and the detailed appraisal and judgements in Supplementary Table 4.

**Internal and external validation**

Ten studies \(14-17,20,23,25,26,30,31\) validated their models by comparing the results with those published by Gail et al. \(22\) three studies \(14,29,32\) compared with Tice et al. \(31\) one \(21\) compared with both Gail et al. \(22\) and Tyrer et al. \(33\) one \(13\) compared with both Gail et al. \(22\) and the results of a Rosner & Colditz model external validation \(17\) and three studies did not report the model validation in the primary articles \(19,22,34\). Six studies assessed internal validation with a sample of the population that generated data for the model, \(15,16,24,29,31,32\) and four with an external population. \(14,20,23,32\) Despite not having reported the external validation in the primary articles, the Rosner & Colditz model \(16,19,27,28\) reported external validation in a subsequent article mentioned before. \(37\) Nine studies used the expected/observed event ratio to measure the calibration accuracy of the model. \(14,16,20,23-25,29,31\)

**Bias due to the study design**

Thirteen studies used a case-control design to obtain breast cancer risk estimates \(14,16,17,20,23-25,26,29,34\) five studies used prospective cohorts \(15,16,19,27,28\) and four models used retrospective cohorts \(24,30,32\). The study of Wang et al. \(29\) and the study of Tyrer et al. \(33\) used risk estimates obtained from a systematic review of the literature.
Limitations of data inputs
Sixteen studies obtained most of the input parameters from self-reported questionnaires. The study of Matsuno et al.25 also imputed ethnicity for women with missing data.

Appropriateness of the model analysis
Thirteen studies used logistic regression to estimate the risk of having breast cancer according to the assessed risk factors, five used proportional hazard Cox models, four used Poisson regression models, and the other two studies used risk estimates obtained from a systematic review of the literature.

Reporting bias
Twenty one studies reported all relevant and necessary information for the model creation. Conversely, a critical lack of information was found in the other three studies.

DISCUSSION
Summary of main results
This systematic review included 24 studies that aimed to estimate the individual risk of developing breast cancer in women in the general population. Twenty studies were based on four specific risk models (the BCRAT, the BCSC, the Rosner & Colditz and the IBIS model) whereas four studies used other original models. The most extensively used were the BCRAT, IBIS and the BCSC models. The number of risk factors included in the models ranged from five to 18. Other than age, which was the only risk factor present in all models, the BCRAT model also included family history, age at first birth, menarche, and previous biopsies. Breast density, benign breast disease, and polygenetic score were predominant in the BCSC model. Although during the last decade the models have shown improvements in their discriminatory accuracy, it remains at best moderate with a maximum AUROC value of 0.71 reported by Eriksson et al. The calibration accuracy was very heterogeneous ranging from 0.85 to 1.52. Furthermore, the quality of the studies was not high due to limitations in the discriminative accuracy, study design, and data inputs.

Agreements and disagreements with other reviews
In this systematic review, we found that the number of individualised breast cancer risk prediction models has increased steadily over the past three decades. This finding is in agreement with the narrative overview published by Cintolo-Gonzalez et al. in 2017, and it updates the results of a previous systematic review published by Anothaisintawee et al. in 2012. In contrast to these reviews, however, our aim was to provide innovative information regarding the quality of the identified prediction models. Thus, we have identified and rigorously analysed the strengths and limitations of 24 individualised models in order to adjust our conclusions to the quality of the evidence.

We have identified two new trends with regards to the use and development of the models, which are the increased use of the BCSC model and the inclusion of common genetic variation in the prediction models. As compared to the information published in the review of Anothaisintawee et al., we found that in contrast to the BCRAT and Rosner & Colditz models that were the most frequently cited models up to 2010 the BCSC model has concentrated the attention of several authors during the last five years, although its discriminatory accuracy has not dramatically
improved. Second, none of the models in the review of Anothaisintawee et al.\textsuperscript{7} included genetic information as a risk factor. By contrast, we have identified four models including genetic information: the IBIS model\textsuperscript{33} that includes genetic phenotype in their updated version, the BCSC model that includes a polygenetic score in both 2015\textsuperscript{12} and 2016\textsuperscript{29} publications, as well as the article by Zhang et al. that added a polygenic risk score to both the BCRAT and the Rosner & Colditz models.\textsuperscript{13}

Most of the included studies reported the AUROC to determine the probability that a randomly chosen woman with disease would be correctly categorised as higher risk compared to a randomly chosen woman without disease. The discriminatory accuracy estimate does not express whether the model is more or less accurate in predicting the risk of specific individuals but measures the capacity of the model to determine which women are at higher/lower risk for developing breast cancer. Thus, both calibration accuracy and discriminatory accuracy should be assessed. Contrary to what is expected, we found that authors reported the E/O ratio only in less than half of the included studies. In addition to the AUROC value, the studies of Zhang et al. reported the polygenic risk score to both the BCRAT and the Rosner & Colditz models.\textsuperscript{13}
and Vachon et al., also reported an improvement in the net reclassification index (NRI) of the BCRA, Rosner & Colditz models, as well as in the BCSC model, respectively.

Overall, the information provided by the AUROC and the E/O ratio was consistent suggesting that the included models have moderate discriminatory accuracy and calibration accuracy when applied to the women in the general population. Nevertheless, it must be taken into account that despite the great importance of validation in terms of AUROC and E/O ratio, the presence of low values of AUROC or clearly different from 1 values of the E/O ratio does not mean that these models are useless. On the contrary, models are clinically useful even with moderate AUROC since they can reclassify individuals at the extremes of risk. Thus, the verdict on risk models should not be based solely on these estimators. Instead, they need to be prospectively evaluated in clinical trials. In fact, there are currently two very large randomised trials assessing risk-based screening strategies. Both of them are using individualised models. Both the IBIS and the BCSC models are being tested in the European trial MyPeBS (My Personalised Breast Screening). Also, the BCSC model is being tested in the US WISDOM trial (Women Informed to Screen Depending On Measures of risk).

Applicability and completeness of evidence
The distribution of risk factors in such different populations may affect the applicability of the models to different contexts. The fact that different subtypes of breast cancer may have different genetic markers is widely accepted. These differences, the nature of breast cancer itself and its low incidence may condition a low discriminatory accuracy of a model. In other words, in the general population, there is a low probability of having breast cancer (even in the highest risk group). This low probability may mean that the discriminatory power of a breast cancer risk model won’t be as high as a risk model targeted to other common diseases such as cardiovascular events, for instance. Another potential limitation in the applicability in the screening context is the completeness and the number of included risk factors, which ranged from five to 18. Nevertheless, some potentially relevant risk factors such as genetic markers have been only included in few models. Recent studies have shown that adding genetic information as a risk factor can increase the discriminative accuracy of the different models which opens the line for further evaluation. An evaluation that should first assess the calibration of these models in prospective cohort studies.

Overall, women are usually screened using mammography. Particularly in Europe, most programmes invite women for screening every 2 years. The presence of some mammographic features in these screening mammograms may be related to the risk of developing breast cancer, as has been recently pointed out by some authors. Only one of the 24 models identified in this systematic review included microcalcifications and masses found at mammography as risk factors in the model. Time-changing variables such as radiological variables may not be as stable as personal history. However, in a screening context, this information is especially relevant because it is easily available from previous screening examinations.

Quality of the evidence
We found variability in the design of the studies that were used to obtain the cancer risk estimates. Notably, the study design used in the BCSC model was a cohort, which is a robust epidemiology design that allows developing and validating prediction models. Another frequently used design was the case-control study, nested or not. Contrary to the cohort study, time-changing variables may not be well obtained in case-control studies.

Regarding the external validation, the models showed some limitations given that few of them were further evaluated in different contexts. As far as we know, there are numerous scientific publications reporting external model validation in different settings and countries. These studies may help to understand the performance of a model in a specific context, but this issue was out of the scope of our review and, therefore, we have not included external validation studies. As an example of the relevance of these studies, we can inform that the BCRA model has more than 50 articles informing the external validation of these models in different countries. The Rosner-Colditz model has also been validated in several studies, one of the most complete validations being the one performed in 2013 by the authors themselves. On the other hand, we found that although the Eriksson et al. model reports the highest AUC (0.71), this model has not been externally validated, which increases the uncertainty about its applicability.

Also, there were limitations in data inputs, mostly due to the fact that in several models the information was provided by self-reported questionnaires that may affect the accuracy of the results. Finally, there is a limitation when comparing the AUROC or E/O ratio across the models given that there is great heterogeneity amongst them. The models were targeted to different populations, included different sets of risk factors, and often used different methodologies. We have taken into account all these variations and presented the results by model categories.

Potential biases in the review process
This systematic review was limited to studies published in English and did not involve an active search for grey literature, which is literature that is not formally published in sources such as books or journal articles. Therefore, some models may not have been identified. However, since we have conducted a comprehensive literature search in Medline, EMBASE and The Cochrane Library, we estimate that the loss of information due to the study selection criteria is low. Some key genetically oriented models, such as BOADICEA and BRACAPRO were not included in this review because they are aimed at high risk women and not useful for women in the general population in the screening context. Full-text screening and data abstraction process were performed by two researchers, which increase the quality of the review process. Moreover, as far as we know, this is the first review assessing the risk of bias of the identified risk prediction models.

CONCLUSIONS
The development of individualised breast cancer risk prediction models has increased over the last three decades, but the improvements in both the discriminatory power and calibration accuracy are still limited. Despite the time that has passed since the first model was published and a large number of available publications, only one model addressed to women attending a population-based screening programme was identified. Currently, it is still a challenge to recommend any of the models as the standard for predicting individual risk in screening context. However, the models have been updated by adding new variables, such as common genetic variation or radiologic variables and have shown improvements in their quality as well as in their discriminative accuracy. These new variables need further evaluation to confirm its promising impact in the prediction capacity to propose personalised strategies for breast cancer screening.

ACKNOWLEDGEMENTS
The authors thank Ms. Lorea Galnara-Cordero for her contributions to the design of the search strategy and the initial retrieval of the citations. The authors also thank Ms. Julieta Politi and Mr. José María Monle-Moraga for their contribution in the screening process. Javier Louro is a Ph.D. candidate at the Methodology of Biomedical Research and Public Health program, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain.
AUTHOR CONTRIBUTIONS
J.L., M.P., M.S. and X.C. designed the study, and J.L. and M.P. wrote the manuscript. J.L., M.P. and M.H.B. performed the screening, data abstraction and quality assessment of included studies. M.S., L.D. and M.R. contributed to the analyses and interpreted the data. M.R., M.H.B. and X.C. collaborated in drafting the manuscript and revising it critically for important intellectual content. All authors read and approved the final manuscript.

ADDITIONAL INFORMATION
Supplementary information is available for this paper at https://doi.org/10.1038/s41416-019-0476-8.

Competing interests: The authors declare no competing interests.

Ethics approval and consent to participate: All procedures performed in this study were in accordance with the ethical standards of the ethics committee of Parc de Salut Mar (CEIC Parc de Salut Mar) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Neither specific patient consent nor ethics committee’s approval were required because we used published articles that were obtained from open access databases.

Data availability: The datasets analysed during the current study are publicly available from the corresponding author.

Consent for publication: Not applicable.

Funding: This work was partially supported by Agència de Qualitat i Avaluació Sanitàries de Catalunya (AQuAS) and by grants from Institut de Salut Carlos III FEDER (grant numbers: PI15/00098 and PI17/00047). J.L is core funded by the Research Network on Health Services in Chronic Diseases (RD12/0001/0015). MHB is core funded by the UK Medical Research Council (funding code: MC_UU_12017/15) and the Scottish Government Chief Scientist Office (funding code: SPH5U15). None of the funders participated in the design of the study, collection, analysis, or interpretation of data, or in writing the manuscript.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES
1. The Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet. 380, 1778–1786 (2012).
2. The European Commission Initiative on Breast Cancer (ECIBC). Recommendations from European Breast Guidelines. 2016. https://ecibc.jrc.ec.europa.eu/recommendations/.
3. U.S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann. Intern. Med. 151, 716–726 (2009).
4. Oeffner, K. C., Fontham, E. T., Etzioni, R., Herzig, A., Michaelson, J. S., Shih, Y. C. et al. Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA 314, 1599–1614 (2015).
5. Mandelblatt, J. S., Stout, N. K., Schechter, C. B., van den Broek, J. J., Miglioretti, D. L., Kroenke, C. et al. Breast cancer screening for women at average risk: an independent review. Breast Cancer Res Treat. 133, 1–10 (2012).
6. Higgins, J. T. P., Green, S. (eds). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. 2011. http://handbook5-1.cochrane.org/.
7. Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M. et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst. Rev. 4, 1 (2015).
8. Deeks, J. J. & Altman, D. G. (eds). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. 2011. http://handbook5-1.cochrane.org/.
9. Stergachis, A., Miller, R., Elston, D., Hsu, Y. C. et al. Breast cancer risk assessment in the Czech female population prognostic models for breast cancer risk in Italy. Eur. J. Cancer Prev. 13, 109–115 (2004).
10. Zhang, X., Rice, M., Twarog, S. S., Rosner, B. A., Elleson, A. H., Tamimi, R. M. 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. 15, e1002664 (2018).
11. Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J. et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 358, j4008 (2017).
36. Brentnall, A. R., Harkness, E. F., Astley, S. M., Donnelly, L. S., Stavrinos, P., Sampson, S. 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.* **17**, 147 (2015).

37. Rosner, B. A., Colditz, G. A., Hankinson, S. E., Sullivan-Halley, J., Lacey, J. V. Jr. & Bernstein, L. Validation of Rosner-Colditz breast cancer incidence model using an independent data set, the California Teachers Study. *Breast Cancer Res Treat.* **142**, 187–202 (2013).

38. Cintolo-Gonzalez, J. A., Braun, D., Blackford, A. L., Mazzola, E., Acar, A., Plichta, J. K. et al. Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications. *Breast Cancer Res Treat.* **164**, 263–284 (2017).

39. Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).

40. MyPeBS. Randomized comparison of risk-stratified versus standard breast cancer screening in European women aged 40–70 (MyPeBS). 2017. http://www.brumammo.be/documents/docs/bmm-my-pebs-clinical-trial-protocol.pdf.

41. Esserman, L. J., Study, W. & Athena, I. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer* **3**, 34 (2017).

42. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA.* **98**, 10869–10874 (2001).

43. Garcia-Closas, M., Gunsoy, N. B., Chatterjee, N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *J. Natl. Cancer Inst.* 2014;106.

44. Maas, P., Barddahl, M., Joshi, A. D., Auer, P. L., Gaudet, M. M., Milne, R. L. et al. Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncol.* **2**, 1295–1302 (2016).

45. Castells, X., Tora-Rocamora, I., Posso, M., Roman, M., Vernet-Tomas, M., Rodriguez-Arana, A. et al. Risk of breast cancer in women with false-positive results according to mammographic features. *Radiology* **280**, 379–386 (2016).

46. Wang, X., Huang, Y., Li, L., Dai, H., Song, F. & Chen, K. Assessment of performance of the Gail model for predicting breast cancer risk: a systematic review and meta-analysis with trial sequential analysis. *Breast Cancer Res.* **20**, 18 (2018).

47. Antoniou, A. C., Pharoah, P. P., Smith, P. & Easton, D. F. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br. J. Cancer* **91**, 1580–1590 (2004).

48. Berry, D. A., Parmigiani, G., Sanchez, J., Schildkraut, J. & Winer, E. Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. *J. Natl. Cancer Inst.* **89**, 227–238 (1997).