The ROSA PROJECT

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

Chapter 4. Risk-based screening protocols (abridged)

4 April 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 4: Risk-based screening protocols (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 Document purpose

Since 1 May 2018, the Australian Department of Health has engaged Cancer Council Australia to undertake a series of activities exploring options for risk-based breast cancer screening in Australia, described collectively as the Breast ROSA (Roadmap to Optimising Screening in Australia) project. The ROSA project has delivered numerous technical reports over 2019-2021, progressing aspects of a Roadmap produced in 2019, with the majority of this technical work completed by Cancer Council NSW (now via the Daffodil Centre, a joint venture between the University of Sydney and cancer Council NSW). This chapter is part of a milestone ROSA report that synthesises the work to date, provides a set of key findings and recommendations, and delivers an updated Roadmap to help achieve risk-based breast cancer screening in Australia.

1.1.2 Breast cancer population screening tests

The BreastScreen Australia program uses mammography as its primary screening test. Alternative breast imaging methods such as digital breast tomosynthesis (DBT), hand-held ultrasound (HHUS), automated breast ultrasound (ABUS) and magnetic resonance imaging (MRI) are now used or being trialled in Australian diagnostic services. Some breast imaging may be suitable for use within the BreastScreen program either for all women or specific subgroups of women, for example, those at higher risk of interval cancers.

An initial scoping review was undertaken in August 2019. This review was then updated and broadened to include contrast-enhanced mammography (CEM) for in May 2020 and updated again in September 2021.

The current summary consolidates these findings and presents a scoping-level review of how emerging breast screening imaging technologies perform in different risk groups within screened populations. Given the high level of interest and activity in the area of emerging breast imaging technologies, we also summarise potentially relevant ongoing systematic reviews and trials of DBT, ultrasound, CEM and MRI.

We note that the Australian Government Department of Health commissioned a 2018 report on emerging breast imaging technologies (Beresford 2018), however this did not address how imaging tests performed in sub-populations as is required for consideration of risk-based screening. It also commissioned an updated review of DBT in 2020 (Grimble 2020) which considered its performance in different age and breast density groups in relation to breast cancer detection rates. The current ROSA review adds a comprehensive overview of various outcomes, for a wider range of breast imaging tools.

1.1.3 Modelling risk-based population breast screening

The net impact of risk-based screening comprises a combination of costs, benefits and harms. Trials of risk-based screening protocols are likely to yield the highest-quality evidence, however, such trials take time, they can evaluate a limited range of screening protocols, and they are unlikely to generate meaningful mortality outcomes in pace with

advances in screening technology and improvements in cancer treatment (and potentially prevention).

Meanwhile, numerous microsimulation modelling studies have estimated the likely costs, benefits and harms of various risk-based screening strategies. This includes modelling of risk-based strategies of potential interest in the Australian setting but not yet being trialled elsewhere. On this basis, the ROSA project undertook a scoping level review of published clinical and health economic models of risk-stratified breast cancer screening.

1.1.4 ROSA modelled evaluation of risk-based breast screening protocols in Australia

Models generally estimate clinical and/or health economic outcomes. The ROSA project has undertaken clinical and health economics modelling of a range of risk-based screening scenarios in the Australian setting, including combinations of risk-based breast imaging technologies, screening intervals and target age ranges. Primary outcomes of interest are mortality rates, tumour characteristics, costs and cost-effectiveness.

1.2 Project activities

The ROSA project has undertaken a range of activities to gain insights into risk-based screening protocols 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 sections and their related ROSA project activities

Chapter section and topic	Approach/methods
2. Alternative screening modalities by risk group (from page 8)	A scoping-level review of breast cancer screening outcomes according to screening imaging modalities.
3. Published modelled evaluations of risk-based screening (from page 9)	A scoping-level review of published clinical and health economic models of risk-stratified breast cancer screening.
4. ROSA modelling of risk-based screening in Australia (from page 39)	A clinical and health economics modelled evaluation of a range of screening protocols in Australia.

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 5.1 (page 48).

In summary, evidence on the value of using other imaging modalities across different risk groups continues to be generally mixed, with evidence of increased cancer detection for higher risk groups (due to breast cancer risk and/or higher breast density) through supplemental ultrasound, digital breast tomosynthesis and MRI. However, uncertainty remains about the extent to which increased cancer detection is due to earlier detection of potential interval cancers versus increased overdiagnosis. Some observed differences between studies are likely to be attributable in part to differences in settings, study groups

and potentially the version of technology used. Australia would benefit from conducting its own studies evaluating screening technologies specifically in the Australian screening population setting, as indicated by the ROSA recommendations.

Our findings add to the Beresford (2018) review of emerging breast imaging technologies, by additionally reporting and comparing the performance of imaging technology according to risk groups. The findings also provide an update to a 2020 review of DBT by Grimble *et al.*(2020) which considered the performance of DBT in different age and breast density groups for evidence up to November 2019; we include evidence to June 2021. The high level of interest and activity in emerging breast imaging technologies for use in population screening is reflected in the number of potentially relevant ongoing systematic reviews (n=14) identified and ongoing collation (and critical appraisal) of emerging evidence is warranted.

Our scoping level review of published clinical and health economic models of risk-stratified breast cancer screening highlights that clinical modelling should incorporate, at a minimum, current screening program protocols and participation rates as well as screening cancer detection rates, interval cancer rates and false positive rates (and thereby other derivable outcome measures, such as recall rates and program sensitivity and specificity). Additionally, modelled estimates should consider the benefits and harms for each risk group as well as for the whole population. Several published models incorporated breast density, which is an important consideration for risk-based breast screening.

The modelled evaluation reported here indicates potential benefits, harms and costs of a range of risk-targeted breast screening protocols, where screening protocols are characterised according to age range, screening technology and screening intervals. Modelled scenarios were selected based on available evidence, consultation with Australian experts and stakeholders, and what can reasonably be estimated through simulation modelling. A shortlist of nineteen screening protocols was identified based on a balance of clinical and cost-effectiveness outcomes. After comparison with an additional set of scenarios as indicated, the resulting shortlist provides a basis from which to design and plan trials of risk-based breast cancer screening in Australia.

1.4 Glossary of terms

ABS	Australian Bureau of Statistics
ADH	Atypical Ductal Hyperplasia
AI	Artificial Intelligence
AIHW	Australia Institute of Health and Welfare
AutoDensity	Image processing software used to automatically measure breast density from mammograms
BAU	Business-As-Usual, used in the ROSA modelling evaluation to describe current BSA protocols
Better prognosis cancers:	A term used in the ROSA modelling evaluation to describe invasive breast cancers that are low grade (grade 1), small (<15mm) and non-nodal at diagnosis.
Bilateral mammography	Mammography of both breasts
BRCA1/2	The genes most commonly affected in hereditary breast (and ovarian) cancer.
Breast Density	The extent (amount and distribution) of radiopaque tissue in the breast. Usually perceived through mammography and described as either the proportion or area of the breast that is dichotomously dense, or through categories such as the BI-RADS breast density categories that combine quantitative and qualitative aspects of the breast density.
BSA	BreastScreen Australia
BSAMR	BreastScreen Australia Monitoring Reports (published regularly by the AIHW)
BSAPMG	BreastScreen Australia Program Management Group
BSV	BreastScreen Victoria
CCA	Cancer Council Australia
CEM	Contrast Enhanced Mammography
Community-detected cancer:	Cancer diagnosed outside the screening program, including interval cancers
Cumulus	Image processing software used to assist a reader measuring breast density from mammograms through adjustment of greyscale thresholds to partition the dense versus non-dense tissue
DBT	Digital Breast Tomosynthesis
DCIS	Ductal Carcinoma *in Situ*, a form of pre-invasive breast disease confined to the breast ducts
DM	Digital Mammography
EAG	The ROSA project Expert Advisory Group
False positive screens	A screening episode recalled to assessment with a benign final outcome after assessment

Family history of breast cancer	Some family history of breast cancer, defined in various ways: refer to context for specific definitions.
FCC	Family Cancer Clinics
HRT	Hormone Replacement Therapy
Hypothetical screening tests	A term used in in the ROSA modelling evaluation describing screening tests modelled for a range of specified sensitivity and specificity values.
ICER	Incremental Cost Effectiveness Ratios; calculated by dividing the difference in costs by the difference in effectiveness.
Interval cancers	In this report defined as breast cancers diagnosed following a negative screen (i.e. not screen-detected) and before the next recommended screen at 24 months (or 12 months if screening annually)
LCIS	Lobular Carcinoma *In Situ*
LYG	Life-years gained
LYS	Lie-years saved
MD	Mammographic Density, used to describe breast density specifically as perceived through mammography.
MHT	Menopausal Hormone Therapy (also known as HRT)
Missed cancers	A term used in the ROSA modelling evaluation, defined as cancers at least 1mm in diameter but not detected at screening.
Mode of detection	Categorical description of how cancers were diagnosed i.e. screen-detected, interval cancer or other (i.e. cancers diagnosed outside the program but after the interval period).
MRI	Magnetic Resonance Imaging
Negative screening episode	A screening round not recalled for further assessment.
Nodal involvement	Breast cancers that involve the lymph nodes.
OOP	Out of pocket (costs)
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 of exposure (C), and the outcomes to be reported and assessed (O)
***Policy1-Breast* Model**	The simulation modelling platform used for ROSA modelling evaluation.
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.
Program specificity	The proportion of non-recalled screening episodes not followed by an interval cancer diagnosis

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 rates	The proportion of screening episodes requiring recall for further assessment.
Recall to assessment	Recall to further investigation by BreastScreen assessment services, following a screening mammogram.
Rescreening rates	In this report defined as the number of women who returned to have a BreastScreen mammogram within 27 months of their most recent screen (or 15 months if annual screening interval) divided by the total number of women who attended the most recent screen, expressed as a percentage
Screen-detected cancer	Cancer detected by a population screening program
Screening test sensitivity	The estimated proportion of cancers present at the time of the screening test that are detected.
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).
US	Ultrasound
Worse prognosis cancers	A term used in in the ROSA modelling evaluation to describe invasive breast cancers that are high grade (grade 3), large (at least 15mm in diameter) and involving the lymph nodes at diagnosis

2 Alternative screening modalities by risk group

2.1 Authors

Suzanne Hughes, Chelsea Carle, Victoria Freeman, Dr Susan Yuill, Dr Louiza Velentzis & A/Prof Carolyn Nickson.

2.2 Background

1.1.1 Rationale

The BreastScreen Australia program uses mammography as the primary screening test. Alternative breast imaging methods such as digital breast tomosynthesis (DBT), hand-held ultrasound (HHUS), automated breast ultrasound (ABUS) and magnetic resonance imaging (MRI) are now used or being trialled in diagnostic services and may be suitable for use within the BreastScreen program either for all women or specific subgroups of women, for example, those at higher risk of interval cancers.

This summary presents a scoping-level review of how emerging breast screening imaging technologies perform in different risk groups within screened populations. Although emerging imaging technologies were reviewed in 2018 (Beresford 2018), their performance in sub-populations was not addressed. An initial scoping review was undertaken in the August 2019. This review was then updated and broadened to include contrast-enhanced mammography (CEM) in May 2020 and subsequently re-updated in September 2021 . The current report presents the consolidated results for this scoping review.

A tomosynthesis review undertaken in 2020 by Grimble *et al*. considered the performance of tomosynthesis in different age and breast density groups but reviewed the evidence only till November 2019. In 2022 a meta-analysis found that increases in cancer detection with tomosynthesis were significantly greater for women with denser breasts (Li 2022).

Given the high level of interest and activity in the area of emerging breast imaging technologies, we also summarise potentially relevant ongoing systematic reviews and trials of tomosynthesis, ultrasound, contrast-enhanced mammography and MRI.

1.1.2 Research question

How do new breast imaging technologies/modalities perform for different breast cancer risk groups?

1.1.3 Aims

1. Identify and summarise the results of studies examining DBT, ultrasound, CEM or MRI as alternatives or adjuncts to mammography for different risk groups in screening populations.

2. Identify potentially relevant ongoing systematic reviews and trials of DBT, ultrasound, CEM or MRI.

Abridgement note: Further detail will be made available through a publication in progress.

3 Published modelled evaluations of risk-based screening

3.1 Authors

Dr Susan Yuill, Suzanne Hughes, Dr Louiza Velentzis & A/Prof Carolyn Nickson

3.2 Background

The net impact of risk-based screening comprises a combination of costs, benefits and harms. Trials of risk-based screening protocols (such as the WISDOM or DENSE trials) are likely to yield the highest-quality evidence however such trials take time, they can evaluate a limited range of screening protocols, and they are unlikely to generate meaningful mortality outcomes in pace with advances in screening technology and improvements in treatment (and potentially prevention).

Meanwhile, numerous microsimulation modelling studies have estimated the likely costs, benefits and harms of various risk-based screening strategies. This includes modelling of risk-based strategies of interest in the Australian setting but not yet being trialled elsewhere.

Models generally estimate clinical and/or health economic outcomes. Models describing only health economic outcomes in settings outside Australia are of limited value because it is difficult to translate such findings to the Australian health setting. However, models describing clinical outcomes (with or without health economic outcomes) provide information that might be translated to the Australian setting.

Not all modelled scenarios are feasible in the Australian health setting. Models without detailed natural histories are limited in quality, based on assumptions about lead time with different screening modalities and missing detail on the variance of tumour natural histories. Models of specific risk groups (e.g. women with high-risk genetic mutations) do not describe population-level risk-based screening as being explored by this project.

On this basis, we undertook a scoping level review of published clinical and health economic models of risk-stratified breast cancer screening, restricted to models that met specific criteria in terms of the population modelled and information on clinical outcomes.

3.3 Research question

What are the relative benefits, harms and costs of risk-based breast cancer screening as estimated by population-level modelling studies relevant to the Australian health setting, and how would their clinical and health economics estimates transpose to an Australian setting?

3.4 Aim

1. Identify modelled estimates of relative benefits, harms and costs of risk-based population-level breast cancer screening that describe clinical outcomes (with or without health economics outcomes)
2. Assess whether the modelled scenarios would be plausible in the Australian setting
3. For plausible scenarios, assess the modelling generalisability to the Australian setting.

3.5 Methods

While this is a scoping review not a systematic review, we used a systematic approach to identify and extract the evidence, by establishing clear PICOs and selection criteria.

3.5.1 PICO protocol

Table 2. The PICO framework for the scoping review of population simulation models of risk-based breast cancer screening.

Population	Intervention/exposure	Comparison	Outcome
Asymptomatic women aged ≥ 40 years of age with average/unknown risk of breast cancer	Risk-stratified breast cancer image-based screening	No screening, or no risk stratification, or an alternative risk stratification.	Clinical outcomes Health economic outcomes

3.5.2 Selection criteria

Selection criteria are shown in Table 29. These selection criteria were specified with the aim of including only high-quality models that simulate risk-based breast cancer screening as an intervention on a population. For transparency and generalisability, models were required to model and report estimated clinical outcomes (not only health economic outcomes). Note that QALYs described quality-adjusted life years, which is 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.

Table 3. More detail on the PICO selection criteria

	Inclusion	***Exclusion***
Study design	Population simulation models of breast cancer screening Outcomes must be modelled to a lifetime horizon or to a specific upper age limit.	Life table models (models applying basic multiplications of assumed outcomes to groups of women, rather than detailed natural histories and interventions)
Population	Asymptomatic women aged ≥ 40 years of age with average/unknown risk of breast cancer	Restricted to a specific risk group.
Intervention or exposure	Risk (including family history, mammographic density and risk prediction tools) stratified breast cancer image-based screening	Personal or clinical breast examination screening Age stratified only
Comparator	Breast cancer screening which is not risk stratified or An alternative risk stratification of breast cancer screening or No screening	No comparator or comparator differs for different risk groups
Outcome	Clinical outcomes (e.g. mortality, tumour characteristics at diagnosis, interval cancer rates, false-positive recall rates) with or without Health economics outcomes (e.g. QALYs, relative and absolute costs per QALY saved)	Models reporting only health economics outcomes e.g. cost per QALYS saved. Models estimating short-term outcomes e.g. outcomes at each screening episode.
Publication date	Published from 2008 onwards	No other publication limits

Publication type	Journal article or report	Conference abstracts, reviews, letters, editorials and comments
Language	English	

3.5.3 Literature searches

To identify relevant articles published from 2008 onwards. Medline and Embase databases were searched in June 2019 and subsequently on the 7th June 2021 (from 2018 onwards), by combining terms for breast risk-based screening and modelling. Details of the complete search strategy are presented in the Appendix. The full text of any articles that might meet the inclusion criteria were collected. Studies that met all the selection criteria were included and those that did not were excluded with reasons for exclusion documented. For details of reasons for exclusion of potentially relevant articles see Table 34.

3.5.4 Data extraction

Pre-determined study details and data were extracted.

3.5.5 Checklist for evaluating risk-based models

Risk-based evaluations were also assessed in terms of their reporting of key parameters. For this purpose, a checklist was created using relevant items from the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist (Husereau 2013) modified into a question from a statement and items from the Drummond framework (Drummond 2005). Additional items were identified from the evaluation of risk-based studies in the ROSA project. A total of 53 items were included covering model inputs and outcomes, model assumptions, data, sensitivity analysis and interpretation and items relevant to economic evaluations. Reporting of items was scored as yes, no, partial, or not applicable.

3.6 Results

3.6.1 Descriptive results

We present consolidated results from all searches conducted. In total, full texts of ninety-one potentially relevant articles were identified of which ten met the inclusion criteria. Excluded studies are listed in Table 34 (page 307) by author and year with PubMed IDs or digital object identifier (DOI).

Summary information of included studies is presented in Table 30 and 31, including the setting, perspective, scenarios modelled, cost considerations, key outcomes and limitations. Five of the included studies modelled scenarios were set in the US, four in Europe (Germany, Spain, two in the Netherlands) and one in Asia (Singapore).

A descriptive summary of each study is provided below:

- The Germany-based modelling by Arnold et al (Arnold 2019) compared scenarios where screening intervals (1, 2, or 3 years) were assigned according to relative risk of breast cancer. Overall, they found that risk-stratified screening could be more efficient, and the optimal approach would require a trade-off between mortality reduction or quality of life. The authors noted that the model was developed on film mammography, that it was imperfectly translated to the German setting from its original US setting, and that it did not include a natural history model, instead assuming stage-specific detection rates.

- The US setting modelling by Trentham-Dietz et al (Trentham-Dietz 2016) compared scenarios where screening intervals (1, 2, or 3 years) were assigned according to relative risk of breast cancer and mammographic density (BIRADS 1/2 versus BI-RADS 3/4). This paper used various CISNET models and estimates varied between the three models used. Overall, setting outcomes for biennial screening of average-risk women as the benchmark, they found that the balance of benefits and harms in other risk groups would meet or exceed that benchmark with triennial screening for average-risk women with BIRADS 1/2 breast density and annual screening of higher-risk women with BIRADS 3/4 breast density.
- The US-setting modelling by Mandelblatt et al (Mandelblatt 2016) compared scenarios based on screening commencement age (40, 45 or 50), screening interval (annual, biennial, or annual for age 40-49 then biennial), and mammographic density (entirely fatty, scattered density, heterogeneously dense and extremely dense). They found that annual screening was 'inefficient' for all levels of mammographic density, and that women with lower breast density had a greater proportion of their cancers detected by screening.
- The US-setting modelling by Stout et al (Stout 2014) evaluated annual screening for women with BI-RADS 3/4 mammographic density and biennial screening for women with BI-RADS 1/2 mammographic density, compared to biennial screening and to no screening. They found that annual screening for women with BI-RADS 3/4 breast density and biennial for other women led to increased benefits but higher false-positives (and higher costs).
- The Spain-based modelling by Vilaprinyo et al (Vilaprinyo 2014) compared screening intervals (1, 2, 3 or 5 years) for various ages of commencement (40, 45 and 50 years), assigned according to combinations of BI-RADS mammographic density categories (1-4), family history of breast cancer, and a personal history of breast biopsy They found that risk-based screening was more efficient and had a better balance of benefits and harms, including protocols with 5-yearly screening for women at lowest risk and annual screening for women at highest risk.
- A Dutch study by Sankatsing et al (Sankatsing 2020) compared scenarios with varying age range and interval of digital mammography in low risk (RR 0.75), average risk (RR 1.0) (total population), and high risk (RR 1.8) groups (based on risk factors other than breast density and BRCA1/2 mutations) to no screening. Triennial screening from 50-71 years was found to be optimal for low-risk women, maintaining the benefits while reducing the harms and costs compared to the current screening schedule. Optimal and cost-effective screening for high-risk women was biennial screening from age 40 to 74 years.
- In another study set in the Netherlands, modelling by Wang et al (Wang 2020) compared screening with digital breast tomosynthesis (DBT) vs digital mammography in high-risk women (BI-RADs 3/4), and in all women, with varying sensitivities for mammography (65-87% depending on breast density) and for DBT (65-100%). More benefits (screen-detected tumours and LYG) and fewer harms (interval tumours) were estimated with DBT screening in women with dense breasts when the DBT sensitivity was ≥72%. Similar findings were estimated for DBT screening in all women when DBT sensitivity reached >85%. DBT was more likely to be a cost-effective alternative to mammography in women with dense breasts. Whether DBT could be cost-effective in general population screening was highly dependant on DBT costs.
- In a study set in the United States, modelling by Shih et al (Shih 2021) compared seven screening strategies. These included no screening, biennial or triennial screening for all women aged 50-75 years, or stratified screening based on baseline mammographic breast density, varying by age range and screening interval, with annual screening in women with high breast density (BI-RADS C/D). A baseline breast density assessment at age 40 years, followed by a strategy of annual screening at age 40 for women with dense breasts, and biennial screening

starting at age 50 for women without dense breasts, was associated with the greatest reduction in breast cancer mortality and was cost-effective but involved the most screening mammograms in a woman's lifetime and higher rates of false-positive results and overdiagnosis compared to other strategies modelled.

- In another study set in the United States using 2 different breast cancer models, Van Broek et al (Van Broek 2021) modelled 47 risk groups based on family history and polygenic risk score (PRS), alone and in combination, with screening strategies varying by starting age and screening interval. Compared to current biennial screening from age 50-74 years, risk-tailored screening was found to prevent more breast cancer deaths for women at high risk based on combined family history and polygenic risk score, but with increases in overdiagnosis and false positives. Screening based on PRS had greater benefits than screening based on family history only but combining PRS and family history maximized improvement in outcomes.
- A study set in Singapore (Wong 2021) compared a risk tailored screening strategy based on individual women's polygenic risk score (PRS) at the age of 35 years with the current biennial screening strategy from age 50-69 years. Women were stratified into 3 risk groups based on score cut-offs: low risk (40-≦60th percentile) (triennial mammogram 40-74 years), intermediate risk (30th-55th percentile) (biennial mammogram from age 40-74 years) and high risk (5-10th percentile) (annual ultrasound from age 35-39 years, then annual mammogram from age 40-74 years). Compared to current screening, the tailored screening strategy was found to be more effective (increase in LYG and QALY gained per women) and cost-effective.

3.6.2 Assessment of evaluations based on checklist

Assessment of articles according to the checklist is presented in Table 8 in the appendix. Included evaluations addressed most checklist items relevant to inputs and outcomes (items 1-21), as well as data (items 28-29), however, partial or no reporting was observed for up to half the studies for tumour sub-types, recurrence, clinical outcomes such as benign breast disease, false negative screens or DCIS, and consideration of lead time for studies estimating life years gained.

In terms of model assumptions (items 22-27), five out of eight evaluations did not model screening participation other than 100% of the target population modelled (e.g. such as observed participation for the setting modelled) and eight out of ten studies did not report whether adherence to treatment was modelled.

For evaluations incorporating cost-effectiveness (items 30-41), up to half of studies did not report details of price adjustments for inflation or current conversion, the type of perspective taken, the willingness to pay threshold, nor explained the choice of discount rate.

In terms of sensitivity analysis (items 42-47) most evaluations did not describe the uncertainty of model parameters, nor considered their correlations or joint distributions while nearly half of the items for interpretation of results (items 48-53) were either partially reported or not reported by the majority of evaluations.

3.7 Tables

Table 30 and *Table 31* summarise the characteristics of studies included in our review.

Table 4. Summary table of characteristics of included risk-based screening studies identified in the 2019 ROSA scoping review.

	Arnold et al, 2019 *(Arnold 2019)*	***Trentham-Dietz et al, 2016*** *(Trentham-Dietz 2016)*	***Mandelblatt et al, 2016^*** *(Mandelblatt 2016))*	***Stout et al, 2014)*** *(Stout 2014)*	***Vilaprinyo et al, 2014*** *(Vilaprinyo 2014)*
Setting	Germany	US	US	US	Spain
Perspective (e.g. healthcare)	Payer of statutory health insurances	Not applicable	Not applicable	Federal payer	NHS (direct healthcare costs)
Base case (e.g. no screening) noting if digital mammography (versus film), and the base case screening interval	No screening. (Analog screen-film mammography performance used in the model).	Biennial screening in women 50-74y. (Model based on digital mammography)	No screening (model was based on digital mammography)	No screening, and biennial film mammography screening from 50-74y	Biennial screening in women aged 50-69 and biennial screening in women aged 45-74. (Model includes estimates from film mammography).
Scenarios	3 uniform screening strategies (annual, biennial, triennial) plus 5 strategies based on varying screening intervals according to W's relative risk of BC. High, medium and low risk categories (with varying RR thresholds) allocated annual, biennial and triennial screening, respectively. E.g. high risk was either >2; >1.5 or >1. RR was based on factors in BCSC risk calculator (breast density, family history, previous biopsy).	Screening intervals tailored to BD and risk. BD was categorised as 'low' for BI-RADS 1/2 and 'high' for BI-RADS 3/4. Risk groups were for RR = {1.0, 1.3, 2.0, 4.0}.	8 strategies varying by screening initiation age (40, 45, or 50y) and screening interval (annual, biennial, and hybrid [annual for women in their 40s and biennial thereafter]). BD was modelled as entirely fatty ("a"), scattered density ("b"), heterogeneously dense ("c") and extremely dense ("d"). Risk levels included 1.3x, 2.0 x, or 4.0 x higher than average.	Annual screening from ages 40-74 y for women with dense breast tissue [BI-RADS 3 or 4] and biennial screening for W with BI-RADS 1 or 2	2,625 screening strategies conducted, 24 uniform and 2,601 risk-based. The risk-based strategies were obtained combining the exam periodicity [annual (A), biennial (B), triennial (T), and quinquennial (Q)], the starting ages (40, 45 and 50 y) and the ending ages (69 and 74 y) in 4 risk groups, Low, Medium-Low, Medium-High and High. Risk groups based on BD measured using BI-RADS 1-4, FH of BC in 1st degree relatives (yes/no) and PH of breast biopsy (yes/no).
Populations	3 million women at age 50	(i) Women aged 50 starting screening (ii) Women aged 65 continuing screening	Cohort of women born in 1970 with average risk and average BD	Cohort of women born in 1960	100,000 women
Outcome horizon	50-100y or until end of life	Lifetime horizon	Lifetime (from 25y to 100y)	Lifetime (from age 40y)	40-79 years of age

	Arnold et al, 2019 *(Arnold 2019)*	***Trentham-Dietz et al, 2016*** *(Trentham-Dietz 2016)*	***Mandelblatt et al, 2016^*** *(Mandelblatt 2016))*	***Stout et al, 2014)*** *(Stout 2014)*	***Vilaprinyo et al, 2014*** *(Vilaprinyo 2014)*
Outcome measures - benefits	BC mortality reduction, QALY.	BC deaths averted, LYS and QALYs gained	BC mortality reduction, BC deaths averted, LY, QALY (accumulated from age 40-100y)	BC mortality reductions, BC deaths averted, LY, QALYs	Number of lives extended; QALYs
Outcome measures - harms	Overdiagnosis, false-positive screening results.	False-positive screening results, benign biopsies, overdiagnosis	False-positive screening results, benign biopsies, and overdiagnosis	False positive screening results	False positive results, false negative results, overdiagnosis of invasive BC; DCIS attributable to screening
Other outcomes	Cost-effectiveness ratios	Cost-effectiveness, harms: benefits ratios (false-positives:BC deaths averted)	Harms: benefits ratios	Cost-effectiveness ratios	Cost-effectiveness ratio, harm: benefit ratios
Costs	Mammographic screening and associated additional procedures (follow-up, core needle, vacuum biopsy), treatments (surgical, chemo, endocrine, radiotherapy, palliative care)	Mammograms, follow-up of positive mammograms, stage-specific cancer treatments	None	Costs of film and digital mammography screens; diagnostic costs; treatment costs;	Costs of screening and diagnosis confirmation, initial treatment, follow-up and advanced care costs
Key outcomes (examples)	Risk-stratified screening programs can be more efficient depending on whether mortality reduction or QALY is more important, the willingness to pay threshold and the adherence assumption (full population adherence or 54% adherence).	Compared to biennial screening of average-risk women, scenarios that would maintain a similar or better balance of benefits and harms in other risk groups were: i) Triennial screening for average-risk / low-BD W; ii) Annual screening for higher-risk / high-BD women.	Women at higher risk have fewer false-positive results per 1000 women screened and higher gains from screening than lower-risk groups. Screening higher-risk women yielded a lower ratio of overdiagnosed cases per BC death averted than screening average-risk women. Annual screening from 50-74y remained inefficient across BD groups.	Compared to either no screening, or biennial screening with film or digital mammography screening for all, annual screening for women with higher BD and biennial for others resulted in more health benefits but also higher false positives and higher costs.	Risk-based screening strategies were more efficient and had lower harm-benefit ratios than uniform strategies. For example, compared to the biennial screening in women 50-69y, the risk-based strategy of 5-yearly screening of low-risk w 50-69y + 5-yearly screening of women of medium-low risk ages 45-74y + 5-yearly screening in women of medium-high risk 45-74y and annual screening of high-risk women 40-74y results in reductions of 8% in costs, 17.2% in false positives and 25% in overdiagnosed cases.

	Arnold et al, 2019 *(Arnold 2019)*	***Trentham-Dietz et al, 2016*** *(Trentham-Dietz 2016)*	***Mandelblatt et al, 2016^*** *(Mandelblatt 2016))*	***Stout et al, 2014)*** *(Stout 2014)*	***Vilaprinyo et al, 2014*** *(Vilaprinyo 2014)*
Limitations (as stated by authors)	i) Model is based on screening performance of analog screen-film mammography; ii) Not all data sources were from Germany; iii) Model originally developed for US setting and not all elements could be adapted to a German setting; iv) Model does not have a natural history component of BC but relies on stage-specific detection rates.	i) Women <50y, BRCA1 /2 carriers and women below average-risk were not modelled; ii) Supplemental or alternative imaging modalities to digital mammography were not considered.; iii) Morbidities were not included in the model; iv) Model assumed a constant risk of BC over time	i) Other imaging technologies, polygenic risk, and non-adherence were not considered; ii) 100% adherence to screening, prompt evaluation of abnormal results, that risk factors influenced only the incidence of disease, but not its natural history, and full use of optimal treatment were assumed; iii) Radiation-induced BCs, owing to more intensive mammography schedules were not modelled; iv) Relative risks were held constant over time even though some are age-dependant	i) Variability across models based on different assumptions; ii) Overdiagnosis was not modelled explicitly; iii) It was assumed that effects on risk were based on BD at ages 40-49y based on cumulative exposure hypothesis, which may overestimate the effects of high BD and screening benefits; iv) 100% screening and treatment adherence was assumed, v) no societal costs	i) Use of data from countries other than Spain when regional or country levels data was not available; ii) Variety of data sources used necessitating careful interpretation of model outputs; iii) BC risk was assumed to influence only the BC incidence and not other factors such as distribution of stages at diagnosis; iv) Risk factors in risk groups assumed to be constant after screening started; v) Age-specific sensitivities for M screening based mostly on film mammography rather than digital; vi) The model does not consider DCIS as one of the BC stages

BC: breast cancer; BCSC: Breast Cancer Surveillance Consortium; BD: breast density; BI-RADS: Breast Imaging Report and Database System; DCSI: ductal carcinoma in-situ; LY: life-years; NHS: national health system; PH: personal history; Q: quinquennial; QALYs: quality adjusted life-years; Y: years

^For Mandelblatt 2016, results are only provided descriptively as this was a subgroup analysis. No tabular results for risk-stratification. The authors refer to a large report generated for the USPSFT.

Table 5: Summary table of characteristics of included *risk-based screening studies identified in the 2021 update to the ROSA scoping review*

	Sankatsing et al, 2020 *(Sankatsing 2020)*	***Wang et al, 2020*** *(Wang 2020)*	***Shih et al, 2021*** *(Shih 2021)*	***Van Broek e al, 2021*** *(Van Broek 2021)*	***Wong et al, 2021*** *(Wong 2021)*
Setting	Netherlands	Netherlands	United States	United States	Singapore
Perspective (e.g. healthcare)	Healthcare system	Direct medical costs	Societal		Healthcare system
Base case (e.g. no screening) noting if digital mammography (versus film), and the base case screening interval	No screening (Model based on digital mammography)	Biennial screening for women aged 50-75y using digital mammography (DM)	No screening (Model based on digital mammography)	No screening (Model based on digital mammography)	Biennial screening 50-69y (current screening strategy) (Model based on mammogram)

	Sankatsing et al, 2020 *(Sankatsing 2020)*	***Wang et al, 2020*** *(Wang 2020)*	***Shih et al, 2021*** *(Shih 2021)*	***Van Broek e al, 2021*** *(Van Broek 2021)*	***Wong et al, 2021*** *(Wong 2021)*
Scenarios	3 risk groups based on risk factors other than breast density: 1. Low (RR 0.75) (based on age at 1st child and number of children, hours of physical activity a week, if ever breastfed, and ages at menopause and menarche) 2. Average (RR 1.0, total population) 3. High (RR 1.8) (based on history of benign breast disease, 1st degree family member with BC, history of proliferative disease without atypia and combination family member with BC and smoking) Screening strategies • Current – biennial 50-74y • Low risk (101 strategies): Start age 50-60y Stop age 64-74y Biennial or triennial • High risk (182 strategies): Start age 40-50y Stop age 74-84y Annual or biennial	1. Digital breast tomosynthesis (DBT) for women with high-density breasts (BI-RADS 3 and 4), (DM for women with non-dense breasts) 2. DBT for all women aged 50–75y Both scenarios were modelled with a range of sensitivities of DBT between 65% and 100% in 5% steps and with the sensitivity of DM varying depending on BD (85% for BI-RAD 1 to 65% for BI-RAD 4).	7 screening strategies based on BD (dense breasts= BI-RADS C and D) 1. No screening 2. Triennial 50-75y 3. Biennial 50-75y 4. Dense breasts at age 50y: annual 50-75y Otherwise: triennial 50-75y 5. Dense breasts at age 50y: annual 50-75y Otherwise: biennial 50-75y 6. Dense breasts at age 40y: annual 40-75y Otherwise: triennial 50-75y 5. Dense breasts at age 40y: annual 40-75y Otherwise: biennial 50-75y	(Two different BC models used) 47 potential risk groups – 5 family history (based on age range in which first-degree relative was diagnosed), 7 polygenic risk score (PRS) and 35 combinations of both Screening strategies varying by age at initiation (30, 35, 40, 45 or 50y), screening interval (annual, biennial, triennial, and hybrid combinations) and modality (ultrasound + mammography vs mammography) Compared to 3 US screening guidelines 1. USPSTF (biennial screening between ages 50-74y) 2. American College of Radiology (annual screening from age 40y) 3. ACS (annual screening from ages 45-54y +/- biennial screening from age 55y)	Risk tailored screening strategy based on individual woman's polygenic risk score (PRS) at age of 35y (buccal swab): 3 risk groups 1. Low 40th -≤60th percentile: triennial mammogram 40-74y 2. Intermediate 30th-55th percentile: biennial mammogram 40-74y 3. High 5-10th percentile: annual ultrasound 35-39y then mammogram 40-74y
Populations	Cohort of women born in 1974.	100,000 women from the Netherlands aged 50-75y	Cohort of 500,000 women born in 1970 who are average risk of BC (no known genetic risk factors or family history of BC)	1985 US birth cohort	Singaporean women aged 35-74y
Outcome horizon	Lifetime (from age 40y)	Lifetime	Lifetime	Lifetime	Time horizon of 40y (age 35-74y)
Outcome measures - benefits	LYG, BC deaths averted	Screen-detected tumours, LYG	BC deaths averted, LYG, QALYs gained	LYG, BC deaths averted	LYG, QALY gained

	Sankatsing et al, 2020 *(Sankatsing 2020)*	***Wang et al, 2020*** *(Wang 2020)*	***Shih et al, 2021*** *(Shih 2021)*	***Van Broek e al, 2021*** *(Van Broek 2021)*	***Wong et al, 2021*** *(Wong 2021)*
Outcome measures - harms	False-positives, overdiagnosis		False positives, benign biopsies, overdiagnosis, additional costs	Overdiagnosis, false positive mammograms	
Other outcomes	Cost-effectiveness ratios, benefit-harms ratios	Cost-effectiveness ratios	Cost-effective ratios	Benefit-harm ratios (LYG/ screens, LYG/ overdiagnoses, BC deaths averted/false positives)	Cost-effectiveness ratios
Costs	Screening, additional diagnostics, stage-specific treatments	Screening DM or DBT, biopsy, treatment (tumour diameter)	Screening: mammography, diagnostic mammography, breast biopsy. Treatment: Tamoxifen, Trastuzumab Annual costs of BC treatment by phase and stage Productivity loss secondary to BC	N/A	Buccal swab Mammogram Ultrasound Questionnaire Stage specific direct medical costs
Key outcomes (examples)	Triennial rather than biennial screening was found to be optimal for low-risk women, maintaining the benefits while reducing the harms and costs. In low-risk women, triennial screening between the ages 50-71y, resulted in 33% reduction in false-positives and costs, and improved harm-benefit ratio compared to the current screening schedule (but 19% fewer LYG). Additional screening between the ages of 40-50y was optimal for high-risk women. Biennial screening between the ages of 40-74y in high-risk women resulted in a screening benefit (additional 25% LYG) but also an increase in false positives (+44%) compared to the current schedule.	The effectiveness of both DBT scenarios increased at increased DBT sensitivity, with more screen-detected tumours, fewer interval tumours and more LYG (in women with dense breasts, with a DBT sensitivity of >=72%; in the whole population with a DBT >85%). DBT is more likely to be a cost-effective alternative to mammography in women with dense breasts. Whether DBT could be cost-effective in a general population highly depends on DBT costs.	A baseline BD assessment at age 40y, followed by annual screening for women with dense breasts, and biennial screening starting at age 50y for women without dense breasts, is associated with the greatest reduction in BC mortality and is cost-effective but involves the most screening mammograms in a woman's lifetime and higher rates of false-positive results and overdiagnosis. For every 1000 women screened, compared to biennial screening from age 50y, annual mammography from age 40y in women with dense breasts (biennial mammography from age 50y in other women) resulted in 3.0 BC deaths averted, 15.1 LYG and 15.1 QALY gained but additionally there were 364.8 false positives, 38.0 benign breast biopsies, and 6.1 overdiagnosed cases	Compared with following general population guideline strategies for women of average risk, risk-tailored screening was associated with more BC deaths averted for women at high risk due to their BC family history and polygenic risk, but with increases in overdiagnosis and false positives. Women with a BC family history who initiated biennial screening at age 40y (vs 50y) had a 36% increase in LYG and 20% more BC deaths averted, but 21% more overdiagnoses and 63% more false positives. Screening tailored to PRS vs biennial screening from 50-74y had smaller positive effects on LYG (20%) and BC deaths averted (11%) but also smaller increases in overdiagnoses (10%) and false positives (26%). Combined use of family history and PRS vs biennial screening from 50-74y had the greatest increase in LYG (29%) and BC deaths averted (18%)	Overall, LYG and QALY gained per woman in the tailored screening program based on cut-offs in score of 60th percentile for low, 35th for intermediate and 5th for high was approximately 0.9720 and 0.9884, respectively. The tailored screening program was also found to be more cost-effective.

	Sankatsing et al, 2020 *(Sankatsing 2020)*	***Wang et al, 2020*** *(Wang 2020)*	***Shih et al, 2021*** *(Shih 2021)*	***Van Broek e al, 2021*** *(Van Broek 2021)*	***Wong et al, 2021*** *(Wong 2021)*
Limitations (as stated by authors)	i). Did not take into account that a women's RR can change over time in base case (reducing mammographic BD with increasing age, altering test sensitivity) ii). Assumed BC risk only affects the incidence of BC and no other parameters including tumour growth rate, sojourn time and stage distribution. iii). Could not calculate the ICERs of current uniform screening as biennial screening between ages 50-74y was close but not on the frontier and hence used biennial screening between ages 48-72y as the cos–effectiveness threshold. iv). No data on distribution of risk groups in the Dutch female population or distribution of BD among risk groups -therefore could not assess the impact of a risk-based screening programme for the whole population.	1. DCIS is not included in the model 2. DM sensitivity generated using meta-analysis that included some studies based on single-reading screening settings (DM sensitivities might be lower compared to double-reading screening settings). 3. Specificity increases with a decrease in BD but there is no reliable data on the dependence of specificity on BD for DBT – therefore used a constant specificity independent of BD in the model. 4. While the study evaluated the effectiveness of a screening program, as there are no data available on the sensitivity of a screening program including DBT, the modality sensitivity of DBT was modelled.	1. Model did not examine alternative screening modalities for women with dense breasts (prior research has shown that for women with dense breasts, digital breast tomosynthesis combined with digital mammography is more cost-effective than digital mammography alone). 2. Although BD changes with age, BD classification was static based on baseline BD, other than differentiating the prevalence of dense breasts for women <50y and those aged ≥ 50y. While both high BD and older age are associated with a higher risk for BC, BD generally decreases with age. 3. The natural history for women with dense breasts may differ from those without dense breasts but due to lack of reliable data, this was not explicitly incorporated into the model.	1. Did not explicitly model the effects of rare, but higher risk variants in genes such as BRCA1, BRCA2, PALB2, CHEK2, or ATM, which are particularly relevant among young women age <50y 2. Accounted for tumour natural history by estrogen receptor and HER2 status, but the models assumed that polygenic risk did not directly affect tumour progression or mode of detection 3. Did not consider screening after age of 74y, costs or QALYs 4. Did not consider risk related to second-degree family members with BC due to data limitations. 5. Assumed perfect uptake of genetic testing, screening, and receipt of treatment 6. Effectiveness of screening in combination with treatment in women age < 40y has been assessed in case-control studies but not in an RCT 7. Considered the effects of BD on mammography performance, but future analyses should consider joint distributions of BD, PRS, and family history.	1. Assumed 100% attendance and compliance with breast cancer screening and follow-ups 2. Scarcity and age of the data for sensitivity of mammography Mammography sensitivity was a limitation in this study, as scarcity and age of the data may impact the results 3. Risk of confounding due to the use of observed screening outcomes rather than natural history data. 4. Did not model for treatment, remission and follow-ups after diagnosis

BC: breast cancer; BD: breast density; BI-RADS: Breast Imaging Report and Database System; LYG: life-years gained; QALYs: quality adjusted life-years; Y: years

3.8 Discussion

3.8.1 Key findings

The modelled evaluations of risk-stratified screening described here suggest that some approaches to risk-based screening may be effective in the settings modelled. We assume that any modelled scenarios would require a full trial or evaluation to generate evidence suitable for widespread implementation in Australia.

In terms of benefits, more intensive screening in higher-risk women, with respect to frequency and commencing at a younger age, predicted lower breast cancer mortality and improved quality of life. For example, Sankatsing et al (Sankatsing 2020) estimated that for higher-risk women, biennial screening commencing at the age of 40 years would result in additional LYG compared to current uniform biennial screening in the Netherlands which commences at age 50. Similarly, Shih 2021 estimated that annual screening from age 40 in women with dense breasts would reduce breast cancer mortality and increase LYG and QALY gained.

In terms of harms, more intensive screening in high-risk women came at the expense of increased false positives and overdiagnoses, and less intensive screening of lower risk women could potentially reduce harms. For example, in an American study (Shih 2021), annual screening in high-risk women from the age of 40 was estimated to increase rates of false positives, benign breast biopsies and overdiagnosis, while less intensive screening for lower-risk women, such as triennial screening, would reduce false positive rates. In a Dutch study, triennial screening of low-risk women (relative risk of breast cancer of 0.75 compared to 1 for average risk women) from the age of 50 to 71 was estimated to lead to reduced false positive rates and an improved harm-benefit ratio compared to biennial screening from 50 to 74 years (Sankatsing 2020).

In terms of cost-effectiveness, risk-based screening was predicted to be more cost-effective than uniform screening in all women in some models (Shih 2021, Wong 2021, Sankatsing 2020), but not in others. Stout *et al.* estimated that annual screening for women with higher breast density and biennial screening for other women was less cost-effective than biennial screening for all (Stout 2014).

Screening higher risk women with an alternative modality (DBT) to mammography could be more cost-effective compared to DM, dependent on the sensitivity and cost of DBT. Directing DBT to higher-risk women was estimated to be more cost-effective than introducing DBT as a standard screening test for all women, however, a higher cost of a DBT screen would require higher sensitivity for DBT to be cost-effective (Wang 2020).

Adherence to screening was an additional consideration which could significantly impact cost-effectiveness. Arnold et al (Arnold 2019) found that reduced screening adherence decreased the cost-effectiveness of risk-based screening strategies.

3.8.2 Model specifications and outcomes

For the 10 studies reported here, various approaches were applied to modelling breast cancer natural history and risk, risk classification, selection of comparison groups, how screening is specified, and reported outcomes. This is to be expected as models are often designed to make best use of available data and to characterise the health setting being modelled. This can create challenges for comparing outcomes from different models, however the methods and findings provide valuable information to inform considerations of risk-based screening in Australia, including

modelled evaluations in the Australian setting and considerations for any large-scale trial or evaluation of risk-based screening.

In this section, we compare examples of various aspects of the modelled evaluations.

Natural history and breast cancer risk

- Wang et al (Wang 2020) did not include DCIS in the model. This is a limitation as the costs and burden of DCIS diagnosis through screening are significant. Vilaprinyo (Vilaprinyo 2014) modelled DCIS but this was not modelled as a natural history stage (e.g. prior to invasive breast cancer). *This is a limitation, as DCIS detection and treatment would be expected to prevent the onset of some invasive breast cancers.*
- Shih et al (Shih 2021) did not examine alternative screening modalities for women with dense breasts. *As indicated in our scoping reviews on this topic, alternative screening modalities for women with dense breasts should be included in considerations of risk-based screening.*
- In the Shih et al model (Shih 2021), although breast density changes with age, other than differentiating the prevalence of dense breasts for women <50 years and those aged ≥ 50 years, breast classification was static, based on the baseline BD assessment. Stout et al (Stout 2014) assumed the effects on risk were based on mammographic density at ages 40-49 years, following the cumulative exposure hypothesis (i.e. that having dense breasts becomes an increasing risk factor over time, as the breast is exposed to the risk effects of its dense tissue). The authors noted that this may have led to overestimation of the effects of high breast density and therefore the screening benefits for this group. *This is a potential limitation for many models incorporating mammographic density (Shih 2021) and should be factored into model interpretation (or better still, examined through sensitivity analyses or a different approach to modelling life-course breast density).*
- Van Broek et al (Van Broek 2021) modelled polygenic risk scores but assumed they did not directly affect tumour progression or mode of detection. *Modelling polygenic risk scores is of some interest in Australia, although the ethical and legal consequences also need careful consideration: and this is currently outside the scope of the ROSA project.*
- Mandelblatt et al (Mandelblatt 2016), Vilaprinyo et al (Vilaprinyo 2014) and Sankasting et al (Sankatsing 2020) assumed that each woman's relative risk did not change over her lifetime, even though some risk factors do vary with age. *This is a reasonable simplifying assumption given the complexity of modelling life-course risk estimates, but it should be considered when interpreting results.*

Risk classification

Approaches to risk classification varied between studies. Seven studies modelled risk groups based on mammographic density, with or without additional risk factors (Shih 2021; Wang 2020; Arnold 2019, Trentham-Dietz 2016, Mandelblatt 2016, Stout 2014, Vilaprinyo 2014), one study combined risk factors excluding mammographic density (Sankatsing 2020) while two studies used a polygenic risk score, alone or in combination with family history (Van Broek 2021; Wong 2021).

Selection of comparison groups

Evaluation of risk-based strategies is based on comparing outcomes of a risk group against outcomes of a comparator.

There were two main comparison group categories:

1. Each risk group receiving risk-based screening was compared to the whole population screened using relevant current screening guidelines. For example, Sankatsing et al (Sankatsing 2020) estimated that triennial screening for low-risk women 50-71 years (based on common breast cancer risk factors) resulted in better harm to benefit ratios compared to uniform biennial screening for all women aged 50-74.
2. Each risk group receiving risk-based screening is compared to the respective group screened under the relevant current screening guidelines or under another risk-based scenario. For example, in Trentham-Dietz et al (Trentham-Dietz 2016) triennial screening for an average-risk and low-density group is estimated to result in fewer false positive mammograms, benign biopsies and over-diagnosed cases compared to biennial screening for the same risk group.

Screening specification

Screening participation

- Stout (Stout 2014), Trentham-Dietz (Trenthem-Dietz 2016), van den Broek (van den Broek 2021), Mandelblatt (Mandelblatt 2016) and Sankatsing (Sankatsing 2020) assumed perfect adherence to screening and Wong (Wong 2021) assumed 100% attendance and compliance with breast cancer screening and follow-ups. *Australian screening participation rates are usually around 55% in the target age range[1] and there are known differences between some population groups. It is important to capture observed screening behaviour, rather than use simplified assumptions about participation, in considerations of risk-based screening in the Australian setting. Assuming full participation in screening can affect the estimated effectiveness and cost-effectiveness of a strategy. For example, a Dutch modelling study (Sankatsing 2020) reported that when observed participation was modelled (80%), the optimal strategy identified under 100% participation for high-risk women became less cost-effective.*

Screening test accuracy

- Wong (Wong 2021) noted that the mammography sensitivity was a limitation in their study, due to the scarcity and age of the available data and Arnold (Arnold 2019) and Vilaprinyo (Vilaprinyo 2014) included specifications (e.g. screening test sensitivity) based on film mammography. *These outcomes may have changed under digital mammography, although there have been limited changes in overall program sensitivity and specificity in Australia with the introduction of digital mammography.*
- Wang (Wang 2020) had no reliable data for the dependence of specificity on breast density for DBT, nor on the sensitivity of a screening program that includes DBT. *As confirmed in our scoping reviews, the program sensitivity and specificity of DBT varies greatly between studies and is likely to depend on factors such as existing recall rates and screening intervals, This should be considered when interpreting evaluations of DBT; ideally, modelled evaluations would incorporate a range of estimates for DBT imaging.*
- Some studies (e.g., Van Broek et al (Van Broek 2021), Mandelblatt et al (Madelblatt 2016), Arnold (Arnold 2019)) considered the effects of breast density on mammography performance while Shih et al (Shih 2021) modelled lower sensitivity for women with dense breasts to account for the masking effect. *This is important, as breast density is well established as being*

[1] Australian Institute of Health and Welfare 2021. BreastScreen Australia monitoring report 2021. Cat. no. CAN 140. Canberra: AIHW.

associated with reduced screening test sensitivity and specificity. This should be incorporated into considerations of risk-based screening in the Australian setting.

- Of note all studies that included breast density (except one – Wang 2020) did not provide full details of the method of assessment, often omitting the version number. Furthermore, these evaluations did not model the addition of alternative screening modalities for women with high breast density which would be of interest in risk-based screening.

Treatment modelling

- Wong (Wong 2021) did not model treatment, remission and follow-up after diagnosis. *This simplifying assumption may underestimate the benefits of earlier diagnosis, and overestimate the harms of overdiagnosis, as more advanced tumours require more intensive treatment.*
- Sankatsing et al (Sankatsing 2020) and van den Broek et al (van den Broek 2021) assumed full adherence to treatment unlike the remaining studies. *This simplifying assumption should be considered in interpreting the modelled results, where partial adherence would reduce the effectiveness and costs of treatment.*
- Vilaprinyo (Vilaprinyo 2014) noted that they relied on data from other countries where Spanish data was not available, so that model outputs should be interpreted with care. *Models often need to draw on data from other settings and this is most appropriate for biological or technological outcomes that are unlikely to vary between settings. However, as a rule, models are most reliable if they use local data.*

Mortality and life years

Many of the studies reported breast cancer-specific mortality including breast cancer deaths averted (Sankatsing 2020, Shih 2021, van Broek 2021, Trentham-Dietz 2016, Mandelblatt 2016, Stout 2014) and breast cancer mortality reduction (Arnold 2019, Mandelblatt 2016, Stout 2014). Other studies did not directly estimate mortality but estimated other outcomes such as number of lives extended (Vilaprinyo 2014), or LYG (Wong 2021, Stout 2014, Sankatsing 2020, Wang 2020). All models reported at least one of LYG, number of lives extended, or QALYs gained). Of note, studies that reported LYG did not report considering lead time in the model. *Mortality outcomes are important as mortality reduction continues to be the key aim of breast cancer screening. Life years gained are also important, but lead time should be considered (as screening can potentially lead to earlier diagnosis with no change to mortality, so that women live with a diagnosis for longer).*

Clinical outcomes

The most commonly reported measures related to benefits were breast cancer deaths averted, breast cancer mortality reduction, LYG, number of lives extended and QALYs gained, as discussed above. Screen-detected cancers compared to interval cancers were reported by Wang (Wang 2020). The most commonly reported clinical outcome measures related to potential harms were false positive screening results (all studies, except Wang 2020, Wong 2021) and overdiagnoses (all studies with the exception of Stout 2014, Wang 2020, Wong 2021). Benign breast biopsies were estimated by three studies ((Trentham-Dietz 2016, Mandelblatt 2016, Shih 2020). Vilaprinyo (Vilaprinyo 2014) estimated false negative results and DCIS attributable to screening.

Interval cancers as a modelled outcome was only considered by 2 studies (Vilaprinyo 2014; Wang 2020). In terms of indicators of tumour staging, e.g. size, nodal involvement, all studies considered these as inputs but not in relation to clinical outcomes.

Such outcomes should all be considered to fully characterise the benefits and harms of different approaches to risk-based screening.

Costs

Eight of the ten models summarised included cost estimates, to varying levels of detail. Health economics modelling is highly context-specific and, as noted by Arnold (Arnold 2019), the willingness-to-pay threshold plays a role in optimal models. Two evaluations did not report a willing-to-pay threshold for their selected settings (Stout 2014; Vilaprinyo 2014). The modelling by Trentham-Dietz (Trentham-Dietz 2016) used a very high cost threshold per quality-adjusted life-year gained (USD100,000) compared to the UK NICE (National Institute for Health and Care Excellence) guidelines (£20,000). In comparison, in Australia the indicative figures used to assess the cost-effectiveness of cancer screening and other prevention strategies are a willingness-to-pay (WTP) threshold of around AUD30,000–50,000 per LYS or QALYS (Lew 2019). These differences are important and should be considered in interpreting findings about the cost-effectiveness of screening in settings outside Australia, noting that US thresholds are vastly different to Australian thresholds and that cost-effectiveness model estimates are highly sensitive to the level of discounting applied to costs and effects.

Other approaches to reporting and comparing costs included using a cost-effectiveness frontier (e.g., Vilaprinyo 2014). Also called the Pareto frontier, this comparison helps to identify the most efficient alternatives to current practice.

We found some published modelled did not report some important information about costs, such as details of price adjustments for inflation or current conversion rates, the type of perspective taken, the willingness to pay threshold, or an explanation of the discount rates chosen. Cost-effectiveness models tend to be very sensitive to discounting assumptions and so this should be a critical component of modelling reports.

Unobservable outcomes

In addition to estimating the benefits, harms and costs of specific health strategies, simulation models can provide important information about estimated changes in important unobservable clinical factors in a way that cannot be done in the real world. For example, Mandelblatt reported in detail about expected outcomes in different risk groups, under current and alternative scenarios, including changes in overdiagnoses in different risk groups under different scenarios and rates of screen-detected cancers attributable to higher screening test sensitivity in women with lower breast density (Mandelblatt 2016). ROSA project researchers generated similar estimates for the Australian setting in a recent paper Bulliard 2021). Models can also potentially estimate rates of true interval cancers (not present at screening) compared to missed cancers (Vilaprinyo 2014, Wang 2020). *Modelling presents an opportunity to generate hypotheses that can sometimes be explored further through epidemiological studies.*

3.8.3 Limitations

All modelling requires some simplified assumptions, which need to be considered in interpreting outcomes. In addition to the limitations discussed in section 3.8.2, some authors also noted limitations such as a lack of observed data to specify model parameters with certainty, and lack of data on the distribution of risk groups among populations to enable the assessment of the impact of a risk-based program for a given population as a whole (Table 30, Table 31).

Only two studies modelled polygenic risk scores in relation to risk-based screening (van den Broek; Wong 2021). Wong et al (2021) assumed 100% attendance and compliance with breast screening

and follow-up but did not specifically refer to perfect uptake of genetic screening. In contrast perfect uptake of genetic testing was reported as an assumption by Van den Broek (Van den Broek 2021). They did not explicitly model the effects of the rare, but higher risk variants in genes such as BRCA1 and BRCA2 which are particularly relevant to women aged <50 years.

Interpreting results in relation to their potential applicability to different settings from the one modelled is important when evaluating risk-based evaluations. The majority of studies did not consider this in their discussion. A more generalisable approach, for example, was taken by Sankatsing et al (2020), who noted that they used relative risk levels in their model to make their results generalisable to other countries.

3.8.4 Translation of these findings to the Australian setting

Australia is well-positioned to use its own data for modelled evaluations of breast cancer screening, given the detail and quality of data collected. As noted elsewhere in this report, data collection could be improved, particularly in terms of current BreastScreen participation and outcomes for some risk groups, and identifiable records of opportunistic screening and high-risk surveillance outside the BreastScreen program.

Considerations for clinical modelling components should include at a minimum current screening program protocols and participation rates as well as screening cancer detection rates, interval cancer rates and false positive rates (and thereby other derivable outcome measures such as recall rates and program sensitivity and specificity).

Several models described here incorporated breast density, which is an important factor to include in evaluations of more risk-based screening. In the Australian setting, breast density is not routinely measured in screening using the visually assessed categories as modelled. For Australia it would be of value to model scenarios that make use of continuous measures of breast density that can be generated by automated measurement tools, as this would allow for more options enabling more targeted high- and low-risk services than achievable with visually assessed breast density.

Excluded studies

Our summary included various models developed by the ‘Cancer Intervention and Surveillance Modeling Network’ (CISNET) *Breast Cancer Working Group*, sponsored by the US National Cancer Institute (with three models included in the paper by Trentham-Dietz et al. (Trentham-Dietz 2016)). Refinements to CISNET models published in 2018 (Alagoz 2018) included detailed clinical model components such as modelling tumours according to their molecular subtypes that could inform best-practice modelling for Australian purposes, as well as improved modelling of DCIS, modelling of breast cancer incidence and other-cause mortality, with some models accounting for co-morbidities, and breast cancer risk modelled according to various risk factors in addition to age (Table 32), modelling of screening participation and specification of mammographic sensitivity and specificity and updated models for treatment effectiveness.

Table 6. Risk factors added to the various CISNET models as reported in 2018.

Model	***Personal risk factors***
Dana-Farber	Breast density
Erasmus*	Breast density, obesity
Georgetown-Einstein*	Breast density
Stanford	Hormone replacement
Wisconsin-Harvard*	Breast density

Our selection criteria aimed to limit our model summaries to those with enough detail on clinical outcomes to enable potential translation or comparison to the Australian setting. On this basis, our summary does not include a 2017 modelling paper from the UK setting that we had reported in our Expert Management Group meetings by Gray et al., 2017 (Gray 2017). The authors described this work as an 'early' model, with a view to an iterative approach to developing economic evidence to inform the introduction of new health care interventions. This modelling simulates screening protocols based on ten-year breast cancer risk and mammographic density, with considerable detail included about cost estimates (including, for example, the cost of risk assessment as estimated from the PROCAS study (£10.57). Key model specifications are summarised in Table 33

Table 7. Key specifications of the modelling exercise in the UK publication (Gray 2017)

Setting	UK
Base case	Current program (3-yearly mammography, mostly 50-70 years)
Scenarios	Screening intervals and/or modalities (supplemental ultrasound or MRI) tailored to breast density and risk. Breast density was categorised as 'low' for Volpara Density Group (VDG) 1/2 and 'high' for VDG 3/4. Risk groups were for absolute ten-year risk (<3.5%, 3.5%-8%, >8%).
Populations	Screening participants only, starting screening at age 47-50.
Outcome horizon	Lifetime horizon
Estimated benefits	Life years and quality-adjusted life years.
Estimated harms	Not clearly stated but false-positive recall rates, biopsy rates and proportion of cancers detected as DCIS have been included as input parameters for all modalities
Costs	Risk assessment, basic diagnosis and treatment costs
Other specifications of interest	Screening test sensitivity was modelled as a function of tumour size and breast density. Recall rates were assumed to vary by test (mammography, mammography with ultrasound, mammography with MRI), but not breast density.

The authors concluded that screening intervals targeted according to ten-year breast cancer risk could be beneficial, but screening modalities (supplemental US or MRI) based on mammographic density and risk would not add net value, mostly due to increased overdiagnosis. Given the similarities between UK and Australian health settings (despite different screening intervals (3-yearly in the UK) and different levels of participation (76% in UK (NHS 2017) (Australasian Society for Breast Physicians 2020), versus 55% in Australia (Australian Institute of Health and Welfare 2018), the ROSA modelling team does use this model to draw on ideas, in addition to the findings of this report.

Identifying optimal scenarios

While modelling can be used to estimate benefits, harms and costs, selecting optimal scenarios requires consensus about how to best balance benefits, harms and costs for different groups of women and in different health settings.

Modelling evaluations can be of greatest utility when they are done for a specific purpose, in consultation with health policy-makers. This is especially important to ensure appropriate health economics evaluations. For example, the UK model authors noted that their estimates were very sensitive to assumed discount rates, and on this basis they recommended that 'decision makers should consider which discounting scenario best reflects the values and preferences of those for

whom they are making a decision' (Gray 2017). Similarly, a German study (Arnold 2019) reported that risk-stratified screening programs could be more efficient than uniform screening, depending on agreed decision criteria such as whether mortality reduction or QALY is more important, how much one is willing to pay for these health outcomes, and model assumptions on levels of screening participation. Frameworks such as 'public health economics' (Edwards 2013) incorporate useful in-depth thinking beyond cost-effectiveness, incorporating careful consideration of existing inequalities in health; this is important for the topic of risk-based breast cancer screening, given the evidence of existing disparities in health service utilisation of breast cancer screening and related services described throughout this report.

Consensus-based activities (such as those in this project's proposed Roadmap) can help support and direct appropriate modelling exercises, generate consensus-based recommendations about the best balance of benefits, harms and costs in the Australian setting, engage appropriate stakeholders, and set agreed standards for the level of evidence and cost-effectiveness required to warrant the implementation of more risk-based screening protocols.

3.9 References

Alagoz, O., D. A. Berry, H. J. de Koning, E. J. Feuer, S. J. Lee, S. K. Plevritis, C. B. Schechter, N. K. Stout, A. Trentham-Dietz and J. S. Mandelblatt (2018). "Introduction to the Cancer Intervention and Surveillance Modeling Network (CISNET) Breast Cancer Models." Med Decis Making 38(1_suppl): 3s-8s.

Arnold, M., K. Pfeifer and A. S. Quante (2019). "Is risk-stratified breast cancer screening economically efficient in Germany?" PLoS ONE 14 (5) (no pagination)(e0217213).

Australasian Society for Breast Physicians. "Find a Breast Physician." Retrieved January, 2020, from https://www.breastphysicians.org/find-a-breast-physician/.

Australian Institute of Health and Welfare (2018). BreastScreen Australia monitoring report 2018. Canberra, AIHW.

Bulliard JL, Beau AB, Njor S, Wu WY, Procopio P, Nickson C, Lynge E. Breast cancer screening and overdiagnosis. Int J Cancer. 2021 Apr 19. doi: 10.1002/ijc.33602. Epub ahead of print. PMID: 33872390.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. New York: Oxford University Press; 2005. p. 55–95.

Edwards, R. T., J. M. Charles and H. Lloyd-Williams (2013). "Public health economics: a systematic review of guidance for the economic evaluation of public health interventions and discussion of key methodological issues." BMC Public Health 13: 1001.

Gray, E., A. Donten, N. Karssemeijer, C. van Gils, D. G. Evans, S. Astley and K. Payne (2017). "Evaluation of a Stratified National Breast Screening Program in the United Kingdom: An Early Model-Based Cost-Effectiveness Analysis." Value Health 20(8): 1100-1109.

Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E; ISPOR Health Economic Evaluation Publication Guidelines-CHEERS Good Reporting Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value Health. 2013;16(2):231-50.

Lew, J. B., E. Feletto, S. Wade, M. Caruana, Y. J. Kang, C. Nickson, K. T. Simms, P. Procopio, N. Taylor, J. Worthington, D. Smith and K. Canfell (2019). "Benefits, harms and cost-effectiveness of cancer screening in Australia: an overview of modelling estimates." Public Health Res Pract 29(2).

Mandelblatt, J. S., N. K. Stout, C. B. Schechter, J. J. van den Broek, D. L. Miglioretti, M. Krapcho, A. Trentham-Dietz, D. Munoz, S. J. Lee, D. A. Berry, N. T. van Ravesteyn, O. Alagoz, K. Kerlikowske, A. N. Tosteson, A. M. Near, A. Hoeffken, Y. Chang, E. A. Heijnsdijk, G. Chisholm, X. Huang, H. Huang, M. A. Ergun, R. Gangnon, B. L. Sprague, S. Plevritis, E. Feuer, H. J. de Koning and K. A. Cronin (2016). "Collaborative Modeling of the Benefits and Harms Associated With Different U.S. Breast Cancer Screening Strategies." Annals of Internal Medicine 164(4): 215-225.

NHS. (2017). "Breast Screening Programme, England - 2015-16." Retrieved January, 2020, from https://digital.nhs.uk/data-and-information/publications/statistical/breast-screening-programme/breast-screening-programme-england-2015-16.

Sankatsing, V. D. V., N. T. van Ravesteyn, E. A. M. Heijnsdijk, M. J. M. Broeders and H. J. de Koning (2020). "Risk stratification in breast cancer screening: Cost-effectiveness and harm-benefit ratios for low-risk and high-risk women." International Journal of Cancer.

Shih, Y. C. T., W. Dong, Y. Xu, R. Etzioni and Y. Shen (2021). "Incorporating Baseline Breast Density When Screening Women at Average Risk for Breast Cancer : A Cost-Effectiveness Analysis." Annals of internal medicine. 09.

Stout, N. K., S. J. Lee, C. B. Schechter, K. Kerlikowske, O. Alagoz, D. Berry, D. S. Buist, M. Cevik, G. Chisholm, H. J. de Koning, H. Huang, R. A. Hubbard, D. L. Miglioretti, M. F. Munsell, A. Trentham-Dietz, N. T. van Ravesteyn, A. N. Tosteson and J. S. Mandelblatt (2014). "Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography." Journal of the National Cancer Institute 106(6): dju092.

Trentham-Dietz, A., K. Kerlikowske, N. K. Stout, D. L. Miglioretti, C. B. Schechter, M. A. Ergun, J. J. van den Broek, O. Alagoz, B. L. Sprague, N. T. van Ravesteyn, A. M. Near, R. E. Gangnon, J. M. Hampton, Y. Chandler, H. J. de Koning, J. S. Mandelblatt, A. N. Tosteson, C. Breast Cancer Surveillance, I. the Cancer and N. Surveillance Modeling (2016). "Tailoring Breast Cancer Screening Intervals by Breast Density and Risk for Women Aged 50 Years or Older: Collaborative Modeling of Screening Outcomes." Annals of Internal Medicine 165(10): 700-712.

Van den Broek, J. J., C. B. Schechter, N. T. van Ravesteyn, A. C. J. W. Janssens, M. C. Wolfson, A. Trentham-Dietz, J. Simard, D. F. Easton, J. S. Mandelblatt, P. Kraft and H. J. de Koning (2021). "Personalizing Breast Cancer Screening Based on Polygenic Risk and Family History." Journal of the National Cancer Institute 113(4): 434-442.

Vilaprinyo, E., C. Forne, M. Carles, M. Sala, R. Pla, X. Castells, L. Domingo, M. Rue and G. Interval Cancer Study (2014). "Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer." PLoS ONE [Electronic Resource] 9(2): e86858.

Wang, J., X. A. Phi, M. J. W. Greuter, A. M. Daszczuk, T. L. Feenstra, R. M. Pijnappel, K. M. Vermeulen, N. Buls, N. Houssami, W. Lu and G. H. de Bock (2020). "The cost-effectiveness of digital breast tomosynthesis in a population breast cancer screening program." European Radiology.

Wong, J. Z. Y., J. H. Chai, Y. S. Yeoh, N. K. Mohamed Riza, J. Liu, Y. Y. Teo, H. L. Wee and M. Hartman (2021). "Cost effectiveness analysis of a polygenic risk tailored breast cancer screening programme in Singapore." BMC Health Services Research 21(1): 379.

3.10 Appendix

3.10.1 Search strategy

Database(s): Embase 1947 to 2019 May 31, Ovid MEDLINE(R) ALL 1946 to May 31, 2019

#	Searches	Results
1	breast.tw.	961548
2	screen*.tw.	1653761
3	(dense breast* or breast densit* or risk or peronalised or personalized or individualised or individualized or tailor* or stratif*).tw.	5182273
4	(model* or simulat*).tw.	6507552
5	1 and 2 and 3 and 4	4781
6	limit 5 to (english language and female and humans)	3507
7	(regression adj2 model*).tw.	354703
8	6 not 7	2866
9	remove duplicates from 8	1943

3.10.2 Excluded studies

Table 8. Potentially relevant articles collected and excluded (shading indicates studies identified in updated 2021 searches)

Article	PubMed ID/DOI	Reason for exclusion
Abrahamsoon 2019	DOI: 10.1177/0962280219832901	No risk-based stratification modelling
Ahern 2014	DOI: 10.1038/bjc.2014.458	Modelling of specific subgroups rather than general population
Arnold 2017	DOI: 10.1186/s12913-017-2766-2	Systematic review
Arnold 2018	DOI: 10.1016/j.jval.2017.12.022	No clinical outcomes
Alagoz 2018	DOI: 10.1177/0272989X17711927	Description of model/model update; No clinical outcomes
Alagoz 2019	DOI: 10.1007/s11606-019-05182-5	No risk-based stratification modelling
Carter 2018	DOI: 10.2214/AJR.17.18484	No risk-based stratification modelling
Chootipongchaivat 2020	DOI: 10.1186/s13058-020-01287-6	Modelling of natural history of DCIS
Dierssen-Sotos 2018	DOI: 10.1038/s41598-018-20832-0	Breast cancer prediction model
Dinh 2011	https://pdfs.semanticscholar.org/af51/ba3dfc9771e6d7c0d36d669fa57301fda874.pdf	No risk-based stratification modelling
Foglia 2020	DOI: 10.1177/0951484819870963	No clinical outcomes
Geuzinge 2020	DOI: 10.1001/jamaoncol.2020.2922	Modelling of specific subgroups rather than general population
Golmakani 2021	DOI: 10.1111/biom.13484	No risk-based stratification modelling
Gray 2017	DOI: 10.1016/j.jval.2017.04.012	No clinical outcomes
Guzauskas 2020	DOI: 10.1001/jamanetworkopen.2020.22874	Intervention not imaged-based screening (screening with genomic testing +/- risk-reducing mastectomy)
Heijnsdijk 2019	DOI: 10.1002/cam4.2476	No risk-based stratification modelling
Hsu 2018	DOI: 10.1177/0962280216682284	Other (no lifetime horizon, descriptive paper only etc)
Hughes 2021	DOI: 10.1200/PO.20.00246	Breast cancer risk prediction model
Jiao 2014	DOI: 10.1016/j.jval.2014.03.780	Modelling of specific subgroups rather than general population
Kaiser 2021	DOI: 10.1016/j.ejrad.2020.109355	Modelling of specific subgroups rather than general population

Kaiser 2021	DOI: 10.1016/j.ejrad.2021.109576	Modelling of specific subgroups rather than general population
Khan 2021	DOI: 10.1002/ijc.33593	Systematic review
Koldehoff 2021	DOI: 10.1016/j.jval.2020.09.016	Systematic review
Koleva-Kolarova 2015	DOI: 10.1016/j.breast.2015.03.013	Systematic review
Le 2020	DOI: 10.1371/journal.pcbi.1008036	No risk-based stratification modelling
Lee 2018	DOI: 10.1177/0272989X17741634	Description of model/model update; No clinical outcomes
Lew 2019	DOI: 10.17061/phrp2921913	No risk-based stratification modelling
Louro 2019	DOI: 10.1038/s41416-019-0476-8	Systematic review
McCarthy 2019	DOI: /10.1093/jnci/djz177	Breast cancer prediction model
Machanda 2015	DOI: 10.1093/jnci/dju380	Modelling of specific subgroups rather than general population
McClintock 2020	DOI: 10.1016/j.mayocp.2020.04.017	Breast cancer prediction models
Machanda 2017	DOI: 10.1016/j.ajog.2017.06.038	No risk-based stratification modelling
Madadi 2018	DOI: 10.1080/24725579.2017.1396512	No risk-based stratification modelling
Manchanda 2020	DOI: 10.3390/cancers12071929	Intervention not imaged-based screening (screening with BRAC testing)
Mandelblatt 2011	DOI: 10.1016/S0960-9776(11)70023-6	No risk-based stratification modelling
Mandelblatt 2018	DOI: 10.1177/0272989X17700624	Summary of CISNET model inputs – no clinical outcomes
Mango 2019	DOI: 10.1002/jmri	No clinical outcomes
Michaan 2021	DOI: 10.1158/1940-6207.CAPR-20-0411	Intervention not imaged-based screening
Myers 2015	DOI:10.1001/jama.2015.13183	Systematic reviews
Muller 2019	DOI: 10.1007/s10198-019-01038-1	Modelling of specific subgroups rather than general population
Neusser 2019	DOI: 10.1080/03007995.2019.1654689	Modelling of specific subgroups rather than general population
Nickson 2019	DOI: 10.17061/phrp2921911	Other (no lifetime horizon, descriptive paper only etc)
Obdejin 2016	DOI: 10.1016/j.ejca.2016.05.012	Modelling of specific subgroups rather than general population
O'Mahony 2015	DOI: 10.1177/0272989X14528380	No risk-based stratification modelling
Onega 2014	DOI: 10.1002/cncr.28771	Other (no lifetime horizon, descriptive paper only etc)
Pataky 2014	DOI: 10.1177/0969141314549758	Modelling of specific subgroups rather than general population
Pashayan 2020	DOI: 10.1038/s41571-020-0388-9	Other (no lifetime horizon, descriptive paper only etc
Petelin 2018	DOI: 10.1038/gim.2017.255	Systematic reviews
Petelin 2019	DOI: 10.1016/j.jval.2019.03.008	Modelling of specific subgroups rather than general population
Petelin 2020	DOI: 10.1038/s41436-020-0751-3	Modelling of specific subgroups rather than general population
Phi 2019	DOI: 10.1016/j.breast.2019.03.004	Modelling of specific subgroups rather than general population
Ripping 2016	DOI: 10.1002/ijc.29912	No risk-based stratification modelling
Roman 2017	DOI: 10.1038/bjc.2017.107	No risk-based stratification modelling
Roman 2019	DOI: 10.1371/journal.pone.0226352	Systematic review
Saadatmand 2013	DOI:10.1093/jnci/djt203	Modelling of specific subgroups rather than general population
Schechter 2018	DOI: 10.1177/0272989X17698685	Description of model/model update; No clinical outcomes
Schiller-Fruhwirth 2017	DOI: 10.1007/s40258-017-0312-3	Systematic review
Schousboe 2011	DOI: 10.7326/0003-4819-155-1-201107050-00003	No clinical outcomes

Seigneurin 2015	DOI: 10.1016/j.canep.2015.08.013	No risk-based stratification modelling
Simoes Correa-Galendi 2021	DOI: 10.1007/s40258-020-00599-0	Modelling of specific subgroups rather than general population
Sprague 2015	DOI: 10.7326/M14-0692	Modelling of specific subgroups rather than general population
Sun 2019	DOI: 10.1001/jamaoncol.2019.3323+	Modelling of specific subgroups (women with breast cancer) rather than general population
Taghipour 2017	DOI: 10.1177/0272989X16660711	No risk-based stratification modelling
Taksler 2021	DOI: 10.1001/jamaoncol.2021.0952	No risk-based stratification modelling
Tehranifar 2021	DOI: org/10.1158/1055-9965.EPI-20-1627	Review of risk prediction models
Terry 2019	DOI: 10.1016/S1470-2045(18)30902-1	Validation of 4 breast cancer risk prediction models
Tessier 2019	DOI: 10.1093/jnci/djz037	Modelling of specific subgroups rather than general population
Tice 2019	DOI: 10.1007/s10549-019-05167-2	Breast cancer prediction model
Tina Shih 2019 (or Shih 2019?)	DOI: 10.1016/j.jval.2018.07.880	No risk-based stratification modelling No clinical outcomes
Tollens 2021	DOI: 10.3390/cancers13061241	Modelling of specific subgroups rather than general population
Tosteson 2008	PMID: 18166758	No clinical outcomes
Ulloa-Perez 2016	PMID: 27623037	No risk-based stratification modelling
van den Broek 2018	PMID: 29554469	Other (no lifetime horizon, descriptive paper only etc)
van Ravesteyn 2012	DOI: 10.7326/0003-4819-156-9-201205010-00002	No risk-based stratification modelling
van Ravesteyn 2018	DOI:10.1177/0272989X17729358	No risk-based stratification modelling
van Ravesteyn 2021	DOI: 10.1093/jnci/djaa218	Modelling of specific subgroups rather than general population (low risk women)
Wu 2013	DOI: 10.1038/bjc.2013.202	Other (no lifetime horizon, descriptive paper only etc)
Wu 2018	DOI: 10.1186/s13058-018-1082-z	No risk-based stratification modelling
Yaffe 2015	PMID: 26676234	No risk-based stratification modelling
Yeh 2020	DOI: 0.7326/M19-3481	Modelling of specific subgroups rather than general population

3.10.3 Evaluation of reporting checklist

Table 9. Checklist for evaluating the reporting of risk-based microsimulation models (Y: yes, N: not stated, P: partial, NA: not applicable). Sources for questions are indicated as follows: Table footnote: C=from CHEERS checklist; C mod= from CHEERS but modified into a question from a statement; D=from Drummond framework; remaining items identified from evaluation of risk-based studies in the ROSA project.

Parameter assessed	**Checklist item**	Stout 2014	Vilaprinyo 2014	Mandelblatt 2016	Trentham-Dietz 2016	Arnold 2019	Sankatsing 2020	Wang 2020	Shih 2021	Van den Broek 2021	Wong 2021
Risk groups	1. Was a rationale reported for the choice of the alternative programmes or interventions or risk groups compared? (D)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	2. Were the alternatives being compared described in a way that was unambiguous and clear? (D)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Comparison group	3. Was the choice of comparison group explained?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Outcomes	4. Were both harms and benefits evaluated? (C)	Y	Y	Y	Y	Y	Y	P	Y	Y	N
Benefits (mortality outcomes)	5. Are mortality outcomes reported? (e.g. breast cancer deaths averted, number of life years extended, life years gained, quality-adjusted life years gained)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	6. If LYG reported, was lead time considered?	N	NA	Y	Y	NA	N	N	Y	Y	N
Other clinical outcomes	7. Were both screen-detected invasive cancers and interval cancers reported as inputs or outputs?	P	Y	P	P	P	P	Y	P	P	P
	8. Was tumour staging or indicators of tumour staging reported (e.g. size, nodal involvement) as inputs or outputs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Parameter assessed	**Checklist item**	Stout 2014	Vilaprinyo 2014	Mandelblatt 2016	Trentham-Dietz 2016	Arnold 2019	Sankatsing 2020	Wang 2020	Shih 2021	Van den Broek 2021	Wong 2021
	9. Were tumour sub-types reported (e.g. hormone receptor status)?	Y	N	Y	Y	Y	N	N	Y	Y	N
	10. Were other clinical outcomes reported (e.g. benign breast disease, false negative screens, DCIS) as inputs or outputs?	P	Y	Y	P	Y	Y	N	Y	P	N
	11. Were recall rates and program sensitivity and specificity reported as inputs or outputs?	P	P	P	P	P	P	P	P	P	P
Harms	12. Were overdiagnosis and false positive screens reported (as inputs of outputs)?	P	Y	Y	Y	Y	Y	P	Y	Y	N
Risk assessment	13. Did the model stratify women by specified risk factors? (e.g. age, breast density, genetic risk as polygenic risk score, high risk mutations, validated risk assessment tools or other)	Y	Y	P	Y	Y	Y	Y	Y	Y	Y
Risk classification	14. Was the approach to risk classification explained in a way that was unambiguous and clear?	Y	Y	P	Y	Y	Y	Y	Y	Y	Y
Screening test accuracy	15. Is screening test accuracy modelled according to age and current conditions (e.g. for digital mammography rather than film mammography)?	Y	P	Y	Y	P	Y	N	Y	Y	P
Breast density	16. Was the way that breast density was defined unambiguously and clearly described (e.g. Volpara version used)? (if version number not provided then it is 'partially addressed')	P	P	P	P	P	NA	Y	P	P	NA

Parameter assessed	**Checklist item**	Stout 2014	Vilaprinyo 2014	Mandelblatt 2016	Trentham-Dietz 2016	Arnold 2019	Sankatsing 2020	Wang 2020	Shih 2021	Van den Broek 2021	Wong 2021
	17. If breast density was included in scenarios were alternative screening modalities used in the interventions?	N	N	N	N	N	NA	Y	N	N	NA
	18. If yes, was program sensitivity and specificity estimated according to breast density?	NA	NA	NA	NA	NA	NA	N	NA	NA	NA
	19. If the model included risk groups according to breast density, were the effects of breast density on mammographic performance considered?	Y	N	Y	Y	Y	NA	Y	P	Y	/A
Treatment	20. Was breast cancer treatment modelled (e.g. surgery, chemo, radio or endocrine therapy mentioned) in terms of costs and/or clinical outcomes? (if only one, state yes)	Y	Y	Y	Y	Y	Y	Y	Y	y	N
Recurrence	21. Was breast cancer recurrence modelled?	N	N	N	N	N	N	N	N	N	N
Assumptions	22. Is adherence to screening participation stated in the model?	N	N	Y	Y	Y	Y	Y	N	Y	Y
	23. If yes, is it stated that adherence to screening participation in the model is the same as observed participation for that setting?	NA	NA	N	N	Y	N	Y	NA	N	N
	24. Are screening participation options other than 100% of the target population modelled?	N	NA	N	N	Y	Y	Y	NA	N	N
	25. Is adherence to treatment modelled?	N	N	N	N	N	Y	N	N	Y	N

Parameter assessed	**Checklist item**	Stout 2014	Vilaprinyo 2014	Mandelblatt 2016	Trentham-Dietz 2016	Arnold 2019	Sankatsing 2020	Wang 2020	Shih 2021	Van den Broek 2021	Wong 2021
	26. If yes, is this assumed to be perfect adherence?	NA	NA	NA	NA	NA	Y	NA	NA	Y	
	27. If polygenic risk scores were used, was this assumed to be available for the whole population modelled?	NA	N/A	NA	NA	NA	NA	NA	NA	Y	Y
Data	28. Were observed data directly used as model input parameters? (Yes if for all parameters, partial if for some, no if for none)	Y	P	P	P	Y	Y	Y	P	P	P
	29. Where observed data was used to model input parameters, was this drawn from the setting being modelled?	Y	Y	Y	Y	Y	Y	P	Y	Y	Y
For economic evaluations	30. Were the primary outcome measures for the economic evaluation clearly stated? (D)	Y	Y	NA	Y	Y	Y	Y	Y	NA	Y
	31. Were the methods for the estimation of quantities and unit costs described?	Y	P	NA	Y	Y	P	Y	Y	NA	P
	32. Were currency and price data recorded? (D)	Y	Y	NA	Y	Y	P	N	Y	NA	Y
	33. Were details of price adjustments for inflation or currency conversion given? (D)	Y	N	NA	N	Y	N	N	Y	NA	N
	34. Was time horizon of cost and benefits stated? (D)	N	Y	NA	Y	Y	Y	Y	Y	NA	Y
	35. Was the type of perspective taken stated? (e.g. payer, provider, society)	Y	Y	NA	N	Y	N	N	Y	NA	Y
	36. Was the discount rate stated? (D)	Y	Y	NA	Y	Y	Y	Y	Y	NA	Y

Parameter assessed	**Checklist item**	Stout 2014	Vilaprinyo 2014	Mandelblatt 2016	Trentham-Dietz 2016	Arnold 2019	Sankatsing 2020	Wang 2020	Shih 2021	Van den Broek 2021	Wong 2021
	37. If more than one discount rate was used, were results presented for all discount rates?	NA	NA	NA	NA	Y	NA	Y	NA	NA	NA
	38. Was the choice of discount rate explained? (D)	Y	P	NA	N	N	Y	Y	N	NA	N
	39. Was an explanation given if cost or benefits were not discounted? (D)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	40. Is a willingness to pay threshold stated?	N	N	NA	Y	Y	N	Y	Y	NA	Y
	41. If yes, is that realistic (i.e. at or below the country's WTP threshold)? (e.g. the WTP in US is $100 000; £20,000 in the UK, AUD$30,000–50,000 per LYS or QALYS)	NA	NA	NA	Y	Y	NA	Y	Y	NA	Y
Sensitivity analysis	42. Was a sensitivity analysis conducted for parameters which had unreliable data?	Y	Y	Y	N	Y	N	Y	Y	N	Y
	43. Was the approach to sensitivity analysis described? (D)	Y	Y	P	N	Y	N	Y	Y	Y	Y
	44. Was the choice of variables for sensitivity analysis described? (D)	Y	Y	P	NA	N	NA	Y	Y	NA	Y
	45. Were the ranges over which the parameters were varied stated? (D)	Y	Y	N	NA	Y	NA	Y	P	NA	Y
	46. Was any account taken of their correlations / joint distributions?	N	N	NA	N	N	N	N	N	NA	N
	47. Was the uncertainty of parameters described? (Cmod)	Y	N	N	N	N	N	Y	P	NA	P

Parameter assessed	**Checklist item**	Stout 2014	Vilaprinyo 2014	Mandelblatt 2016	Trentham-Dietz 2016	Arnold 2019	Sankatsing 2020	Wang 2020	Shih 2021	Van den Broek 2021	Wong 2021
Effect of uncertainty on outcomes	48. Was the uncertainty in relation to model structure and assumptions described? (Cmod)	Y	Y	P	Y	N	P	N	N	Y	Y
	49. Was the effect of uncertainties on cost, outcome or cost effectiveness described? (Cmod) (For example, if CIs or ranges of costs or cost effectiveness planes are provided, accept as 'yes')	Y	N	NA	N	Y	P	Y	Y	NA	Y
	50. Were results interpreted in relation to discounting? (If only discounted results presented then 'No'. If commentary in discussion then 'Y')	N	N	NA	N	N	N	Y	Y	NA	N
	51. If simplified assumptions (e.g. participation rates) were made, were these considered when outcomes were interpreted?	Y	N	Y	Y	Y	P	Y	N	Y	Y
Limitations	52. If there was a lack of observed data to model parameters with certainly was that considered in the discussion?	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
	53. Was the generalisability of the results in relation to applying findings to different settings discussed? (D) ('Y' if there was any commentary in the discussion)	N	N	N	N	N	Y	Y	N	N	N

Abbreviations: N: no; NA: not applicable; P: partially addressed; Y: yes.

4 ROSA clinical and health economics modelling

4.1 Executive summary

4.1.1 Authors

Dr Pietro Procopio, Dr Sabine Deij, Dr Lara Petelin, Dr Louiza Velentzis, Dr Saima Islam, Dr Jennifer Cauchi, A/Prof Carolyn Nickson.

We would also like to acknowledge contributions to various stages of model development from Prof Karen Canfell, A/Prof Dennis Petrie, Prof Bruce Mann, Ms Karinna Saxby, Prof Anne Kavanagh, Dr Graham Byrnes, A/Prof Ray Watson, Ms Elizabeth Korevaar, Dr Hannah Bromley and Ms Sarah Carr.

4.1.2 Modelling to evaluate options for risk-based breast screening

Risk-based breast screening is likely to involve routine risk assessment for all BreastScreen participants, with screening protocols tailored according to breast cancer risk to reduce the likelihood of interval cancers and false positive screening outcomes. Consistent with other programmatic public health interventions, policy decisions in relation to risk-based breast cancer screening would need to consider benefits in relation to risk, harms and costs. As outlined in Chapter 1, the ROSA project has devised an agreed framework for how to measure these potential benefits and harms.

Trials of risk-based screening protocols are likely to yield the highest-quality evidence. However, as highlighted in our overview and critical appraisal of risk-based breast cancer screening trials elsewhere in Chapter 4, such trials take time, can evaluate a limited range of screening protocols, and are unlikely to generate meaningful mortality outcomes in pace with advances in screening technology and improvements in cancer treatment (and potentially prevention).

Meanwhile, as outlined in our review of published clinical and health economic evaluations of modelled risk-stratified breast cancer screening options, numerous microsimulation modelling studies have estimated the likely costs, benefits and harms of various risk-based screening strategies.

Modelling is a form of data analysis that enables consideration of more complex questions and factors in the distribution of and interaction between multiple factors of interest.

Modelling is helpful for consideration of risk-based breast screening, given the complex relationship between breast cancer risk, screening behaviour, screening tests and breast density. Modelling also enables costs to be applied to various simulated events, so that the cost-effectiveness of different scenarios can be compared.

Modelling is not a replacement for clinical trials, and it will always be limited by the assumptions and simplifications required and the quality of the observed data and estimates used. However, it can provide indicative outcomes for a wide range of interventions, providing insights that can help isolate interventions that are likely to be most effective and cost-effective, subject to confirmation in real-world trials and evaluations.

To help evaluate the likely benefits, harms and costs of various risk-based population screening protocols in Australia, the ROSA project has undertaken clinical and health economics modelling of a range of risk-based screening scenarios in the Australian setting. These scenarios include

combinations of current risk-based breast imaging technologies, screening intervals and target age ranges. Primary outcomes of interest were mortality rates, tumour characteristics, costs and cost-effectiveness.

4.1.3 Contracted activities

Contracted ROSA activities for clinical and health economics modelling are specified as follows:

a) Select feasible and promising risk-based screening protocols for review.
b) Include work on improved precision of tumour subtypes and expected treatment costs, burden and prognosis.
c) Collect and assemble clinical and health economic data.
d) Model selected screening protocols.
e) Report generated estimates of the benefits, harms and costs of various screening protocols.
f) Expert Advisory Group to advise on which risk-based screening protocols to evaluate through clinical and health economics modelling.
g) Expert Advisory Group to provide model input data where feasible/possible.
h) Expert Advisory Group to review and discuss modelled costs, benefits and harms of various screening protocols.

4.1.4 Consultations

The project has undertaken several consultations. Specifically:

- The project has provided overviews and updates of the modelling approach through various presentations to the Expert Advisory Group (EAG).
- In March 2020, the project held a ‘face-to-face’ all-day modelling workshop with the BreastScreen Australia Program Management Group (BSAPMG).
- In May 2021, a ROSA ‘Modelling Update’ report was distributed to the project EAG and the BSAPMG, seeking feedback on the proposed methods and specifications through a series of directed questions. The resulting feedback and our responses and planned changes were subsequently distributed to the EAG and BSAPMG, with their requested changes incorporated where feasible into this report.
- In June 2022, the project provided to the EAG modelling methods and interim findings for review (excluding cost-effectiveness estimates).

As described in the Methods section, this feedback is incorporated into the report.

4.1.5 Research question

What are the likely clinical benefits and harms of various risk-based population screening protocols in the Australian setting, compared to the current BreastScreen Australia (BSA) program, for a range of assumed screening sensitivity and specificity values according to breast density?

4.1.6 Aim

To evaluate the likely benefits, harms and costs of various risk-based population screening protocols in the Australian setting, compared to the current BreastScreen Australia (BSA) program.

4.1.7 Summary of methods

The PICO protocol used to address the research question is shown in Table 36.

Table 10. PICO components and key considerations for modelled evaluation of risk-based screening. Only metrics that are generated by the simulation are included as outputs. False positive recall rates were included in development documents as a secondary outcome, however observed values of this metric are used to specify screening test specificity and so these values are an input to, rather than an output from the model.

PICO component	Approach
Population	Australian women aged 40-74, with a primary focus on women aged 50-74 Risk classification based on *a priori* group size allocation
Intervention	Interventions including changes to screening participation under the current program, and risk-based screening protocols (screening tests and screening intervals)
Comparator	Current BreastScreen protocols and participation, and current imaging and diagnostic services outside the program, i.e. business as usual (BAU)
Outcomes	<u>Primary</u> • Mortality • Population cancer diagnoses by tumour characteristics and mode of detection (screen-detected, interval, other) • Cost and cost-effectiveness <u>Secondary</u> • Program sensitivity • Recall rates • Treatment intensity • Overdiagnosis (cancers diagnosed before becoming symptomatic that would not have otherwise been diagnosed in a woman's lifetime) • Estimated 'missed' cancers (cancers present but not detected at screening).

The scenarios included in this evaluation are summarised in Table 37. Scenarios are defined according to risk-stratification, screening intervals, screening technologies, BSA target age range and participation rates. There were 160 scenarios in total, comprising the comparator scenario (BAU), plus 52 intervention scenarios for each modelled target age range for screening (50-74, 45-74 and 40-74), plus an additional 3 scenarios estimating the impact of changes to participation only. The modelled scenarios were selected based on available evidence, consultation with Australian experts and stakeholders, and what can reasonably be estimated through simulation modelling. All interventions were assumed to commence at full capacity on 1 Jan 2025. The impact of the COVID-19 pandemic was not incorporated.

Table 11. Modelled scenarios.

	BAU	Risk-based screening scenarios	Participation
Risk stratification	Annual screening policies[3]	Population-level risk groups Lower (RG1), Average (RG2), Higher (RG3)	As for BAU
Screening intervals	1 or 2 years	1, 2 or 3 years Assigned to risk groups [RG1, RG2, RG3] as either [2-2-2] or [3-2-2] or [3-2-1]^	As for BAU

Screening technologies	Mammography	Mammography, DBT or hypothetical* tests	As for BAU
BSA target age range	50-74 years	50-74, 45-74 or 40-74 years	As for BAU
Participation rate	Current BSA (approx. 55%)	As for BAU	60%, 65%, and 70%

^Annotations for screening intervals show the target screening intervals (in years) for each risk group in the order [RG1-RG2-RG3]. For example, [3-2-1] indicates triennial screening for RG1, biennial screening for RG2, and annual screening for RG3.
*Defined according to screening test sensitivity and specificity.

Screening technologies included mammography, digital breast tomosynthesis (DBT) and various hypothetical screening tests defined according to their screening test sensitivity and specificity. The approach using hypothetical screening tests was used to capture the range of published estimates for supplemental ultrasound and magnetic resonance imaging (MRI), for which there is such significant variation in the literature (as reported elsewhere in Chapter 4) that single point estimates could not reasonably be modelled. For example, screening sensitivity was modelled according to breast density (Figure 1),modelling a 'space' of screening outcomes that could then be mapped to different technologies.

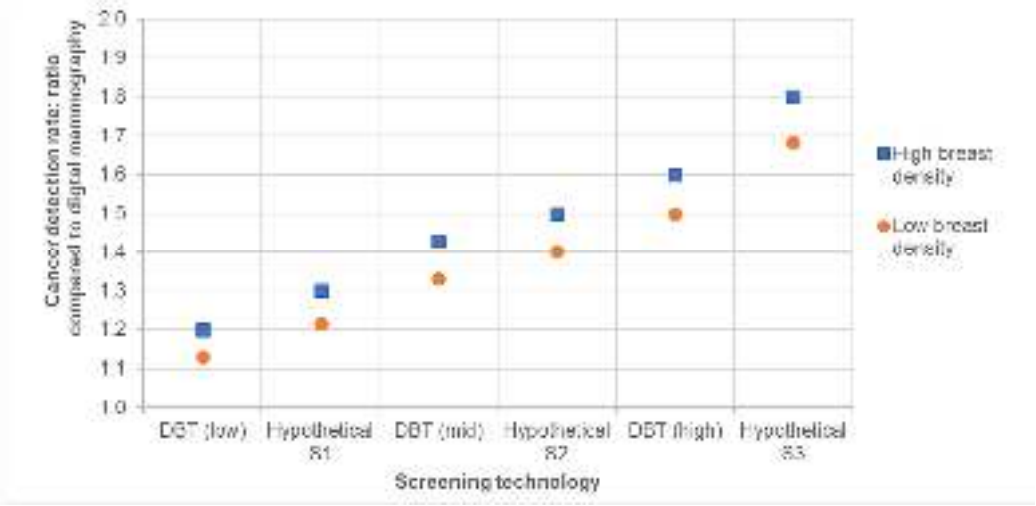

Figure 1. Modelled cancer detection rates (CDR) relative to digital mammography for screening technologies as specified in the modelled scenarios, according to breast density. The probability of a cancer being detected by screening also depends on tumour size, for all scenarios modelled.

Primary outcomes for this analysis were breast cancer mortality, tumour characteristics at diagnosis and cost-effectiveness. Secondary outcomes comprised interval cancer rates, overdiagnoses, and false-positive rates. We also report outcomes of great interest to programmatic breast cancer screening, such as program sensitivity, recall rates and overdiagnoses. Outcomes are generally reported at a population level to enable comparison of all scenarios and to focus on the population-level effectiveness of risk-based breast cancer screening.

The screening protocols modelled were tailored to population level risk groups with around 30% of women in a 'lower-risk' group, 50% of women in an 'average' risk group, and 20% of women in a 'higher-risk' group, based on a risk assessment conducted at their first screen. For a cohort of established screening participants, these risk groups would approximately correspond to a group-level 5-year breast cancer risk of 1.6% in the lower-risk group, 1.9% in the average-risk group, and

2.3% in the higher-risk group, respectively. Women in the highest risk group tend to have higher breast density. Outcomes are reported at a population level using the modelled values of lifetime breast cancer risk.

4.1.8 Summary of findings

Primary clinical outcomes

The modelled estimates indicate that some risk-based breast cancer screening scenarios could improve outcomes for the higher-risk group, while also improving outcomes at a population level. As would be expected in the real world, many reported outcomes are correlated, so that scenarios that tend to improve benefits also tend to increase harms (and vice versa). For example, we estimate that scenarios that tend to prevent more deaths also tend to increase overdiagnoses.

In terms of breast cancer mortality, we estimate that introducing a risk-based approach to screening for the current target age range of 50-74 years from 1 Jan 2025 could, in the first 10 years of implementation, reduce population level breast cancer mortality by up to 7%, saving up to 873 lives. Extending risk-based screening to younger age groups (40-74 or 45-74) could further reduce population-level breast cancer mortality, with the greatest differences noted for women in the highest risk group. Scenarios involving 3-yearly screening of the lower-risk group are expected to increase mortality in that risk group.

In Australia's contemporary treatment setting, finding breast cancers early continues to have a major impact on both survival and quality of life after diagnosis. Our findings indicate that risk-based breast screening could substantially decrease the proportion and rates of 'worse prognosis' invasive breast cancers (large, nodal, grade 3 breast cancers), with shortlisted scenarios resulting in estimated reductions of up to 20% in the rates of these 'worst prognosis' cancers in the higher-risk group. This is important not only in terms of reducing the risk of women dying from breast cancer, but also in minimising the extent of treatment required and long-term sequalae of more intensive treatment. Scenarios involving 3-yearly screening of the lower-risk group estimated a potential increase in later-stage tumours among that group; while the projected trends indicate that this effect may diminish over time, this is an important potential harm of increasing screening intervals in lower-risk women.

Some outcomes, such as screen-detected cancers rates, could fluctuate markedly in the first 7-8 years after introducing risk-based screening, peaking at first and then settling into a steady state over time, while other outcomes, such as the stage of cancers at diagnosis, are expected to improve in the short-term and demonstrate sustained improvement over time.

Secondary clinical outcomes

In terms of secondary clinical outcomes, for the scenarios modelled, we estimate that risk-based breast screening, depending on the scenario, could:

- Increase the proportion of invasive screen-detected cancers that are overdiagnosed by up to 50%, noting that overdiagnoses under the current program are estimated to be lowest in the higher-risk group, and some scenarios modelled would lead to this group having the highest rates of overdiagnosis.
- Reduce interval cancer rates in the higher-risk group, with some scenarios leading to rates comparable to the current rates for the average-risk group. Scenarios involving reduced screening intensity for the lower risk group are estimated to increase interval cancer rates among screened women, sometimes exceeding average rates across all risk groups.

- Reduce or increase DCIS diagnoses, recall rates, and 'missed' cancers.
- Reduce population-level treatment intensity in terms of extent of surgery (breast conserving vs mastectomy), chemotherapy and radiotherapy.

Abridgment note: Results on costs and cost-effectiveness are withheld as they including sensitive information not for public distribution.

Shortlisted scenarios

We identified a shortlist of 19 risk-based breast screening protocols which compared to current practice are expected to reduce breast cancer deaths, find more advanced breast cancers earlier when they have a better prognosis, and ensure a balance of costs and impacts on quality of life at a population level. All shortlisted scenarios involve digital mammography for lower-risk and average-risk women, and a targeted screening technology for the higher-risk group. While we describe these targeted tests as 'hypothetical', they are within the bounds of what might be expected from existing screening technologies such as supplemental ultrasound and MRI. Ten of the 19 shortlisted scenarios do not involve earlier entry to the screening program, while three would start screening at age 45 and six at age 40.

Shortlisted scenarios involving annual screening are expected to prevent the greatest number of deaths, while generally costing more than biennial screening scenarios and reducing quality-adjusted life-years at a population level (which incorporates reduced quality of life due to screening participation). As expected with modelled scenarios involving annual screening, the rate of invasive cancers with a worse prognosis (large, grade 3 and involving the lymph nodes) is reduced for all scenarios, but particularly for the higher-risk group. The estimated additional screens required depended largely on the entry age for screening and screening intervals, with some influence from modelled rescreening behaviour and cancer diagnoses.

Ten shortlisted scenarios involved extending the screening interval for the lower risk group to 3 years. While these scenarios met our shortlisting criteria in terms of population-level benefits and harms, they are estimated to involve increase mortality in this lower risk group. This is an important consideration.

For comparison: increased screening participation

For the secondary research question of increased participation using current screening protocols, outcomes reported for a cohort of women aged 50 in 2025 indicate modest increases in costs compared to most risk-based scenarios, with cost-effectiveness planes showing more favourable outcomes for increased participation compared to many risk-based scenarios.

We estimate that current outcomes for the higher-risk group and the whole population would be improved through increased screening participation alone, but greater improvements could be achieved for the higher-risk group through targeted, risk-based screening. This is an important consideration given the higher cancer rates and currently lower-than-average screening outcomes for women in the higher risk group.

4.1.9 Discussion (summary)

Generalisability

The *Policy1-Breast* simulation model is specified to evaluate health interventions in the Australian population and health services, with the intention of assessing risk-based screening scenarios incorporating both breast cancer risk and mammographic breast density, to a level of detail that

accounts for factors such as false positive outcomes, overdiagnoses and changes in treatment patterns. This is an improvement on models designed for settings outside of Australia, due to marked differences in populations and health service resourcing and delivery models, and an improvement on many models that do not incorporate the same level of detail. However, as noted by BreastScreen stakeholders during consultation in the design phase, the generalisability of national Australian estimates to individual state and territory programs is limited by the different service delivery and funding models in place and, to some extent, the different population profiles.

Alignment with stakeholder perspectives

We included 3-yearly screening intervals in our modelling scenarios as requested by some BSA stakeholders. We report in detail on the estimated outcomes for scenarios including this option for lower-risk women, with outcomes reported both for the risk group and at a population level. Ten of these scenarios are included in the shortlist presented however, as described above, we suggest, as future work, an additional evaluation of more '[2-2-1]' scenarios, scenarios targeting DBT only to higher-risk group and scenarios selectively recruiting women aged 40-49 to screening before concluding that the '[3-2-1]' scenarios are among the best options to investigate further.

Additionally, it was outside the scope of the current evaluation to model specific screening protocols used in current or recent clinical trials.

4.1.10 Strengths and limitations

Strengths

Key strengths of the current evaluation include our use of a model developed using extensive observed Australian data combined with published estimates relevant to the analysis, selection of scenarios and outcomes in consultation with key BSA stakeholders and experts specific to the topic of consideration of risk-based breast cancer screening in Australia, detailed modelling of breast density including its association with both breast cancer risk and screening test accuracy, and modelling of breast cancer treatment based on age, mode of detection and tumour characteristics at diagnosis (size, grade, nodal involvement). Sensitivity analyses on the clinical components of the model indicate that model estimates are generally stable to the parameters explored.

Limitations

All modelling is subject to limitations in comparison to a real-world trial due to the various simplifying assumptions and parametrisations required. Key simplifying assumptions include: no other major disruptors to population level patterns of breast cancer risk, progression, diagnosis or treatment; no major changes to screening behaviour over time or in response to the introduction of risk-based screening; a single risk assessment on entry to the screening program and stable cost estimates over time.

Additionally, it was assumed that individual-level lifetime risk does not vary, although this is likely to some degree given known risk factors such as being overweight and alcohol consumption; this assumption could over-estimate the effectiveness of risk-based screening as the risk group allocation could be less accurate over time. We assume that quality of life would not be adversely affected by undergoing risk assessment; given the range of perspectives about how risk advice

would be valued,[2,3] this assumption could either over- or under-estimate the benefits and harms of risk-based screening.

In the context of highly variable or sparse evidence about the accuracy of specific screening technologies, and the lack of Australian population-specific studies, our interim estimation using a matrix of screening test accuracy values provides a useful reference map as this evidence becomes more certain.

Cost and cost-effectiveness estimates incorporated detailed adjustments for the benefits and harms of screening participation, reflected in the quality-adjusted life-year outcomes and the costs. These estimates would be improved by including costs for establishing risk assessment in the program, and coordination, planning and evaluation costs to support implementation. As is common in health economics, cost-effectiveness estimates are highly sensitive to specified discount rates. We adhered to the Australian standard value of 5% discounting per year, however this could differentially impact outcomes for younger or older women when comparing scenarios with different entry ages for screening. It is difficult to circumvent this limitation, and so we confined our approach to shortlisting to include, but not be solely guided by, the quality-adjusted life-year measures.

Additional information about strengths and limitations of this evaluation are included in the detailed report.

4.1.11 Future work

While we modelled 160 scenarios (a significant modelling exercise given the complexity of the task), we modelled a very limited number of scenarios where annual screening was offered to the higher-risk group without triennial screening for the lower-risk group; this was only done for scenarios using either digital mammography or digital breast tomosynthesis as the screening test for the whole program. We also did not include scenarios offering DBT only to the higher-risk group, nor scenarios where only some women aged 40-44 and/or 45-49 are invited to BSA screening. The range of estimates reported in the current evaluation suggest that such scenarios should be ideally evaluated and compared to the current modelled estimates, as they may yield a better balance of benefits, harms and cost-effectiveness than the current shortlist, noting also that some stakeholders are concerned about extending screening intervals and that DBT is now in common use in some BSA assessment services. The current shortlist provides some insights about priority protocols to help design and plan trials of risk-based breast cancer screening in Australia, but it may be improved through firstly evaluating these additional scenarios, which may or may not change the shortlist.

Evaluation of scenarios changing the eligibility of screening for women aged 75+ could also be added as an option for future evaluation, as well as evaluations that incorporate risk assessment outside the program or repeated risk assessments e.g., at every screen event or at regular intervals, and scenarios incorporating the use of Artificial Intelligence (AI) incorporating prior and current clinical information to guide future screening protocols. Various approaches to risk assessment were discussed with stakeholders during development of the modelled scenarios but it was not feasible to incorporate these into the current evaluation. Additional reporting and analysis focussing on specific population groups (e.g. age groups, breast density groups, women who do or don't attend screening) would also be of value.

[2] Lippey J, Keogh LA, Mann GB, Campbell IG, Forrest LE. "A Natural Progression": Australian Women's Attitudes About an Individualized Breast Screening Model. Cancer Prev Res (Phila). 2019 Jun;12(6):383-390. doi: 10.1158/1940-6207.CAPR-18-0443. Epub 2019 Apr 19. PMID: 31003994.
[3] Dolan H, McCaffery K, Houssami N, Cvejic E, Brennan M, Hersch J, Dorrington M, Verde A, Vaccaro L, Nickel B. Australian Women's Intentions and Psychological Outcomes Related to Breast Density Notification and Information: A Randomized Clinical Trial. JAMA Netw Open. 2022 Jun 1;5(6):e2216784. doi: 10.1001/jamanetworkopen.2022.16784. PMID: 35708691; PMCID: PMC9204548.

While it would be preferable to report estimates for specific technologies, as outlined in our evidence reviews of screening technologies (Chapter 4) and summarised in this chapter, evidence on the sensitivity and specificity of DBT, ultrasound, MRI and contrast-enhanced mammography in screening populations is widely varied and sometimes sparse, and dependent on study designs and settings. Evidence about how specific screening technologies would be expected to perform in the Australian screening setting and population would be highly valuable. That evidence could then be used to generate modelled estimates of the benefits and harms of specific technologies, including re-applying the shortlisting criteria to help identify the most effective screening intervals and age ranges in the Australian population screening setting.

In addition to alternative or supplemental screening technologies, there is increasing interest in screening protocols that systematically combine prior and current clinical and computer-generated information to guide future screening protocols. This is expected to enhance and standardise current clinical practice, which takes clinical histories and current clinical information into account most often in a qualitative, clinical decision-making framework. AI approaches could potentially be added to the model scenarios as quantitative evidence about the relative benefits, harms and costs of AI approaches emerges.

Finally, as noted the generalisability of national modelling is limited due to differences between state and territory programs.

4.1.12 Conclusion

The modelled evaluation reported here indicates potential benefits, harms and costs of a range of risk-targeted breast screening protocols, where screening protocols are characterised according to age range, screening technology and screening intervals. Modelled scenarios were selected based on available evidence, consultation with Australian experts and stakeholders, and what can reasonably be estimated through simulation modelling. A shortlist of nineteen screening protocols was identified based on a balance of clinical and cost-effectiveness outcomes. After comparison with an additional set of scenarios as indicated, the resulting shortlist provides a basis from which to design and plan trials of risk-based breast cancer screening in Australia.

Abridgement note: Additional detail on methods and results will become available through peer-reviewed publications in preparation. Key findings are summarised in Section 5.1 (Q3, from page 50).

5 Appendices

5.1 Summary of 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. How do alternative or supplemental breast imaging technologies/modalities perform for different breast cancer risk groups, compared to digital mammography?

Digital breast tomosynthesis

Key evidence

1a. For digital breast tomosynthesis (DBT) when used in a population screening setting, all reviewed studies [randomised controlled trials (RCTs), fully paired or cohort studies] assessed DBT used in conjunction with 2D imaging (as either digital mammography or a synthetic 2D image), rather than DBT alone, with outcomes compared to screening using digital mammography.

1b. Following from (1a):

a) These studies showed that DBT combined with 2D imaging increased cancer detection rates across all risk groups based on age.
b) Findings on interval cancer outcomes were mixed. DBT with 2D imaging may decrease interval cancer rates in women with higher breast density and increase interval cancer rates in women with lower breast density, but the evidence is not consistent.
c) Findings on screening program sensitivity are mixed, with some studies finding no differences and others finding increased program sensitivity for some age groups with inconsistent outcomes according to breast density.
d) Program specificity was increased similarly across risk groups based on age and breast density.
e) All outcomes vary markedly between populations and settings, particularly in terms of false positive recall rates.

Ultrasound

Key evidence

2a. For supplemental ultrasound used in population breast screening, adding ultrasound to mammography (whether hand-held; HHUS or automated breast ultrasound; ABUS), compared to digital mammography alone can increase cancer detection rates and false positive rates for women with dense breasts and/or women at very high risk of breast cancer.

2b. For supplemental ultrasound used in population breast screening, adding ultrasound to mammography (whether hand-held; HHUS or automated breast ultrasound; ABUS) compared to digital mammography alone, the increases in cancer detection rates and false positive rates appeared consistently greater for women with denser breasts.

Magnetic Resonance imaging (MRI)

Key evidence

3a. For magnetic resonance imaging (MRI) compared to digital mammography in a population screening setting, all reviewed studies compared supplemental MRI for high-risk women, with outcomes compared to screening using digital mammography alone.

3b. Following from (3a), supplemental MRI increases cancer detection and false positive recall rates in high-risk women, compared to screening using digital mammography. The increase in cancer detection is lower for women who are mutation carriers compared to those who are negative or untested for any predisposing mutations and is possibly greater for younger women (40-49 years).

Contrast enhanced mammography

Key evidence

4. No studies of contrast-enhanced mammography (CEM) used in population breast screening were identified with risk-stratified results.

Considerations for implementation

1. Breast imaging technologies are rapidly evolving and expected to improve over time due to advances in technologies and incorporation of AI systems.

Priority evidence gaps

1. Evaluation of breast imaging technologies used in population screening in the Australian setting.

Q2. What are the relative benefits, harms and costs of risk-based breast cancer screening as estimated by population-level modelling studies relevant to the Australian health setting, and how would their clinical and health economics estimates translate to an Australian setting?

Key evidence

1. Published modelled evaluations of risk-based breast screening indicate that some risk-based scenarios may improve the balance of benefits, harms and cost-effectiveness compared to current approaches to population breast screening.
2. Assessing which modelled scenarios are optimal requires consensus about how to best balance benefits, harms and costs for different groups of women and in different health settings.
3. Clinical modelling components should include, at a minimum, current screening program protocols and participation rates as well as screening cancer detection rates, interval cancer rates and false positive rates.
4. Modelled estimates should include the benefits and harms for each risk group as well as the whole population.
5. Breast density is an important consideration for risk-based breast screening and should be incorporated into modelled evaluations.

Q3. What are the likely benefits, harms and costs of various risk-based population screening protocols in the Australian setting, compared to the current BreastScreen program?

Key evidence

1a. The ROSA modelled evaluation of risk-based screening (stratified to around 30% of women in a lower-risk group, 50% of women in an average risk group, and 20% of women in a higher risk group) indicates that risk-based screening could, in the first 10 years of implementation, reduce population level breast cancer mortality by up to 7%, saving up to 873 lives.

Following from (1a), this evaluation indicates that:

1b. Risk-based screening is expected to have a greater impact on mortality for the higher-risk group for scenarios where alternative screening technologies are used.

1c. Less frequent (triennial) screening of 30% of the population (women at lowest risk of breast cancer) may lead to small increases in breast cancer mortality in that risk group.

1d: Some outcomes, such as screen-detected cancers rates, could fluctuate markedly in the first 7-8 years of risk-based screening, while other outcomes, such as the stage of cancers at diagnosis, are expected to improve in the short-term and demonstrate sustained improvement over time.

1e. Estimated costs and cost-effectiveness of modelled scenarios indicate a cost-effectiveness frontier preferencing scenarios involving either (i) digital mammography for all women combined with targeted screening technologies for higher-risk women or (ii) screening technologies other than mammography for all screened women.

Following from (1a), this evaluation indicates that, depending on the scenario, risk-based screening could:

1f. Reduce interval cancer rates in the higher-risk group, with some scenarios leading to rates comparable to the current rates for the average-risk group.

1g. Decrease the proportion of large, nodal, grade 3 breast cancers at a population level by up to 25%.

1h. Increase the proportion of screen-detected cancers that are overdiagnosed by up to 50%, noting that overdiagnoses under the current program are estimated to be lowest in the higher-risk group, and some scenarios modelled would lead to this group having the highest rates of overdiagnosis.

1i. Reduce interval cancer rates in the higher-risk group, with some scenarios leading to rates comparable to the current rates for the average-risk group

1j. Reduce or increase DCIS diagnoses, recall rates, and ‘missed’ cancers.

1k. Reduce population-level treatment intensity in terms of extent of surgery (breast conserving vs mastectomy), chemotherapy and radiotherapy.

2a. The modelled evaluation of 156 scenarios identified a shortlist of 19 risk-based breast screening protocols which were most promising when compared to current practice in terms of reducing breast cancer deaths, finding more advanced breast cancers earlier when they

have a better prognosis, and ensuring a balance of costs and impacts on quality of life at a population level.

2b. Following from 2(a), all shortlisted scenarios involve digital mammography for lower-risk and average-risk women, and a targeted screening technology for the higher-risk group.

2c. Following from 2(c), 10 of the 19 shortlisted scenarios would be for the current target age range of 50-74 years, while three scenarios would target screening from age 45 and six scenarios from age 40.

2d. *Abridgment note: This finding (on costs and cost-effectiveness) is withheld due to sensitive information not for public distribution.*

Considerations for implementation

1. The ROSA modelled evaluation indicates that some risk-based screening protocols are expected to improve the clinical effectiveness of population breast cancer screening for the Australian population.
2. Perspectives on potentially extending screening intervals for lower risk-women are highly varied between senior BreastScreen state and territory personnel.
3. Modelled estimates for life-years and quality-adjusted life-years incorporated the impacts of screening, diagnosis and treatment. This meant that some more intensive screening protocols that were expected to improve population-level life-years compared to the current screening program (through saving lives) could also reduce quality-adjusted life-years at a population level through factors such as increased population-level exposure to screening, increased overdiagnosis, and living longer with a breast cancer diagnosis.

Priority evidence gaps

1. Accurate estimates of the sensitivity and specificity of different screening technologies in the Australian screening setting, for different risk groups.
2. Modelled estimates of scenarios offering annual screening to the higher-risk group and biennial screening to the lower-risk and average-risk group, targetted use of digital breast tomosynthesis, and risk-based recruitment of women aged 40-49.
3. Modelled estimates of scenarios changing the eligibility of screening for women aged 75+
4. Modelled estimates of scenarios that incorporate risk assessment outside the program or repeated risk assessments.
5. Modelled estimates of outcomes for specific population groups (e.g. age groups, breast density groups, women who do or don’t attend screening).