The ROSA PROJECT

‘Roadmap for Optimising Screening in Australia – Breast’, investigating risk-based breast cancer screening.

Chapter 3. Risk assessment (abridged)

20 March 2023, abridged 1 May 2024

The Daffodil Centre

A partnership between

Cancer Council | The University of Sydney

Produced by the Daffodil Centre for Cancer Council Australia.

The Daffodil Centre

153 Dowling St, Woolloomooloo, NSW 2011

PO Box 572 Kings Cross NSW 1340

info@daffodilcentre.org

daffodilcentre.org

The project 'Roadmap for Optimising Screening in Australia – Breast' was funded by the Australian Government Department of Health and Aged Care (formerly the Department of Health). The project lead was Associate Professor Carolyn Nickson, with contributions from an extensive team of project personnel and various advisory and reference groups as described herein.

Suggested citation: The Daffodil Centre (2024). The ROSA Project: Roadmap for Optimising Screening in Australia – Breast. Chapter 3: Risk assessment (Abridged). 20 March 2023, abridged 1 May 2024. Produced by the Daffodil Centre on behalf of Cancer Council Australia.

The Daffodil Centre acknowledges the Traditional Custodians of Country throughout NSW and recognises the continuing connection to lands, waters, and communities. We pay our respect to Aboriginal and Torres Strait Islander cultures and to Elders past, present, and emerging.

CONTENTS

1 Executive summary

1.1 Background

1.1.1 Risk assessment

Risk assessment is central to any population-level approach to risk-based breast cancer screening. This includes assessment of breast cancer risk and specific risk factors related to screening effectiveness, such as mammographic breast density. Risk tools can produce risk estimates for individual-level advice and be used to assign women to risk groups. The size of each risk group is an important consideration for population-level risk-based screening, to enable planning of resources for risk-based screening protocols, and to help ensure relatively stable and accurate risk assessment and advice over time.

1.1.2 Levels of evidence for population screening applications

Development and validation of risk assessment tools to support population risk-based breast cancer screening is presently an active area of research. While risk assessment tools can be developed and reported using various methods (e.g. case-control studies), as recommended by the ROSA Expert Management Group in 2018, validation studies reporting observed outcomes by risk group are required to provide a level of evidence that confirms the tools are suitable for Australian population-level application.

1.1.3 Which risks?

Population level risk-based screening would require a risk assessment that can stratify the population into different levels of breast cancer risk, based on the average risk of each stratum (i.e. risk group). In addition to breast cancer risk, interval cancer rates (i.e. cancers diagnosed following a negative screening episode and prior to the next scheduled screen) are a key performance indicator for BreastScreen Australia and are known to be higher for some population groups. Interval cancers are a combination of new and missed cancers and reflect both reduced screening test sensitivity and the natural history of breast cancers in the screened population. Interval cancers may also arise through surveillance of asymptomatic women outside of the BreastScreen Australia program. Breast screening program sensitivity describes the proportion of breast cancers detected by screening rather than as interval cancers. Improving program sensitivity is expected to improve the effectiveness of screening, however, with some potential concomitant increase in overdiagnosis. Breast screening program specificity (how well screening identifies women without breast cancer) is estimated through false-positive recall rates and negative predictive values (NPVs).

1.1.4 Breast density

Higher mammographic breast density is an established risk factor for reduced breast cancer screening program sensitivity and specificity and increased rates of interval cancer, through the combined effects of reducing mammographic screening test sensitivity and increasing breast cancer risk. Breast density is, therefore, likely to be central to effective risk-based breast screening. Breast density can be routinely assessed either visually (usually by radiologists) or using automated image analysis tools.

1.2 Contracted activities

The ROSA project has undertaken a range of activities to gain insights about current health services as part of considering options for risk-based breast cancer screening in Australia. The topics covered in this chapter and the general approach/methods used is outlined in Table 1.

Table 1. Chapter 3 sections and their related ROSA project activities

Chapter section/s	ROSA activity
3.1 Risk assessment overview	Analysis and synthesis included in previous technical reports
3.2 Risk assessment tools: comparison of tools within populations	A systematic review of validated risk assessment tools compared within populations (between-tool comparisons) in terms of breast cancer risk
3.3 Risk assessment tools: comparison across populations	A scoping-level review of validated risk assessment tools compared between populations (within-tool comparisons)
3.4 Mammographic breast density and screening outcomes	A systematic review of breast cancer screening outcomes (i.e. program sensitivity, interval cancer rates and false-positive screening outcomes) according to mammographic breast density
3.5 Risk assessment for the Australian screening population	An epidemiological comparison of an established questionnaire-based risk assessment tool (the Gail model) and risk assessment using information routinely collected by BreastScreen with the addition of breast density on a cohort of Australian screened women

1.3 Summary of findings

Drawing from the detailed analyses and results described throughout this chapter, the project generated an itemised set of key findings which were reviewed by the ROSA Expert Advisory Group over May to July 2022, accompanied by summaries of the evidence outlined here. The final set of EAG-endorsed key findings is shown in Appendix 6.1 (page 91).

In summary, for breast screening populations, some risk assessment tools can identify groups of women at higher or lower risk, depending on the study setting and population. We found that no tool fitted well in multiple studies, that tool precision appears to depend on the population and setting, and tools that involve calibration to the target population prior to validation (with reference to e.g. the target population profile and cancer incidence) can show a better fit to observed outcomes. Risk assessment tools can also vary between versions, e.g. we found that BCRAT version 3 consistent in distinguishing women in the lowest risk group whereas the same was not observed for versions 2 and 1. This aligns with improvements to tools with updated versions. The risk factors specified as inputs for the different tools varied considerably. For example, race or ethnicity was included as a risk factor in BCRAT and its variant AABCS as well as the Tyrer-Cuzick v7 and v8 tools, however, not all studies collected or had complete information on this factor. This is an important consideration for policy makers and health services when selecting the most suitable tool for a specific application in terms of resources and costs required for data collection.

Mammographic breast density did not improve the accuracy of breast cancer risk assessment tools based on self-reported information usually including family history and prior breast biopsies. We did not review evidence on the accuracy of breast density alone as a risk assessment tool, with an

equivalent assessment of whether other risk factors improve the accuracy of risk assessment when added to breast density. This is a very active research area, and ongoing review of high-quality evidence is warranted. Regardless, breast density remains an important tool for assessing risk of interval cancer rates, program sensitivity, and false positive rates in population breast screening programs.

Polygenic risk scores did not improve the accuracy of tools based on self-reported information usually including family history and prior breast biopsies. As for breast density, we did not review evidence on the accuracy of polygenic risk scores alone as a risk assessment tool. Of note, risk assessment incorporating genetic test results may have ethico-legal consequences for individual women that should be well understood as part of considering genetic testing as a routine part of population-level risk assessment to support risk-based breast cancer screening and surveillance.

As demonstrated by this analysis, there are several metrics to consider in evaluating and comparing risk assessment tools aiming to predict breast cancer risk, depending in part on how the risk assessment tool might be used. For example, tools demonstrating a good fit based on expected over observed ratios are suitable for informing women about their estimated individual breast cancer risk (noting that this is always a group-level average risk); however, tools that are effective in assigning women to risk groups based on observed breast cancer incidence rates, irrespective of fit, are potentially suitable for triaging women to risk-based interventions. Alternative study designs may be required to assess risk tool prediction for invasive breast cancer incidence according to prognostic indicators (e.g., tumour subtype, grade, size, nodal) and for prediction of *in situ* cancers.

No Australian studies comparing tools on a single cohort were included in our analysis. However, we note a recent study (Li 2021) that validated and compared six risk assessment tools on a cohort of 7,608 Australian women within the target age range for screening (50-65 years), finding only one model (BOADICEA) calibrated well across the spectrum of 15-year breast cancer risk (p-value < 0.03). This study did not incorporate breast density information in the risk assessment tools.

Our analysis of a simplified risk approach to risk assessment on a large cohort of BreastScreen Victoria participants (Section 5, starting page 76) finds that, for women aged 50-69 attending subsequent round screening, combinations of family history and breast density may be comparable to the established US National Institutes of Health Breast Cancer Risk Assessment Tool in terms of estimating risk of future invasive breast cancer, screen-detected invasive breast cancer or interval cancer. Larger studies are required to verify this finding; the current analysis indicates that more simplified approaches to risk assessment should be included in consideration of options for risk-based breast screening in Australia, mindful of the resources and imposts involved in undertaking detailed risk assessment, and stakeholder interest in informing women about their breast density.

We found that breast density alone can stratify breast screening populations into groups according to interval cancer rates, program sensitivity, and false positive rates, although the accuracy of this risk stratification varies between studies. In addition, there appears to be a general trade-off between accurate discrimination of either the lowest breast density group or the highest breast density group; no measurement tools perform well at both ends of the spectrum.

Drawing from this summary of evidence, our key findings to support recommendations in relation to risk assessment tools are shown above. Combined with findings from other ROSA activities they collectively help support project recommendations for actions that will help progress the roadmap to more risk-based screening in Australia.

1.4 Glossary of terms

Calibration	As used in this report, describes the agreement between predictions from a risk assessment tool and observed outcomes.
Community-detected cancer	Cancer diagnosed outside the screening program, including interval cancers.
DBT	Digital breast tomosynthesis.
DCIS	Ductal carcinoma in situ.
Discrimination	As used in this report, refers to how well a risk assessment tool differentiates those at higher risk of having an event from those at lower risk.
External validation	As used in this report, refers to studies that aim to assess the predictive performance of existing risk assessment tools using data external to the development sample (i.e. using data from different participants).
False positive screen	A screening episode recalled for further assessment with a benign final outcome after assessment.
Family history of breast cancer	Some family history of breast cancer. Can be defined in various ways.
FCC	Family Cancer Centre or Family Cancer Clinic.
Higher-risk groups	As used in this report, groups of women estimated to be at higher risk of breast cancer. The definition and size of this group depends on the risk assessment tool and/or guidelines used.
Internal validation	As used in this report, using the same population sample to develop and validated a risk assessment tool.
Interval cancer	Cancer diagnosed following a negative screening episode, within a defined period of the screen (usually 12 or 24 months)
LYG	Life-years gained.
LYS	Life-years saved.
Mode of detection	Categorical description of how cancers were diagnosed e.g. screen-detected, interval cancer or other.
Negative screening episode	A screening round not recalled for further assessment.
Overdiagnosis	Cancers detected by screening that would not have otherwise been found in a woman's lifetime.
PICO/PECO framework	A framework to define an approach to a research question in terms of the population of interest (P), the intervention (I) or exposure (E) being assessed, the comparator intervention or exposure (C), and the outcomes to be reported and assessed (O).

Positive predictive value (PPV)	The proportion of recalled screens that result in a screen-detected cancer. Can report either invasive breast cancers or invasive breast cancers combined with DCIS diagnoses.
Program sensitivity	The proportion of cancers diagnosed by screening rather than as interval cancers. Can be reported for a period and/or a cohort.
Prospective study design	A study that follows outcomes subsequent to a specific intervention or exposure. Most often applied to prospective cohort studies, where outcomes in a cohort are followed over time.
QALY	Quality-adjusted life year. A composite measure of quality of life and quantity of life; QALYs are the number of life years saved adjusted for any reduction in quality of life (including morbidity), such as a temporary decrease after receiving a false positive screening result, or a prolonged decrease due to a breast cancer diagnosis
QALYS	Quality-adjusted life-year saved.
Recall to assessment	Recall for further investigation by BreastScreen assessment services, following a screening mammogram.
Risk assessment tool	As used in this report, a tool for estimating the risk of breast cancer in the future, sometimes to specific cancer types (e.g. invasive breast cancers) or modes of detection (e.g. interval cancers).
Risk categories	Ranges of estimates of risk for a future event as predicted by a risk assessment tool.
Risk predictor	As used in this report, a risk factor included in a risk prediction tool such as age, height, body mass index, mammographic density, etc
Screen-detected cancer	Cancer detected by a population screening program
SES	Socioeconomic status
Strong family history of breast cancer	A strong family history of breast cancer, defined in various ways, often according to whether the family member/s with breast cancer are/were first- or second-degree relatives, and/or the age at which their breast cancer was diagnosed (so that diagnosis at a younger age is more likely to be interpreted as a strong family history).

2 Risk assessment overview

2.1 Risk assessment and stratification

2.1.1 Risk stratification

Risk stratification is a broad concept that requires assumptions or specifications about:

- **which risks** are estimated
- **how** risks are estimated
- **how often** risks are assessed
- how risk is classified according to **risk groups**.

2.1.2 Risk estimation

Various tools are available to estimate the risks of developing invasive breast cancer and/or interval cancers. All validated tools (except for tools identifying high-risk genetic mutations) involve some self-reported information, and tools that estimate risk of interval cancers incorporate mammographic breast density.

Risk assessment tools were initially summarised in the earlier ROSA '*Summaries of Evidence*' report in August 2019 and then updated as a systematic review in the final report. We found that various tools are capable of providing good high-level estimates that are validated in terms of estimated and observed outcomes at a risk-group level. This is a very active area of research and development, with increasing use of machine learning and artificial intelligence methods to help maximise the information available at the time of risk assessment and in clinical records (including prior mammograms); these methods will not produce a panacea for the question of risk-based screening but may improve the accuracy of risk estimation.

Mammographic breast density is assessed visually or using automated methods. While some qualitative information is expected to be missed by image processing algorithms, automated methods do not suffer from low inter- and intra-reader reliability, they are fully repeatable, and increasingly well-validated on large cohorts. On this basis we assume that risk-based screening in Australia that incorporates breast density assessment would use automated imaging processing methods to produce measurements.

In terms of how risk is enumerated in the context of risk-based breast cancer screening, risk can be based on:

- future risk of invasive breast cancer (e.g. 5-year, 10-year risk, or lifetime risk, usually expressed as a probability e.g. a 5% risk of occurrence)
- future risk of an interval cancer (e.g. in the next 12 months, or before the next scheduled screen (determined in part by the time to the next screen))
- current risk of invasive breast cancer (i.e. to help determine the best technologies to use at the time of risk assessment); or
- a combination of the three above.

2.1.3 Projected versus short-term risks

Based on current risk tools, women could be *ranked* against the population in the same way for any projected period (e.g. 5-year, 10-year or lifetime), however their shorter-term risk will depend in part

on their age at the time of their risk estimation, so if specific risk values are important (e.g. as thresholds for risk groups), shorter-term risk estimates may be more appropriate.

For age-targeted programmatic screening such as BreastScreen, shorter-term risk may be most valuable for managing resources in the shorter-term, but longer-term risk is also relevant for planning life-course screening participation, as some proportion of cancers arising in women aged 75+ may be detectable by screening but not be symptomatic while women are within the target age range for screening (50-74 years).

2.1.4 Risk of invasive breast cancer

Ideally, risk of invasive breast cancer would include estimated risk of particularly aggressive or difficult-to-treat breast cancers. At this stage, based on available tools, we assume risk estimation would be for any type of invasive breast cancer.

2.1.5 Risk of an interval cancer

Some trials and modelled evaluations of risk-based screening focus on reducing interval cancers. Interval cancers are a combination of missed and new cancers and arise through a combination of reduced sensitivity at the screening test (which can be due to both breast density and the appearance of the tumour), rate of tumour progression, and whether women seek investigation of symptoms arising after a negative screening result. Some interval cancers are likely to arise due to women seeking extra surveillance tests between BreastScreen screens, and this could affect the accuracy of tools aiming to estimate risk of interval cancer based on the likelihood of missed or aggressive tumours.

2.1.6 Risk of DCIS

Breast imaging of asymptomatic women may also lead to the diagnosis of breast disease other than invasive breast cancer, with identification of ductal carcinoma in situ (DCIS) being a significant outcome of the BreastScreen program. As DCIS can be a precursor to invasive breast cancer, all DCIS is currently treated in Australian women. Trials and studies are underway to investigate potential surveillance of low-grade, small DCIS. Estimated risk of invasive breast cancer will correlate with estimated risk of DCIS, to some degree.

2.1.7 Current practices within and outside BreastScreen

BreastScreen services currently use risk factors such as family history of breast cancer, personal history of breast cancer and, in some jurisdictions, known high-risk gene mutations, to broadly identify a group of women at higher risk of invasive breast cancer. Women at higher risk are offered annual screening or, in some cases, are referred to higher-risk management clinics such as Family Cancer Clinics (FCCs). Annual screening will logically reduce interval cancers in any risk group by reducing the opportunity for interval cancers to develop that might arise in the second year between biennial screens.

Outside BreastScreen, primary care and specialised clinics providing risk-based surveillance services usually focus risk assessments on risk of invasive breast cancer, although alternate specific breast imaging tests may be selected according to a woman's known or estimated breast density (e.g.. ultrasound for younger women, or supplemental ultrasound or MRI once breast density is observed through a primary test such as mammography or DBT). This is effectively based on an estimate of the risk of a false negative result at that imaging test, aiming to reduce that risk. Although this test may occur without the next surveillance test scheduled, this may be broadly considered equivalent to estimating risk of an interval cancer in the BreastScreen program.

2.1.8 Current practices

BreastScreen does not estimate risk numerically, but women flagged for annual screening are broadly considered to be at higher risk of invasive breast cancer. GPs, specialist clinics and FCCs use tools that generate numerical risk estimates, although it is not clear how this information is recorded or communicated to women across different services. There is a concerning number of differing guidelines in use, so that women may be provided with different risk information depending on where they live and who/which provider they see.

2.1.9 Frequency of risk assessment

For BreastScreen clients, risk could potentially be assessed once on entry to the program, then at every screen, or at another option in between (e.g. every 3rd screen, or approximately every 6 years, or after specific birthdays e.g. 50, 60 and 70 years). The feasibility of each approach will depend on how onerous the risk assessment is for both clients and the program. The value of each approach will depend on how accurate the risk assessment is, whether the risk assessment includes short-term outcomes (e.g. risk of an interval cancer prior to the next screen). It would also be important to consider and manage expectations among BreastScreen clients, for example an expectation that risk-based screening should use the most up-to-date information, or alternatively that screening should be simple and easy.

Decisions would also be required about the threshold for changing an individual woman's risk classification. Otherwise, women with an estimated risk near the threshold between two categories may oscillate between two risk categories from screen to screen due to slight changes in their risk factors, slight changes in the risk tool used, or slight changes in the selected thresholds for different risk categories. This is likely to require a policy overlay involving agreed strategies for managing repeated changes in risk classification (e.g., a requirement that women would only be upgraded or downgraded in terms of risk group allocation after two consecutive assessments indicating a change is required).

BreastScreen risk assessment is currently done at each screening episode, with data collected through questionnaires completed by the client with or without assistance from BreastScreen personnel. These questions asked vary between jurisdictions.

Outside BreastScreen, current practices in primary care are unclear. Medicare Benefits Schedule items are available for GPs to request imaging for women with specific risk factors (based on personal or family history of breast cancer, known high-risk gene mutations and age), but the trigger for GPs receiving this information and then requesting appropriate services is unclear. This is presumably a clinical decision by each GP, sometimes in response to a patient request. We have not identified guidelines that specify or recommend how often risk should be assessed.

2.1.10 Risk classification

There are various options for mapping individual risk estimates to risk groups (risk classification). For example:

1. Relative to the population average (e.g. 1.5 times higher than average)
2. Based on specific risk thresholds (e.g. 20% lifetime risk)
3. Based on *a priori* group size allocation (e.g. in the top 10% (decile) of estimated risk values); or
4. Based on specific risk factors such as family or personal history of breast cancer.

The various guidelines and policies in place for the assessment and management of breast cancer risk in asymptomatic women use three of these four options, as follows:

Option 1 (risk relative to the population average) is used by clinical guidelines such as the Cancer Australia 'Advice about familial aspects of breast and epithelial ovarian cancer'[1], the RACGP 'Guidelines for preventive activities in general practice' [2] and the RACGP National guide to a preventive health assessment for Aboriginal and Torres Strait Islander people [3] to identify and manage women at higher risk of breast cancer. This is the approach used in the iPrevent tool[4] and the UK PROCAS study[5]. This approach is appropriate given these services draw on 'open-ended' resources such as Medicare Benefits Schedule items for medical imaging, although it is important to maintain the evidence base for how these estimates are made, and to ensure they are calibrated to the Australian population, accounting for changes in breast cancer risk between birth cohorts and potential changes in breast cancer incidence through prevention strategies [26].

Option 2 (based on specific risk thresholds) is the approach described in the eviQ guidelines used by FCCs[6] to identify and manage women at higher risk of breast cancer; it is similar to Option 1, except that the thresholds use different values and risk is described differently.

Option 3 (*a priori* group size allocation) is not currently used systematically in Australia, to our knowledge.

Option 4 (risk based on specific risk factors such as family or personal history of breast cancer) is the general approach currently used by BreastScreen to allocate women to annual screening, with different jurisdictions variously considering strong family history of breast cancer, personal history of breast cancer, history of ovarian cancer, personal history of LCIS, ADH and/or ALH[7] and known high-risk gene mutations. This was described in detail in the ROSA Clinical Services report in August 2019 (see Appendix, Chapter 2).

2.1.11 Other considerations

Risk classification has been a considered in detail in the development of a protocol for routine risk assessment and advice in the BreastScreen Victoria program (the BRAVO project). The BRAVO protocol includes specifying, for the first time in Australia, a group of women at lower-than-average risk of breast cancer. To develop these categories, we specified a new category of women at less than half the population average risk, and otherwise followed the categories used in RACGP guidelines (see footnote 2). We then described these categories in terms of Option 1 and Option 2 as described above, based on outcomes in the lifepool cohort[8] (**Figure 1**). This is the general working model for the BRAVO protocol intended to be piloted soon, however the categories are subject to revision after incorporation of further population-level cancer data and modelling of the various options via the Policy1-Breast model.

---

1 Cancer Australia. Advice about familial aspects of breast cancer and epithelial ovarian cancer: A guide for health professionals. Cancer Australia. 2015. Third edition.; Retrieved from https://canceraustralia.gov.au/system/tdf/publications/advice-about-familial-aspects-breast-cancer-and-epithelial-ovarian-cancer/pdf/2015_bog_familial_aspects_int.pdf?file=1&type=node&id=2878.

2 Royal Australian College of General Practitioners. Guidelines for preventive activities in general practice. RACGP. 2018. Retrieved from https://www.racgp.org.au/FSDEDEV/media/documents/Clinical%20Resources/Guidelines/Red%20Book/Guidelines-for-preventive-activities-in-general-practice.pdf.

3 National Aboriginal Community Controlled Health Organisation & The Royal Australian College of General Practitioners. National guide to a preventive health assessment for Aboriginal and Torres Strait Islander people. RACGP. 2018. Retrieved from https://www.racgp.org.au/FSDEDEV/media/documents/Clinical%20Resources/Resources/National-guide-3rd-ed-Sept-2018-web.pdf.

4 Breast cancer trials. iPrevent. [Internet]. Retrieved 11/02/19, from https://www.breastcancertrials.org.au/iprevent

5 Evans DG, Astley S, Stavrinos P, et al. Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and family history clinics: a dual cohort study. Southampton (UK): NIHR Journals Library; 2016 Aug. (Programme Grants for Applied Research, No. 4.11.) Chapter 4, PROCAS: Predicting Risk of Breast Cancer at Screening. Available from: https://www.ncbi.nlm.nih.gov/books/NBK379493/

6 Cancer Institute NSW. Referral guidelines for breast cancer risk assessment and consideration of genetic testing. [Internet]. 2015. Retrieved 10/12/18, from https://www.eviq.org.au/cancer-genetics/refe rral-guidelines/1620-referral-guidelines-for-breast-cancer-risk-as.

7 LCIS: lobular carcinoma in situ; ALH: atypical lobular hyperplasia; ADH: atypical ductal hyperplasia,

8 Lifepool.org.au

		Lower than average **<0.5 x population risk**	***Average*** **0.5 to 1.5 x population risk**	***Moderately increased*** **1.5 to 3 x population risk**	***Potentially high risk*** **>3 x pop risk**
50-59	Population average ten-year risk	29 cancers/10,000 person-years x 10 years = 2.9%			
	Ten-year risk range	<1.5%	1.5% - 4.4%	4.4% - 8.8%	>8.8%
60-69	Population average ten-year risk	38 cancers/10,000 person-years x 10 years = 3.8%			
	Ten-year risk range	<1.9%	1.9% - 5.7%	5.7% - 14.4%	>14.4%

Figure 1. *A template for risk categorisation used in the development of the BRAVO protocol.*

2.1.12 Stakeholder perspectives on risk assessment and advice

The Australian Government Department of Health commissioned the report 'A literature review, stocktake and stakeholder insights about Australian women's attitudes to participating in population-based breast screening' (Allen and Clarke 2020).[9] This report followed an earlier report on breast density[10] commissioned by the Department of Health, which indicated likely complexities in communicating clinical information about breast density.

The 2020 report aimed to explore how women make informed decisions about participating in breast screening, to undertake a stocktake of materials that Australian women might use to inform themselves of the benefits and risks of participating in population-based breast screening and gather related insights from stakeholders from each BreastScreen jurisdiction (all but ACT) through semi-structured interviews. The report addressed the following research questions:

- What do women know and understand about the risks of breast cancer, and the benefits, risks and limitations of participating in breast screening?
- What are women's attitudes towards and perceptions of the risks of breast cancer, and the benefits, risks and limitations of participating in breast screening?
- How would women like to be informed about their risks of breast cancer, and the benefits, risks and limitations of participating in breast screening?
- What is the relationship between women's understanding, attitude and perceptions of the risks of breast cancer, and the benefits, risks and limitations of participating in breast screening and their participation in population-based breast screening?

The literature review of 46 peer-reviewed articles returned limited information about Australian women's knowledge about breast cancer risk factors (in general) nor their knowledge about specific risk factors (including age, family history, breast density, etc.). The review found that the evidence base on Australian women's understanding of breast cancer risk/protective factors and breast

[9] Allen and Clarke. A literature review, stocktake and stakeholder insights about Australian women's attitudes to participating in population based breast screening. 18 June 2020. Downloaded from www.health.gov.au/resources/publications/breastscreen-australia-understanding-informed-decision-making-a-literature-review-about-australian-womens-attitudes-to-participating-in-population-based-breast-screening

10 Breast Density A literature review to inform BreastScreen Australia's position statement on breast density and screening Final report: 28 September 2018

screening is limited, however there was an alignment between the findings from qualitative research and feedback provided to the project team during semi-structured interviews.

Key conclusions relevant to ROSA from the report are summarised below:

Current understanding about women's preferences

The report concluded that little is known about the range of information Australian women want to know or how they want to be informed about breast cancer risks and population-based breast screening.

Detail on risk information

Women indicated that they wanted to know about a range of risks in addition to breast density, especially age as a risk factor and how that may change over time. There was some confusion surrounding the importance of breast density as a risk factor, whether women should know about it and what they should do about it if they are told.

The report concluded that '*We do know that providing education to women on breast cancer, risk factors and breast screening can increase knowledge, reduce misperceptions about cancer and increase participation in population-based breast screening. We also learnt that many women want full, balanced information on screening, including issues associated with overdiagnosis/over-treatment but other women may be more concerned that changing the 'pro' screening message to include more balanced information could result in confusion.'*

Risk factor information

The need for visual imagery was mentioned several times throughout the report. Some interview participants thought that complex information about risk and statistics is best depicted in pictures (rather than words); they also went on to note some key statistics that they thought women knew but perhaps did not understand completely. There were also differing views regarding preferences for the use of concepts like "nine out of ten" and "most".

The report concluded that '*Key messages could potentially focus on describing:*

- *mammography as a test and how the procedure will be implemented*
- *breast cancer incidence and that breast screening saves lives through early detection and wider treatment options/choices*
- *risk factors and what is known (especially regarding increasing age and why screening is most appropriate for women aged over 50 years)*
- *including information about overdiagnosis/over-treatment but also acknowledging what we do not know (i.e., that screening finds some cancers that would never cause harm but we do not know which ones, so we treat everything and we are working to better identify non-harmful lesions), and*
- *present statistical data in icon arrays.*

Effective communication formats when advising women on risk

Interview participants offered a wealth of information about different approaches that they had found to be effective (or not). Key principles identified in the report included:

- *Ensuring simple, short evidence-based information is easily available, in the places where women are looking and provide that information multiple times, in different ways*
- *Asking women about the information they want and how they would like to receive it: never make assumptions*

- *Recognising that different motivators will resonate with different women (even within one cohort such as a cultural group or age-band), and*
- *Communicating with care, dignity and concern at all points of the screening engagement.*

Communication about risk with screening participants

This report identified a number of opportunities for discussing risk with women. This includes during the screening invitation, at pre-screening, and during and after the examination, but the screening appointment was considered the most effective opportunity for transmitting information, noting that it is time-limited (and some women need more time to receive and understand information provided).

From their analysis the report proposed:

- *Communicating with screeners between appointments to discuss advances in breast imaging or other program changes*
- *Using multiple touchpoints with a strong digital presence (online information, testimonials, letters, texts, emails, social media, opportunistic reach-outs, etc.) across a long timespan (although interview participants were keen to ensure that women were not put-off from the program by spamming)*
- *Using simple, plain English language (in resources and clinical interactions) that is evidence-based: information should also be transparent to support informed decision making but not delivered in a patronising way*
- *Creating messages that are built around hope, rather than fear (e.g, do it for your family, by screening I am hoping that if I have cancer it will be detected early and I will survive) and which take a well-woman's approach to managing health*
- *Providing for a low health literacy without making assumptions about what women do or do not know*
- *Being aware of the state/territory legislative context in which screening programs operate as this can restrict what information can be shared, with whom and when*
- *Considering sharing information with local politicians about screening rates in their areas (a different kind of local champion), and*
- *Starting communications early and continuing through a woman's screening life-long journey: all information, particularly the limitations of screening, should be provided at the outset when women provide informed consent for a mammogram but that this should be repeated at all subsequent engagements as well.*

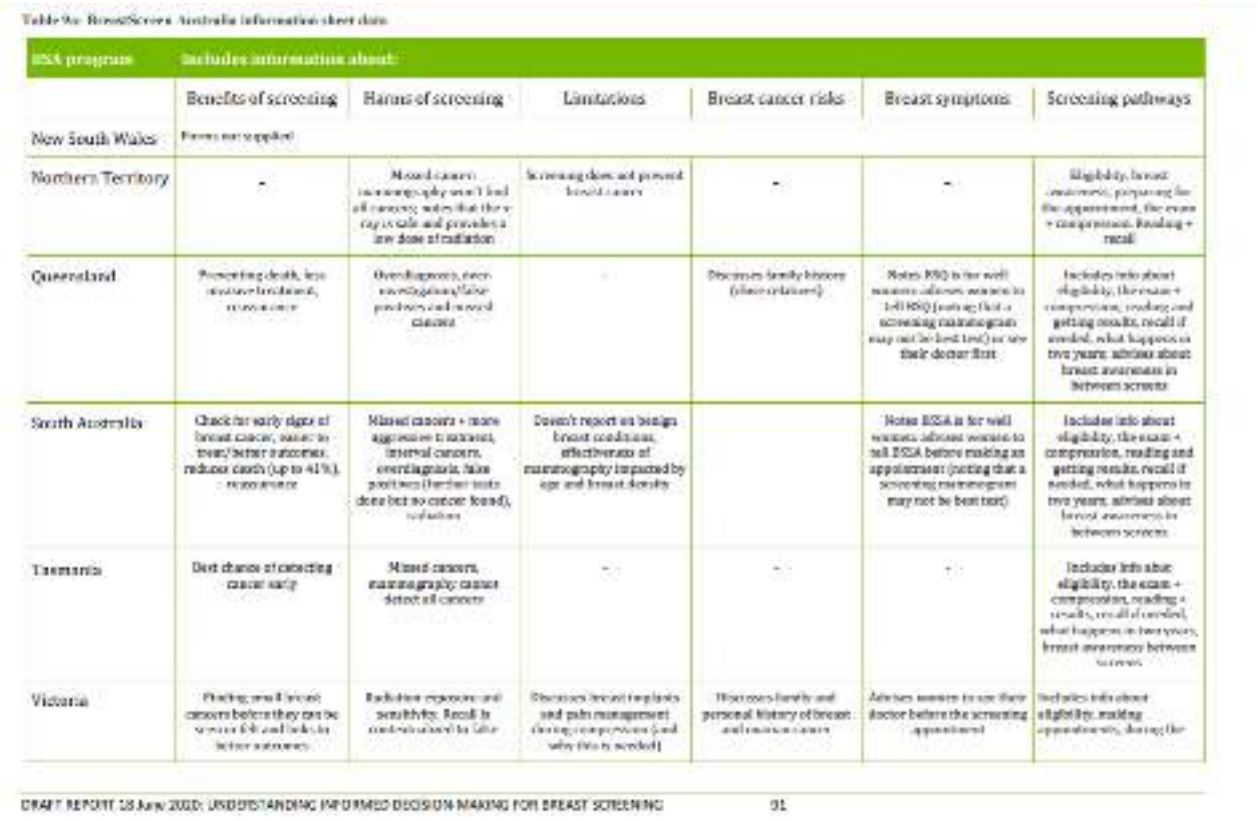
Health professional knowledge & agreement of risk

Stakeholder interviews highlighted the issue of health professional knowledge of screening. Some GPs send symptomatic women to BreastScreen or don't understand why there isn't a diagnostic report from BreastScreen.

> *"An area consistently identified as challenging by stakeholders was communicating complex clinical information when the science may not provide settled evidence on the direction to take or when there is no clear consensus on what to do. This is problematic for both women who are deciding on whether to participate in breast screening as well as clinical staff who are providing advice." (p 116)*

The report also noted that 'trusted health advisors (including GPs) also probably require further information about the benefits, harms and limitations of breast-screening in order for them to support women's informed decision-making'. The report summarised BreastScreen Australia

information sheet data (Figure 2 taken from Table 9a of report) that highlighted the variability of information between jurisdictions in terms of the information provided to women.



BSA program	Includes information about:					
	Benefits of screening	Harms of screening	Limitations	Breast cancer risks	Breast symptoms	Screening pathways
New South Wales	[illegible]					
Northern Territory	-	[illegible]	[illegible]	-	-	[illegible]
Queensland	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]
South Australia	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]
Tasmania	[illegible]	[illegible]	-	-	-	[illegible]
Victoria	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

DRAFT REPORT 18 June 2020: UNDERSTANDING INFORMED DECISION-MAKING FOR BREAST SCREENING 91

Figure 2. BreastScreen Australia information sheet data by jurisdiction, as summarised in the Allen and Clarke report.

Gap analysis

Some interview participants noted that it would be useful to have some consistent language regarding screening and that this could be developed through a program-wide style guide. Related to this, some interview participants said that some jurisdictions do not have the funding available to develop resources. They suggested that a better approach could be to develop resources at the federal level, with the opportunity to co-brand for each state/territory. However, participants also noted that more resources will not necessarily make a big difference to non-attenders and groups who are traditionally under-screeners.

Implications for ROSA

The 2020 report[11] provides many useful insights about how to communicate information related to risk-based screening to different groups of women. The findings indicate that education on risks is of value, that visual imagery is a helpful tool for communicating risks, that balanced information on benefits and harms would be welcomed, and that different approaches will work best for different women. While the screening test is a clear opportunity for providing risk information, other opportunities can also be used, mindful that the information should be clear and accessible to all

[11] Allen and Clarke. A literature review, stocktake and stakeholder insights about Australian women's attitudes to participating in population based breast screening. 18 June 2020. Downloaded from www.health.gov.au/resources/publications/breastscreen-australia-understanding-informed-decision-making-a-literature-review-about-australian-womens-attitudes-to-participating-in-population-based-breast-screening

women. Health professionals, and potentially public figures, can help convey information at many different touchpoints, and national coordination could markedly improve the consistency and efficiency of delivery of risk information.

These insights are highly relevant for the ROSA project. For example, for screening scenarios being compared using modelled evaluations, scenarios where risk is more likely to be explained to a majority of women should be recognised as holding some higher value. Nationally coordinated and resourced efforts to develop standard risk information for all Australian women would be of value to support any changes towards more risk-based screening.

3 Breast cancer risk assessment tools

3.1 Background

To understand the potential for risk-based screening strategies based on breast cancer risk estimates from multivariable risk assessment tools, we first need to understand how well these tools perform.

The development and validation of risk assessment tools and the combination of different risk measures (such as health questionnaires, mammographic density and genetic information) into single risk tools are very active areas of research.

An initial scoping level review was undertaken in August 2019, to assess how accurately risk assessment tools, based on questionnaire data and/or mammographic density and/or genetic information, could stratify women into groups according to their risk of breast cancer. The review was restricted to external validation cohort studies and the key measure of interest was expected versus observed (E/O) rates of cancers.

The scoping review highlighted the diversity of risk assessment tools reported in the literature and a mix of approaches used to validate those tools. Being a scoping review, the literature searches were not comprehensive and potential sources of bias were not assessed. Thus, it is possible that some studies were not identified, and the reliability of the evidence was uncertain. These issues can be addressed by performing a systematic rather than scoping review, as systematic reviews ensure that all the available evidence is identified and an objective assessment of the certainty of the evidence undertaken. This is a resource-intensive task, however given the potential importance of risk assessment tools in risk-based screening, the initial scoping review was subsequently extended into a systematic review.

Two key questions of interest were:

(i) How do different breast cancer risk assessment tools compare to one another in a single setting?
(ii) How does a given breast cancer risk assessment tool perform in various settings?

3.2 Systematic review

For the first question, we conducted a comprehensive systematic review (Velentzis, L.S.; Freeman, V.; Campbell, D.; Hughes, S.; Luo, Q.; Steinberg, J.; Egger, S.; Mann, G.B.; Nickson, C. Breast Cancer Risk Assessment Tools for Stratifying Women into Risk Groups: A Systematic Review. Cancers 2023, 15, 1124. https://doi.org/10.3390/cancers15041124. This publication, in the journal Cancers (impact factor 6.575) is attached.

3.3 Scoping review

For the second question, we conducted a scoping review as reported below.

3.3.1 Authors

Victoria Freeman, Dr Denise Campbell, Suzanne Hughes, Dr Julia Steinberg, Dr Louiza Velentzis & A/Prof Carolyn Nickson.

3.3.2 Background and aims

As outlined in the previous section, to understand the potential for risk-based screening strategies based on breast cancer risk estimates from multivariable risk assessment tools, we first need to understand how well these tools perform in various settings.

Two key topics of interest are: (i) how different breast cancer risk assessment tools compare to one another in a single setting and (ii) how a given breast cancer risk assessment tool performs in various settings. For the first topic we conducted a comprehensive systematic review, reported in Section 3. This review required significant methodological detail due to the complex differences between studies such as how the risk tools had been applied and the screening and population settings in which the studies were conducted.

In this report we describe our work addressing the second question, for which we directed remaining project resources at a scoping review for all risk assessment tools based on questionnaire data either with or without the inclusion of mammographic density and/or genetic information, and a systematic review confined to a single but commonly evaluated risk assessment tool (i.e. the Tyrer-Cuzick tool). This systematic review is of interest in its own right and also provides a demonstration of how other tools could be fully evaluated across different settings.

The research question for this scoping review was as follows: For asymptomatic women, how does a given breast cancer risk assessment tool based on questionnaires, genetic information and/or mammographic density perform in predicting breast cancer risk across the risk groups determined by the tool (i.e. within tool comparisons)?

3.3.1 Methods

The methods for this review are included in systematic review protocol (CRD42020159232) submitted to PROSPERO on the 28th February 2020, and subsequently updated on 8th September 2021 (see Appendix 3.4.1 (page 56) for full protocol).

A scoping review of studies summarising evidence on published tools was first conducted in 2019. Subsequently, from the systematic review conducted in 2021, we focused on the Tyrer-Cuzick risk assessment tool (referred to hereafter as the Tyrer-Cuzick tool), selected due to the availability of studies for review and comparisons. The PICO protocol for the scoping review and systematic review are shown in Table 2.

Table 2. PICO: Within tool comparisons – compares different risk categories of a given breast cancer risk assessment tool between different populations.

Population	*Intervention/exposure*	*Comparison*	*Outcomes*	*Study design*
Asymptomatic women aged ≥ 18 years of age undergoing mammogram screening	Scoping review: Breast cancer risk assessment tool, where a risk category is ascertained using a specific risk assessment based on questionnaire data with or without information on genetic information and/or mammogram density Systematic review: Tyrer-Cuzick tool	Another risk category ascertained using the same risk assessment tool	Ratio of expected to observed (E/O) for the following outcomes by tool determined risk category: • Breast cancer mortality • Breast cancer incidence (invasive, in situ) • Breast cancer incidence (invasive, in situ) according to prognostic indicators (e.g., tumour subtype, grade, size, nodal involvement) • Interval cancer rates • Breast cancer incidence according to age	Cohort studies or systematic reviews thereof

Selection criteria and definitions

Detailed selection criteria are shown in Table 3.

Table 3. Study selection criteria for this review.

Selection criteria	*Inclusion*	*Exclusion*
Population	Asymptomatic women aged ≥ 18 years Scoping review: •Part of a population-based screening program OR •Participants in large cohort studies e.g. Nurses' Health Study, Women's Health Initiative Study and European Prospective Investigation into Cancer and Nutrition Not restricted to general population i.e. can include high-risk populations e.g. women with family history of breast cancer or previous positive mammogram and low-risk populations Systematic review: population sample is different from which the tool was developed i.e. External validation	Restricted to women undergoing breast imaging as follow-up for breast cancer, in situ, or for breast abnormalities Restricted to specific ethnic groups (African American or Hispanic American populations) Scoping review: Entire development population is used for validation of the risk tool without cross-validation Systematic review: population or proportion of population used to develop the tool is also used for validation of the risk tool i.e. Internal validation Population restricted to high-risk women
Intervention / Exposure	A breast cancer risk assessment tool based on questionnaire, genetic information and/or mammographic density. Risk assessment tools developed in high-risk populations can be included Risk assessment tools may be in original form or abridged	Risk assessment tool involving any subjective input (i.e. Requiring clinician input)
Comparator	Another risk category ascertained using the same risk assessment tool Risk assessment tools developed in high-risk populations can be included. Risk assessment tools may be in original form or abridged	Risk assessment tool involving any non-standardised input (i.e. Requiring clinician input)
Outcome	Breast cancer mortality Breast cancer incidence (invasive and/or in situ) Breast cancer incidence according to prognostic indicators e.g. tumour subtype, grade, size, nodal involvement Breast cancer incidence according to age Interval cancer rates	When outcome remains unclear (e.g. whether DCIS vs DCIS + invasive or 5-year vs 10-year risk) after attempting to contact author for clarification Studies that specifically state use of film mammography only When risk determined by a tool is projected beyond the period for which the tool was developed When a tool only predicts 1-year risk for an outcome
Outcome metrics	Expected/observed (E/O) ratio (including estimates obtained from goodness of fit images) reported by tool determined risk category	Results of goodness of fit test/ Hosmer-Lemeshow test Odds ratios Only overall E/O reported
Study design	Cohort studies or Systematic review thereof	Cross-sectional studies, case-control studies, case-cohort studies and nested case-control studies
Publication type	Peer reviewed journal article or report	Conference abstracts, reviews, letters, editorials and comments
Publication date	2008 onwards	
Language	English	

DCIS = ductal carcinoma in situ; E = expected; O = observed.

Definitions

The following definitions apply for the purposes of this systematic review:

- ***Internal validation*** occurs when the population sample used to develop a risk assessment tool is also used to test the fit of the tool (Moons and Wolff 2019)
- ***External validation*** refers to studies that aim to assess the predictive performance of existing risk assessment tools using data external to the development sample (i.e. using data from different participants) (Moons and Wolff 2019)
- ***Calibration*** reflects the agreement between predictions from the risk assessment tool and observed outcomes (measured as the E/O ratio) (Moons and Wolff 2019). For the purposes of this review, calibration refers to E/O measured across different risk categories
- ***Discrimination*** refers to how well a risk assessment tool differentiates those at higher risk of having an event from those at lower risk (Alba 2017)
- ***Risk assessment tool*** refers to a tool for estimating the probability that a currently healthy individual (i.e. asymptomatic) with specific risk factors will develop a condition in the future (modified from Meads 2012)
- ***Risk categories*** refer to ranges of estimates of risk for a future event as predicted by a risk assessment tool
- ***Risk predictor*** refers to a risk factor included in a risk prediction tool such as age, height, body mass index, mammographic density, etc.
- ***New risk assessment tools*** are defined as follows: If an existing risk assessment tool is extended (e.g. with addition of new risk predictors) or updated (e.g. adjustment of tool coefficients) this is considered to be the development of a **new** risk assessment tool (Moons & Wolff 2019; Moons 2012). If this new risk assessment tool is not developed in a separate population to that in which it is being validated, for the purposes of this review, it is not considered to be an external validation study
- A ***prospective study design*** is defined based on the timing of the data collection i.e. risk predictors collected prior to the outcome occurring.

Terminology

Throughout this review we use the term ‘risk assessment tool’. Elsewhere in the literature, synonymous terms are used such as: risk prediction tool, prognostic model, risk prediction model, risk model, breast cancer prediction model.

The term ‘risk predictors’ may also refer to covariates, risk indicators, prognostic factors, determinants or independent variables (Moons and Wolff 2019).

The Gail risk assessment tool is also known as the Breast Cancer Risk Assessment Tool (BCRAT) and the Tyrer-Cuzick tool is also known as the IBIS tool. Throughout this review we uniformly refer to these tools as BCRAT and Tyrer-Cuzick tools, respectively.

Literature searches

Scoping review: To identify relevant articles published from 2008 onwards both Medline and Embase databases were searched in January 2019, combining terms for breast cancer, risk assessment tools/breast density and expected-to-observed rate ratio. In April 2019, the Cochrane Database of Systematic Reviews (CDSR) and Health Technology Assessments (HTA) databases were searched by combining the terms “breast cancer” and “risk” and the Canadian Agency for Drugs and Technologies in Health (CADTH) database was searched for reports using the term

"breast cancer". For details of the complete search strategy see Table 15 in the Appendix 3.4.2 (page 62). Monthly citation alerts up until 1st April 2019 were used to identify relevant articles subsequently published.

To identify published or ongoing systematic reviews the Cochrane Database of Systematic Reviews (CDSR) was searched by combining text terms "breast cancer" and "risk" and PROSPERO (the International Prospective Register of Systematic Reviews) was also searched using the term "breast cancer risk" (excluding Cochrane reviews and animal studies).

Registered, potentially relevant clinical trials of breast cancer risk assessment tools were identified by searching the following clinical trial registries for ongoing or recruiting trials (April 2019):

- Clinicaltrials.gov with two searches performed: one using the term "breast cancer" under condition and "risk assessment" under other terms and a second search using the term "breast cancer" under condition and "breast screening" under other terms. Recruitment status selected for both searches were "not yet recruiting", "recruiting", "enrolling by invitation" and "active, not recruiting"
- World Health Organisation International Clinical Trials Registry Platform (WHO ICTRP) using the condition as "breast cancer" and the terms "risk assessment" and "breast screening" in the same search
- National Institutes of Health (NIH) National Cancer Institute (https://www.cancer.gov/about-cancer/treatment/clinical-trials/search) using the terms "breast cancer" under condition, "risk assessment" under keywords/phrases and in a separate search, changing the keywords/phrases to "breast screening"
- Australian New Zealand Clinical Trials Registry (ANZCTR) (http://www.anzctr.org.au/TrialSearch.aspx?isBasic=false) using the condition as "breast cancer" and the term "risk assessment" for one search using the term "breast screening" in a separate search.

Relevant withdrawn or terminated trials were noted but not summarised.

Systematic review (Tyrer-Cuzick risk tool): To identify relevant peer reviewed articles published from 2008 onwards, Medline and Embase databases were searched on 10th December 2019, combining text terms and database-specific subject headings where available for breast cancer, risk assessment and calibration. To identify published or ongoing systematic reviews from 2008 onwards, the Cochrane Database of Systematic Reviews (CDSR) and PROSPERO were also searched on the same date. The CDSR was searched by combining text terms "breast cancer" and "risk" and PROSPERO was searched using the term "breast cancer risk". All searches were limited to articles published in English. Updated searches of all databases were performed on the 5 March 2021 and the 1st July 2021. For Medline and Embase databases, the final searches covered the literature up until the 29th June 2021. Searches were broad to capture all relevant records reporting on risk assessment tools. Search results were then used to identify records reporting only on the Tyrer-Cuzick tool, as this tool was subsequently selected as the risk tool of focus (see introduction). For details of the complete search strategy see Table 14 (page 62) in Appendix 3.4.2. Subsequently, the reference lists of all relevant articles and systematic reviews were checked for potential additional contributing articles.

Application of selection criteria

Scoping review: Titles and abstracts of the articles identified by the literature searches were screened against pre-specified inclusion criteria by one systematic reviewer (VF). Full text articles of

potential relevance or unclear relevance were subsequently retrieved to be assessed by the same reviewer (VF) for inclusion. If there was uncertainty regarding inclusion, this was resolved by checking with a second reviewer (SH).

Systematic review (Tyrer-Cuzick model): Titles and abstracts of the articles identified by literature searches were screened against pre-specified inclusion criteria and were split equally between two systematic reviewers (VF, DC); 20% of these were assessed by both reviewers to facilitate review concordance. Full text articles of potential relevance or unclear relevance were subsequently retrieved to be assessed for inclusion using a form incorporating the pre-specified selection criteria. Reviewers were not blinded to journal titles or study authors/institutions.

Authors were contacted and queried with open-ended questions when there was a lack of clarity around criteria for inclusion (e.g. whether the outcome was invasive breast cancer only or invasive breast cancer and DCIS). If no response was received, these studies were excluded.

Disagreements between reviewers were resolved by discussion, and where consensus could not be reached, they were resolved by adjudication by a third reviewer (SH).

Data extraction

Scoping review: Prespecified study details and data were extracted. Relevant data was plotted for expected/observed ratios (E/O) according to risk group. To improve comparison of different tools, we plotted risk categories according to their reported distribution within the study group (by category midpoint percentiles). An example of how these percentile values were calculated is shown in Table 4 below:

Table 4. An example of calculating midpoint percentile for risk groups, as done for Brentnall et al. 2018

Risk group	% participants in risk group	Cumulative % in risk group	Midpoint cumulative % for risk group	Percentile
<2%	19	0.19	0.095	10
2% - <3%	42	0.61	0.400	40
3% - <5%	28	0.89	0.750	75
5% - <8%	9	0.98	0.935	94
≥8%	2	1	0.990	99

Systematic review (Tyer-Cuzick model): Extraction of pre-specified study characteristics and results were split equally and arbitrarily between the two reviewers and followed by accuracy checks. Disagreements were resolved by third reviewer adjudication (SH). The registered PROSPERO protocol (see Appendix 3.4.1) details the specific study characteristics and result items that were pre-specified for extraction. It should be noted, that in regard to the 95% confidence intervals for calibration outcomes, if these were not reported, where possible they were calculated using the shortcut method (Vandenbroucke 1982) and then inverted to obtain E/O estimates. Authors were contacted to request calibration data when the study only reported this in graphical form.

Risk of bias appraisal

Scoping review: Not conducted as this level of quality assessment is not typically required for a scoping review.

Systematic review (Tyrer-Cuzick model): The risk of bias for risk assessment tool studies was assessed independently by two reviewers (VF, DC) using PROBAST (Prediction model Risk Of Bias

ASsessment Tool) (Wolff and Moons 2019; Moons and Wolff 2019). This tool was designed to specifically assess the risk of bias for prognostic or diagnostic prediction risk assessment tool studies. Rulings were developed where necessary to account for reviewer judgements that required topic-specific knowledge. These rulings were initially trialled independently over several studies by the same two reviewers, with third reviewer input where required.

PROBAST assesses potential sources of bias over the following four domains; (i) participant selection, (ii) predictor measurement, (iii) outcome measurement and (iv) analysis, using pre-specified criteria (Moons and Wolff 2019). It was decided a priori that i) risk of bias domains which contained signalling items relating only to model development would be omitted[12] as the primary interest of this review concerns risk assessment tool validation and ii) the applicability of a study would not be formally assessed by the PROBAST tool; concerns would instead be highlighted where necessary in the discussion.

The overall risk of bias for each study was categorised as either low (i.e. of low risk for all domains assessed), high (i.e. of high risk for at least one domain examined) or unclear (i.e. of unclear risk for any domain and of no high-risk rating for any domain). Differences in opinion were resolved by consensus discussion or adjudication by a third reviewer.

For each study, PROBAST specifies that a separate risk of bias assessment should be conducted for each distinct risk assessment tool being validated as well as for each individual outcome (Wolff and Moons 2019).

3.3.2 Results

Scoping review

Published relevant systematic reviews

One health technology assessment report (MaHTAS 2015) was identified as potentially relevant from the HTA database search while one systematic review (Anothaisintawee 2012) and another systematic review including a meta-analysis (Meads 2012) were identified through searching Medline and Embase. These articles were used to snowball for further articles as they did not focus on risk categories and contained articles that met our exclusion criteria; this also prevented use of between tool comparison from the meta-analysis information. One primary study was also snowballed from the HTA report and is reported below.

Published relevant primary studies

Nine studies (Brentnall 2018; Arrospide 2013; Nickson 2018; Petracci 2011; Powell 2014; Quante 2012; Rosner 2013; Tice 2008) examined the prediction of invasive breast cancer for different risk categories of a given breast cancer risk assessment tool. Two additional articles were excluded: one (namely Min 2014) because it was not clear if its setting included population screening and another (Matsuno 2011) because the study group assessed comprised of postmenopausal women self-selected to be at higher risk of breast cancer. (A list of excluded studies from the scoping review are presented in Appendix Table 16).

The Risk assessment tools included different versions or variations of the Tyrer-Cuzick, Gail, Chen, Barlow, Breast Cancer Surveillance Consortium (BCSC), Petracci, Rosner-Colditz, BRCAPRO, Vermont tools. The risk factors included in each of these tools are summarised in Table 5 (page 30). Studies examining invasive breast cancer outcomes are summarised in Table 6 (page 31), studies examining invasive breast cancer and DCIS combined are summarised in Table 7 (page 40) and

[12] Specifically, questions 4.5, 4.8 and 4.9

one study where the outcome was not clear (i.e. unclear if invasive or invasive with DCIS) is summarised in Table 8 (note additional information was sought from the authors, but was not provided). Follow-up tended to be reported by 5- or 10- year risk except for one study (Vacek 2011) which reported by person-year risk.

Gail model (NCI Risk Assessment Tool)

Selected results for the Gail (NCI) model are shown in Figure 3 (page 27), for four different studies.

Key findings:

- The Gail model tended to predict outcomes well across population risk groups (as indicated by E/O ratios close to 1), with a tendency to overestimate risk in higher-risk groups. The Marin Women's Study was an exception, with a poor fit between estimated and observed outcomes
- Ranked risk groups assigned using the Gail model had a clear positive association with observed incident invasive breast cancers following risk assessment. For example, in the Australian Lifepool cohort ('lifepool'), there were approximately twice as many invasive cancers in the highest decile compared to the lowest quintile.

Of note, the study populations differed slightly for these four studies. For example, the Marin Women's Study comprised a relatively young US screening population (46% younger than 50 years) with high rates of breast cancer, nulliparity, and delayed childbirth. The NHS and WHI-ES cohorts were large cohorts from the US, and the Lifepool cohort comprised Australian BreastScreen Victoria participants observed from baseline data collected at subsequent round screening. The authors of the Australian study (Nickson et. al., 2018) noted that the Gail model was developed for the US population and a modification of the ethnicity variables may improve the model fit for the Australian population setting.

Tyrer-Cuzick model

Outcomes for the Tyrer-Cuzick model (v6/7.02) are shown in

Key findings:

- Results were mixed for the two studies identified that met our selection criteria
- Ten-year risk fitted well for the midrange of the population but over-estimated risk by around 30% for women above the 95th percentile of risk in the Kaiser Permanente Washington Breast Cancer Surveillance Consortium
- Five-year risk fitted well for deciles 5 and 7-9 but was a poor fit (generally overestimating risk) in the Marin Women's Study
- Ranked risk groups assigned using the Tyrer-Cuzick model had a clear positive association with observed incident invasive breast cancers, for both studies.

Risk assessment tools including breast density

Outcomes for models including breast density for three identified studies are shown in Figure 5 (page 29).

Key findings:

- The BCSC model showed a very good fit between observed and expected outcomes, however this result needs to be interpreted with high caution as the model was developed in 60% of the reported study cohort

- Estimated and observed risk did not match for other models shown (Tyrer-Cuzick v7.02 and Chen v1), although Tyrer-Cuzick v7.02 consistently overestimated risk (suggesting that it could be improved by scaling the predicted rates for all categories)
- Ranked risk groups assigned using the models shown had a clear positive association with observed incident invasive breast cancers, for all three studies.

Systematic review (Tyrer-Cuzick focussed):

The literature searches were undertaken (Appendix 3.4.2, page62) with the final update conducted on the 1st July 2021 (covering literature up until 29 June 2021).

There were a total of 5020 records identified across the Medline, Embase and Cochrane Database of Systematic Reviews (CDSR) databases. A search of the PROSPERO database for ongoing systematic reviews yielded no relevant systematic review protocols. Reference lists of potentially relevant systematic reviews were checked and one additional reference was identified. Seven studies (namely Brentnall 2018; Choudhury 2020; Glynn 2019; Jantzen 2021; McCarthy 2020; Powell 2014; Terry 2019) satisfied the inclusion criteria, which comprised of 8 validation cohorts. The main reason for exclusion of articles were ineligible study population (Appendix Table 16).

Study characteristics for included studies are presented in Table 9 (page 45). The majority of studies were conducted in the USA (n=5/7) (Choudhury 2020; McCarthy 2020; Glynn 2019; Brentnall 2018; Powell 2014), one in Canada (Jantzen 2021) and one was based on combined data from the USA, Canada and Australia (Terry et al, 2019). Study cohorts varied in (i) size (from 10,200 to >132,000), (ii) in the age range of included women (from 40-69y, 20-70 to <40->80y) and(iii) in the length of follow-up (5-24 years).

Risk predictors reported by the included studies are presented in Table 10 (page 47). Four different versions of the Tyrer-Cuzick tool was used (i.e. v7.0; v7.02; v8.0; v8.0b). As expected, there were differences in risk predictors between different versions of the Tyrer-Cuzick tool but differences in risk predicators was observed even between studies using the same version of the tool. For example, Choudhury et al (2020) did not include data for ovarian cancer, use/duration of HRT, prior number of breast biopsies and other history of breast pathology such as ADH or LCIS, nor number of second-degree relatives with breast cancer in the Tyrer-Cuzick v8 model, whereas these risk predictors were reported by Jantzen et al (2021) which used the same tool version. It should also be noted that the methodology used by authors of individual studies to calculate risk predictors in a tool differed from the recommended methodology, especially when data was not collected for a particular risk predictor. Where this data was unclear, the authors of the studies were contacted for further information or clarification.

An overall risk of bias assessment is presented in Table 11 (page 48) and a record of how each source of bias was assessed across the four domains of PROBAST for each study is presented in Table 12 (page 49). All studies were determined at high risk of bias for the domain of analysis, followed by high risk of bias for the domain of predictors (4 of 7 studies) and an unclear risk of bias for the domain of outcome (6 of 7 studies). The overall risk of bias ruling was high for all studies.

The expected over observed (E/O) ratios and their corresponding 95% confidence intervals are presented in Table 13.

3.3.3 Discussion

Scoping review

This review included nineteen relevant studies which contained 21 cohorts. In total, there were 27 risk assessment tools reported, including different variations or versions of models. No identified risk assessment tools incorporated genetic risk, although this was in scope. Overall, it was especially difficult to compare results across studies for a given breast cancer risk assessment tool, not only due to the different populations and breast screen settings (e.g. population-based biennial screening in European studies and institution-based annual screening or unreported screening intervals in US studies) but also the age ranges for which the data was reported (ranging from 20-70 years or 30-60 to 50-69 years) and whether risk was predicted for 5 or 10 years. However, several risk assessment tools have been validated in terms of demonstrating a reasonable fit between observed and expected outcomes in prospective cohort studies. The Gail and Tyrer-Cuzick tools have been evaluated in numerous studies, as have some tools incorporating breast density.

The value of incorporating breast density in risk assessment was unclear; several studies suggested that when breast density was added to risk models, risk tended to be over-estimated at 5- or 10-years. However, consistent overestimation could be addressed through straightforward model calibration (i.e. reducing the risk rates across all risk groups); results are more concerning when the expected to observed ratio is very high for some groups and very low for others, and particularly when these differences do not follow a clear gradient across risk groups.

Only one study involved an Australian population, in which the Lifepool cohort was used for validation of the Gail model (Nickson 2018). The model fit was very good, despite some overestimation in higher risk groups. This was the only model validated at a population level in Australia, and the authors noted that it would most likely improve with adjustment to suit local ethnic profiles. No validation of established risk assessment tools incorporating breast density on the Australian target population for breast screening was identified.

Plots of observed incident cancer rates by baseline risk categories confirmed that various risk tools can stratify women into different levels of risk, even if the predicted absolute risk of cancer is not accurate. This level of validation would be suitable for risk-based interventions, and women could be advised about their rank of risk on this basis (e.g. ‘You are in the top 10% of risk’), but absolute risk estimates should be modified before used to describe absolute risk (e.g. it may take some model refinement to safely advise ‘You have a 3% risk of breast cancer in the next 5 years’).

In addition, this analysis identifies a risk assessment tool (namely Gail v2) that calibrated well to a large cohort of BreastScreen Australia participants. We found no validation of established risk assessment tools incorporating breast density on the Australian target population for breast screening.

Systematic review

Overall, it was especially difficult to compare results across studies for a given breast cancer risk assessment tool, not only due to the different populations and breast screen settings (e.g. population-based biennial screening in European studies and institution-based annual screening or unreported screening intervals in US studies) but also as the age ranges for which the data was reported varied (ranging from 20-70 years or 30-60 to 50-69 years) and whether risk was predicted for 5- or 10-years.

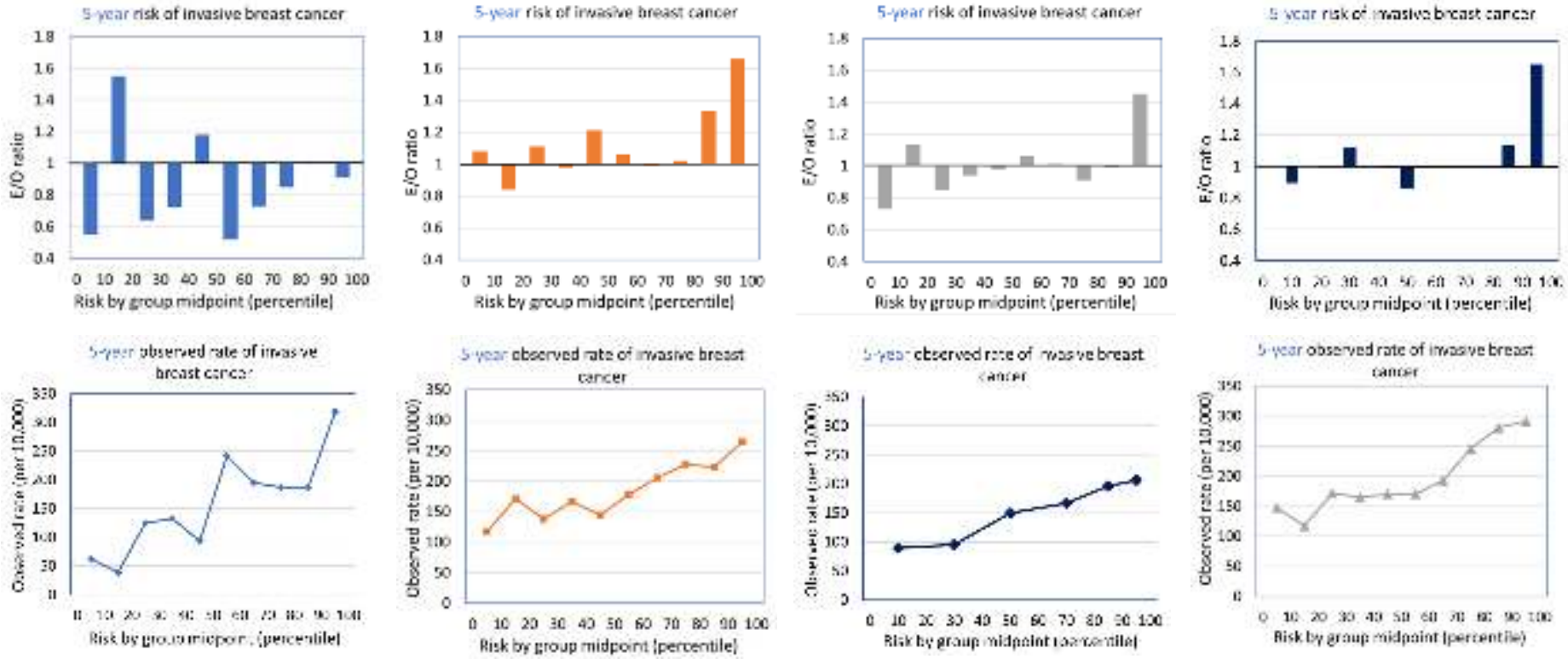
Combining findings from research question 1 and 2, BCRAT version 3 and 4 consistently distinguished women in the lowest and highest risk groups based on 5-year predicted risk but was

inconsistent in terms of its goodness-of-fit, depending on the setting. Similarly, the Tyrer-Cuzick version 8.0b tool (which incorporates breast density data) consistently distinguished women in the lowest risk group based on 5-year predicted risk and in the highest risk group for 5-year risk. However, in most settings this tool showed evidence of miscalibration (poor fit between estimated and observed outcomes). There was insufficient evidence to draw conclusions about the other tools assessed except for the iCARE tools which, based on two studies, performed similarly to BCRAT and Tyrer-Cuzick.

These results further support the findings from Research Question 1 that in breast screening populations, some risk assessment tools can confidently identify groups of women at higher or lower risk, depending on the study setting and population.

We found no validation of the Tyrer-Cuzick tool in the Australian target breast screening population.

3.3.4 Figures



a) Marin Women's Study cohort (US, 40y+) (Powell et al. 2014)

b) Nurses' Health Study cohort (US, 55-74y) (Schonberg 2016)

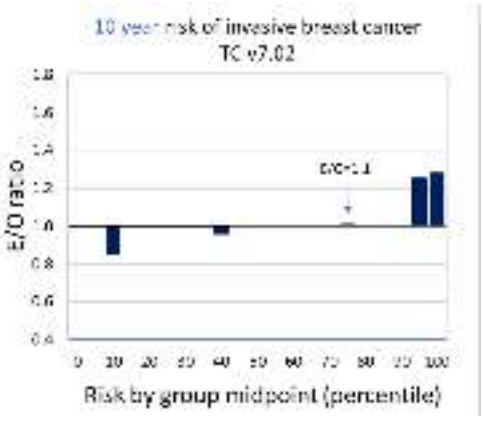
c) lifepool cohort (50-69y) (Australia, Nickson 2018)

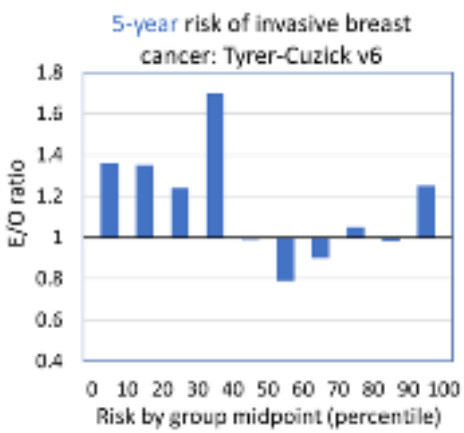
d) Women's Health Inititive - Extension Study cohort (US, 55-74y) (Schonberg 2016)

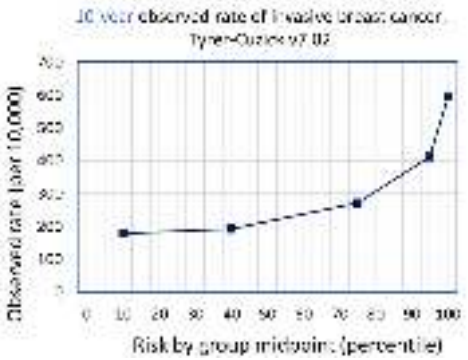
Figure 3. Expected to Observed ratio (E/O) and observed cancer diagnoses for the Gail v2 risk assessment tool, various studies. Detailed in Table 5.

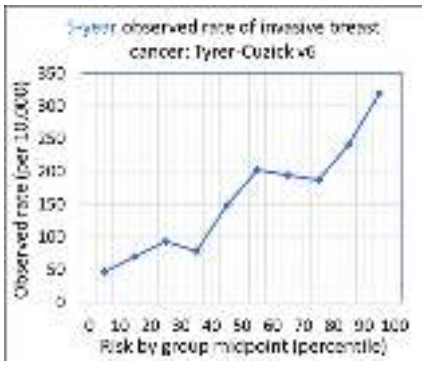
a) Kaiser Permanente Washington Breast Cancer Surveillance Consortium cohort (50-59y) (Brentnall et al. 2018) (23)

b) Marin Women's Study cohort (US, 40y+) (Powell et al. 2014) (38)









Corresponding risk prediction group:	
Risk prediction group	Percentile
<2%	10
2% - <3%	40
3% - <5%	75
5% - <8%	95
≥8%	99

Figure 4. Expected to Observed ratio (E/O) and observed cancer diagnoses for the Tyrer-Cuzick v6/7.02 assessment tool, various studies. Detailed in Table 5.

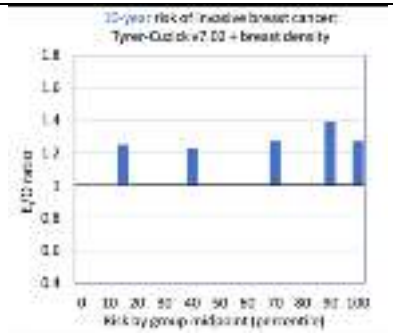
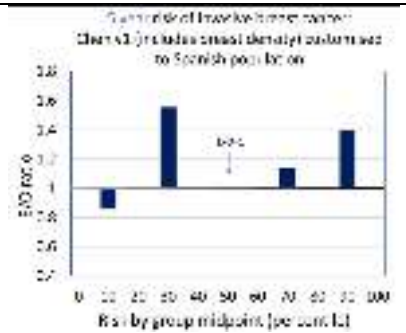
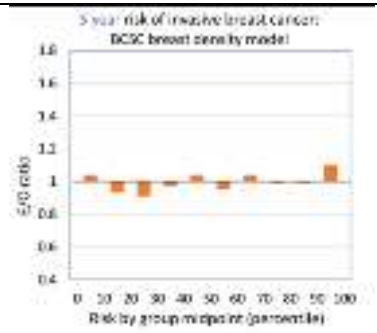
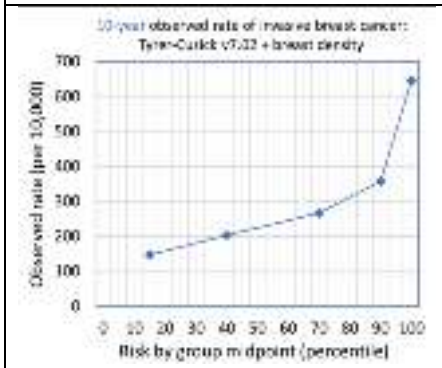
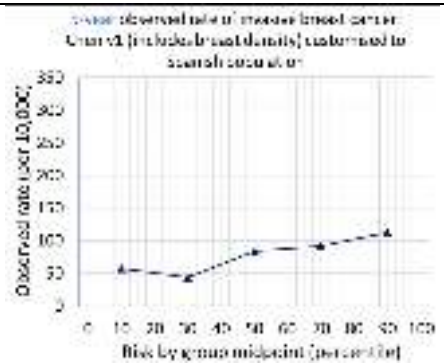
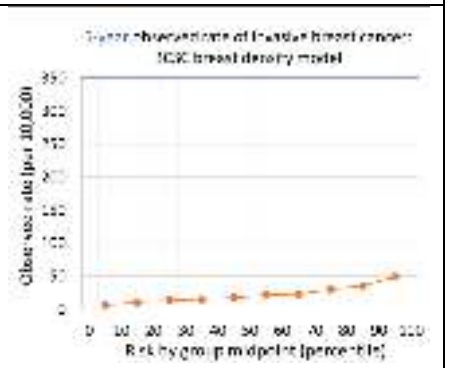
a) Tyrer-Cuzick with breast density. Kaiser Permanente Washington Breast Cancer Surveillance Consortium cohort (US, 50-59y) (Brentnall et al. 2018)	*b) Chen v1 risk assessment tool, Breast cancer early-detection program in the Sabadell-Cerdanyola (EDBC-SC) area in Catalonia (Spain, 50-69y) (Arrospide et al. 2013)*	*c) BCSC breast density/Tice model, Breast Cancer Surveillance Consortium (US, 41% 50-69y; 46% <50y) (Tice et al. 2008)*
		
		
Corresponding risk prediction group: Risk prediction group / Percentile <2% / 14 2% - <3% / 42 3% - <5% / 70 5% - <8% / 90 ≥8% / 97		

Figure 5. Expected to Observed ratio (E/O) and observed cancer diagnoses for various models including breast density, various studies (see Table 5)

3.3.5 Tables

Table 5. Risk factors in selected risk assessment tools corresponding to Figures 1-3

Risk model	Age	Age of menarche	Age of 1st live birth	Age of menopause	Parity	HRT	Breast density	BMI or weight	Height	Race and /or ethnicity including Ashkenazi Jewish ancestry	Prior/ number of breast biopsies	Other path	Atypical ductal hyperplasia	Lobular carcinoma in situ	Degree relatives assessed	Age of breast cancer	Bilateral breast cancer	Male breast cancer	Ovarian cancer (personal/ familial)	Family history of other cancers	Prior mastectomy and oophorectomy in relatives	Study
Gail v1 (BCRAT) customised to Spanish population	>35 years	X	x								x		x		1st							Arrospide 2013
Gail v2 (BCRAT) SEER 1983-7	>35 years	X	x							x	x		x		1st							(Nickson 2019, Powell 2014, Schonberg 2016)
Chen v1 customised to Spanish population	x		x				x	x			x				1st							Arrospide 2013
Tyrer-Cuzick (IBIS) v6.0.0	x	X	x	x	x	x		x		x	x	x	x	x	1st-2nd	x	x		x			Powell 2014
Tyrer-Cuzick (IBIS) v7.02	x	X	x	x	x			x	x	x	x	x	x	x	1st-2nd	x	x	x	x (+ age affected)			Brentnall 2018
Tyrer-Cuzick (IBIS) v7.02 + breast density	x	X	x	x	x		x	x	x	x	x	x	x	x	1st-2nd	x	x	x	x (+ age affected)			Brentnall 2018
BCSC/Tice model	x						x			x	x				1st							Tice 2008
BRCAPRO	x											x			1st-2nd	x	x	x	x		x	Powell 2014

Table 6. Within-tool comparisons: Studies comparing expected/observed invasive breast cancer ratios for different risk categories of a given breast cancer risk assessment tool (scoping review).

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (Incidence data)	Follow-up	Estimated risk category	Invasive breast cancer (% in risk category) Expected/observed ratio (95%CI)
Powell 2014, (USA), Retrospective cohort, 2003-2007	Marin Women's Study (MWS), a mammography-based study of women in Marin County, California where the study was performed at all major screening centres in the county, including those associated with Kaiser Permanente, Marin General Hospital, and Novato Community Hospitals. The screening centres contribute to the San Francisco Mammography Registry (SFMR) which participates in the Breast Cancer Surveillance Consortium (BCSC). All women were invited to participate no matter their history, however, the population of interest was restricted to women who were breast cancer free at baseline. Note. Women in this County had higher rates of breast cancer, nulliparity and delayed childbirth. N = 12,843 Age at entry ranged <40 - ≥80 y; % <50 y = 46.3% % ≥70 y = 7.3% Previously screened: those with screening history allowed	NR	**Tyrer-Cuzick v6.0.0** (Cancer Statistics Registrations, England and Wales 1994) Risk factors ascertained from self-reported questionnaire as well as from linkage with SFMR for breast density, demographic data, BMI and family history. Questionnaire data were collected from 2006-2009 with the reference baseline set at beginning of 2003	Five-year follow-up period with confirmation of breast cancer between 2003-2007 either self-reported on the questionnaire or included in the SFMR cancer registry data. Median follow-up = NR	**5-year risk^ by decile** D1	(NR %) 1.36 (NR)
					D2	(NR %) 1.35 (NR)
					D3	(NR %) 1.24 (NR)
					D4	(NR %) 1.70 (NR)
					D5	(NR %) 0.99 (NR)
					D6	(NR %) 0.79 (NR)
					D7	(NR %) 0.90 (NR)
					D8	(NR %) 1.05 (NR)
					D9	(NR %) 0.98 (NR)
					D10	(NR %) 1.25 (NR)
Brentnall 2018, (USA) , Prospective cohort, 1996-2014	Women in Kaiser Permanente Washington Breast Cancer Surveillance Consortium who attended ≥ one mammogram screening between 1996-2013	Annual screening for women aged 50-75 years and for high-risk women	**Tyrer-Cuzick v7.02** (Thames cancer registry first breast cancer rates,	6 months from entry until invasive breast cancer diagnosis or censoring at DCIS diagnosis, death,	**10-year risk** < 2%	All ages at entry (36%) 0.82 (0.76-0.89) Aged < 50 years (62%) 1.08 (0.99-1.19) Aged 50-59 years (19%) 0.95 (0.81-1.12) Aged ≥ 60 years (8%) 0.79 (0.55-1.15)
					2% to < 3%	All ages at entry (32%) 0.99 (0.92-1.06) Aged < 50 years (24%) 1.39 (1.22-1.59)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (Incidence data)	Follow-up	Estimated risk category	Invasive breast cancer (% in risk category) Expected/observed ratio (95%CI)
	Excluded those with prior diagnosis of breast cancer or DCIS or diagnosed with within 6 months of initial mammogram N = 132,139 Age at entry (median) = 50 years Previously screened: NR Subgroups Age at entry < 50 years N = 60,185 50-59 years N = 43,759 ≥ 60 years N = 28,195	aged 40-49 years However, 62% of women aged <50 years at entry were low risk	United Kingdom, 2005-2009) Risk factors self-reported at time of entry mammogram	age 75 years or December 2014 Median follow-up = 5.2 years		Aged 50-59 years (42%) 1.19 (1.17-1.32) Aged ≥ 60 years (36%) 0.75 (0.65-0.87)
					3% to < 5%	All ages at entry (22%) 0.98 (0.91-1.05) Aged < 50 years (11%) 1.30 (1.52-1.12) Aged 50-59 years (28%) 1.16 (1.04-1.30) Aged ≥ 60 years (38%) 0.80 (0.71-0.91)
					5% to < 8%	All ages at entry (7%) 1.15 (1.03-1.28) Aged < 50 years (2%) 1.54 (1.16-2.08) Aged 50-59 years (9%) 1.14 (0.97-1.33) Aged ≥ 60 years (14%) 1.08 (0.90-1.30)
					≥ 8%	All ages at entry (2%) 1.27 (1.08-1.49) Aged < 50 years (1%) 1.49 (1.03-2.27) Aged 50-59 years (2%) 1.37 (1.05-1.82) Aged ≥ 60 years (4%) 0.97 (0.76-1.27)
Brentnall 2018, (USA) , Prospective cohort, 1996-2014	Women in Kaiser Permanente Washington Breast Cancer Surveillance Consortium who attended ≥ one mammogram screening between 1996-2013 Excluded those with prior diagnosis of breast cancer or DCIS or diagnosed with within 6 months of initial mammogram N = 132,139 Age at entry (median) = 50 years Previously screened: NR Subgroups Age at entry < 50 years N = 60,185 50-59 years N = 43,759 ≥ 60 years N = 28,195	Annual screening for women aged 50-75 years and for high-risk women aged 40-49 years However, 62% of women aged <50 years at entry were low risk	**Tyrer-Cuzick v7.02** (Thames cancer registry first breast cancer rates, United Kingdom, 2005-2009) Risk factors self-reported at time of entry mammogram + breast density (BI-RADS) on initial mammogram adjusted for BMI	6 months from entry until invasive breast cancer diagnosis or censoring at DCIS diagnosis, death, age 75 years or December 2014 Median follow-up = 5.2 years	10-year risk < 2%	All ages at entry (40%) 0.85 (0.79-0.93) Aged < 50 years (61%) 1.25 (1.14-1.39) Aged 50-59 years (27%) 1.25 (1.08-1.45) Aged ≥ 60 years (16%) 1.15 (0.85-1.56)
					2% to < 3%	All ages at entry (25%) 0.96 (0.89-1.04) Aged < 50 years (22%) 1.33 (1.18-1.54) Aged 50-59 years (29%) 1.23 (1.09-1.39) Aged ≥ 60 years (26%) 0.90 (0.76-1.08)
					3% to < 5%	All ages at entry (22%) 1.01 (0.93-1.08) Aged < 50 years (12%) 1.34 (1.19-1.59) Aged 50-59 years (28%) 1.27 (1.12-1.41) Aged ≥ 60 years (34%) 0.83 (0.74-0.94)
					5% to < 8%	All ages at entry (8.6%) 1.25 (1.12-1.39) Aged < 50 years (3%) 1.49 (1.19-1.89) Aged 50-59 years (11%) 1.39 (1.20-1.61) Aged ≥ 60 years (16%) 1.10 (0.93-1.32)
					≥ 8%	All ages at entry (3.5%) 1.28 (1.14-1.45) Aged < 50 years (1%) 1.59 (1.19-2.17) Aged 50-59 years (4%) 1.27 (1.05-1.52) Aged ≥ 60 years (7%) 1.09 (0.90-1.33)
Arrospide 2013, (Spain), Prospective cohort, 1995-2010	Women attending for the first time the breast cancer early-detection program in Sabadell-Cerdanyola	Biennial mammograph	**Gail v1** (BCDDP 1973-1980) customised	6 months from entry to 3 and 5 years. Followed for vital	**5-year risk*** **By quintile** Q1	Ages 50-69 entry (19.9%) 1.06 (NR)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (Incidence data)	Follow-up	Estimated risk category		Invasive breast cancer (% in risk category) Expected/observed ratio (95%CI)
	(EDBC-SC) area in Catalonia, Spain between 1995-1998. No personal history of breast cancer. Risk factors collected via questionnaire at time of entry mammogram and breast density collected (BI-RADS) from initial mammogram. N initial =13,760 N final =13,709 Aged 50-69 Previously screened: No, first time participants	y for women aged 50-69.	(Customised using an estimated incidence function of invasive breast cancer in Catalonia) Risk factors collected via questionnaire at time of entry mammogram and breast density collected (BI-RADS) from initial mammogram	status or possible diagnosis of breast cancer until 2010. Invasive cancer diagnosis was recorded regardless of whether made within or outside of the program. Median follow-up = 13.3 years Note. In this review, only 5-year data is extracted	Q2		Ages 50-69 entry (20.2%) 0.95 (NR)
					Q3		Ages 50-69 entry (19.4%) 1.38 (NR)
					Q4		Ages 50-69 entry (20.4%) 1.50 (NR)
					Q5		Ages 50-69 entry (20.2%) 1.06 (NR)
Nickson 2018, (Australia), Prospective cohort, 2010-2016	Women from the Lifepool cohort participating in Australia's BreastScreen population-based mammography programme. This analysis was restricted to women aged 50-69 years who had a reference screen with a benign final outcome within ±60 days of completing baseline questionnaire, had no personal history of invasive breast cancer, DCIS or LCIS. N= 40,158 Aged 50-69 Previously screened: Yes, recruitment primarily via appointment letters for women attending subsequent rounds of BreastScreen programme screening. Subgroups	Biennial mammography for women aged 50–74 years (extended from 50 to 69 years in mid-2015)	**Gail v2** (SEER 1983-7) Self-reported questionnaire completed on enrolment.	Censoring occurred at diagnosis (invasive or in-situ), death or 31 December 2016 , whichever occurred first. Median follow-up = 4.3 years	**5-year risk by quintile and upper two deciles**		
					Q1	0.6–1.1%	Aged 50-69 years (20%) 0.90 (0.71–1.15) Aged 50-59 years (20%) 0.83 (0.60–1.19)
						0.9-1.1%	Aged 60-69 years (20%) 0.96 (0.71–1.34)
					Q2 1.1–1.4%		Aged 50-69 years (20%) 1.12 (0.89–1.42) Aged 50-59 years (20%) 1.14 (0.81–1.66) Aged 60-69 years (20%) 0.88 (0.68–1.17)
					Q3 1.4–1.7%		Aged 50-69 years (20%) 0.86 (0.72–1.04) Aged 50-59 years (20%) 1.10 (0.81–1.52) Aged 60-69 years (20%) 0.92 (0.72–1.19)
					Q4 1.7–2.3%		Aged 50-69 years (20%) 0.99 (0.84–1.19) Aged 50-59 years (20%) 0.87 (0.68–1.12) Aged 60-69 years (20%) 1.02 (0.81–1.30)
					Q5	2.3-13.9%	Aged 50-59 years (20%) 1.40 (1.20–1.64)
						2.3–22.0%	Aged 50-69 years (20%) 1.40 (1.10–1.80) Aged 60-69 years (20%) 1.47 (1.19–1.84)
					D9	2.3–3.0%	Aged 50-69 years (50% of Q5 for corresponding age group) 1.13 (0.91–1.43)
						2.1-2.5%	Aged 50-59 years (50% of Q5 for corresponding age group) 1.28 (0.89–1.89)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (Incidence data)	Follow-up	Estimated risk category		Invasive breast cancer (% in risk category) Expected/observed ratio (95%CI)
	Age at entry 50-59 years N= 20,216 60-69 years N= 19, 942					2.8-3.3%	Aged 60-69 years (50% of Q5 for corresponding age group) 1.24 (0.91–1.74)
					D10	3.0–22.0%	Aged 50-69 years (50% of Q5 for corresponding age group) 1.65 (1.33–2.07)
						2.5-13.9%	Aged 50-59 years (50% of Q5 for corresponding age group) 1.49 (1.08–2.12)
						3.3-22.0%	Aged 60-69 years (50% of Q5 for corresponding age group) 1.66 (1.25–2.27)
Schonberg 2016, (USA), Two prospective cohorts:, the Nurses' Health Study (NHS) and Women's Health, Initiative – extension study (WHI-ES)., NHS: 2004-2010, WHI-ES: 2005-2010	Women from the NHS and WHI-ES cohorts. NHS cohort contains female nurses, age 30-55 years starting in 1976, from 11 US states with baseline and then biennial follow-up in the form of questionnaires via mail (starting from 2004) to obtain detailed lifestyle and medical history. WHI-ES cohort is an extension arm of a multicentre study including postmenopausal US women aged 50-79 years in up to three clinical trials (WHI-CT) or observational study (WHI-OS) from 1993-1998 whereby majority of these participants agreed to the extension study (82% of WHI-CT participants and 73% of WHI-OS). WHI-ES participants were chosen due to many ≥75 years and most had stopped using hormone therapy, which is typical of current practice. Exclusion criteria were personal history of any cancer (except for non-melanoma skin cancers), did not return 2004 NHS questionnaire or chose not to participate in WHI-	Note. Higher intensity breast cancer screening in WHI participants than general population.	**Gail v2** (SEER 1983-7) Two separate cohorts used for validation – NHS data (96% non-Hispanic white) and WHI-ES (around 85% non-Hispanic white) with self-reported questionnaire at baseline and annual (WHI-ES) or biennial (NHS) follow-up questionnaires	Followed for up to five years or until invasive breast cancer or death, whichever came first. All WHI breast cancer cases were confirmed by pathology report while for NHS, self-reported breast cancer cases (12% of cases) were included due to previous report of accuracy when compared to medical records. Median follow-up = NR	**5-year risk^ by decile**		
					D1	NHS	Aged 55-74 years (10%) 1.08 (0.84-1.38) Aged 75+ years (10%) 1.22 (0.79-1.87)
						WHI-ES	Aged 55-74 years (10%) 0.73 (0.59-0.91) Aged 75+ years (10%) 1.00 (0.68-1.48)
					D2	NHS	Aged 55-74 years (10%) 0.84 (0.68-1.04) Aged 75+ years (10%) 1.17 (0.79-1.73)
						WHI-ES	Aged 55-74 years (10%) 1.13 (0.89-1.44) Aged 75+ years (10.1%) 0.96 (0.68-1.35)
					D3	NHS	Aged 55-74 years (10%) 1.11 (0.88-1.40) Aged 75+ years (10%) 0.83 (0.60-1.14)
						WHI-ES	Aged 55-74 years (10%) 0.85 (0.70-1.04) Aged 75+ years (10%) 0.80 (0.59-1.07)
					D4	NHS	Aged 55-74 years (10%) 0.98 (0.80-1.21) Aged 75+ years (10%) 1.69 (1.09-2.62)
						WHI-ES	Aged 55-74 years (10%) 0.94 (0.77-1.16) Aged 75+ years (10%) 1.11 (0.79-1.55)
					D5	NHS	Aged 55-74 years (10%) 1.21 (0.97-1.52) Aged 75+ years (10%) 0.93 (0.68-1.27)
						WHI-ES	Aged 55-74 years (10%) 0.98 (0.80-1.20) Aged 75+ years (10%) 0.99 (0.73-1.35)
					D6	NHS	Aged 55-74 years (10%) 1.06 (0.87-1.30) Aged 75+ years (10%) 1.20 (0.85-1.69)
						WHI-ES	Aged 55-74 years (10%) 1.06 (0.86-1.29) Aged 75+ years (10%) 1.07 (0.79-1.46)
					D7	NHS	Aged 55-74 years (10%) 0.99 (0.82-1.20) Aged 75+ years (10.1%) 1.52 (1.05-2.21)
						WHI-ES	Aged 55-74 years (10%) 1.01 (0.84-1.22) Aged 75+ years (10%) 1.14 (0.84-1.55)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (Incidence data)	Follow-up	Estimated risk category		Invasive breast cancer (% in risk category) Expected/observed ratio (95%CI)
	ES, died before 2004 (NHS) or before the WHI-ES and those with missing data on the final model's risk factors. NHS N = 73, 066 WHI-ES N = 74, 887 Subgroups Age at entry NHS: 55-74 years, N= 52,111, 75+ years, N= 19,182 WHI-ES: 55-74 years, N= 57,009, 75+ years, N= 22,602				D8	NHS	Aged 55-74 years (10%) 1.02 (0.86-1.23) Aged 75+ years (10%) 1.42 (1.02-1.97)
						WHI-ES	Aged 55-74 years (10%) 0.91 (0.77-1.07) Aged 75+ years (10%) 1.04 (0.79-1.36)
					D9	NHS	Aged 55-74 years (10%) 1.33 (1.11-1.60) Aged 75+ years (10%) 1.40 (1.04-1.88)
						WHI-ES	Aged 55-74 years (10%) 0.99 (0.85-1.15) Aged 75+ years (10%) 1.13 (0.88-1.46)
					D10	NHS	Aged 55-74 years (10%) 1.66 (1.41-1.97) Aged 75+ years (10%) 1.68 (1.29-2.18)
						WHI-ES	Aged 55-74 years (10%) 1.45 (1.25-1.69) Aged 75+ years (10%) 1.47 (1.16-1.86)
					Q4		Ages 20-70 at entry (25%) 0.92 (CD)
Powell, 2014, (USA), Retrospective cohort, 2003-2007	Marin Women's Study (MWS), a mammography-based study of women in Marin County, California where the study was performed at all major screening centres in the county, including those associated with Kaiser Permanente, Marin General Hospital, and Novato Community Hospitals. The screening centres contribute to the San Francisco Mammography Registry (SFMR) which participates in the Breast Cancer Surveillance Consortium (BCSC). All women were invited to participate no matter their history, however, the population of interest was restricted to women who were breast cancer free at baseline. Note. Women in this County had higher rates of breast cancer, nulliparity and delayed childbirth.	NR	**Gail v2** (SEER 1983-7) Risk factors ascertained from self-reported questionnaire as well as from linkage with SFMR for breast density, demographic data, BMI and family history. Questionnaire data were collected from 2006-2009 with the reference baseline set at beginning of 2003	Five-year follow-up period with confirmation of breast cancer between 2003-2007 either self-reported on the questionnaire or included in the SFMR cancer registry data. Median follow-up = NR	**5-year risk^ by decile**		
					D1		All ages at entry (NR%) 0.55 (NR)
					D2		All ages at entry (NR%) 1.55 (NR)
					D3		All ages at entry (NR%) 0.64 (NR)
					D4		All ages at entry (NR%) 0.72 (NR)
					D5		All ages at entry (NR%) 1.18 (NR)
					D6		All ages at entry (NR%) 0.52 (NR)
					D7		All ages at entry (NR%) 0.73 (NR)
					D8		All ages at entry (NR%) 0.85 (NR)
					D9		All ages at entry (NR%) 1.01 (NR)
					D10		All ages at entry (NR%) 0.91 (NR)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (Incidence data)	Follow-up	Estimated risk category	Invasive breast cancer (% in risk category) Expected/observed ratio (95%CI)
	N = 12,843 Age at entry ranged <40 - ≥80 years; % <50 years = 46.3% % ≥70 years = 7.3% Previously screened: those with screening history allowed					
Arrospide 2013, (Spain), Prospective cohort, 1995-2010	Women attending for the first time the breast cancer early-detection program in Sabadell-Cerdanyola (EDBC-SC) area in Catalonia, Spain between 1995-1998. No personal history of breast cancer. Risk factors collected via questionnaire at time of entry mammogram and breast density collected (BI-RADS) from initial mammogram. N initial =13,760 N final =13,709 Aged 50-69 Previously screened: No, first time participants	Biennial screening mammography for women aged 50-69.	**Chen v1** (BCDDP 1973-1980) customised (Customised using an estimated incidence function of invasive breast cancer in Catalonia) (includes breast density and weight but not age at menarche or interactions) Risk factors collected via questionnaire at time of entry mammogram and breast density collected (BI-RADS) from initial mammogram	6 months from entry to 3 and 5 years. Followed for vital status or possible diagnosis of breast cancer until 2010. Invasive cancer diagnosis was recorded regardless of whether made within or outside of the program. Median follow-up = 13.3 years[a] Note. In this review, only 5-year data is extracted	**5-year risk^ by quintile** Q1 Q2 Q3 Q4 Q5	 Ages 50-69 entry (19.5%) 0.86 (NR) Ages 50-69 entry (20.3%) 1.55 (NR) Ages 50-69 entry (20.1%) 1.00 (NR) Ages 50-69 entry (20.2%) 1.13 (NR) Ages 50-69 entry (20.0%) 1.39 (NR)
Tice 2008 , (USA), Prospective cohort, 1994-CD	Women aged ≥35 years who had ≥1 mammogram with BI-RADS in any of 7 mammography registries in the National Cancer Institute–funded Breast Cancer Surveillance Consortium (BCSC). Exclusion	NR	**BCSC/Tice model** (Breast density+ Breast biopsy + family history)	Women were entered into the model 6 months after initial mammogram and were censored at	**5-year risk by decile** D1 D2 D3 D4 D5	 All ages at entry (10%) 1.04 (0.92–1.18) All ages at entry (10%) 0.93 (0.84–1.02) All ages at entry (10%) 0.91 (0.83–0.99) All ages at entry (10%) 0.97 (0.89–1.05) All ages at entry (10%) 1.04 (0.97–1.12)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (Incidence data)	Follow-up	Estimated risk category	Invasive breast cancer (% in risk category) Expected/observed ratio (95%CI)
	criteria were diagnosis of breast cancer prior to first eligible mammography examination, breast cancer diagnosis within the first 6 months of follow-up and those with breast implants. Women diagnosed with DCIS were censored. When several mammograms were available, analysis was based on first mammogram only. Age at entry ranged 35-84y % <50 y = 45.9% % ≥70 y = 12.8% N= 629, 229 Note. The tool was developed in 60% of this population		(SEER 1992-2002) Risk factors obtained from self -report at mammography	death, DCIS diagnosis, or end of follow-up. Median follow-up = 5.3 years	D6 D7 D8 D9 D10	All ages at entry (10%) 0.95 (0.89–1.02) All ages at entry (10%) 1.04 (0.98–1.11) All ages at entry (10%) 0.99 (0.93–1.05) All ages at entry (10%) 0.99 (0.94–1.05) All ages at entry (10%) 1.10 (1.05–1.15)
Arrospide 2013, (Spain), Prospective cohort, 1995-1998	Women attending for the first time the breast cancer early-detection program in Sabadell-Cerdanyola (EDBC-SC) area in Catalonia, Spain between 1995-1998. No personal history of breast cancer. Risk factors collected via questionnaire at time of entry mammogram and breast density collected (BI-RADS) from initial mammogram. N initial =13,760 N final =13,709 Aged 50-69 Previously screened: No, first time participants	Biennial mammography for women aged 50-69.	**Barlow v1[c]** (BCSC 1996-2002) Risk factors collected via questionnaire at time of entry mammogram and breast density collected (BI-RADS) from initial mammogram	6 months from entry to 3 and 5 years. Followed for vital status or possible diagnosis of breast cancer until 2010. Invasive cancer diagnosis was recorded regardless of whether made within or outside of the program. Median follow-up = 13.3 years[a] Note. In this review, only 5-year data is extracted	**5-year risk by quintile** Q1 Q2 Q3 Q4 Q5	 Ages 50-69 at entry (19.8%) 1.82 (NR) Ages 50-69 at entry (20.8%) 2.83 (NR) Ages 50-69 at entry (19.0%) 3.44 (NR) Ages 50-69 at entry (20.7%) 4.20 (NR) Ages 50-69 at entry (19.9%) 3.29 (NR)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (Incidence data)	Follow-up	Estimated risk category	Invasive breast cancer (% in risk category) Expected/observed ratio (95%CI)
Petracci 2011, Prospective validation cohort within case-control study, 1998-2004	Validation cohort consisted of those from the Florence- European Prospective Investigation into Cancer and Nutrition cohort, (EPIC) cohort study aged 35-64 years. Exclusion criteria were women with prevalent breast cancer at recruitment, diagnosis of breast cancer within 6 months of recruitment, incomplete covariate data N= 8426 women with complete data for all risk factors (validation cohort) Age at entry ranged 35-64 years Previously screened: NR	NR	**Petracci model** (Florence Cancer Registry 1989-1993) Risk factors ascertained from self-reported standardised questionnaire at study entry.	Follow-up began 6 months after recruitment and continued until December 31, 2004 and did not end on date of breast cancer incidence or death. Final age was considered last age at lost to follow-up or age on December 31, 2004. Median follow-up = NR	**5-year risk^ by quintile** Q1 0-1.57	Aged 35-64 years at entry (20%) 0.79 (0.54-1.16)
					Q2 1.57-2.10	Aged 35-64 years at entry (20%) 1.00 (0.70-1.42)
					Q3 2.10-2.69	Aged 35-64 years at entry (20%) 0.93 (0.69-1.25)
					Q4 2.69-3.53	Aged 35-64 years at entry (20%) 0.99 (0.75-1.30)
					Q5 ≥3.53	Aged 35-64 years at entry (20%) 1.52 (1.16-1.98)
Rosner 2013, (USA), Prospective cohort study, 1995-2009	Women in the California Teachers Study (CTS) validation population were limited to those who were postmenopausal at baseline. Age range of women in CTS was limited to 47-74 years so that they could be compared to NHS participants. Only baseline data was used from the CTS due to variable questionnaire follow-up collection. Age at entry ranged 47-74 years; N=22641[b] Previously screened: NR	NR	**Rosner-Colditz** Refit using Nurses' Health Study data (1994-2008) and validated in CTS (1995-2009) cohorts	Five-year follow-up period (no more detail available). Median follow-up = NR	**5-year risk by decile** D1	Aged 47-74 years at entry (10%) 0.93 (NR)
					D2	Aged 47-74 years at entry (10%) 0.94 (NR)
					D3	Aged 47-74 years at entry (10%) 0.92 (NR)
					D4	Aged 47-74 years at entry (10%) 0.97 (NR)
					D5	Aged 47-74 years at entry (10%) 1.02 (NR)
					D6	Aged 47-74 years at entry (10%) 1.12 (NR)
					D7	Aged 47-74 years at entry (10%) 1.05 (NR)
					D8	Aged 47-74 years at entry (10%) 1.21 (NR)
					D9	Aged 47-74 years at entry (10%) 1.06 (NR)
					D10	Aged 47-74 years at entry (10%) 1.09 (NR)
Powell 2014, (USA), Retrospective cohort, 2003-2007	Marin Women's Study (MWS), a mammography-based study of women in Marin County, California where the study was performed at all major screening centres in the county, including those associated	NR	**BRCAPRO** (Age-dependent penetrance and prevalence based on systematic	Five-year follow-up period with confirmation of breast cancer between 2003-2007 either self-reported	**5-year risk^ by decile** D1	All ages at entry (10%) 0.51 (NR)
					D2	All ages at entry (10%) 0.53 (NR)
					D3	All ages at entry (10%) 0.88 (NR)
					D4	All ages at entry (10%) 0.67 (NR)
					D5	All ages at entry (10%) 0.58 (NR)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (Incidence data)	Follow-up	Estimated risk category	Invasive breast cancer (% in risk category) Expected/observed ratio (95%CI)
	with Kaiser Permanente, Marin General Hospital, and Novato Community Hospitals. The screening centres contribute to the San Francisco Mammography Registry (SFMR) which participates in the Breast Cancer Surveillance Consortium (BCSC). All women were invited to participate no matter their history, however, the population of interest was restricted to women who were breast cancer free at baseline. Note. Women in this County had higher rates of breast cancer, nulliparity and delayed childbirth. N = 12,843 Age at entry ranged <40 - ≥80 y; % <50 y = 46.3% % ≥70 y = 7.3% Previously screened: those with screening history allowed		review of the literature) Risk factors ascertained from self-reported questionnaire as well as from linkage with SFMR for breast density, demographic data, BMI and family history. Questionnaire data were collected from 2006-2009 with the reference baseline set at beginning of 2003	on the questionnaire or included in the SFMR cancer registry data. Median follow-up = NR	D6	All ages at entry (10%) 0.92 (NR)
					D7	All ages at entry (10%) 0.61 (NR)
					D8	All ages at entry (10%) 0.48 (NR)
					D9	All ages at entry (10%) 0.40 (NR)
					D10	All ages at entry (10%) 0.64 (NR)

Table 7. Within tool comparisons: Studies comparing expected/observed breast cancer (DCIS included) ratios for different risk categories of a given breast cancer risk assessment tool (scoping review)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (incidence data)	Follow-up	Estimated risk category	Breast cancer (DCIS included) (% in risk category) Expected/observed ratio (95%CI)
Evans 2016[a], (United Kingdom), Prospective cohort, 1987-2013	Women referred to University Hospital of South Manchester with high risk family history clinic based on family history of breast/other cancers as well as hormonal and reproductive factors. Women underwent breast examination and mammography. Excluded if had prior breast cancer, prevalent cancers also excluded. N = 9527 Median age at entry = 39 years Previously screened: NR	NR	**Tyrer-Cuzick v6** (Cancer Statistics Registrations, England and Wales 1994) Comprehensive breast cancer risk assessment completed at entry. (Note. DCIS accounted for in about 20% of cases)	Date of last follow-up was usually 1 August 2011. Censoring if women had left the area at diagnosis, risk-reducing mastectomy or last mammogram if this was after 1 August 2011. Median follow-up = 10.2 years	**10-year risk Overall**	All ages at entry (100%) 0.92 (0.83-1.02)
					0-1%	All ages at entry (12.3%) 0.52 (0.34-0.84)
					1-2%	All ages at entry (17.3%) 0.89 (0.65-1.25)
					2-3%	All ages at entry (17.5%) 0.97 (0.74-1.30)
					3-4%	All ages at entry (15.8%) 0.86 (0.68-1.10)
					4-5%	All ages at entry (11.6%) 0.83 (0.65-1.09)
					5-8%	All ages at entry (18.2%) 0.99 (0.82-1.22)
					8-12%	All ages at entry (5.4%) 1.08 (0.79-1.49)
					≥12%	All ages at entry (1.8%) 1.03 (0.69-1.61)
Evans 2016[a], (United Kingdom), Prospective cohort, 1987-2013	Women referred to University Hospital of South Manchester with high risk family history clinic based on family history of breast/other cancers as well as hormonal and reproductive factors. Women underwent breast examination and mammography. Excluded if had prior breast cancer, prevalent cancers also excluded. N = 9527 Median age at entry = 39 years Previously screened: NR	NR	**Gail v2** (SEER 1983-1987)[b] Comprehensive breast cancer risk assessment completed at entry. (Note. DCIS accounted for in about 20% of cases)	Date of last follow-up was usually 1 August 2011. Censoring if women had left the area at diagnosis, risk-reducing mastectomy or last mammogram if this was after 1 August 2011. Median follow-up = 10.2 years	**10-year risk Overall**	All ages at entry (100%) 0.99 (0.90-1.10)
					0-1%	All ages at entry (12.8%) 0.50 (0.32-0.81)
					1-2%	All ages at entry (16.8%) 0.64 (0.48-0.87)
					2-3%	All ages at entry (15.9%) 0.75 (0.58-0.98)
					3-4%	All ages at entry (11.2%) 0.77 (0.58-1.04)
					4-5%	All ages at entry (17.7%) 1.07 (0.86-1.35)
					5-8%	All ages at entry (16.1%) 1.07 (0.88-1.32)
					8-12%	All ages at entry (5.9%) 1.30 (0.97-1.79)
					≥12%	All ages at entry (3.5%) 2.10 (1.39-3.33)
Vacek 2011, (USA), Retrospective cohort, 1996-2010	Women aged 70 years or older undergoing at least one mammogram in the Vermont Breast Cancer Surveillance System (VBCSS; one of the participating registries in the NCI's BCSC) between 1996-2001, had not been previously diagnosed with breast cancer	NR	**Gail v1** (BCDDP 1973-1980)	Follow-up started one year after entry mammogram. Censoring for those without cancer	**100,891 person year risk by Quintile** (as calculated by review team from age adjusted data	

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (incidence data)	Follow-up	Estimated risk category	Breast cancer (DCIS included) (% in risk category) Expected/observed ratio (95%CI)
	and accepted use of their data for research. Cancers diagnosed or lost to follow-up within a year of entry to mammogram were not included in the cohort. Women were included in analysis only if they had complete data for each risk model. Note. Some women in this cohort had mammograms for diagnostic purposes N = 12,721/ 19779 (100,891/141,034 person years) had complete data for Gail. Age at entry ranged 70 - ≥85y; For the total cohort: % 70-74 y = 54.6% % 75-79 y = 24.5% % 80-84 y = 13.0% % ≥85y = 7.9% Previously screened: NR		Risk factor information obtained by questionnaire at time of mammography	at either the last mammogram or benign biopsy recorded in the VBCSS or the last Medicare claims record, whichever came later. Mean follow-up = 7.1 years	presented in Fig 1 in the study) Q1 Q2 Q3 Q4 Q5	 Age ≥70 (age adjusted) at entry (27%) 1.0 (CD) Age ≥70 (age adjusted) at entry (16%) 1.2 (CD) Age ≥70 (age adjusted) at entry (21%) 1.0 (CD) Age ≥70 (age adjusted) at entry (16%) 1.5 (CD) Age ≥70 (age adjusted) at entry (20%) 2.1 (CD)
Vacek 2011, (USA), Retrospective cohort, 1996-2010	Women aged 70 years or older undergoing at least one mammogram in the Vermont Breast Cancer Surveillance System (VBCSS; one of the participating registries in the NCI's BCSC) between 1996-2001, had not been previously diagnosed with breast cancer and accepted use of their data for research. Cancers diagnosed or lost to follow-up within a year of entry to mammogram were not included in the cohort. Women were included in analysis only if they had complete data for each risk model. Note. Some women in this cohort had mammograms for diagnostic purposes N = 11,002/19,779 (79,599/141,034 person years) had complete data for Barlow. Age at entry ranged 70 - ≥85y; For the total cohort: % 70-74 y = 54.6% % 75-79 y = 24.5% % 80-84 y = 13.0% % ≥85y = 7.9% Previously screened: NR	NR	**Barlow**[c] (BCSC 1996-2002) Risk factor information obtained by questionnaire at time of mammography	Follow-up started one year after entry mammogram. Censoring for those without cancer at either the last mammogram or benign biopsy recorded in the VBCSS or the last Medicare claims record, whichever came later. Mean follow-up = 7.1 years	**79,599 person year risk by Quintile** (as calculated by review team from age adjusted data presented in Fig 1 in the study) Q1 Q2 Q3 Q4 Q5	 Age ≥70 (age adjusted) at entry (19%) 1.0 (CD) Age ≥70 (age adjusted) at entry (21%) 1.0 (CD) Age ≥70 (age adjusted) at entry (21%) 1.2 (CD) Age ≥70 (age adjusted) at entry (20%) 1.3 (CD) Age ≥70 (age adjusted) at entry (20%) 1.5 (CD)
Vacek 2011, (USA),	Women aged 70 years or older undergoing at least one mammogram in the Vermont Breast Cancer	NR	**BCSC/Tice model**	Follow-up started one year after entry	**78,128 person year risk by Quintile**	

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (incidence data)	Follow-up	Estimated risk category	Breast cancer (DCIS included) (% in risk category) Expected/observed ratio (95%CI)
Retrospective cohort, 1996-2010	Surveillance System (VBCSS; one of the participating registries in the NCI's BCSC) between 1996-2001, had not been previously diagnosed with breast cancer and accepted use of their data for research. Cancers diagnosed or lost to follow-up within a year of entry to mammogram were not included in the cohort. Women were included in analysis only if they had complete data for each risk model. Note. Some women in this cohort had mammograms for diagnostic purposes N = 9,900/ 19779 (78,128/141,034 person years) had complete data for BCSC/Tice model. Age at entry ranged 70 - ≥85y; For the total cohort: % 70-74 y = 54.6% % 75-79 y = 24.5% % 80-84 y = 13.0% % ≥85y = 7.9% Previously screened: NR		(SEER 1992-2002) Risk factor information obtained by questionnaire at time of mammography	mammogram. Censoring for those without cancer at either the last mammogram or benign biopsy recorded in the VBCSS or the last Medicare claims record, whichever came later. Mean follow-up = 7.1 years	(as calculated by review team from age adjusted data presented in Fig 1 in the study) Q1 Q2 Q3 Q4 Q5	 Age ≥70 (age adjusted) at entry (20%) 1.0 (CD) Age ≥70 (age adjusted) at entry (19%) 1.3 (CD) Age ≥70 (age adjusted) at entry (22%) 1.5 (CD) Age ≥70 (age adjusted) at entry (20%) 1.8 (CD) Age ≥70 (age adjusted) at entry (20%) 2.8 (CD)
Vacek 2011, (USA), Retrospective cohort, 1996-2010	Women aged 70 years or older undergoing at least one mammogram in the Vermont Breast Cancer Surveillance System (VBCSS; one of the participating registries in the NCI's BCSC) between 1996-2001, had not been previously diagnosed with breast cancer and accepted use of their data for research. Cancers diagnosed or lost to follow-up within a year of entry to mammogram were not included in the cohort. Women were included in analysis only if they had complete data for each risk model. Note. Some women in this cohort had mammograms for diagnostic purposes. Development population most likely included validation participants, however, women above 70 years only consisted of 13.4% of the development cohort. N = 11,390/19,779 (81,811/141,034 person years) had complete data for Vermont. Age at entry ranged 70 - ≥85y; For the total cohort: % 70-74 y = 54.6%	NR	**Vermont** (VBCSS 1996-2001) Risk factor information obtained by questionnaire at time of mammography	Follow-up started one year after entry mammogram. Censoring for those without cancer at either the last mammogram or benign biopsy recorded in the VBCSS or the last Medicare claims record, whichever came later. Mean follow-up = 7.1 years	**81,811 person year risk by Quintile** (as calculated by review team from age adjusted data presented in Fig 1 in the study) Q1 Q2 Q3 Q4 Q5	 Age ≥70 (age adjusted) at entry (21%) 1.0 (CD) Age ≥70 (age adjusted) at entry (20%) 1.3 (CD) Age ≥70 (age adjusted) at entry (20%) 1.1 (CD) Age ≥70 (age adjusted) at entry (20%) 1.5 (CD) Age ≥70 (age adjusted) at entry (19%) 2.6 (CD)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (incidence data)	Follow-up	Estimated risk category	Breast cancer (DCIS included) (% in risk category) Expected/observed ratio (95%CI)
	% 75-79 y = 24.5% % 80-84 y = 13.0% % ≥85y = 7.9% Previously screened: NR					

N = number of participants; CI = confidence interval; NR = not reported; CD = cannot determine; DCIS = ductal carcinoma in-situ; y = years; v = version; BCDDP = Breast Cancer Detection and Demonstration Project; SEER = Surveillance, Epidemiology, and End Results; NCI = National Cancer Institute; BCSC = Breast Cancer Surveillance Consortium; VBCSS = Vermont Breast Cancer Surveillance System;

a Manual model was excluded due to clinician input. The BOADICEA model was also excluded due to the population in which the model is validated

b Assumed dates from year study published and reference to NCI website

c The Barlow model was initially intended to predict invasive cancer or DCIS within 1-year of screening, this study has used this model over an extended period of time

Table 8. Within tool comparisons: Studies comparing expected/observed breast cancer (unclear if DCIS included) ratios for different risk breast cancer risk assessment tools (scoping review)

Study, (Country), study design, time period	Population	Screening protocol	Risk assessment tool (incidence data)	Follow-up	Estimated risk category	Breast cancer (unclear if DCIS included) Expected/observed ratio (95%CI) (% in risk category)
Quante 2012, (USA), Prospective cohort, 1995-NR	Women recruited to the New York site of the Breast Cancer Family Registry (BCFR) who had ≥ one subsequent update on cancer and vital status. Eligible ages upon entry were 20-70 years. Excluded those with history of invasive or in situ breast cancer and history of bilateral prophylactic mastectomy. Contains higher risk population: Eligible participants had ≥2 relatives with history of breast or ovarian cancer; diagnosis of breast or ovarian cancer at <45 years, personal history of both breast and ovarian cancer; an affected male with breast cancer in the family or known BRCA1 or BRCA2 mutation carriers. N=1857 Age at entry = 20-70 y 66.8% aged 20-49 y Previously screened: NR Cohort includes those with very high risk (mutation carriers) and at lower risk (mutation negative and/or with more distant relatives with cancer)	NR	**Tyrer-Cuzick v6.0.0** (Cancer Statistics Registrations, England and Wales 1994) Baseline collection of epidemiologic, multi-generational pedigree and genetic data (Note for TC, not explicitly stated in-text that outcome was only invasive breast cancer)	Followed up to 10 years or until invasive breast cancer or death. Censoring also occurred when there was incomplete 10-year follow-up. 76% of the cohort were observed for ≥5 years, and 4% were observed for ≤ 1 year. Mean follow-up = 8.1 years	**10-year risk by Quartile** (as calculated by review team from data presented in Fig 2) Q1	Ages 20-70 at entry (25%) 0.60 (CD)
					Q2	Ages 20-70 at entry (25%) 0.66 (CD)
					Q3	Ages 20-70 at entry (25%) 0.69 (CD)
					Q4	Ages 20-70 at entry (25%) 1.19 (CD)

Table 9. Characteristics of cohort studies using the Tyrer-Cuzick tool included in systematic review

First author and year of publication, study design[a]	***Country***	***Risk assessment tools validated (risk factor collection relevant to tool)***	***Validation population***	***Cohort size and age at entry (mean, median or range)***	***Screening protocol***	***Follow-up***
Invasive breast cancer assessed						
Jantzen 2021 Prospective	Canada	TC v8.0b	CARTaGENE. Women were enrolled between 2009 and 2010.	N=10,200 Age range: 40-69 years. Median: 53.1 years	2 yearly, 50-69y	5 years
Choudhury 2020 Prospective	USA	TC v.8 (self-report)	Validation performed on two cohorts of women from the UK-based Generations Study (GS) and the US-based Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO). The PLCO cohort was used to validate the TC tool Women in the GS were enrolled between 2003-2012 and were from the general population. The PLCO was a multicentre study (10 centres) enrolling women aged 55-74 between 1993-2001. Age range in the current study was restricted to 35-75 years. Previous screening exposure: NR	N = 64,874 <u>GS</u> Aged 42 years (median) for cohort <50 years; aged 58 years (median) for cohort ≥50 years <u>PLCO</u> Aged 61 years (median) for cohort ≥50 years Subgroups <u>GS:</u> <50 years N = 28,232 ≥50 years N = 36,642 <u>PLCO:</u> ≥50 years N = 48, 279	NR	For both cohorts: from entry until 5y after entry, time of last contact or linkage to cancer or death registries, whichever came first. Confirmation of cancer diagnosis for PLCO cohort by annual questionnaire, followed by medical record and periodic cancer registry check as available.GS cohort self-reports confirmed by National Health Service Central Registers, hospital or pathology records. Median follow-up = 5 years
McCarthy 2020 Prospective	USA	TC v7 and TC v8.0b	Newton-Wellesley Hospital (Newton, Massachusetts). Women were recruited between 2007-2009	N=35,921 Age range: 40-84; mean: 53.9 years	NR	6.7 years (mean)
Glynn 2019 Prospective	USA	TC v8[c] (self-report)	Nurses' health study (NHS) cohort with baseline and then biennial follow-up in the form of questionnaires via mail. Women were included if they had	N = 76,922 Aged <50- ≥70 years	NR	Follow-up was every 2y as long as women were alive, free of breast cancer and continued to report the latest risk factor information. Women who

First author and year of publication, study design[a]	***Country***	***Risk assessment tools validated (risk factor collection relevant to tool)***	***Validation population***	***Cohort size and age at entry (mean, median or range)***	***Screening protocol***	***Follow-up***
			complete data on baseline risk factors from 1980-2006. Previous screening exposure: NR			developed another type of cancer (except non-melanoma skin cancer) at diagnosis date were censored. Women with invasive breast cancer from 1980-2008 who had a pathology report were included in the analysis. Median follow-up = 24 years
Terry 2019	USA, Canada, Australia	TC v8.0b	Breast Cancer Prospective Family Study Cohort (ProF-SC). Women were recruited between 1992 and 2011	N=15,732 Age range: 20-70 years (no mean or median reported)	NR	11.1 years
Brentnall 2018 Prospective	USA	TC v7.02 (self-report) TC v7.02 + BD (self-report + BD (BIRADS) adjusted for BMI)	Women in Kaiser Permanente Washington BCSC attending ≥ one mammogram screening between 1996-2013 Previous screening exposure: NR	N =132,139 Aged 50 years (median) Subgroups < 50 years N = 60,185[b] 50-59 years N = 43,759 ≥ 60 years N = 28,195	Annual screening for women aged 50-75 years and for high-risk women aged 40-49 years	6m from entry until invasive breast cancer diagnosis confirmed by SEER tumour registry and pathology databases or censoring at DCIS diagnosis, death, age 75 y or December 2014. Median follow-up = 5.2 years
Powell 2014	USA	TC v6.0	Marin Women's Study (MWS). Women recruited between 1999-2004	N= 40,229 <40 to ≥80; (NR)	NR	NR

BD = breast density; BCSC = Breast Cancer Surveillance Consortium; BI-RADS = Breast Imaging and Reporting Data System; BMI = body mass index; DCIS = ductal carcinoma in-situ; E = expected; m = months; N = number of participants; NR = not reported; O = observed; TC = Tyrer-Cuzick; v = version; y = years.

[a] Study design is defined based on the timing of the data collection i.e. if the predictors are collected prior to outcome occurring; [c] Risk assessment tool version as indicated by author correspondence;

Table 10.Risk predictors included in published versions of the Tyrer-Cuzick tool assessed in included studies

Risk assessment tool	***Age***	***Age of menarche***	***Age of 1st live birth/ FTP***	***Age of menopause***	***Parity***	***Number of FTP***	***Menopausal status***	***HRT use***	***Oral contraceptive use***	***Breast density***	***Hysterectomy***	***BMI or weight***	***Height***	***Race and/ or ethnicity***	***Prior/ no. of breast biopsies***	***Other pathology***	***ADH***	***LCIS***	***History/ no. of degree relatives assessed***	***Age of onset of breast cancer***	***Bilateral breast cancer***	***Male breast cancer***	***Ovarian cancer (personal/ familial)***	***Alcohol intake***	***Smoking status***	***Time spent breastfeeding***	***HRT type/ duration of use***	***BRCA1/ BRCA2 status***	***Polygenic SNP score***	***StuNeed tody ID***
TC v6	x	x	x	x	x*			x				x					x	x	1st 2nd	x	x		x							Powel et al, 2014
TC v7	x	x	x	x	x			x				x	x						1st	x			x				x			Glynn et al 2019
TC v7	x	x	x	x	x			x						x	x		x	x	1st 2nd 3rd				X							McCarthy et al 2020
TC v7.02	x	x	x	x	x		x					x	x		x	x	x		1st 2nd											Brentnall et al 2018
TC v7.02 + BD	x	x	x	x	x		x			x		x	x		x	x	x		1st 2nd											Brentnall et al 2018
TC v8	x	x	x	x	x		x	x				x	x						1st											Choudhury et al 2020 (GS)
TC v8	x	x	x	x	x		x	x				x	x			x	x	x	1st 2nd				x				x			Jantzen et al 2021 (CC)
TC v8.0b	x	x	x	x	x			x		x				x	x		x	x	1st 2nd 3rd				x							McCarthy et al 2020
TC v8.0b	x	x	x		x		x	x				x	x		x				1st 2nd	x			x					x		Terry et al 2019

*Only age at first live birth was included.
Abbreviations: ADH = atypical ductal hyperplasia; FTP: full term pregnancy; HRT: hormone replacement therapy; LCIS = lobular carcinoma in situ;

Table 11. Summary of risk of bias for included prognostic risk assessment tool studies in systematic review.

Study	RAT	Cohort	Year	Outcome	Participants	Predictors	Outcome	Analysis[a]	Overall RoB
Jantzen 2021	TC v8	CARTaGENE	5	Invasive	Low	Low	Unclear	High	High
McCarthy 2020	TC v7	NWH	6	Invasive	High	Low	Unclear	High	High
McCarthy 2020	TC v8.0b	NWH	6	Invasive	High	Low	Unclear	High	High
Choudhury 2020	TC v8	GS	5	Invasive	Low	Unclear	Unclear	High	High
Choudhury 2020	TC v8	PLCO	5	Invasive	Low	Unclear	Unclear	High	High
Terry 2019	TC v8.0b	ProF-SC	5	Invasive	Low	High	High	High	High
Terry 2019	TC v8.0b	ProF-SC	10	Invasive	Low	High	High	High	High
Glynn 2019	TC v8	NHS	2	Invasive	High	High	Unclear	High	High
Brentnall 2018	TC v7.02	KPW-BCSC	10	Invasive	Low	High	Unclear	High	High
Brentnall 2018	TC v7.02 + BD	KPW-BCSC	10	Invasive	Low	High	Unclear	High	High
Powell 2014	TC v6.0.0	MWS	5	Invasive	Low	High	Unclear	High	High

a. Items 4.5, 4.8 and 4.9 omitted as they are signalling questions for model development and not validation; Key to domain and overall rating: High risk of bias – PN/N for any signalling question, high risk of bias in any domain; Low risk of bias – PY/Y for all signalling questions with no PN/N/NI ratings, low risk of bias in all domains with no moderate or high risk domains; Unclear risk of bias – NI in any signalling question with no N/PN ratings, unclear risk of bias in any domain with no high risk domains.

Key to overall rating

High	High risk of bias in any domain
Low	Low risk of bias in all domains, no moderate or high risk domains
Unclear	Unclear risk of bias in any domain, no high risk domains

BD = breast density; BCSC= Kaiser Permanente Washington Breast Cancer Surveillance Consortium; GS = Generations Study; MWS: Marin Women's Study; NHS: Nurses Health Study; N-W Hospital= Newton-Wellesley Hospital; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; ProF-SC= Breast Cancer Prospective Family Study Cohort; RAT = risk assessment tool; RoB = risk of bias; TC = Tyrer-Cuzick; v = version;
[a]Shading indicates studies included only under PICO1; [b]Risk assessment tool version as indicated by author correspondence.

Table 12. Detailed assessment of risk of bias of included prognostic risk assessment tool studies for systematic review

Study	RAT	Cohort	Year	Outcome	Participants			Predictors				Outcome							Analysis[a]							Overall
					1.1	1.2	RoB	2.1	2.2	2.3	RoB	3.1	3.2	3.3	3.4	3.5	3.6	RoB	4.1	4.2	4.3	4.4	4.6	4.7	RoB	RoB
Jantzen 2021	TC v8	CARTaGENE	5	Invasive	Y	Y	Low	Y	Y	Y	Low	PN	Y	Y	NI	PY	Y	Unclear	N	PY	N	N	Y	Y	High	High
McCarthy 2020	TC v7	NWH	6	Invasive	Y	N	High	PY	Y	Y	Low	Y	NI	Y	NI	NI	Y	Unclear	N	NI	Y	Y	Y	Y	High	High
McCarthy 2020	TC v8.0b	NWH	6	Invasive	Y	N	High	PY	Y	Y	Low	Y	NI	Y	NI	NI	Y	Unclear	N	NI	Y	Y	Y	Y	High	High
Choudhury 2020	TC v8	GS	5	Invasive	Y	Y	Low	NI	Y	NI	Unclear	Y	NI	Y	Y	NI	Y	Unclear	N	NI	Y	NI	Y	Y	High	High
Choudhury 2020	TC v8	PLCO	5	Invasive	Y	Y	Low	NI	Y	NI	Unclear	NI	NI	Y	PY	NI	Y	Unclear	N	NI	N	NI	Y	Y	High	High
Terry 2019	TC v8.0b	ProF-SC	5	Invasive	Y	Y	Low	PN	NI	NI	High	N	N	Y	PN	NI	NI	High	N	NI	N	N	Y	Y	High	High
Terry 2019	TC v8.0b	ProF-SC	10	Invasive	Y	Y	Low	PN	NI	NI	High	N	N	Y	PN	NI	N	High	N	NI	N	N	Y	Y	High	High
Glynn 2019	TC v8	NHS	2	Invasive	Y	N	High	PN	Y	Y	High	NI	PY	Y	NI	NI	Y	Unclear	Y	NI	N	N	Y	Y	High	High
Brentnall 2018	TC v7.02	KPW-BCSC	10	Invasive	Y	Y	Low	NI	Y	N	High	Y	NI	Y	NI	NI	NI	Unclear	Y	PY	Y	N	Y	Y	High	High
Brentnall 2018	TC v7.02 + BD	KPW-BCSC	10	Invasive	Y	Y	Low	NI	Y	N	High	Y	NI	Y	NI	NI	NI	Unclear	Y	PY	Y	N	Y	Y	High	High
Powell 2014	TC v6.0.0	MWS	5	Invasive	Y	Y	Low	PY	N	Y	High	PY	PY	Y	NI	PY	PY	Unclear	N	NI	Y	PN	PY	Y	High	High

a. Items 4.5, 4.8 and 4.9 omitted as they are signalling questions for model development and not validation; Key to domain and overall rating: High risk of bias – PN/N for any signalling question, high risk of bias in any domain; Low risk of bias – PY/Y for all signalling questions with no PN/N/NI ratings, low risk of bias in all domains with no moderate or high risk domains; Unclear risk of bias – NI in any signalling question with no N/PN ratings, unclear risk of bias in any domain with no high risk domains.

BD = breast density; GS = Generations Study; MWS: Marin Women's Study; N = No; NI = No Information; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; PN = Probably No; PY = Probably Yes; RAT = risk assessment tool; RoB = risk of bias; TC = Tyrer-Cuzick; UC = Unclear; v = version; Y = Yes.
[a]Shading indicates studies included only under PICO1; [b]Items 4.5, 4.8 and 4.9 omitted as they are signalling questions for model development and not validation; [c]Risk assessment tool version as indicated by author correspondence.

Key to domain and overall rating:

High	PN/N for any signalling question, high risk of bias in any domain;
Low	PY/Y for all signalling questions with no PN/N/NI ratings, low risk of bias in all domains with no moderate or high risk domains;
Unclear	NI in any signalling question with no N/PN ratings, unclear risk of bias in any domain with no high risk domains.

Table 13. Studies comparing calibration for different risk categories of a given breast cancer risk assessment tool (systematic review)

Study	*Risk assessment tool*	*Estimated risk category*	*Age groups (% in risk category)*	*E/O ratio (95% CI)*
Invasive breast cancer assessed				
Jantzen 2021	TC v8.0b CARTaGENE	**5-year risk by decile** Low risk <1%	40-69 years (32%)	2.38 (1.35-4.19)
		Intermediate risk ≥1% - <1.66%	40-69 years (50%)	0.78 (0.63-0.97)
		Average risk ≥1.66% - <3%	40-69 years (16%)	1.24 (0.84-1.84)
		High risk ≥3%	40-69 years (2%)	0.71 (0.38-1.32)
Choudhury 2020[b]	TC v8: GS cohort	**5-year risk by decile** D1	< 50 years (10%) ≥ 50 years (10%)	1.29 (0.62-2.72) 0.93 (0.64-1.35)
		D2	< 50 years (10%) ≥ 50 years (10%)	0.99 (0.59-1.68) 1.25 (0.88-1.78)
		D3	< 50 years (10%) ≥ 50 years (10%)	1.02 (0.64-1.64) 1.15 (0.85-1.56)
		D4	< 50 years (10%) ≥ 50 years (10%)	1.06 (0.68-1.64) 1.40 (1.01-1.93)
		D5	< 50 years (10%) ≥ 50 years (10%)	0.95 (0.65-1.39) 0.99 (0.76-1.29)
		D6	< 50 years (10%) ≥ 50 years (10%)	1.70 (1.04-2.77) 0.92 (0.72-1.16)
		D7	< 50 years (10%) ≥ 50 years (10%)	0.91 (0.66-1.26) 0.83 (0.67-1.02)
		D8	< 50 years (10%) ≥ 50 years (10%)	1.32 (0.91-1.90) 1.20 (0.95-1.52)
		D9	< 50 years (10%) ≥ 50 years (10%)	1.18 (0.87-1.62) 1.39 (1.10-1.74)
		D10	< 50 years (10%) ≥ 50 years (10%)	1.18 (0.92-1.51) 1.30 (1.08-1.56)
	TC v8: PLCO cohort	**5-year risk by decile** D1	50-75 years (NR)	0.67 (0.52-0.87)
		D2	50-75 years (NR)	0.62 (0.50-0.78)
		D3	50-75 years (NR)	0.90 (0.70-1.16)
		D4	50-75 years (NR)	0.61 (0.50-0.75)
		D5	50-75 years (NR)	0.61 (0.50-0.74)
		D6	50-75 years (NR)	0.63 (0.52-0.76)
		D7	50-75 years (NR)	0.62 (0.52-0.74)
		D8	50-75 years (NR)	0.71 (0.59-0.85)
		D9	50-75 years (NR)	0.70 (0.59-0.83)
		D10	50-75 years (NR)	0.89 (0.77-1.03)
McCarthy 2020	TC v7 N-W Hospital	**6-year risk by decide** D1	40-84 years (10%)	0.97 (0.68-1.38)
		D2	40-84 years (10%)	0.88 (0.66-1.18)
		D3	40-84 years (10%)	0.82 (0.63-1.07)
		D4	40-84 years (10%)	1.19 (0.88-1.60)
		D5	40-84 years (10%)	1.38 (1.02-1.87)
		D6	40-84 years (10%)	1.14 (0.88-1.47)
		D7	40-84 years (10%)	0.93 (0.74-1.16)
		D8	40-84 years (10%)	1.13 (0.90-1.41)
		D9	40-84 years (10%)	1.30 (1.05-1.62)
		D10	40-84 years (10%)	1.21 (1.02-1.43)
McCarthy 2020	TC v8.0b N-W Hospital	**6-year risk by decide** D1	40-84 years (10%)	0.70 (0.50-0.96)
		D2	40-84 years (10%)	1.02 (0.74-1.42)
		D3	40-84 years (10%)	0.91 (0.69-1.20)
		D4	40-84 years (10%)	1.04 (0.79-1.37)
		D5	40-84 years (10%)	1.38 (1.02-1.86)
		D6	40-84 years (10%)	1.13 (0.88-1.46)
		D7	40-84 years (10%)	1.29 (1.00-1.66)

Study	*Risk assessment tool*	*Estimated risk category*	*Age groups (% in risk category)*	*E/O ratio (95% CI)*
		D8	40-84 years (10%)	1.16 (0.93-1.44)
		D9	40-84 years (10%)	1.16 (0.95-1.41)
		D10	40-84 years (10%)	1.48 (1.24-1.76)
Terry 2019	TC v8.0b ProF-SC	**5-year risk** Q1	20-70 years (30%)	0.51 (0.36-0.72)
		Q2	20-70 years (22%)	0.93 (0.70-1.25)
		Q3	20-70 years (18%)	0.80 (0.63-1.01)
		Q4	20-70 years (30%)	1.16 (1.01-1.33)
		10-year risk Q1	20-70 years (24%)	0.65 (0.48-0.88)
		Q2	20-70 years (23%)	0.90 (0.73-1.11)
		Q3	20-70 years (20%)	0.89 (0.74-1.06)
		Q4	20-70 years (33%)	1.16 (1.05-1.28)
Glynn 2019	TC v8[e] Nurse's Health Study	**2-year risk by decile** D1	All ages (NR%) < 50 years (NR%) 50-59 years (NR%) 60-69 years (NR%) ≥ 70 years (NR%)	0.66 (0.53-0.84) 0.81 (0.70-0.95) 1.14 (0.37-3.55) 0.34 (0.05-2.44) 0.28 (0.07-1.13)
		D2	All ages (NR%) < 50 years (NR%) 50-59 years (NR%) 60-69 years (NR%) ≥ 70 years (NR%)	1.20 (0.95-1.53) 1.04 (0.89-1.21) 1.04 (0.79-1.35) 0.81 (0.42-1.55) 0.85 (0.44-1.64)
		D3	All ages (NR%) < 50 years (NR%) 50-59 years (NR%) 60-69 years (NR%) ≥ 70 years (NR%)	1.18 (0.96-1.46) 1.01 (0.84-1.22) 1.07 (0.90-1.27) 1.43 (0.91-2.24) 0.86 (0.59-1.26)
		D4	All ages (NR%) < 50 years (NR%) 50-59 years (NR%) 60-69 years (NR%) ≥ 70 years (NR%)	1.16 (0.95-1.41) 1.34 (1.01-1.78) 1.10 (0.95-1.29) 1.30 (0.97-1.73) 0.83 (0.63-1.10)
		D5	All ages (NR%) < 50 years (NR%) 50-59 years (NR%) 60-69 years (NR%) ≥ 70 years (NR%)	1.02 (0.86-1.21) 0.99 (0.72-1.38) 0.93 (0.81-1.06) 1.35 (1.08-1.69) 1.05 (0.81-1.36)
		D6	All ages (NR%) < 50 years (NR%) 50-59 years (NR%) 60-69 years (NR%) ≥ 70 years (NR%)	1.16 (0.98-1.38) 1.08 (0.70-1.68) 1.01 (0.88-1.16) 1.02 (0.86-1.19) 1.14 (0.90-1.44)
		D7	All ages (NR%) < 50 years (NR%) 50-59 years (NR%) 60-69 years (NR%) ≥ 70 years (NR%)	1.06 (0.91-1.25) 0.73 (0.48-1.11) 1.15 (0.98-1.34) 1.16 (1.00-1.35) 1.03 (0.84-1.27)
		D8	All ages (NR%) < 50 years (NR%) 50-59 years (NR%) 60-69 years (NR%) ≥ 70 years (NR%)	0.96 (0.83-1.1) 1.33 (0.79-2.25) 0.94 (0.81-1.09) 1.02 (0.91-1.15) 1.01 (0.84-1.21)
		D9	All ages (NR%) < 50 years (NR%) 50-59 years (NR%) 60-69 years (NR%) ≥ 70 years (NR%)	0.98 (0.87-1.12) 1.38 (0.88-2.16) 0.92 (0.80-1.07) 0.98 (0.88-1.08) 1.02 (0.86-1.21)
		D10	All ages (NR%) < 50 years (NR%) 50-59 years (NR%) 60-69 years (NR%)	1.48 (1.31-1.67) 1.49 (0.74-2.98) 1.35 (1.16-1.56) 1.32 (1.20-1.44)

Study	*Risk assessment tool*	*Estimated risk category*	*Age groups (% in risk category)*	*E/O ratio (95% CI)*
			≥ 70 years (NR%)	1.40 (1.22-1.60)
Brentnall 2018	TC v7.02 Kaiser Permanente Washington	10-year risk < 2%	All ages (36%) < 50 years (62%) 50-59 years (19%) ≥ 60 years (8%)	0.82 (0.76-0.89) 1.08 (0.99-1.19) 0.95 (0.81-1.12) 0.79 (0.55-1.15)
		2% to < 3%	All ages (32%) < 50 years (24%) 50-59 years (42%) ≥ 60 years (36%)	0.99 (0.92-1.06) 1.39 (1.22-1.59) 1.19 (1.17-1.32) 0.75 (0.65-0.87)
		3% to < 5%	All ages (22%) < 50 years (11%) 50-59 years (28%) ≥ 60 years (38%)	0.98 (0.91-1.05) 1.30 (1.52-1.12) 1.16 (1.04-1.30) 0.80 (0.71-0.91)
		5% to < 8%	All ages (7%) < 50 years (2%) 50-59 years (9%) ≥ 60 years (14%)	1.15 (1.03-1.28) 1.54 (1.16-2.08) 1.14 (0.97-1.33) 1.08 (0.90-1.30)
		≥ 8%	All ages (2%) < 50 years (1%) 50-59 years (2%) ≥ 60 years (4%)	1.27 (1.08-1.49) 1.49 (1.03-2.27) 1.37 (1.05-1.82) 0.97 (0.76-1.27)
	TC v7.02 + BD Kaiser Permanente Washington	10-year risk < 2%	All ages (40%) < 50 years (61%) 50-59 years (27%) ≥ 60 years (16%)	0.85 (0.79-0.93) 1.25 (1.14-1.39) 1.25 (1.08-1.45) 1.15 (0.85-1.56)
		2% to < 3%	All ages (25%) < 50 years (22%) 50-59 years (29%) ≥ 60 years (26%)	0.96 (0.89-1.04) 1.33 (1.18-1.54) 1.23 (1.09-1.39) 0.90 (0.76-1.08)
		3% to < 5%	All ages (22%) < 50 years (12%) 50-59 years (28%) ≥ 60 years (34%)	1.01 (0.93-1.08) 1.34 (1.19-1.59) 1.27 (1.12-1.41) 0.83 (0.74-0.94)
		5% to < 8%	All ages (8.6%) < 50 years (3%) 50-59 years (11%) ≥ 60 years (16%)	1.25 (1.12-1.39) 1.49 (1.19-1.89) 1.39 (1.20-1.61) 1.10 (0.93-1.32)
		≥ 8%	All ages (3.5%) < 50 years (1%) 50-59 years (4%) ≥ 60 years (7%)	1.28 (1.14-1.45) 1.59 (1.19-2.17) 1.27 (1.05-1.52) 1.09 (0.90-1.33)
Powell 2014	TC v6.0 Marin Women's Study	**5 year risk by decile** D1	>40 years (10%)	1.09 (0.95-1.25)
		D2	>40 years (10%)	1.36 (0.61-3.02)
		D3	>40 years (10%)	1.35 (0.70-2.60)
		D4	>40 years (10%)	1.24 (0.70-2.18)
		D5	>40 years (10%)	1.70 (0.91-3.15)
		D6	>40 years (10%)	0.99 (0.63-1.55)
		D7	>40 years (10%)	0.79 (0.54-1.16)
		D8	>40 years (10%)	0.90 (0.61-1.34)
		D9	>40 years (10%)	1.05 (0.70-1.57)
		D10	>40 years (10%)	0.98 (0.69-1.39)

AUC = area under the curve; BD = breast density; BMI = body mass index; CI = confidence interval; D = decile; DCIS = Ductal carcinoma in-situ; E = expected; GS = Generations Study N-W Hospital= Newton-Wellesley Hospital; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; m = months; NA = not applicable; NR = not reported; O = observed; TC = Tyrer-Cuzick; v = version; [b]PLCO cohort calibration data contained in figures across tools was provided by authors upon request; [e]Risk assessment tool version as indicated by author correspondence.

3.3.6 References

Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, McGinn T, Guyatt G. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. JAMA. 2017 Oct 10;318(14):1377-1384. doi: 10.1001/jama.2017.12126. PMID: 29049590.

Arrospide A, Forne C, Rue M, Tora N, Mar J, Bare M. An assessment of existing models for individualized breast cancer risk estimation in a screening program in Spain. BMC Cancer 2013; 13: 587.

Brentnall AR, Cuzick J, Buist DSM, Aiello Bowles EJ. Long-term accuracy of breast cancer risk assessment combining classic risk factors and breast density. JAMA Oncology 2018; 4(9): e180174.

Chay WY, Ong WS, Tan PH, Leo NQJ, Ho GH, Wong CS 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 Research 2012; 14: R19.

Choudhury PP, Wilcox AN, Brook MN, Zhang Y, Ahearn T, Orr N et al. Comparative validation of breast cancer risk prediction models and projections for future risk stratification. Journal National of the Cancer Institute 2020; 112(3): djz113.

Gao F, Machin D, Chow KY, Sim YF, Duffy SW, Matchar DB, et al. Assessing risk of breast cancer in an ethnically South-east Asia population (results of a multiple ethnic groups study). BMC Cancer 2012; 12: 529.

Glynn RJ, Colditz GA, Tamimi RM, Chen WY, Hankinson SE, Willett WW, et al. Comparison of questionnaire-based breast cancer prediction models in the Nurse's Health Study. Cancer, Epidemiology Biomarkers & Prevention 2019; 28(7): 1187-94.

Han Y, Lv J, Yu C, Guo Y, Bian Z, Hu Y, et al. Development and external validation of a breast cancer absolute risk prediction model in Chinese population. Breast Cancer Research. 2021; 23: (62).

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from: www.training.cochrane.org/handbook].

Hurson AN, Choudhury PP, Gao C, Husing A, Eriksson M, Shi M, et al. Prospective evaluation of a breast-cancer risk model integrating classical risk factors and polygenic risk in 15 cohorts from six countries. International Journal of Epidemiology. 2021; Mar 23: dyab036.

Husing A, Quante AS, Chang-Claude J, Aleksandrova K, Kaaks R, Pfeiffer RM. Validation of two US breast cancer risk prediction models in German women. Cancer Causes and Control. 2020; 31(6): 525-36.

Jantzen R, Payette Y, de Malliard T, Labbe C, Noisel N, Broet P. Validation of breast cancer risk assessment tools on a French-Canadian population-based cohort. BMJ Open. 2021; 11(4): e045078.

Jee YH, Gao C, Kim J, Park S, Jee SH, Kraft P. Validating breast cancer risk prediction models in the Korean Cancer Prevention Study-II Biobank. Cancer Epidemiology Biomarkers and Prevention. 2020; 29(6): 1271-7.

Lakeman IMM, Rodriguez-Girondo M, Lee A, Ruiter R, Stricker BH, Wijnant SRA, et al. Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort. Genetics in Medicine. 2020; 22(11): 1803-11.

Li K, Anderson G, Viallon V, Arveux P, Kvaskoff M, Fournier A et al. Risk prediction for estrogen receptor-specific breast cancers in two large prospective cohorts. Breast Cancer Research 2018; 20: 147.

MacInnis RJ, Bickerstaffe A, Apicella C, Dite GS, Dowty JG, Aujard K, et al. Prospective validation of the breast cancer risk prediction model BOADICEA and a batch-mode version BOADICEACentre. British Journal of Cancer. 2013; 109(5): 1296-301.

Matsuno RK, Costantino JP, Ziegler RG, Anderson GL, Li H, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. Journal of the National Cancer Institute. 2011; 103(12): 951-61.

McCarthy AM, Guan Z, Welch M, Griffin ME, Sippo DA, Deng Z et al. Performance of breast cancer risk assessment models in a large mammography cohort. Journal of the National Cancer Institute 2020; 112(5): djz177.

Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat. 2012 Apr;132(2):365-77. doi: 10.1007/s10549-011-1818-2. Epub 2011 Oct 22. PMID: 22037780.

Min JW, Chang MC, Lee HK, Hur MH, Noh DY, Yoon JH, et al. Validation of risk assessment models for predicting the incidence of breast cancer in Korean women. Journal of Breast Cancer. 2014; 17(3): 226-35.

Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012 May;98(9):691-8. doi: 10.1136/heartjnl-2011-301247. Epub 2012 Mar 7. PMID: 22397946.

Moons KGM*, Wolff RF*, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019 Jan 1;170(1):W1-W33. doi: 10.7326/M18-1377. PMID: 30596876.

Nickson C, Procopio P, Velentzis LS, Carr S, Devereux L, Mann GB, et al. Prospective validation of the NCI Breast Cancer Risk Assessment Tool (Gail model) on 40,000 Australian women. Breast Cancer Research. 2018; 20(1): (155).

Pastor-Barriuso R, Ascunce N, Ederra M, Erdozain N, Murillo A, Ales-Martinez JE, et al. Recalibration of the Gail model for predicting invasive breast cancer risk in Spanish women: a population-based cohort study. Breast Cancer Research and Treatment. 2013; 138(1): 249-59.

Petracci E, Decarli A, Schairer C, Pfeiffer RM, Pee D, Masala G, et al. Risk factor modification and projections of absolute breast cancer risk. Journal of the National Cancer Institute. 2011; 103(13): 1037-48.

Powell M, Jamshidian F, Cheyne K, Nititham J, Prebil LA, Ereman R. Assessing breast cancer risk models in Marin County, a population with high rates of delayed childbirth. Clinical Breast Cancer. 2014; 14(3): 212-20.

Rosner BA, Colditz GA, Hankinson SE, Sullivan-Halley J, Lacey Jr JV, Bernstein L. Validation of Rosner-Colditz breast cancer incidence model using an independent data set, the California Teachers Study. Breast Cancer Research and Treatment. 2013; 142(1): 187-202.

Schonberg MA, Li VW, Eliassen AH, Davis RB, LaCroix AZ, McCarthy EP, et al. Performance of the Breast Cancer Risk Assessment Tool among women aged 75 years and older. Journal of the National Cancer Institute. 2016; 108(3): djv348.

Terry MB, Liao Y, Whittemore AS, Leoce N, Buchsbaum R, Zeinomar N, et al. 10-year performance of four models of breast cancer risk: a validation study. The Lancet Oncology. 2019; 20(4): 504-17.

Vandenbroucke JP. A shortcut method for calculating the 95 per cent confidence interval of the standardized mortality ratio. American Journal of Epidemiology. 1982 Feb 1;115(2):303-4.

Vacek PM, Skelly JM, Geller BM. Breast cancer risk assessment in women aged 70 and older. Breast Cancer Research and Treatment. 2011; 130(1): 291-9.

Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S; PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019 Jan 1;170(1):51-58. doi: 10.7326/M18-1376. PMID: 30596875.

3.4 Appendix

3.4.1 Registered PROSPERO Protocol

CRD42020159232. Available from: link
https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020159232

1. Review title

How well do breast cancer risk assessment tools stratify adult women into population-level breast cancer risk groups? A systematic review.

2. Original language title

N/A

3. Anticipated or actual start date

9 December 2019

4. Anticipated completion date

31 December 2021

5. Stage of review at time of this submission

The review has not yet started: No, the review has started.

Review stage	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	Yes	Yes
Data extraction	Yes	No
Risk of bias (quality) assessment	Yes	No
Data analysis	No	No

Funded proposal: Funding has been provided by the Commonwealth Department of Health, Australia.

6. Named contact

Victoria Freeman

7. Named contact email

victoria.freeman@nswcc.org.au

8. Named contact address

N/A

9. Named contact phone number

N/A

10. Organisational affiliation of the review

Daffodil Centre

11. Review team members and their organisational affiliations

Ms Victoria Freeman, The Daffodil Centre, The University of Sydney, a joint venture with Cancer Council NSW

Dr Denise Campbell, The Daffodil Centre, The University of Sydney, a joint venture with Cancer Council NSW

Ms Suzanne Hughes, The Daffodil Centre, The University of Sydney, a joint venture with Cancer Council NSW

Dr Julia Steinberg, The Daffodil Centre, The University of Sydney, a joint venture with Cancer Council NSW

Dr Qingwei Luo, The Daffodil Centre, The University of Sydney, a joint venture with Cancer Council NSW

Dr Louiza Velentzis, The Daffodil Centre, The University of Sydney, a joint venture with Cancer Council NSW; Melbourne School of Population & Global Health, University of Melbourne, Victoria, Australia

Associate Professor Carolyn Nickson, The Daffodil Centre, The University of Sydney, a joint venture with Cancer Council NSW; Melbourne School of Population & Global Health, University of Melbourne, Victoria, Australia

12. Funding sources/sponsors

Commonwealth Department of Health, Australia.

13. Conflicts of interest

None

14. Collaborators

Professor G Bruce Mann – Clinical advisor

Breast Service, Royal Women's and Royal Melbourne Hospital, Australia; Department of Surgery, University of Melbourne, Australia

Paul Vardon – Advisor on research translation

Queensland Health, Cancer Screening Unit, Preventive Health Branch, Australia

15. Review question

How well do breast cancer risk assessment tools stratify asymptomatic adult women into population-level breast cancer risk groups?

PICO1: For asymptomatic women, how does a given breast cancer risk assessment tool perform in predicting breast cancer risk across the risk groups determined by the tool (within tool comparisons)?

PICO2: For asymptomatic women, how do different breast cancer risk assessment tools compare in their ability to predict breast cancer risk across the risk groups determined by each of the tools (between tool comparison)?

16. Searches

Medline (including MEDLINE Epub Ahead of Print, I-Process & Other Non-Indexed Citations), Embase and Cochrane Database of Systematic Reviews (CDSR) databases will be searched. Database-specific subject headings and text terms will be combined where available for terms related to breast cancer, risk assessment, and calibration. Searches will be limited to articles published in English from the 1st January 2008 onwards. Reference lists of full text articles identified for consideration will be checked for additional potentially relevant studies.

17. URL to search strategy

Uploaded, not publicly available until the review is complete.

18. Condition or domain being studied

Newly diagnosed breast cancer

19. Participants/population

Inclusion: Asymptomatic women aged ≥18 years.

Exclusion: Populations restricted to women undergoing breast imaging as follow-up for breast cancer or after an abnormal screen, or African American or Hispanic American populations, or groups deemed to be at high risk for breast cancer.

20. Intervention(s), exposure(s)

Inclusion: Risk assessment tools, including abridged tools and tools developed for high-risk populations (if the study population is the general population).

Exclusion: Risk assessment tools that include subjective inputs (such as clinical judgement).

21. Comparator(s)/control

For PICO1, no comparator.

For PICO2, different risk assessment tools where those tools assessed have comparable outcomes and periods of risk.

22. Types of study to be included

Inclusion: For PICO1, external validation cohort studies; for PICO2, RCTs or paired cohort studies in which multiple risk assessment tools are externally validated in the same study group. For the purposes of this review, external validation refers to studies that assess the predictive performance of existing risk assessment tools using data external to the development sample (ie. from different participants). (Moons and Wolff 2019)

Exclusion: Case-cohort, nested case-control and case-control studies; internal validation studies where the same study group or subset of the study group was used for both tool development and validation.

23. Context

Inclusion: Studies that report outcomes (breast cancer) for women in specific tool-determined risk groups.

The review was limited to studies published from 2008 onwards in order to select studies likely to use more relevant imaging methods and more recent versions of risk assessment tools, while not excluding studies with long follow-up (which likely use earlier versions of tools and superseded imaging methods).

Exclusion: Studies that do not report outcomes (breast cancer) for women in specific tool-determined risk groups ie. report outcomes (such as a goodness of fit statistic) across the entire population of women regardless of the individuals' risk scores.

24. Main outcome(s)

The primary outcome will be incidence of breast cancer (invasive with or without in-situ) for each tool-determined risk group.

Timing and effect measures

To assess calibration performance, the effect measure will be the ratio of the expected (E) number of breast cancers (based on the risks for individuals predicted by the tool) to the observed (O) number of breast cancers (E/O) and corresponding 95% confidence interval (CI) reported for tool-determined risk groups over 5 years, 10 years or lifetime follow-up. Studies to be included will need to provide the E/O ratio and 95% CI or include sufficient data to calculate E/O and the 95% CI.

25. Additional outcome(s)

Breast cancer mortality for tool-determined risk groups.

Incidence for different types of breast cancer defined by characteristics such as tumour subtype, grade, size, nodal involvement for each tool-determined risk group.

Interval breast cancers for tool-determined risk groups.

26. Data extraction (selection and coding)

Titles and abstracts of identified articles will be screened against pre-specified inclusion criteria. This will be split equally between two reviewers with 20% assessed by both reviewers to ensure concordance. Full-texts of potentially relevant articles will be independently assessed for inclusion by two reviewers using pre-specified selection criteria. Disagreements will be resolved by third reviewer adjudication. Authors will be contacted for further information if reviewers are unclear whether a study meets inclusion criteria. Reviewers will not be blinded to journal titles or study authors/institutions. Reasons for exclusion will be recorded for all excluded full-text articles.

Extraction of study characteristics and results will be split equally and arbitrarily between two reviewers followed by accuracy checks. Disagreements will be resolved by third reviewer adjudication.

The following characteristics will be extracted for each included study:

1. Study information (first author, publication year, country, study design, setting, recruitment period, participant inclusion/exclusion criteria, screening protocol).
2. Participants' characteristics (number, age and percentage previously screened).
3. Risk assessment tool information (tool and version, incidence data used for predictions, discrimination (c-statistic, AUC) if reported, whether an abridged tool was used and risk factors not included/assessed, when and how tool risk factors were collected, data collection years)
4. Follow-up (period, censoring mechanisms used, median follow-up)
5. Reported relevant outcomes stratified by estimated risk and any subgroups analysed separately

6. Other relevant information including statistical methods used and factors potentially affecting risk of bias.

The following data will be extracted for each included study:

1. Description of each tool-determined risk group and participant numbers and/or percentages per group.
2. For breast cancer incidence E, O and E/O +/- 95% CI and observed breast cancer rates for each tool-determined risk group, and if reported, assessment of discrimination performance (c-statistic, AUC).
3. For other outcomes, the outcome rates for each tool-determined group.

27. Risk of bias (quality) assessment

Two reviewers will independently assess the risk of bias for each included study. Differences will be discussed and if no consensus is reached, a third reviewer will adjudicate.

The risk of bias for RCTs will be assessed using the Cochrane Collaboration Risk of Bias-II tool. (Higgins 2019).

The risk of bias for cohort studies will be assessed using the Prediction model study Risk Of Bias Assessment Tool (PROBAST) designed to specifically assess the risk of bias of prediction model studies. (Moons and Wolff 2019) Authors will be contacted for details of the risk assessment tool that are required for the risk of bias assessment if not readily available.

High risk of bias will be considered when assessing the strength of the evidence.

For each study, the PROBAST specifies that a separate assessment should be conducted for each distinct risk assessment tool being evaluated. Therefore, for PICO2, a single study will have an overall risk of bias rating for each tool evaluated.

28. Strategy for data synthesis

For PICO1, we will compare E/O ratios for the different breast cancer risk groups determined by a specific tool; For PICO2, E/O ratios for tool determined breast cancer risk groups will be compared between different risk assessment tools.

A narrative synthesis and graphs of risk assessment tool performance by tool-determined risk group will be provided.

The narrative synthesis and graphs will be presented for each PICO question. For each PICO question results will be analysed separately for each specific outcome e.g. 5-year invasive breast cancer incidence, 10-year invasive breast cancer or in-situ incidence.

For PICO1, results will be grouped by tool; comparison of results for a specific outcome from different studies will enable comparison of the tool's performance in different populations; comparison of 5-year and 10-year outcomes for the same tool in the same population will enable analysis of tool performance over different lengths of follow-up.

For PICO2, the performance of different tools across risk groups within each study will be compared for specific outcomes.

For both PICOs, calibration results (E/O) will be visually synthesised using bar graphs with risk groups plotted on the horizontal axis and E/O ratio with 95% CI on the vertical axis.

A sub-analysis will be undertaken of those studies that report discrimination performance in addition to calibration performance.

29. Analysis of subgroups or subsets

Analysis of subgroups defined by risk factors such as age will be undertaken if data are available.

30. Type and method of review

Systematic review

Health area of review: Cancer, Public Health

31. Language

English

32. Country

Australia

33. Other registration details

None

34. Reference and/or URL for published protocol

None

35. Dissemination plans

A paper will be submitted to a peer-reviewed journal in this field for publication. The results will also be communicated to key stakeholder groups including the Australian Government Department of Health, BreastScreen Australia, oncologists, general practitioners, cancer centres and consumer groups.

36. Keywords

Breast neoplasms; risk assessment; risk prediction; mammography; mass screening; breast density; calibration; discrimination

37. Details of any existing review of the same topic by the same authors. (50 words)

None.

38. Current review status

Ongoing.

39. Any additional information

This review is being undertaken as part of the Roadmap to Optimising Screening in Australia (ROSA; formerly Optimising Early Detection of Breast Cancer Australia) project funded by the Australian Department of Health and led by Cancer Council Australia.

40. Details of final report/publication(s)

N/A

3.4.2 Search strategy

Systematic review

Table 14. Search strategy used for systematic review. Database(s) searched included Embase Classic+Embase 1947 to 2021 June 30, Ovid MEDLINE(R) ALL 1946 to June 30, 2021

#	Searches	Results
1	exp Breast Neoplasms/	896844
2	(breast adj3 (cancer* or tumour* or tumor* or carcinoma* or neoplasm* or screen*)).tw.	848153
3	1 or 2	1076644
4	exp risk assessment/ or breast cancer risk assessment tool/	907727
5	((risk* or susceptib* or predict*) adj6 (tool* or score* or model* or questionnaire* or instrument* or appraisal* or calculation* or calculator* or algorithm*)).tw.	973839
6	risk factor* calculat*.tw.	119
7	(risk adj2 (assess* or predict*)).tw.	358216
8	(observed* adj4 (expected* or predict* or assigned)).tw.	51471
9	observed-to-expected.tw.	12105
10	expected-to-observed.tw.	3138
11	("E/O" or "O/E").tw.	16563
12	(validat* or calibrat* or area under the curve or c-statistic or AUC or AUROC or receiver operat*).tw.	1993304
13	(clinical conference or conference or comment or editorial).tw.	415278
14	(neoadjuvant or recurrent or recurrence or chemotherapy).tw.	2326698
15	4 or 5 or 6 or 7	1916826
16	8 or 9 or 10 or 11 or 12	2048868
17	3 and 15 and 16	9199
18	17 not 13	9166
19	18 not 14	6397
20	limit 19 to (english language and humans and yr="2008 -Current")	4900
21	remove duplicates from 20	3463

Scoping review

Table 15. Search strategy used for scoping review. Databases searched included Embase Classic+Embase 1947 to 2019 January 21, Ovid MEDLINE(R) ALL 1046 to December 13, 2018

#	Searches	Results
1	(breast adj3 (cancer* or tumour* or tumor* or carcinoma* or neoplasm*)).tw.	685084
2	((risk* or susceptib*) adj6 (tool* or score* or model* or questionnaire*)).tw.	261361
3	(risk assessment or risk prediction).tw.	137304
4	(density adj3 (mammogr* or breast)).tw.	6864
5	2 or 3 or 4	368892
6	observed*.tw.	5542165
7	expected*.tw.	814205
8	predicted*.tw.	868350
9	6 and 7	146092
10	6 and 8	126749
11	(valid* or calibrat*).tw.	1678955
12	9 or 10 or 11	1920463
13	1 and 5 and 12	2556
14	limit 13 to English language	2505
15	limit 14 to human	2207
16	limit 15 to yr=”2008 -Current”	1871
17	(prognos* or neoadjuvant or recurrent or recurrence*).tw.	2428174
18	16 not 17	1225
19	remove duplicates from 18	825

3.4.3 Excluded studies

Table 16. Potentially relevant articles collected and excluded for the scoping and systematic review.

Article	*Scoping review*	*Systematic review*	*Reason for Exclusion*
Anothaisintawee 2012	Exclude	N/A	E/O not reported by risk category (only overall)
Arrospide 2013	Exclude	N/A	TC tool not used
Banegas 2011	Exclude	N/A	E/O not reported by risk category (only overall)
Brentnall 2015	Exclude	Exclude	No outcome metric of interest
Chay 2012	Exclude	N/A	TC tool not used
Dartois 2015	Exclude	N/A	Ineligible population (internal validation)
Eriksson 2017	Exclude	N/A	Ineligible study design (nested case-control)
Evans 2016	Exclude	Exclude	Ineligible population (high risk population)
Fung 2019	Exclude	N/A	No outcome metric of interest
Gao 2012	Exclude	Include	Ineligible study type (risk assessment tool development)
Hughes 2019	Exclude	N/A	No outcome metric of interest
Hung 2019	Exclude	N/A	E/O not reported by risk category (only overall)
Kerlikowske 2015	Exclude	N/A	Ineligible study type (risk assessment tool development)
Kurian 2009	Exclude	N/A	Ineligible population (breast cancer patients)
Lakeman 2020	Exclude	N/A	Ineligible study type (risk assessment tool development)
Li 2018	Exclude	N/A	TC tool not used
Lo 2018	Exclude	N/A	Ineligible population (LCIS patients at baseline)
Lophatananon 2017	Exclude	N/A	Ineligible study design (case-control)
Louro 2019	Exclude	N/A	Inclusion of ineligible study design and/or no risk stratified E/O estimates
MacInnis 2019	Exclude	N/A	E/O not reported by risk category (only overall)
McCarthy 2013	Exclude	Include	N/A No outcome metric of interest
Meads 2012	Exclude	N/A	E/O not reported by risk category (only overall)
Min 2014	Exclude	N/A	TC tool not used
Pastor-Barriuso 2013	Exclude	N/A	TC tool not used
Pfeiffer 2013	Exclude	N/A	E/O not reported by risk category (only overall)
Phillips 2019	Exclude	N/A	Ineligible population (high risk population)
Pu 2014	Exclude	N/A	Ineligible publication type (Review)
Quante 2012	Exclude	Exclude	Ineligible population (high risk population)
Quante 2015	Exclude	Exclude	Ineligible population (high risk population)

Article	***Scoping review***	***Systematic review***	***Reason for Exclusion***
Roman 2019	Exclude	N/A	Inclusion of ineligible study design and/or no risk stratified E/O estimates
Schonfeld 2010	Exclude	N/A	E/O not reported by risk category (only overall)
Shieh 2016	Exclude	N/A	Ineligible study design (nested case-control)
Sontag 2011	Exclude	N/A	No outcome metric of interest
Stegeman 2012	Exclude	N/A	E/O not reported by risk category (only overall)
Stevanato 2019	Exclude	N/A	Inegible study design (cross-sectional)
Taghipour 2012	Exclude	N/A	Ineligible study type (risk assessment tool development)
Tice 2015	Exclude	N/A	Ineligible study type (risk assessment tool development)
Tice 2019	Exclude	N/A	E/O not reported by risk category (only overall)
Tworoger 2014	Exclude	N/A	Ineligible study type (risk assessment tool development)
Ulusoy 2010	Exclude	N/A	Ineligible study design (case-control)
Vachon 2015	Exclude	N/A	Ineligible study design (case-control)
Van Veen 2018	Exclude	N/A	Ineligible study design (case-cohort)
Viallon 2009	Exclude	N/A	E/O not reported by risk category (only overall)
Wang 2014	Exclude	N/A	No outcome metric of interest
Wang 2018	Exclude	N/A	E/O not reported by risk category (only overall)
Yala 2019	Exclude	N/A	No outcome metric of interest

4 Mammographic density assessment tools

4.1 Authors

Suzanne Hughes, Chelsea Carle, Dr Qingwei Luo, Prof Dianne O'Connell, Dr Louiza Velentzis & A/Prof Carolyn Nickson.

4.2 Background

4.2.1 Breast density and breast cancer screening

Mammographic breast density (mammography density, or MD) is a risk factor for breast cancer and it can visually mask or camouflage breast cancers so they are more difficult to detect on mammography. This contributes to increased interval cancer rates and decreased mammographic screening program sensitivity and specificity for women with more dense breasts. Identifying women most at risk of interval cancers and offering them suitable supplemental or alternative imaging may reduce the risk of interval cancers, however, this would need to be balanced against potential increases in false positive outcomes.

4.2.2 ROSA scoping review (2019)

A scoping level review was undertaken in August 2019 to assess the ability of different MD assessment methods to identify women most at risk of interval cancers, with additional information collected concerning other screening outcomes such as false positive screens. This scoping review differed from a 2018 literature review commissioned by the Australian Government Department of Health on breast density (Cording 2018) as the ROSA review focussed instead on screening performance outcomes for each breast density category determined by different breast density assessment tools and in different settings and was restricted to digital mammography screening.

4.2.3 Systematic review

The 2019 ROSA scoping review identified evidence for the performance of two MD tools, BI-RADS and Volpara. BI-RADS is a visually estimated measure which depends on the assessor whereas Volpara is an objective automated measure. Both demonstrated that higher breast density in large populations of screened women was associated with higher invasive interval cancer rates and lower program sensitivity for invasive cancers.

Following its pre-specified criteria, our scoping review found no eligible evidence for MD tools other than BI-RADS and Volpara, and no eligible comparisons of different tools in the same population, so it was not possible to draw any conclusions as to the comparative performance of different tools. Furthermore, none of the studies found were in biennially screened populations aged 40-74 years (as for the Australian breast screening program). However, being a scoping review, the literature searches were not comprehensive and potential sources of bias not assessed. It was possible that some studies may not have been identified and it is not known which of the evidence, if any, was more reliable (i.e. at lower risk of bias).

Thus on this basis, the ROSA project subsequently undertook a systematic review to help ensure that all the available evidence was identified and an objective assessment of the certainty of the evidence undertaken. This was a resource-intensive task, but important given the potential role of breast density measurement in risk-based breast screening.

4.3 Research question

How well do MD measurement tools applied to screening mammograms stratify asymptomatic breast screening participants according to their risk of a subsequent interval cancer?

4.4 Methods

4.4.1 Systematic review requirements

A systematic review protocol (CRD42021238396) outlining details of the systematic review was registered on PROSPERO on 23rd March 2021. This protocol is provided in the Appendix (Section 4.5 (page 69)).

Undertaking a best-practice systematic review (rather than a scoping review) requires a number of additional steps (Higgins 2019), including:

- the development of a protocol and its registration, which is a requirement for publication in most journals
- more sensitive and comprehensive literature searches
- independent assessment of full text articles for inclusion by 2 reviewers
- Independent extraction of results by 2 reviewers
- risk of bias assessments by 2 reviewers for each of the included outcomes.

4.4.2 PICO questions and protocols

We used the PICO model to define the clinical questions of interest in terms of MD in the context of the ROSA project. PICO stands for population, intervention, comparison and outcomes. PICO questions and protocols were used to specify the scope of the review.

Two PICO questions/protocols were specified for this systematic review activity:

> PICO 1: How accurately does a given MD measurement tool stratify women according to their risk of a subsequent interval breast cancer?
>
> PICO 2: How do different MD measurement tools compare in their ability to stratify women according to their risk of a subsequent interval breast cancer?

The protocols used for these PICOs are shown in Table 17 and Table 18 below.

Table 17. PICO 1 protocol, MD assessment – within tool comparisons.

Population	***Intervention/exposure***	***Comparison***	***Outcomes***	***Study design***

Asymptomatic women aged ≥ 40 years of age undergoing mammographic screening	Categories of breast density ascertained using a specific mammographic density assessment tool	Another category of breast density ascertained using the same MD assessment tool	Interval invasive cancer rates Interval DCIS rates Screening program sensitivity for invasive cancer Screening program sensitivity for DCIS Screening program specificity False positive screening rates "Missed" invasive cancers or DCIS detected by radiological review of prior images	Cohorts including individual arms of RCTs Systematic reviews

RCTs = randomised controlled trials

Table 18. PICO 2 protocol, MD assessment – between tool comparisons

Population	***Intervention/exposure***	***Comparison***	***Outcomes***	***Study design***
Asymptomatic women aged ≥ 40 years of age undergoing mammographic screening	Categories of breast density ascertained using a specific MD assessment tool	Categories of breast density ascertained using another specific MD assessment tool	Interval invasive cancer rates Interval DCIS rates Screening program sensitivity for invasive cancer Screening program sensitivity for DCIS Screening program specificity False positive screening rates "Missed" invasive cancers or DCIS detected by radiological review of prior images	RCTs Paired cohort studies Systematic reviews

RCTs = randomised controlled trials

4.4.3 Selection criteria and definitions

Table 19. Study selection criteria for breast density assessment tools. These are the same for PICO 1 and PICO 2 unless otherwise specified.

	Inclusion	***Exclusion***
Population	Asymptomatic women aged ≥ 40 years of age undergoing mammographic screening with DM, or DM+DBT including programs that offer different numbers of mammographic view based on density Not restricted to general population i.e., can include high-risk populations e.g., women with family history of breast cancer or previous positive mammogram and low-risk populations DM is either CR or DR or DR only or type of DM not reported	Study population restricted to women who have had breast cancer or DCIS, or undergoing breast imaging as follow-up for a suspect mammogram or who have ever had a recall Screening program film mammography only i.e., no digital mammography Film mammograms that have been digitised Restricted to women with negative mammogram who undergo supplemental screening Restricted to women undergoing tomosynthesis screening Screening is mammogram + clinical examination not mammogram alone Screening is either DM or DBT and allocation not described as random Screening program offers more frequent or additional imaging to women with dense breasts
Intervention/Exposure PICO 1	A category of breast density or texture ascertained using a specific MD or texture measurement tool	Risk assessment tools that include breast density
Comparator PICO 1	Another category of breast density or texture ascertained using the same MD or texture measurement tool	
Intervention/Exposure PICO 2	Tool assessing MD or texture	Risk assessment tools that include breast density as one of several factors considered e.g., Tyrer Cuzick v.8 tool
Comparator PICO 2	Alternative tool assessing MD or texture	
Outcome	**By density category for:** Interval invasive cancer rates or hazard ratios 12-month interval invasive cancer rates Interval DCIS rates	Interval breast cancer includes interval DCIS as well as interval invasive cancer where invasive cancer and DCIS cannot be separately identified

	Inclusion	***Exclusion***
	Screening program sensitivity for invasive cancer Screening program sensitivity for DCIS Screening program specificity False positive screening rates "Missed" invasive cancers or DCIS detected by radiological review of prior images (rates preferred but can also include % interval cancer or subsequent SDC at next screen)	Program sensitivity includes DCIS as well as invasive cancer Predicted program sensitivity Outcomes not reported by tool-determined density category eg risk per 25% increase in density or mean density Overlapping cohorts with same outcome – exclude those that offer DM or FM-superseded by studies using DM only
Study design PICO 1	Cohorts including individual arms of an RCT, cases in case control studies or case series Or Systematic review thereof	Diagnostic accuracy studies
Study design PICO 2	RCTs or paired cohort studies Or Cohort studies in which groups being assessed with different mammographic measurement tools are matched by age, risk and screening location, and are screened in the same program and over the same time period Or Systematic review thereof	Diagnostic accuracy studies, non-paired cohort studies in which groups are not matched by age, risk and screening location, and/or are not screened in the same program and/or over the same time period.
Publication type	Journal article or peer-reviewed report	Conference abstracts, reviews, letters, editorials and comments
Publication date	2008 and onwards	
Language	English	

DCIS = ductal carcinoma in situ; RCTs = randomised controlled trials

For the purposes of this review:

- Studies comparing MD assessments based on digital mammography with MD assessments based on film mammography, using the same MD assessment tool, were included if they provided relevant results for the study arm receiving digital mammography that could be included in the review
- **Invasive interval cancers** were defined as interval cancers diagnosed following a negative screen and before the next recommended screen
- **Invasive interval cancer rates** were defined as the number of invasive cancers diagnosed following a negative screen (i.e. the reference screening round) and before the next recommended screen divided by total number of reference screens in the analysis
- **Screening program sensitivity** was defined as the percentage of screen-detected and interval invasive cancers that were screened detected ((screen-detected invasive cancers / screen-detected invasive cancers + invasive interval cancers) x 100)
- **Screening program specificity** was defined as the percentage of screens that were negative screens with no interval cancer before the next scheduled screen.
- **False positive screening rates** were defined as the number of positive screens with a benign outcome, divided by total number of screens.

Abridgment note: Detailed methods and results will become available through a publication in process. *Also refer to the summary of findings, section 6.1, from page 91 (questions 4 and 5).]*

4.5 Appendix

4.5.1 PROSPERO protocol

CRD42021238396 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021238396

1. Review title.

How well do mammographic density measurement tools stratify breast screening participants according to risk of a subsequent interval cancer? A systematic review.

2. Original language title.

N/A

3. Anticipated or actual start date.

1st March 2020

4. Anticipated completion date.

31st December 2021

5. Stage of review at time of this submission.

The review has not yet started: No, the review has started.

Review stage	Started	Completed
Preliminary searches	Yes	No
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	No	No
Data extraction	No	No
Risk of bias (quality) assessment	No	No
Data analysis	No	No

Funded proposal: Funding has been provided by the Commonwealth Department of Health, Australia.

6. Named contact.

Suzanne Hughes

7. Named contact email.

suzanneh@nswcc.org.au

8. Named contact address

Cancer Research Division, Cancer Council NSW, PO Box 572, Kings Cross NSW 1340, Australia

9. Named contact phone number.

+61 2 9334 1722

10. Organisational affiliation of the review.

Cancer Council NSW

11. Review team members and their organisational affiliations.

Ms Suzanne Hughes, Cancer Council NSW, Australia; School of Public Health, University of Sydney, Australia

Ms Chelsea Carle, Cancer Council NSW, Australia

Dr Louiza Velentzis, Cancer Council NSW, Australia; Melbourne School of Population and Global Health, University of Melbourne, Australia; Sydney School of Public Health, University of Sydney, Australia

Professor Dianne O'Connell, Cancer Council NSW, Australia; School of Public Health, University of Sydney, Australia; School of Medicine and Public Health, University of Newcastle, Australia

Dr Carolyn Nickson, Cancer Council NSW, Australia; Melbourne School of Population and Global Health, University of Melbourne, Australia; Sydney School of Public Health, University of Sydney, Australia

12. Funding sources/sponsors.

Commonwealth Department of Health, Australia.

13. Conflicts of interest.

None

14. Collaborators.

Dr Bruce G Mann – Clinical advisor

Breast Service, Royal Women's and Royal Melbourne Hospital, Australia; Department of Surgery, University of Melbourne, Australia

Paul Vardon – Advisor on research translation

Queensland Health, Cancer Screening Unit, Preventative Health Branch, Australia

15. Review question.

How well do mammographic density measurement tools applied to screening mammograms stratify asymptomatic breast screening participants according to their risk of a subsequent interval cancer?

Question 1: How accurately does a given mammographic density measurement tool stratify women according to their risk of a subsequent interval breast cancer?

Question 2: How do different mammographic density measurement tools compare in their ability to stratify women according to their risk of a subsequent interval breast cancer?

16. Searches.

Medline (including MEDLINE Epub Ahead of Print, I-Process & Other Non-Indexed Citations), Embase, Cochrane Database of Systematic Reviews (CDSR) and Health Technology Assessment (HTA) databases will be searched. Database-specific subject headings and text terms will be combined where available for terms related to screening, breast density and screening performance outcomes. Searches will be limited to articles published in English from 1 January 2008 onwards. Reference lists of full text articles identified for consideration will be checked for additional potentially relevant studies.

17. URL to search strategy.

Diagnosis of invasive breast cancer following a screening episode with a clear final result in an organised screening program and within the screening interval for that program.

18. Condition or domain being studied.

Newly diagnosed invasive breast cancer following a screening episode with a clear result.

19. Participants/population.

Inclusion: Asymptomatic women aged ≥ 40 years undergoing population digital mammographic screening.

Exclusion: Populations restricted to women (i) undergoing breast imaging as follow-up for breast cancer or after an abnormal screen, (ii) undergoing tomosynthesis or film mammography, or (iii) participating in a screening program that offers supplemental imaging to women with higher breast densities.

20. Intervention(s), exposure(s).

Mammographic density measurement or categorisation tools.

21. Comparator(s)/control.

For Question 1, no comparator.

For Question 2, different mammographic density measurement tools.

22. Types of study to be included.

Inclusion: For Question 1, cohort studies and individual arms of randomised controlled trials (RCTs); for Question 2, RCTs with women randomised to different mammographic density measurement tools, or paired cohort studies in which every woman is assessed with each of the different mammographic density measurement tools being compared, or cohort studies in which groups being assessed with different mammographic measurement tools are matched by age, risk and screening location, and are screened in the same program and over the same time period.

Exclusion: For Question 1, diagnostic accuracy studies; for Question 2, diagnostic accuracy studies and non-paired cohort studies in which groups are not matched by age, risk and screening location, and/or are not screened in the same program and/or over the same time period.

Diagnostic accuracy studies are excluded as this systematic review is restricted to screening performance outcomes for mammographic density measurement tools in the context of population screening.

23. Context.

Inclusion: Studies that report interval cancers or other screening performance outcomes breast in organised screening programs by mammographic density category assessed from any previous screening mammogram.

The review is limited to studies published from 2008 onwards in order to select studies likely to use more relevant imaging methods and more recent versions of mammographic density measurement tools.

Exclusion: Studies where interval cancer rates among screened women are not reported or cannot be derived from reported figures. Studies that do not differentiate invasive breast cancers from diagnoses of ductal carcinoma in situ (DCIS).

24. Main outcome(s).

The primary outcome will be interval invasive breast cancer rates for each tool-determined mammographic density category as assessed from previous screening mammograms. For the purposes of this review interval cancers are defined as “primary breast cancers that are diagnosed in women after a screening examination which has yielded a negative result, defined as no recommendation for recall or negative further assessment after recall, and before any subsequent screen is performed or within a time period equal to the screening interval (2 years)” (Van Bommel 2017 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5420149/)

Timing and effect measures

To assess interval cancer rates, the minimum follow-up for negative mammographic screens will be 12 months from the preceding screening mammogram. Outcomes will also be compared where available for the screening program’s standard interval. The rate of interval invasive cancers per total number of screens will be compared for each mammographic density category. For Question 2, for each tool being assessed, the mid-point percentile value of each density category will be used to help compare tools according to population percentile ranks of breast density values. Where possible, the highest and lowest mammographic density categories will be compared to the interval rate for women in the average or median mammographic density category. Where possible this will be reported by age group.

25. Additional outcome(s).

For each tool-determined mammographic density category:

- Interval DCIS rates.
- Screening program sensitivity for invasive breast cancer
- Screening program sensitivity for DCIS
- Screening program specificity for a negative result
- False positive rates (women recalled to assessment with a benign final outcome for the screening episode)
- Rates of “missed” invasive cancers or DCIS as determined by radiological review of prior images.

Measures of effect

The rate of false positives and "missed" invasive cancers or DCIS per total number of screens, sensitivities and specificities will be compared for each mammographic density category. For Question 2, for each tool being assessed, the mid-point percentile value of each density category will be used to help compare tools according to population percentile ranks of breast density values. Where possible, the highest and lowest mammographic density categories will be compared to the outcomes for women in the average or median mammographic density category. Where possible this will be reported by age group.

26. Data extraction (selection and coding).

Titles and abstracts of identified articles will be screened against pre-specified inclusion criteria. This will be split equally between two reviewers with 20% assessed by both reviewers to ensure concordance. The full texts of potentially relevant articles will be independently assessed for inclusion by two reviewers using pre-specified selection criteria. Disagreements will be resolved by third reviewer adjudication. Reasons for exclusion will be recorded for all excluded full-text articles.

Extraction of study characteristics and results will be split equally and arbitrarily between two reviewers followed by the other reviewer checking the work of the assigned reviewer. Disagreements will be resolved by third reviewer adjudication.

The following characteristics will be extracted for each included study:

1. Study information (first author, publication year, country, study design, setting, recruitment period, participant inclusion/exclusion criteria, screening protocol).
2. Participants' characteristics (number, age, screening round (first or subsequent), percentage previously screened).
3. Mammographic density measurement tool and categories, and radiologist experience where applicable.
4. Interval cancer definition and ascertainment.
5. Follow-up since mammographic density measurement (period, median follow-up).
6. Reported relevant outcomes stratified by mammographic density and any subgroups (i.e. age groups or outcomes by screening round) analysed separately.
7. Funding and reported potential conflicts of interest.

The following data will be extracted for each included study:

Description of each mammographic density tool and how mammographic density measurements were allocated to reported groups.
Description of each tool-determined mammographic density group and participant numbers and/or percentages per group.
Interval invasive cancer rates, interval DCIS rates, screening program sensitivity for invasive cancer, screening program sensitivity for DCIS, screening program specificity for a negative result, false positive rates, rates of "missed" invasive cancers or DCIS detected by radiological review of prior images for each tool-determined mammographic density group.

27. Risk of bias (quality) assessment.

Two reviewers will independently assess the risk of bias for each included study. Differences will be discussed and if no consensus is reached, a third reviewer will adjudicate.

The risk of bias for cohort studies included for either question and single arms of RCTs for question 1 will be assessed using QUADAS-2 (Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JD, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM, QUADAS-2 Group. (2011) QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Annals of Internal Medicine 155(8):529-36) which assesses the sources of bias most relevant to screening performance studies.

The risk of bias for RCTs included for question 2 will be assessed using the Cochrane Collaboration Risk of Bias-II tool [Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from: www.training.cochrane.org/handbook].

28. Strategy for data synthesis.

For Question 1, we will compare outcomes for the different mammographic density groups determined by a specific mammographic density measurement tool. For Question 2, outcomes across different mammographic density groups will be compared between different mammographic density measurement tools.

A narrative synthesis and graphs of outcomes by tool-determined mammographic density group will be provided.

A narrative synthesis and graphs will be presented for each PICO question. For each PICO question results will be analysed separately for each specific outcome.

For Question 1, results will be grouped by mammographic density tool; comparison of results for a specific outcome from different studies will enable comparison of the tool's performance in different populations.

For Question 2, the performance of different tools across risk groups within each study will be compared for specific outcomes.

For both research questions, results will be visually synthesised by plotting either mammographic density groups or mammographic density percentiles in the study group on the horizontal axis, and outcomes on the vertical axis.

29. Analysis of subgroups or subsets.

Analysis of subgroups defined by risk factors such as age and screening round will be undertaken if data are available.

30. Type and method of review.

Systematic review

Health area of review: Cancer, Public Health

31. Language.

English

32. Country.

Australia

33. Other registration details.

None

34. Reference and/or URL for published protocol.

None

35. Dissemination plans.

A paper will be submitted to a peer-reviewed journal in this field for publication. The results will also be communicated to key stakeholder groups including the Australian Government Department of Health, BreastScreen Australia, oncologists, general practitioners, cancer centres and consumer groups.

36. Keywords.

Breast neoplasms; mammographic density; mass screening; breast density

37. Details of any existing review of the same topic by the same authors.

None.

38. Current review status.

Ongoing.

39. Any additional information.

This review is being undertaken as part of the Optimising Early Detection of Breast Cancer Australia (Breast-ROSA) project funded by the Australian Department of Health and is part of the *Pathways to a Cancer-free Future* program sponsored by Cancer Council NSW.

40. Details of final report/publication(s).

N/A

5 Potential simplified risk assessment for the Australian screened population

5.1 Authors

A/Prof Carolyn Nickson, Dr Pietro Procopio, Dr Louiza Velentzis, Sam Egger

5.2 Background

Routine risk assessment and advice at BreastScreen services would be central to any risk-based breast cancer screening program.

The lifepool cohort (lifepool.org) as recruited over 50,000 women mainly through BreastScreen Victoria (BSV) and collected their BreastScreen mammograms and client service data as well as cancer outcomes registered at the Victorian Cancer Registry, and deaths recorded on the National Death Index. All participants completed a detailed questionnaire at baseline which included questions about risk factors not usually collected by BreastScreen. A majority of the cohort was recruited through BSV screening appointments, limited (at BSV's request) to clients attending subsequent round screening. This cohort provides an invaluable resource for exploring options for risk stratification of BSV clients.

Our previous analysis of this cohort [1] showed that the widely-validated US National Institutes of Health BCRAT risk assessment tool (the 'Gail' model) was an effective tool for stratifying subsequent round BSV participants aged 50–69 years to groups according to risk of invasive breast cancer diagnosed up to 5 years following risk assessment. This version of the BCRAT tool used the following information:

- Family membership
- Personal history of breast cancer (which excludes eligibility)
- Current age
- Age at menarche
- Age at first live birth
- Number of first-degree relatives with breast cancer
- Personal history of a breast biopsy, and the number of breast biopsies
- Personal history of hyperplasia
- Race (using categories to suit the US population; this is difficult to translate to the Australian setting)

Of note, that version of the tool did not incorporate breast density. However, as reported in our systematic review of risk assessment tools validated on large screening populations [2], breast density did not improve calibration of questionnaire-based risk assessment.

Scores were generated for a specified period into the future (e.g. 5-year risk), or to a specified age (e.g. by age 74 years).

This analysis found clear accordance between expected and observed incidence invasive breast cancers for all but the highest risk group (as expected in a screening population, since the tool was developed on the general population (Figure 6).

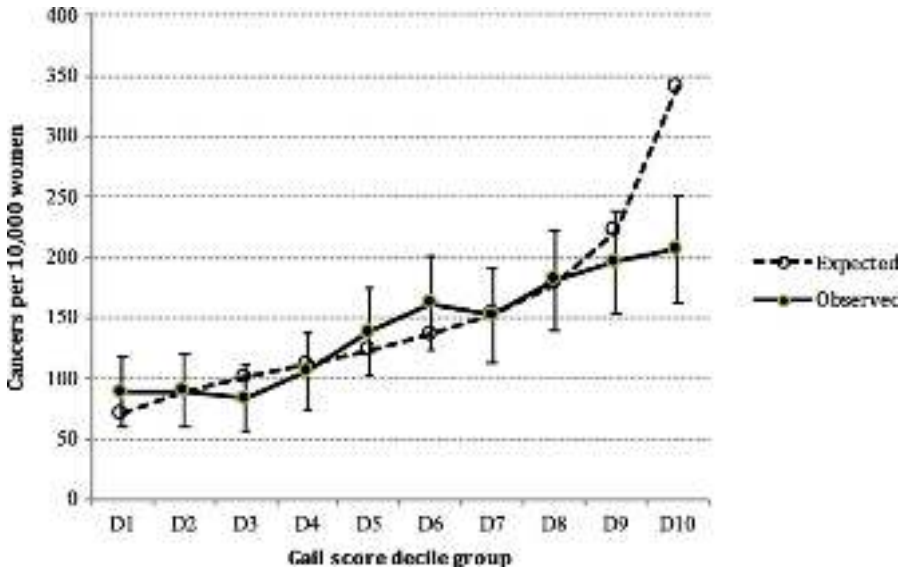


Figure 6. Expected and observed outcomes according to BCRAT scores generated by baseline questionnaires. Overall chi-squared test, $p < 0.0001$ (D1–D9 categories only; $p = 0.57$). D=Decile [1]

For population-level applications, it is important to consider which information is essential for accurate allocation of women to risk groups for risk-adjusted screening protocols. As indicated by the ROSA stakeholder consultation, substantial resources and adaptations would be required to implement detailed routine risk assessment, particularly in a way that ensures that clients and staff feel equipped and confident to engage in the process. Questionnaire-based established risk assessment tools (described elsewhere in this chapter) were generally developed from large epidemiological datasets with a view to facilitating individual-level risk assessment and advice, often with the assistance of a clinician. They require women to recall and share personal and often sensitive information such as their reproductive history, ethnicity, and height and weight.

With growing evidence that breast density information (and potentially genetic testing) may supersede the need for detailed questionnaires [3], it is possible that a simplified approach to risk assessment would be similarly effective, while being substantially more feasible. Our earlier analysis of the BCRAT tool on the lifepool cohort also found that a simplification of this version of the BCRAT model would be similarly effective on this cohort. [1] Specifically, using machine learning, we identified that the most important risk factors to include (in order from most to least important) were age, age at first live birth, age at menarche, number of first-degree relatives and the number of previous breast biopsies. As indicated in Figure 7 (where the ranking of groups with each added variable is shown from left to right), risk stratification was stable without the need to add information on ethnicity or a history of hyperplasia from breast biopsy. We reasoned that ethnicity may have appeared to be less important due to the US population profile and ethnicity definitions used to develop the BCRAT tool being considerably different to the Australian setting.

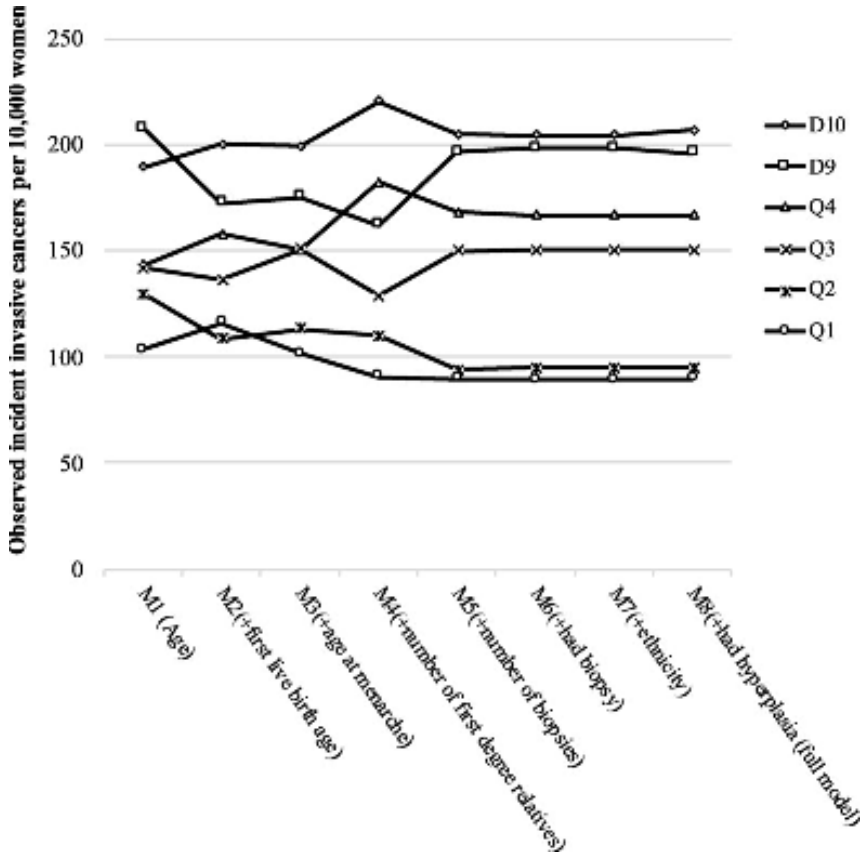


Figure 7. Observed incident cancers per 10,000 women according to quantile groups for the reduced BCRAT models (M1–8) assessed. M=model, Q=quintile

In this context, we compare here the BCRAT tool in the lifepool cohort with a simplified approach to risk assessment incorporating breast density and data routinely collected by BSV. The general aim was to assess the extent to which a simplified approach could identify a relatively small proportion of women at either higher than average risk or lower than average risk, for potential risk-based screening protocols.

5.3 Aims

For women aged 50-69 attending subsequent round screening, what is the association between risk group and risk of future invasive breast cancer, screen-detected invasive breast cancer or interval cancer when risk is assessed through:

a) Combinations of family history and breast density ('FH-MD'); compared to
b) The BCRAT risk assessment tool.

5.4 Methods

5.4.1 Study group

The study group comprises 46,385 women who enrolled in the lifepool cohort during the period 1 July 2010 to 6 Oct 2014 with a BreastScreen Victoria screen within +/- 60 days of completing the baseline questionnaire.

Women were excluded from the analysis if they met any of the following criteria:

- Age at baseline questionnaire outside the historical target age range of 50-69 years
- Mammogram at baseline captured through computed radiography (CR) rather than a full-field digital (DR) mammogram. This exclusion aimed to ensure generalisability of the findings to digital mammography.
- Baseline screen was a first-round episode. This was due to insufficient sample size to separately assess this group, who are expected to have markedly different outcomes; the group size was small due to lifepool recruiting women primarily from subsequent round screening.
- A screen-detected cancer at their baseline screen, since the aim is to assess prediction of a cancer in the future.

5.4.2 Risk factors

Age was determined from the lifepool questionnaire date of birth field, and verified against linked BSV data (no conflicts were found).

Screening round was obtained from BSV data linked to the lifepool cohort.

BCRAT scores were drawn from our previous analysis, using 5-year risk estimates.

Family history was defined according to having a first-degree family member. BSV definitions of family history changed in 18 November 2016, so that BSV family history categories as recorded at lifepool enrolment would differ to the categories applied now. This means that some lifepool cohort participants will have had different family history categories assigned at BSV over the course of the prospective outcomes analysed here. More recent BSV data has been linked to the lifepool cohort, however this is available only for women who have continued BSV participation, and family history can change over time as additional relatives have a cancer diagnosis. This analysis aims to report prospective outcomes according to family history. Therefore, family history was determined from the lifepool baseline questionnaire. In the current analysis it is categorised as ‘yes’ for women reporting a first-degree family member with breast cancer and ‘no’ for all other women.

Mammographic breast density (MD) was measured using AutoDensity [4] from the baseline screening mammogram. Measurements were available for all women in the sample.

5.4.3 Weighting

The lifepool cohort is expected to have a different age distribution to current BreastScreen participants, mostly due to recruitment methods and the mid-2014 extension of the target age to include women aged 70-74. Lifepool participants may also have a different family history of breast cancer to all BSV participants, as women with a family history may have been more motivated to join the lifepool cohort.

Therefore, this analysis applied weights to the study group to better reflect the age and family history distribution of BreastScreen participants. This was done by applying the Stata ‘pw’ sampling weights function, according to the representation of each age/family history group in the study group, based on BSV subsequent-round participation by 5-year age group and family history category for 2019. Further information is provided in the Appendix.

5.4.4 Risk stratification

The weighted cohort was stratified into lower, average and higher risk groups. To aid with interpretation, AutoDensity percent density percentiles were mapped by age group to approximately match the distribution of Volpara Density Grade (ed. 5) for women aged 50-74 as reported from a BSV screening site in Bell et al. (A 11.0%, B 48.5%, C 30.0%, D 10.5%) [5]. *Percent density* (the

proportion of the breast area that was distinctly more radio-opaque) was used because a previous analysis has suggested that using dense area or adjusting percent density for age makes little difference to the association between AutoDensity measures and outcomes of cancer risk and interval cancers [6-7].

For both the FH-MD and the BRCAT scores, the size of the higher risk group was determined by the proportion of the weighted sample with a strong family history and MD in the highest two categories (C/D). The size of the lower risk group was determined by the proportion of the weighted sample with no family history and MD in the lowest two categories (A/B). The remaining group (comprising women with A/B breast density and a strong family history, and women with C/D breast density and no strong family history) was classified as average risk. The continuous BCRAT scores were then partitioned to match these proportions.

5.4.5 Outcomes

The following outcomes were available the analysis:

- Invasive breast cancers (as recorded at the Victorian Cancer Registry, up to end 2016)
- Interval cancers (as recorded by BreastScreen Victoria, up to end 2015)
- Screen-detected cancers (as recorded by BreastScreen Victoria, up to end 2016)
- DCIS (as recorded at the Victorian Cancer Registry, up to end 2016)

The analysis focuses on three categories of outcomes:

1. Invasive breast cancers, irrespective of mode of detection.
2. Screen-detected invasive breast cancers.
3. Interval cancers (invasive or DCIS, reasoning that any interval outcomes should be minimised)

5.4.6 Data analysis

We used descriptive statistics to report the distribution of risk factors in the study group and the numbers and rates of outcomes during the follow-up period, with outcomes censored to 31 December 2016 (the most recent date for which outcomes except interval cancer data were complete), with analyses repeated to a censor date of 31 December 2015 as a sensitivity analysis (this being the period for which all outcomes were complete).

To account for different times to diagnosis and duration of follow-up within the study group, we used the risk-group-specific hazards to estimate outcomes for each risk group, including confidence intervals. These hazards models were weighted to match the age/family history profile of BSV 2019 subsequent round screens. For FH-MD hazards were assessed with and without adjustment for age (BCRAT already incorporates age as a risk factor).

To assist with interpretation, we plotted cumulative observed rates of outcomes according to different FH-MD and BCRAT risk groups, using the lifepool baseline screen as the reference point, noting that these outcomes were not adjusted for age.

5.5 Results

5.5.1 Study group

After exclusions, 35,576 women remained in the sample for analysis. A further 193 women were excluded due to missing data on MD and 1 woman due to a missing BCRAT score, leaving 35,382 women in the analysis. The mean follow-up time for the cohort was 4.2 years (median 4.2 years, range 0.3 to 6.5 years).

A total of 569 incident cases (invasive or DCIS) were available for analysis, as detailed in Table 20.

Table 20. Numbers of outcomes available for analysis. IBC = invasive breast cancer, DCIS = ductal carcinoma in situ. 'Other' includes all cancers diagnosed outside the program not defined as interval cancers.

	Mode of detection			
	Screen-detected	**Interval**	**Other**	**Total**
IBC	291	66	123	480
DCIS	68	6	15	89
Total	359	72	138	569

The weighting required to match the sample to the 2019 age and family history distribution of BSV subsequent round participants is shown in Table 21. Larger weighting numbers indicate less certainty about the estimates generated. This uncertainty is reflected in the confidence intervals around hazards model estimates (which are effectively powered by the lifepool sample size, before any weighting is applied).

Table 21. Population sampling weights used to adjust statistical models.

Age group	**Family history category**	
	No	**Yes**
50-54	5.16	2.28
55-59	5.47	2.55
60-64	5.59	1.96
65-69	6.58	2.56

5.5.2 Risk classification

The distribution of breast density groups and family history was as shown in Table 22. As indicated, 40% of the group were assessed as breast density C/D, and 23% had a strong family history. To meet the aim of identifying relatively small groups of clients as 'higher risk' and 'lower risk', clients with no strong family history and breast density category A (denoted in bold italics) were classified as 'lower risk', and clients with a strong family history and breast density category C/D (denoted in bold) were classified as 'higher risk'.

Table 22. Distribution of family history and breast density in the study group.

	Strong family history		
Breast density	No	Yes	Total
A	***2,951 (8.3%)***	0,857 (2.4%)	3,808 (10.8%)
B	13,294 (37.6%)	3,921 (11.1%)	17,215 (48.7%)
C	8,288 (23.4%)	**2,472 (7.0%)**	10,760 (30.4%)
D	2,723 (7.7%)	**876 (2.5%)**	3,599 (10.2%)
Total	27,256 (77.0%)	8,126 (23.0%)	35,382 (100.0%)

BCRAT scores were then partitioned to the same distribution of risk groups, as closely as possible. The resulting risk categories are shown in Table 23.

Table 23 Distribution of risk groups in lifepool participants (subsequent round screeners) aged 50-69 years at baseline (weighted to BSV participant profile from 2019). The risk group distribution was first determined by the FH-MD categories as indicated, then the BCRAT scores were partitioned to the same proportions as closely as possible.

Risk classifier	**Risk group**		
	Lower	**Average**	**Higher**
FH-MD	8.3% (No strong family history, MD category A)	82.2% (Remainder)	9.5% (Strong family history, MD categories C/D)
BCRAT	8.4% (5-year risk 0.6% to 0.9%)	81.6% (5-year risk 0.9% to 3.0%)	10.0% (5-year risk 3.0% to 22.0%)

The difference between risk classification using these scores is indicated in Table 24. Based on measures of agreement between the two risk classifiers (Stata 'kap'), we would expect 69% of clients to be assigned to the same risk group by either classifier ($p<0.0001$) (and, conversely, 31% of clients to be assigned to different risk groups).

Table 24. Agreement between the two risk classifiers.

		BCRAT		
		Lower	**Average**	**Higher**
FH-MD	**Lower**	278 (0.8%)	2656 (7.5%)	17 (0.05%)
	Average	2,702 (7.6%)	24,187 (68.4%)	2,194 (6.2%)
	Higher	0 (0.0%)	2,027 (5.7%)	1,321 (3.7%)

5.5.3 Outcomes

Rates by risk group

The observed number and rate of outcomes for each risk classifier and category is shown in Table 25. These figures indicate, for example, highest rates of invasive breast cancers in the higher risk groups by either FH-MD or BCRAT risk classification, and lowest rates in the lower risk groups. This pattern also holds for screen-detected invasive breast cancers, but only for interval cancers when

using the FH-MD classifier. Cancer diagnosis rates (CDRs) with confidence intervals based on the combined person-years observed in each sub-group indicate some statistically significant differences; these differences are tested in the competing risks hazards modelling reported below.

Table 25. N (%) outcomes by risk category. Proportions show the proportion of women in each group with the outcome in the cohort of 35,382 lifepool participants analysed. CDR = cancer diagnosis rate (per 10,000 person-years observed in the group reported).

Outcome	Risk group	FH-MD		BCRAT	
		N (%)	CDR	N (%)	CDR
Invasive breast cancers	Lower risk	27 (0.092%)	27.2 (18.9-40.8)	21 (0.71%)	20.5 (13.6-32.6)
	Average risk	386 (1.33%)	38.6 (34.9-42.9)	397 (1.38%)	39.5 (35.7-43.8)
	Higher risk	67 (2.01%)	58.6 (46.4-75.2)	62 (1.76%)	52.2 (40.5-68.4)
Screen-detected invasive breast cancers	Lower risk	18 (0.61%)	18.0 (11.5-29.9)	10 (0.34%)	9.8 (5.4-19.9)
	Average risk	233 (0.81%)	23.0 (20.2-26.4)	240 (0.84%)	23.8 (20.9-27.2)
	Higher risk	40 (1.21%)	35.7 (26.4-49.5)	41 (1.17%)	35.5 (26.0-49.9)
Interval cancers (invasive or DCIS)	Lower risk	1 (0.03%)	1.1 (N/A)	4 (0.14%)	3.8 (1.4-13.7)
	Average risk	60 (0.21%)	6.4 (5.0-8.4)	62 (0.22%)	6.3 (4.9-8.3)
	Higher risk	11 (0.34%)	9.2 (5.2-18.1)	6 (0.17%)	4.5 (2.0-11.9)

Outcomes over time

To indicate the time between risk assessment and cancer diagnoses, outcomes are presented below by risk classifier and risk group according to time since baseline screen.

Invasive breast cancers

Figure 8 shows the observed invasive breast cancers over time by risk group, for FH-MD risk classification. As expected in this cohort of active screening participants, many cancers are diagnosed at around 2 years since baseline, in line with the next scheduled biennial screen.

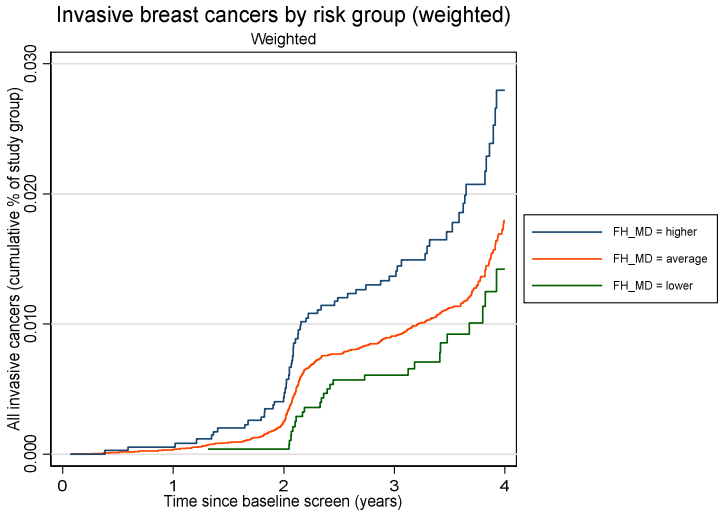


Figure 8. Observed (weighted) incident invasive breast cancers according to time since baseline and risk category, for risk classified using FH-MD.

Observed invasive breast cancers over time by risk group using the BCRAT classifier are shown in Figure 9.

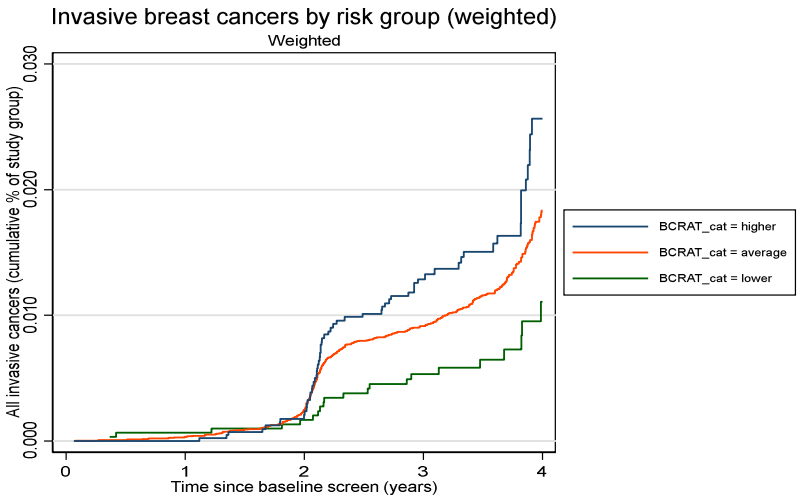


Figure 9. Observed (weighted) incident invasive breast cancers according to time since baseline and risk category, for risk classified using BCRAT scores.

Of note, for both approaches to risk assessment, from year 2 the cumulative plots maintain their rank according to the assigned risk group (lower, average, higher).

Screen-detected invasive breast cancers

Figure 10Figure 10 shows screen-detected invasive cancers by MD-FH risk group, according to time since baseline. The highest risk group appears to maintain its rank for both risk classifiers, with negligible difference between average and lower risk groups.

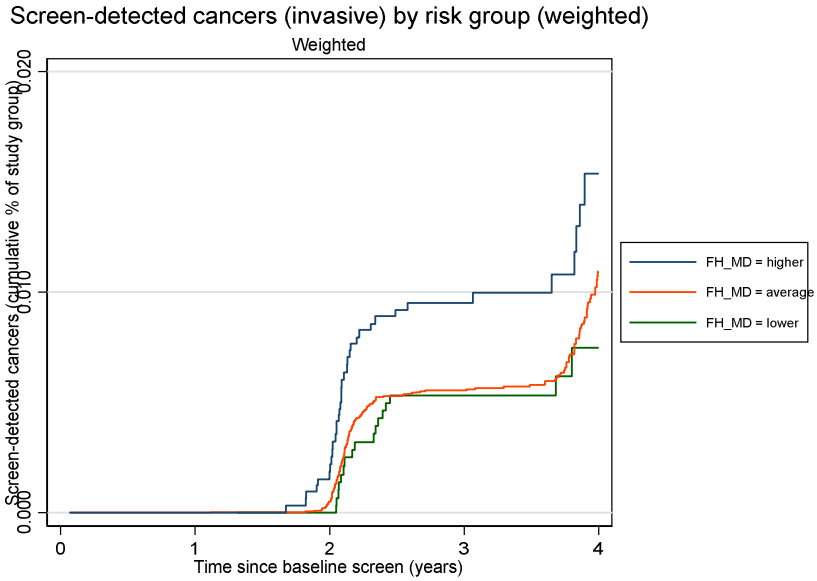


Figure 10. Observed (weighted) incident invasive screen-detected breast cancers (invasive or DCIS) according to time since baseline and risk category, for risk classified using FH-MD.

Figure 11 shows screen-detected invasive cancers by BCRAT risk group, according to time since baseline, indicating a clearer differentiation between lower and average risk groups.

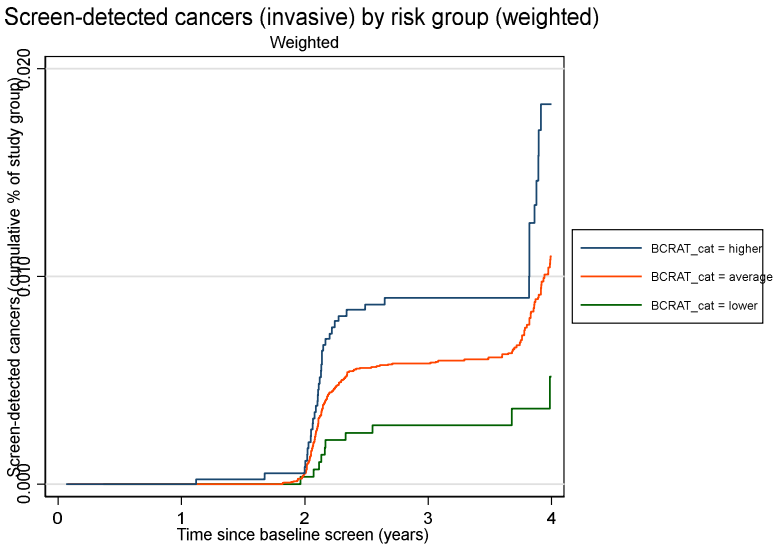


Figure 11. Observed (weighted) incident invasive screen-detected breast cancers (invasive or DCIS) according to time since baseline and risk category, for risk classified using BCRAT scores.

Interval cancers

Figure 12 shows interval cancers (either invasive or DCIS) by MD-FH risk group, according to time since baseline. The cumulative plots maintain their rank according to the assigned risk group (lower, average, higher) (noting that there was only one interval cancer in the FH_MD lower risk group).

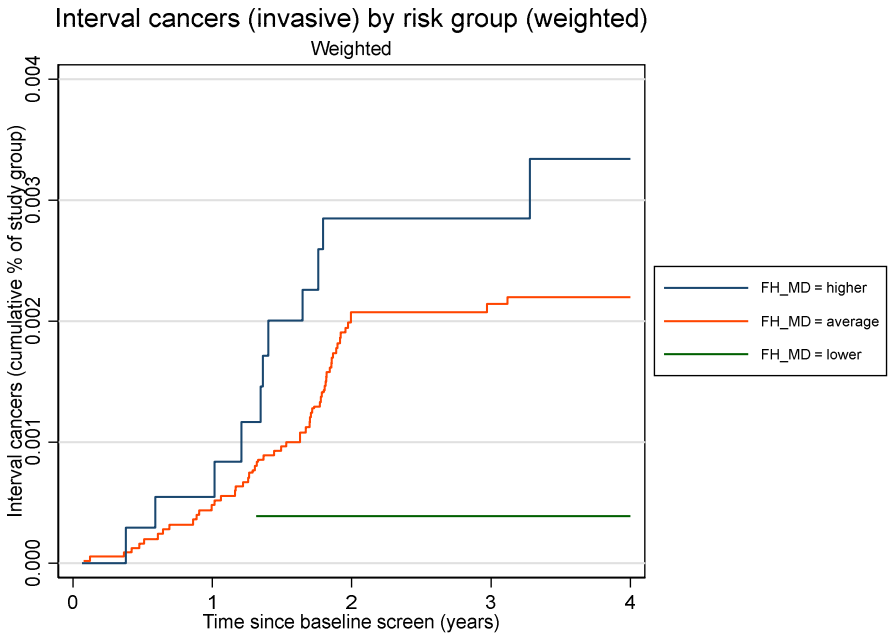


Figure 12. Observed (weighted) incident interval breast cancers (invasive or DCIS) according to time since baseline and risk category, for risk classified using FH-MD.

Figure 13 shows interval cancers (either invasive or DCIS) by BCRAT risk group, according to time since baseline. The groups do not maintain their ranks over time, with the average risk group having the highest interval cancer rates by year 4.

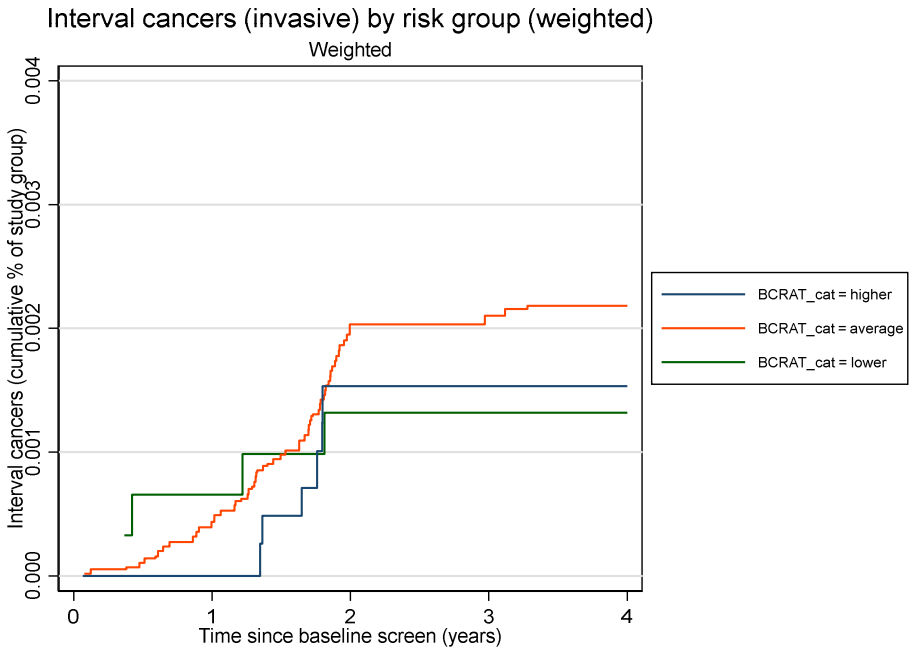


Figure 13. Observed (weighted) incident interval breast cancers (invasive or DCIS) according to time since baseline and risk category, for risk classified BCRAT scores.

5.5.4 Statistical analysis

Statistical comparisons for each outcome and risk classifier are reported in Table 26. These estimates are generated from a competing-risks regression model (Stata 'stcrreg') accounting for time since baseline and competing risks (other cancer diagnoses or deaths). The high p-values and wide confidence intervals including the null value (1)) indicate that the differences between groups are not statistically significant for any outcome or classifier based on the data available for analysis. For the FH-MD risk classifier, adjusting the regression model for age did not significantly alter the model fit or estimates. As noted in the methods, age adjustment was not appropriate for the BCRAT model since age is already incorporated as a risk factor in the BCRAT risk assessment tool.

Table 26. Hazards ratios (HRs) for outcomes of interest, for lower and higher risk groups compared to the average risk group (HR (95% confidence interval). DCIS = ductal carcinoma in situ.

Outcome	Risk group	FH-MD	BCRAT
Invasive breast cancers	Lower risk	0.75 (0.51-1.10), p=0.139	0.97 (0.62-1.52), p=0.898
	Average risk	REF	REF
	Higher risk	1.07 (0.82-1.41), p=0.612	0.91 (0.69-1.19), p=0.483
Screen-detected invasive breast cancers	Lower risk	0.93 (0.58-1.51), p=0.773	0.72 (0.38-1.37), p=0.324
	Average risk	REF	REF
	Higher risk	1.10 (0.78-1.57), p=0.587	1.08 (0.77-1.52), p=0.642
Interval cancers (invasive or DCIS)	Lower risk	0.21 (0.03-1.49), p=0.117	1.16 (0.41-3.27), p=0.785
	Average risk	REF	REF
	Higher risk	0.98 (0.51-1.89), p=0.959	0.49 (0.21-1.14), p=0.099

5.6 Discussion

Descriptive reporting of outcomes following subsequent round screening in 35,382 BreastScreen clients (namely, rates and plots of cumulative outcomes by risk group) indicates that simplified risk assessment using only family history and breast density categories may classify clients according to their risk of invasive breast cancer, screen-detected invasive breast cancer and interval cancers in a way that is comparable to or an improvement on risk assessment using the BCRAT risk assessment tool. This is an important indication, given the markedly different work involved in ascertaining (and updating) BCRAT risk assessment compared to the FH-MD classifier.

However, after accounting for sample size, time-to-event and competing risks, no statistically significant differences were identified between risk groups for any outcome under either approach to risk assessment.

Statistical power for these comparisons may be improved through more equal partitioning of the sample into risk groups; on this basis as an exploratory analysis we also assessed risk groups based on *percent density* breast density tertiles and BCRAT score tertiles (three equally-sized groups for each score). This yielded statistically significant results only for interval cancer outcomes and only when these were compared between the lower and middle tertiles of *percent density* scores (HR 0.29 (95% CI 0.14-0.64, p=0.002), suggesting that breast density alone could identify one third of BreastScreen participants at lower risk of interval cancers. The episodic nature of the outcomes (where most events align with screening schedules) may mean that larger sample sizes are required to capture statistically significant differences between groups in analyses that (quite rightly) factor in the time from baseline to the outcome of interest.

Outcomes may be influenced by more intensive screening or surveillance of study group members with a strong family history, so that this group is more likely to have cancers diagnosed with the follow-up period and these cancers are less likely to be interval cancers. This may explain some unusual patterns in the point estimates for interval cancer outcomes (such as lower rates for women with the highest BCRAT scores) but this would require verification with a larger sample.

As reported in our systematic review of risk assessment tools (Section 3, page 16), although detailed questionnaire-based tools do not appear to improve greatly with the addition of breast density, tools that are firstly calibrated to the risk profiles of the population in which they were applied demonstrated a better fit. This may be worth exploring in the future considerations of questionnaire-based risk assessment tools, to fully explore their utility.

Overall, our analysis indicates that further investigation of simplified models combining family history and breast density is warranted using larger sample sizes and including outcomes from the first screening rounds. Although the vast majority of screens provided by BreastScreen are subsequent round screens (87% of all screens in 2019 [8] and 89% in 2020 [9]), assessing outcomes following the first screening round is important given that program sensitivity (the proportion of cancers diagnosed by screening rather than as interval cancers) is higher at first round screening [8-9] and particularly among women with higher breast density [10], and that risk-based screening protocols would ideally be established at the start of each client's BreastScreen enrolment.

Such an analysis could be done for the FH-MD classifier as part of any large-scale implementation or evaluation of routine breast density assessment in BreastScreen services. BreastScreen WA has routinely reported breast density for women not recalled to assessment, and recently published rates of screen-detected cancers and interval cancers according to various risk factors including breast density (visually assessed as either dense (heterogeneously or extremely dense), or other) and family history (first degree relative with breast cancer versus other), for over one million screens provided over 2007-2017. [11] Comparing outcomes according to the number of events per 10,000 woman-years observed, they found significantly higher rates of screen-detected cancers and interval cancer for clients with a family history of breast cancer or dense breasts. It may be helpful to analyse those data further, accounting for competing risks and combining breast density and family history variables as done in this analysis, reporting findings separately for first and subsequent round screening, and perhaps according to tumour stage. Interval cancers in this setting may well differ to other jurisdictions given that breast density advice is estimated to lead to 20% of clients with dense breasts having an ultrasound, [12] and the findings could not be directly compared to BCRAT scores as done in this analysis. However, this analysis should yield more certain estimates for the FH-MD classifier due to the larger sample size.

Such large-scale analyses would be valuable not only in terms of understanding current outcomes for different risk groups, but also for how risk groups might best be identified as part of a more personalised approach to breast cancer screening in Australia.

5.7 Conclusion

For women aged 50-69 attending subsequent round screening, combinations of family history and breast density ('FH-MD') may be comparable to the BCRAT risk assessment tool in terms of estimating risk of future invasive breast cancer, screen-detected invasive breast cancer or interval cancer. Larger studies are required to verify this finding; the current analysis indicates that more simplified approaches to risk assessment should be included in consideration of options for risk-based breast screening in Australia, mindful of the resources and imposts involved in undertaking detailed risk assessment, and stakeholder interest in informing women about their breast density.

5.8 References

1. Nickson C, Procopio P, Velentzis LS, et al. Prospective validation of the NCI Breast Cancer Risk Assessment Tool (Gail Model) on 40,000 Australian women. *Breast Cancer Res*. 2018;20(1):155. Published 2018 Dec 20. doi:10.1186/s13058-018-1084-x
2. Velentzis, L.S.; Freeman, V.; Campbell, D.; Hughes, S.; Luo, Q.; Steinberg, J.; Egger, S.; Mann, G.B.; Nickson, C. Breast Cancer Risk Assessment Tools for Stratifying Women into Risk Groups: A Systematic Review. Cancers 2023, 15, 1124. https://doi.org/10.3390/cancers15041124
3. 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 [published online ahead of print, 2020 May 1]. Eur Radiol. 2020;10.1007/s00330-020-06901-x. doi:10.1007/s00330-020-06901-x
4. Nickson, C., Arzhaeva, Y., Aitken, Z. et al. AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes. Breast Cancer Res 15, R80 (2013). https://doi.org/10.1186/bcr3474
5. Bell RJ, Evans J, Fox J, Pridmore V. Using an automated measure of breast density to explore the association between ethnicity and mammographic density in Australian women. J Med Imaging Radiat Oncol. 2019 Apr;63(2):183-189. doi: 10.1111/1754-9485.12849. Epub 2019 Jan 8. PMID: 30623584.
6. Nickson C, Kavanagh AM. Tumour size at detection according to different measures of mammographic breast density. J Med Screen. 2009;16(3):140-146. doi:10.1258/jms.2009.009054
7. Kavanagh AM, Byrnes GB, Nickson C, et al. Using mammographic density to improve breast cancer screening outcomes. Cancer Epidemiol Biomarkers Prev. 2008;17(10):2818-2824. doi:10.1158/1055-9965.EPI-07-2835
8. Australian Institute of Health and Welfare 2021. BreastScreen Australia monitoring report 2021. Cat. no. CAN 140. Canberra: AIHW.
9. Australian Institute of Health and Welfare (2022) BreastScreen Australia monitoring report 2022, catalogue number CAN 150, AIHW, Australian Government
10. Puliti D, Zappa M, Giorgi Rossi P, Pierpaoli E, Manneschi G, Ambrogetti D, Ventura L, Mantellini P; DENSITY Working Group. Volumetric breast density and risk of advanced cancers after a negative screening episode: a cohort study. Breast Cancer Res. 2018 Aug 9;20(1):95. doi: 10.1186/s13058-018-1025-8. PMID: 30092817; PMCID: PMC6085631.
11. Noguchi N, Marinovich ML, Wylie EJ, Lund HG, Houssami N. Screening outcomes by risk factor and age: evidence from BreastScreen WA for discussions of risk-stratified population screening. Med J Aust. 2021 Oct 18;215(8):359-365. doi: 10.5694/mja2.51216. Epub 2021 Aug 9. PMID: 34374095;
12. Darcey E, Hunt EJ, Keogh L, McLean K, Saunders C, Thompson S, Woulfe C, Wylie E, Stone J. Post-mammographic screening behaviour: A survey investigating what women do after being told they have dense breasts. Health Promot J Austr. 2021 Oct;32 Suppl 2:29-39. doi: 10.1002/hpja.396. Epub 2020 Sep 10.

5.9 Appendix

5.9.1 BSV family history categories

Table 27. Family history categories as defined in the BSV 'Family History of Breast Cancer Policy and Procedure', version 2 (19 July 2017).

ATTACHMENT ONE: ALLOCATION OF FAMILY HISTORY RISK

Risk Rating	Definition
1. Average risk	Approximately 95% of the population can be classified as having an average risk of breast cancer based on their family history. A woman is classified as having an average risk of breast cancer, based on her family history of breast and ovarian cancer if she has one of the following: • No family history of breast cancer • One 1 degree relative diagnosed with BrCa > 50 y.o. • One 2 degree relative diagnosed with breast cancer at any age • One 2 degree male relative diagnosed with breast cancer any age • Two 2 degree relatives on diff sides of the family diagnosed with breast cancer > 50 • Two 2 degree relatives on the same side of the family diagnosed with breast cancer > 50 • No family history of ovarian cancer • One 1 degree or 2 degree relative with ovarian cancer at any age, and no relatives with breast cancer • Two 1 degree or 2 degree relatives, on diff sides of family if both 2 degree, with ovarian cancer at any age
2. Moderately increased risk	Approximately 4% of the population can be classified as having a moderately increased risk of breast cancer based on their family history. A woman is classified as having a moderately increased risk of breast cancer, based on her family history of breast and ovarian cancer if she has one of the following: • One 1 degree female relative diagnosed with BrCa < 50 y.o. • One 1 degree male relative diagnosed with BrCa at any age • One 1 degree female relative with bilateral breast cancer > 50 y.o. • One 2 degree female relative with bilateral breast cancer any age • Two 1 degree female relatives diagnosed with breast cancer > 40 y.o. • Two 2 degree female relatives on the same side diagnosed with breast cancer, at least one with BrCa < 50 y.o. • Two relatives (1 degree or 2 degree) with breast cancer on different sides of the family with at least one < 50 y.o. • One 1 degree or 2 degree relative with ovarian cancer at any age and one 1 degree or 2 degree female relative with breast cancer on the same side >50

3. Potentially high risk	Approximately 1% of the population can be classified as having a potentially high average risk of breast cancer based on their family history.
	
	A woman is classified as having a potentially high risk of breast cancer, based on her family history of breast and ovarian cancer if she has one of the following:
	
	• Ovarian cancer
	• One 1 degree male relative with bilateral breast cancer
	• One 1 degree female relative with bilateral breast cancer < 50 y.o.
	• Two 1 degree or 2 degree relatives on the same side diagnosed with breast cancer at any age who also have one or more of the following on the same side of the family:
	  ◦ Relative with ovarian cancer
	  ◦ Breast cancer diagnosed before 40 y.o.
	  ◦ Breast cancer in a male relative
	  ◦ Bilateral breast cancer
	• Three or more 1 degree or 2 degree relatives on the same side diagnosed with breast cancer at any age
	• Two relatives (1 degree or 2 degree) with ovarian cancer on the same side of the family at any age
	• One 1 degree or 2 degree relative with ovarian cancer at any age who also have one or more of the following on the same side of the family
	  ◦ Breast cancer diagnosed before 50 y.o.
	  ◦ Breast cancer in a male relative
	  ◦ Bilateral breast cancer

6 Appendices

6.1 Key findings

Drawing from the detailed analyses and results described throughout the chapter, the project generated a set of key findings which were reviewed by the ROSA Expert Advisory Group over May to July 2022. The final set of key findings is outlined below.

Q1. Breast cancer risk tools (between tool comparisons). For asymptomatic women, how do different breast cancer risk assessment tools compare in their ability to predict breast cancer risk across the risk groups determined by each of the tools?

Key evidence

1. For breast screening populations, some risk assessment tools based on self-reported information usually including family history and prior breast biopsies can identify groups of women at higher or lower risk.

2a. The precision of breast cancer risk assessment tools depends on the population and setting.

2b. The precision of breast cancer risk assessment tools can be improved with calibration to the target population.

3. Mammographic breast density assessments have not been demonstrated in the reviewed external validation cohort studies to improve the accuracy of breast cancer risk assessment tools based on self-reported information usually including family history and prior breast biopsies.

4. Polygenic risk scores have not been demonstrated in external validation cohort studies to improve the accuracy of breast cancer risk assessment tools based on self-reported information usually including family history and prior breast biopsies.

Considerations for implementation

1. Breast cancer risk assessment tools are expected to improve over time due to advances in technologies, image analysis and incorporation of AI systems.
2. Breast cancer risk assessment incorporating genetic test results may have ethico-legal consequences for individual women. These consequences should be well-understood before any introduction of population-level risk assessment incorporating genetic testing, with any implementation being on an opt-in basis and supported by an informed decision-making process.
3. .While the contribution of breast density assessment to breast cancer risk assessment tools was not demonstrated in this review, breast density remains an important tool for assessing risk of reduced sensitivity and specificity of mammographic screening tests.
4. Breast cancer risk assessment tools of equal accuracy that rely on limited or no self-reported information may be more reliable and easier to implement than more detailed questionnaire-based tools, once suitable information systems are established.

Priority evidence gaps

1. The accuracy of breast cancer risk assessment tools where input data is missing, compared to risk assessment with complete information.
2. The accuracy of breast density alone as a risk assessment tool, with an assessment of whether other risk factors improve the accuracy of risk assessment when added to breast density.
3. The accuracy of risk assessment tools for predicting breast cancer incidence according to prognostic indicators (e.g. tumour subtype, grade, size, nodal) and in situ breast cancer incidence.
4. Further information on the performance of breast cancer risk assessment tools in the Australian breast screening population, noting that risk assessment tools can perform differently in different settings and populations.

Q2. Breast cancer risk tools (within tool comparisons): For asymptomatic women, how does a given breast cancer risk assessment tool perform in predicting breast cancer risk across the risk groups determined by the tool?

Key evidence

1. In the Australian setting, the Gail risk assessment tool (version 2), which does not include breast density, can identify groups of BreastScreen Australia participants at higher or lower risk of breast cancer.

Q3. Simplified risk assessment using breast density: For BreastScreen participants, how does risk assessment using family history and breast density compare to risk assessment using family history alone?

Key evidence

1. For women aged 50-69 attending subsequent round screening, combinations of family history and breast density may be comparable to the BCRAT questionnaire-based risk assessment tool in terms of estimating risk of future invasive breast cancer, screen-detected invasive breast cancer or interval cancer.

Considerations for implementation

1. More simplified approaches to risk assessment should be included in consideration of options for risk-based breast screening in Australia, mindful of the resources and imposts involved in undertaking detailed risk assessment, and stakeholder interest in informing women about their breast density.

Priority evidence gap

1. Larger studies to validate the findings indicated by our analysis.

Q4. Breast density as a risk tool (within tool comparisons): How accurately does a given mammographic density measurement tool stratify women according to their risk of a subsequent interval cancer and other screening outcomes?

Key evidence

1. Breast screening populations can be stratified into groups according to interval cancer rates, program sensitivity, and false positive rates. *Abridgment note: Some detail from this statement is withheld and is expected become available through a publication in process.*

2. For each breast density assessment tool assessed, the accuracy of this risk stratification varied between studies (which varied in terms of settings and screening program design).
3. For each breast density assessment tool assessed, interval cancer risk stratification is often accurate either for higher risk groups or lower risk groups, but rarely both.

4a. For population mammographic screening, while breast density does not universally improve breast cancer risk assessment tools, it is a critical risk factor for estimating expected program sensitivity, program specificity, interval cancer rates and false positive rates.

4b. Breast density assessment tools and other potential tools to identify groups of women according to BreastScreen Australia program sensitivity, program specificity, interval cancer rates and false positive rates are expected to improve over time due to advances in technologies, image analysis and incorporation of AI systems.

Q5. Breast density as a risk tool (between tool comparisons). How do different mammographic density measurement tools compare in their ability to stratify women according to their risk of a subsequent interval breast cancer and other screening outcomes?

Key evidence

1. There is some evidence that the performance of different breast density assessment tools in the same population is very similar.

Priority evidence gaps

1. Further evaluation of how different approaches to breast density as a risk tool compare on the same population, in an Australian screening setting.