

RESEARCH ARTICLE

How should artificial intelligence be used in breast screening? Women's reasoning about workflow options

Diana Popic^{ID}¹, M. Luke Marinovich^{ID}², Nehmat Houssami^{ID}^{2,3}, Julie Hall¹, Stacy M. Carter^{ID}^{1,*}

1 Australian Centre for Health Engagement Evidence and Values, School of Social Sciences, Faculty of Arts, Social Sciences and Humanities, University of Wollongong, Wollongong, New South Wales, Australia,

2 The Daffodil Centre, The University of Sydney, Sydney, New South Wales, Australia, a Joint Venture with Cancer Council NSW, Sydney, New South Wales, Australia, **3** School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia

* stacyc@uow.edu.au



Abstract

Studies show that breast screening participants are open to artificial intelligence (AI) in breast screening, but hold concerns about AI performance, governance, equitable access, and dependence on technology. Little is known of consumers' views on how AI should be used in breast screening practice. Our study aims to determine what matters most to women regarding AI use in the workflow of publicly funded breast screening programs, and how women choose between workflow options. We recruited forty women of screening age to learn about AI, the Australian breast screening program, and four possible workflows that include AI – one where AI works alone, and three different combinations of humans and AI. Participants then joined one of eight 90-minute dialogue groups to discuss their normative judgements on workflow options. Women proposed four conditions on AI deployment: preserving human control, evidence to assure AI performance, time to become familiar with AI, and clearly justifying the need for implementation. These informed women's unified rejection of AI working alone, and divided preferences across the other three workflows, as they traded off workflow attributes. Current evidence on AI performance convinced some women, but not others. Most women believed humans mitigate risk the best, so workflows should continue to be designed around them. Public breast screening services are trusted and valued by women, so significant changes require careful attention to outcomes relevant to women. Our results – women's detailed judgements on workflow design options – are new to the research literature. We conclude that women expect that AI only be deployed to do tasks it can do well, only where necessary, and only to fill gaps that radiologists cannot meet. Advancements in AI accuracy alone are unlikely to influence all women to accept AI making final decisions, if clinicians are available to perform the same task.

OPEN ACCESS

Citation: Popic D, Marinovich ML, Houssami N, Hall J, Carter S (2025) How should artificial intelligence be used in breast screening? Women's reasoning about workflow options. PLoS One 20(5): e0323528. <https://doi.org/10.1371/journal.pone.0323528>

Editor: Wenjie Shi, Universitätsklinikum Magdeburg: Universitätsklinikum Magdeburg, GERMANY

Received: October 11, 2024

Accepted: April 10, 2025

Published: May 30, 2025

Copyright: © 2025 Popic et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Ethics approval for our study was provided by the Human Research Ethics Committee of the University of Wollongong, Australia (2021/067). Our ethics approval prohibits public sharing of our dataset - participants did not consent to public

data sharing. Our processed data has been de-identified and is shared in our manuscript – we have included excerpts of discussion transcripts (the raw data for this project) within our paper. We do not have the consent of research participants to share transcripts, and cannot seek this retrospectively. The Human Research Ethics Committee of the University of Wollongong, Australia can be contacted at uow-humanethics@uow.edu.au. This does not alter our adherence to PLOS ONE policies on sharing data and materials, however we are unable to share data as this is restricted under the terms of our Human Research Ethics Approval.

Funding: The authors disclose support for the research of this work from the National Breast Cancer Foundation, IIRS-20-011. SMC declares research funding from the Australian National Health and Medical Research Council (1181960), National Breast Cancer Foundation (IIRS-20-011) and Australian Commission on Safety and Quality in Health Care for work on AI in healthcare, membership of the NSW Health AI Taskforce, and support for the costs of travel for conference presentations on AI in healthcare from medical indemnity insurers and non-government organisations. NH declares competitive funding from the Australian National Health and Medical Research Council (1194410), and from the National Breast Cancer Foundation for AI-related research (EC-21-001). MLM declares competitive, independent funding from the National Breast Cancer Foundation for work on AI in breast cancer screening (IIRS-20-011). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. National Breast Cancer Foundation can be found at www.nbcf.org.au. Australian National Health and Medical Research Council can be found at www.nhmrc.gov.au.

Competing interests: SMC, DP, MLM and JH declare no competing interests. NH declares membership of the TGA (Therapeutic Goods Administration) Advisory Committee on Medical Devices.

Introduction

As machine learning and the digital health environment enable greater automation via medical artificial intelligence (AI), a question arises: Is this what service users want? The ethical and system challenges in delegating clinically-related tasks to AI systems [1] suggest a need for systematic engagement with the patients and publics who depend on health services. Recent systematic and scoping reviews of the relatively new literature on patient and public views [2–6] provide some insights. Knowledge of AI is variable [2,4–6]. Knowledge about healthcare AI may be less than in other contexts [2,4]. While consumers are mostly open to healthcare AI and its potential benefits [2–6], this is contingent on AI performance [2,6], use in low-risk settings [2], human oversight [5], equitable access [6], and familiarity [6]. People may accept AI if their clinicians consider it trustworthy [5,6]. When these conditions are not met, public cautiousness [5,6] or opposition [2] may result, due to concerns about performance, privacy, governance, equity, choice, alienation or dehumanisation, deskilling, and undermining patient-professional relationships [2–6]. Patients and publics call for education and consultation [2,4], remaining patient-centred [5,6], and addressing performance, governance and ethics issues before and during implementation [3–6].

One well-advanced use-case for AI in healthcare is mammographic screening for breast cancer [7]. Screen-reading AI is deployed in some breast screening contexts, including some private practices, but was not used in Australian public breast screening at the time of this study. All Australian women aged 50–74 are invited biennially to attend publicly-funded mammographic screening for breast cancer through Breast-Screen (women aged 40–49 and older than 74 years are able to attend). In Breast-Screen Australia, as in many organised screening programs, images are read by two radiologists; a third reader arbitrates in the case of disagreement (some programs use a consensus approach). Published studies [8–12] exploring women's views show understandings, hopes, and concerns about AI in breast screening are similar to those expressed about AI in healthcare generally. This includes evidence that first-time screeners may hold stronger positive and negative views of AI [11], and that women are concerned about dependence on AI [8]. Women appear to have stronger expectations regarding responsibility, accountability and governance in breast screening AI than in AI in general [8,10–12]; equitable access is also important to women [8]. We have shown women are especially concerned about how AI and radiologists are combined in breast screening systems, and about AI performance [12].

Elsewhere, we published the first stage of analysis of these data [13]. This includes: 1. Women's self-reported knowledge; 2. Women's expressed support for the use of AI in breast screening at three timepoints; and 3. A qualitative analysis showing four conditions that women said should be met before AI was deployed: preserving human control, evidence to assure performance, time to become familiar with the technology, and clearly justifying the need for implementation, summarised in Table 1 [13].

In this paper, we build on this earlier work to examine a key issue for those considering the implementation of AI in breast screening programs: how AI should be included into the screening workflow. We aimed to answer the questions:

Table 1. Four conditions women imposed on AI implementation in breast cancer screening – a summary [13].

Human control and decision making is preserved	Women valued human attributes, which meant the current workflow in many programs of 2–3 radiologists screen-reading independently was highly valued as an assurance of system performance. Women wanted radiologists to continue reading mammograms, retaining oversight, control and skills. This would ensure accountability for decisions, and meaningful explanations to patients. Investment in a radiology workforce should be sustained. This finding was consistent regardless of women's overall positive or negative sentiment towards AI in breast screening.
Relevant, high-quality evidence shows improved accuracy prior to implementation	Regardless of overall sentiment towards AI, women highly valued AI system performance. Many were unimpressed by the (2022) evidence of performance. Radiologists, in their view, were tried and tested; AI was not. Evidence should be relevant to practice, including trials and global case studies; some women emphasised they did not want to be the trial.
Women have time to get informed and familiar with AI	AI was generally seen as unfamiliar, and sometimes futuristic, or even unrealistic. Women sometimes ascribed this to their age or generation, and sometimes worried it might be a barrier to participation. To address unfamiliarity, women suggested transparency about AI use, slow, staged implementation, and the potential for women to opt out of AI use in their care.
Reasons for change outweigh the costs	We told women that AI was of interest to breast screening services because of a shortage of radiologists. Responses to this premise varied. Some women were suspicious, suggesting that cost-cutting, or simply the availability of AI, were more likely reasons. Others accepted the premise, but believed screening services weren't adequately exploring other solutions, e.g., workforce development. Some women accepted AI as a viable solution to a radiologist shortage; even they stressed that policy environments should not over-emphasise AI or grow dependent on it, and should not permit automation or radiologist job loss simply because they could. AI should not be used just because it was available: it should be used only when there were adequate reasons to support service change, and for some, only if it was necessary to keep screening services functioning. Retaining radiologists could bring future improvements in breast screening, while AI could introduce numerous downsides: it was critical that the benefits outweighed the risks, harms and losses.

<https://doi.org/10.1371/journal.pone.0323528.t001>

1. What matters most to women when considering the use of AI in the workflow of publicly-funded breast screening programs?
2. How do women choose between options for including AI in breast screening workflows?

Materials and methods

Ethics approval

Ethics approval was provided by the Human Research Ethics Committees of the University of Wollongong (2021/067).

Approach

We used online dialogue groups, a qualitative data collection method developed by empirical bioethics researchers. Dialogue groups engage participants in discussing scenarios, asking them to make normative judgements about options, consider the potential to change their views, and provide reasons [13].

Participant recruitment and selection

Consistent with qualitative methodologies, we aimed to recruit diverse participants whose experience and perspectives were relevant to our research questions: in this case, those eligible to be invited to public breast cancer screening. Recruitment and selection was undertaken by Tavner Research [13], via social media and random digit dialling (see [S1 Appendix](#) and [S2 Appendix](#)). Social media advertising was on Facebook, to diverse geolocated ‘community’ pages to ensure national reach. To understand the views of screening participants, we excluded those with personal/close family member/close friend experience of breast cancer (ever), or employment in breast screening, breast cancer care or cancer control (last five years). Women with a direct experience of breast cancer are frequently excluded from studies about screening (e.g., [14]) because the experience of breast cancer changes women’s experience and perception of screening

[15]. We included women who had, and had not, participated in screening in the prior four years, allocating them to different groups. Women required conversational English skills and access to a computer/tablet and secure internet connection to participate.

Informed consent was recorded orally to reduce burden on participants to print, sign and scan their written consent. After participants received the Participant Information Statement, their oral consent was recorded using a digital recorder and stored on secure University of Wollongong data systems. Oral consent was approved by the Human Research Ethics Committee at the University of Wollongong. The committee did not require consent to be witnessed. Each woman spent approximately 2.5 hours over two weeks on consent, support for IT skills, engaging with information, and a 90-minute online discussion, and were compensated \$AUD150.

Information provision and data collection

Data were collected from 12 August 2022 to 21 October 2022. To support informed participation in dialogue groups, we shared three 5–10 minute videos [16–18] on specialist research-only bulletin board service VisionsLive, one every two days for six days. The videos explained: AI, including its uses; screening and breast screening, including current and potential workflows; and evidence on AI performance (as of 2022) in each workflow [19,20]. Workflows and evidence on AI performance in each is presented in Fig 1. We encouraged participant engagement and interaction. Questions women posted on the bulletin board were answered by investigators. Women were required to view all videos and post comments on at least two before participating in a group. Video content and discussion group tasks were developed between researchers and content experts in ethics, AI and breast screening.

Senior researcher SMC moderated dialogue groups by recapping information from the three videos and asking participants' judgements on the alternative workflows, with reasons (see [S3 Appendix](#)). Audio-recordings of discussions were transcribed verbatim by a professional transcription service.

Data analysis

Participant demographics were compared to the population of Australian women aged 50–74 using the Australian Bureau of Statistics' Tablebuilder [21]. Dialogue group analysis focused on women's judgements about the four workflows. We used reflexive thematic analysis (RTA) [22–26], a proven approach in qualitative health policy research offering both robustness and flexibility. DP developed codes inductively from the data, largely coded semantically and summarised patterns and insights for each group. Records of the analysis process include a spreadsheet of positive and negative judgements about the four workflows, by group and cohort, and memos to SMC for feedback. Bulletin board comments did not add to the analysis so were not included. For transparency, [S4 Appendix](#), details our method against agreed quality standards for reporting qualitative analysis [27]. We did not measure or report theme frequency, as the authors of the method have specified that this is inconsistent with the tenets of RTA [24].

Results

Participant characteristics

We recruited eight groups of 4–6 women, totalling 40 participants. We achieved diversity across age groups, residential location, levels of education (from early school leaver to postgraduate), most states and territories, and birthplace in or outside of Australia. Groups contained proportionally more younger and university-educated women, and more women born in Australia, compared to the Australian population. More demographic details are available in our earlier paper [13].

Openness to AI comes with conditions

Women commonly expressed the following: 'yes [I'm open to the development of AI in breast screening], *but* [on conditions]'. As we will show, women's judgments about the four workflows relied heavily on the four conditions reported in our

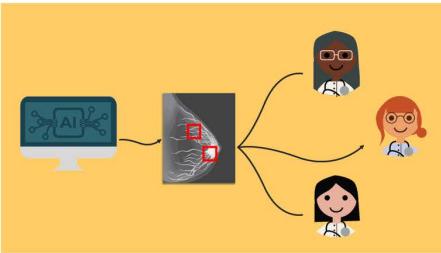
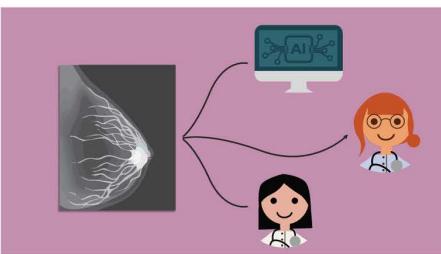
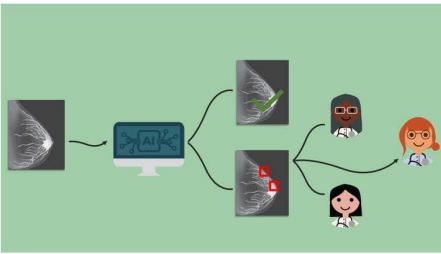
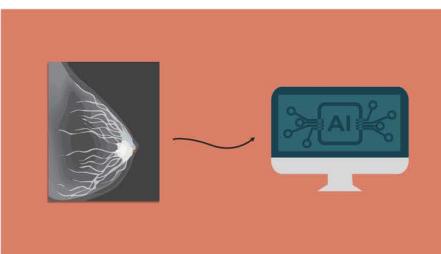
	Video transcript	Evidence summary
COMPUTERISED DECISION SUPPORT (presented to participants as "AI HELPS SPECIALISTS MAKE THEIR DECISIONS")	<p>Here the mammogram would be taken as usual. The AI could then provide a recommendation about the risk of cancer shown on the mammogram or could mark areas of the mammogram to help the specialist to make their decision. The information from the AI would go to the two human specialists, they would provide the decision and if they disagreed, a different specialist would make the final decision - the arbitration process.</p> 	Small increases in cancer detection but higher recall and false alarms, and specialists may have difficulty disagreeing with the AI.
REPLACING ONE RADIOLOGIST (presented to participants as "AI REPLACES ONE SPECIALIST")	<p>AI can replace one specialist. Instead of two human specialists examining a woman's mammograms, a specialist and the AI examine them. If they disagree, a different specialist will make the final decision – the arbitrator.</p> 	Could increase or decrease the number of women recalled for more tests. Arbitration by specialists would reduce the number of recalls.
TRIAGE (presented to participants as AI AS TRIAGE)	<p>AI can be used as Triage. Meaning that AI examines all mammograms. And if a mammogram is below a 'very low risk' threshold, it's marked as very low risk by the AI and is not examined by a specialist. Any mammogram above this 'very low risk' threshold would be sent on to be examined by two specialists. If they disagreed, a different specialist would make the final decision – the arbitrator.</p> 	Can safely identify about half of "normal" mammograms - these mammograms would never be reviewed by a specialist.
AI ALONE	<p>AI could be used alone. In this option the AI examines a person's mammograms without input from specialists. And this would mean that the AI alone would decide if that person was invited back for further investigation.</p> 	Currently does not perform well enough to be used on its own but may in the future. At present would likely increase the number of recalls.

Fig 1. Workflow options and evidence presented in bulletin boards and dialogue groups.

<https://doi.org/10.1371/journal.pone.0323528.g001>

introduction [13]. Women were open to potential benefits, but also recognised that contemporary life featured extensive, disappointing automation. Women not screened in the last four years appeared overall to be more cautious about AI than women who had screened. However, their substantive concerns were similar, so they are not reported separately in our analysis.

Women's judgements about four potential workflows for AI in breast screening

The four conditions in [Table 1](#) informed women's judgements about the four potential workflows they considered. Below we show how women explained their negative and positive judgements about these workflows.

Different women preferred different workflows

Women consistently rejected using AI alone. However, judgements about the other three workflows – Computerised Decision Support (CDS), replacing one radiologist, and triage—were divided. Women sometimes changed their views, had difficulty deciding, or disagreed. [Figs 2](#) and [3](#) illustrate women's most common responses in relation to their four conditions.

Computerised decision support (CDS)

Women who preferred CDS emphasised the importance of human control and familiarity. They supported CDS because two radiologists continued to read every mammogram, with AI relegated to being a complementary tool used by humans. Familiarity—resemblance to the current workflow—was also reassuring. These women were mostly unconvinced by the evidence presented. When pushed on what they thought about the potential for CDS to increase false positives, some said the risk would be mitigated by the two or three humans.

The women who rejected CDS, conversely, were concerned it would undermine human decision-making: *independent* human decision-making was valued, to preserve human control and accountability. They were more concerned about evidence CDS could increase false positives, and worried that CDS would not address the workforce shortage.

Replacing one radiologist

Women who preferred this workflow also reasoned from familiarity—because this workflow appeared similar to the current one—and human control, because every mammogram was seen by at least one human. Women saw this workflow as a reasonable trade-off: every mammogram was seen by a radiologist, but with fewer radiologists than the current workflow, thus addressing workforce shortages. With respect to evidence, some of these women were unmoved, others were reassured by evidence that arbitration could reduce false positives. This workflow was seen to combine the best attributes of AI and humans.

Some women—mostly those preferring CDS—said there was insufficient evidence to warrant AI making any decisions independent of humans. Using AI to read independently would require evidence that it was 100% accurate.

Triage

Triage generated the widest range of views. This turned in part on what it means to retain human control and human decision making. Women who favoured triage said this need not mean all mammograms being read by a person: it could instead entail directing human expertise and resources to complex, high-risk mammograms. These women valued the capacity for programs to set a flexible, performance-based threshold, ensuring only very low-risk mammograms would be read by AI alone, while human expertise could be applied to higher risk mammograms, where it was most needed.

Women who preferred triage also appeared more likely to reason from the evidence, and while they accepted and gave weight to evidence on all four workflows, they favoured the evidence for triage. They also preferred triage as a solution for workforce shortages.

		FOUR CONDITIONS WOMEN IMPOSED ON AI IMPLEMENTATION			
		Human control is preserved	Current evidence on workflow	Familiarity	Addresses workforce shortage
WORKFLOW OPTIONS	1 Computerised decision support	1		1	
	2 Replacing one radiologist	2	2	2	2
	3 Triage	3	3		3
	4 AI alone				
Icon	Meaning of icon	Supporting quote			
	HUMAN CONTROL IS PRESERVED				
1	Two humans read and review every mammogram	"I believe that [AI] should not replace anybody and that it should be an addition by itself.... I think it's too risky otherwise to replace anyone, but it can be used as an additional pair of eyes is what I think." (DG8)			
2	Humans read and review every mammogram, radiologist and AI make independent decisions	"What I like about [Replacing one radiologist] is that a specialist still actually looks at the scans... Moving forward progressively with AI, but also having that safety in knowing that somebody with a trained eye is going to be looking at it." (DG2)			
3	Humans read and review complex, high-risk mammograms	"...if we're going to use it we should be using it to the best advantage. And in my opinion the triage model gives that by narrowing the number of scans that need that human eye check... But I'm a realist in terms of the fact that need far outweighs supply of service, and I want the expert eyes focused where the need is greatest, and that's why I think that triage model is the best utilisation of resources." (DG7)			
	CURRENT EVIDENCE ON WORKFLOW				
2	Evidence on arbitration reducing recalls reassures women	"I chose [replacing one radiologist], because I think overall with the arbitration that it really evens out everything, it looks just like the best option... It could increase or decrease the number of women recalled, but that's offset by the fact that if arbitration was there that would reduce the number of recalls." (DG1)			
3	Evidence accepted and welcomed, flexible threshold can respond to AI performance	"[Triage], I also think is okay in the sense that the AI is only putting out mammograms that are very, very clearly normal. And that's only going to be accepted if the studies show that that is a suitable thing to do. Okay. If you've got the evidence to back a particular level that is considered to be "normal," then they're the only ones that are going to be excluded from being seen by a human. Then you getting the questionable ones seen by your two or three humans if there's an issue." (DG6)			
	FAMILIARITY				
1	Three radiologists continue to read mammograms like the current workflow	"I'd stick to [CDS], and I've got faith in that system... Going through all those other processes, what about if one of them fails?... You know, it's all that anxiety and worry to contend with, isn't there? That's why [CDS] I liked. You've got two people, humans, and a backup with a machine. All three of them can't be wrong. Someone has got to be right." (DG5)			
2	Women thought the current workflow included only one radiologist reading mammograms	Participant 1: AI's going to come in for sure and I think if we're talking about what process to start it with, [replacing one radiologist] would be great because you're not going to – Participant 2: It's not as confronting. Participant 1: It's not as confronting and it helps with so many things. (DG6)			
	ADDRESSES WORKFORCE SHORTAGE				
2	Two radiologists required, reduced pressure on existing workforce	"[AI replaces one radiologist] is actually doing what I'm guessing your intention. It's freeing up the specialists. Well, one of them anyway, which is going to reduce the load, enabling more to be read." (DG6)			
3	Fewer mammograms require radiologist's expertise, reduced pressure on existing workforce	"...Really all [triage] is doing is essentially making it easier for workforce after that. It's reducing the number of essentially patients by taking out the ones that are clear and safely clear, and then you've got the workforce handling a smaller number but a smaller number that need their attention. So it's weighing risk versus bearing in mind that one of the issues is of course workforce and access to screening, things like that." (DG3)			

Fig 2. Women's positive judgements about potential workflows for AI in breast screening.

<https://doi.org/10.1371/journal.pone.0323528.g002>

		FOUR CONDITIONS WOMEN IMPOSED ON AI IMPLEMENTATION			
WORKFLOW OPTIONS		Human control is preserved	Current evidence on workflow	Familiarity	Addresses workforce shortage
	1 Computerised decision	1	1		1
	2 Replacing one radiologist	2	2		
	3 Triage	3	3	3	
	4 AI alone	4	4	4	4
Icon	Meaning of icon	Supporting quote			
	HUMAN CONTROL IS PRESERVED				
	Radiologists may over-rely on AI and not read mammograms carefully, accountability for final decision on result unclear	"I couldn't fully see how [CDS] could really work... I would have thought that, if you're sending something to specialists, and saying, "Look at these spots, essentially," helping them make their decisions, aren't you really therefore encouraging them to focus on the AI's emphasis? So you're actually depriving them of really the opportunity and the responsibility of fully checking the data. See, I couldn't see how that actually added to the quality of the result. (DG3)"			
	Independent AI decisions without human review, radiologists may feel pressure making decisions in parallel with AI	"I feel [replacing one radiologist] would put a lot of pressure on the specialist to get it right. If there was only one person there, and you've got the AI, I reckon if the AI kept getting it right, and the specialist getting it wrong, it could cause a bit of a friction happening there. I know it's not a human, but, yeah." (DG2)			
	Humans don't read or review every mammogram	"I don't like [AI] working alone. That's very dangerous and open for malpractice, medical negligence because no human eye has looked at the scan and so who are you going to sue, a machine?" (DG8)			
	CURRENT EVIDENCE ON WORKFLOW				
	Current evidence shows AI does not perform adequately in CDS	"I just don't think [the evidence for CDS] is good. Because there's all these false alarms, and people are getting worried..." (DG5)			
	Insufficient evidence to warrant independent AI decision-making and current evidence shows AI does not perform adequately in replacing one radiologist	Participant: I think what sort of turns me off [replacing one radiologist] is it could increase or decrease the number of women recalled. I think that the trauma of being recalled unnecessarily is something that puts me off [it]. Moderator: So minimising those unnecessary recalls is important to you? Participant: Yeah. And any sort of intervention. So further needle biopsies, you know, that sort of stuff, I think is – can be quite traumatic, if it's not really required. So often you find those balls start rolling, it's hard to get it to stop. (DG2)			
	Insufficient evidence to warrant AI working alone	"The only reason I'm not leaning towards the triaging is because, again depending on how accurate the data is, if it's eliminating low risk it would want to be 100 per cent accurate. 'Cause there could be someone – what if someone's borderline and they get missed out?" (DG7) "Not in the near future. No. I think there'd have to be quite a few years of one of the other options first, see how that went. Some sort of proof that [AI alone] is not going to miss anything before I'd be happy or comfortable about it being completely AI." (DG4)			
	FAMILIARITY				
	Workflow described as surreal and science-fiction because AI makes decisions about human health alone	"[Triage] and [AI alone], it's still – I don't know. Little bit iffy about [AI alone] because that's just too Jetson style at the moment. And sort of [triage] as well, I think." (DG6)			
	ADDRESSES WORKFORCE SHORTAGE				
	Three radiologists still required to read every mammogram	"Now [CDS] to me is still using two specialists. It defeats the purpose, I think. It's not free – it's not doing its job of freeing up humans." (DG6)			
	Unjustified risk of job loss, deskilling of radiologists and loss of radiology knowledge in breast cancer	Participant 1: Never okay to let the human specialists go [in AI alone]. I would always want someone who's skilled and able to read it as a backup. I don't think I'd like to trust AI 100% without any human decisions.. Participant 2: You wouldn't want lose everyone and not have any human that would be able to do it, I can't see that that would be a good option at all (DG1)			

Fig 3. Women's negative judgements about potential workflows for AI in breast screening.

<https://doi.org/10.1371/journal.pone.0323528.g003>

In contrast, some women disliked or strongly opposed an AI system triaging mammograms out of radiology review. To these women, maintaining human control meant preserving human decision making over every mammogram. They saw the evidence as insufficient to warrant AI making any final decisions and tended to reject or question the current evidence. For quality assurance or to improve acceptability, some women suggested human review of a subset of AI-read low-risk mammograms. Triage was also rejected as too unfamiliar—even “sci-fi”—by some women.

AI alone

Women rejected AI working alone. The evidence suggested it did not yet perform well enough for this use; it was also ‘too out there’, futuristic and unfamiliar. Some said AI alone may be acceptable in the distant future (10 or 20 years away); others said it would never be acceptable or justified. This was because of loss of human control, radiologist job loss, loss of radiology skills and knowledge, and loss of human accountability.

Patterns of judgments across the workflows

Across the four workflows, patterns of intuition and judgement help explain women’s choices.

Women trade workflow attributes against each other

Women chose workflows by weighing up how well each met their four conditions. Each woman weighted each condition differently. When multiple conditions were equally important, choosing a preferred workflow became more difficult.

Some women perceived AI implementation as inevitable, thus chose workflows that they thought health services would consider ‘realistic’, even if they were unfamiliar, discomforting, or non-preferred. This is a reminder that women may express acceptance of a workflow but not be enthusiastic for the workflow, instead reluctantly conceding industrial, cultural, and technological change.

Moderator: What’s your thought about [AI replaces one specialist]?

Participant: Look, I go back to that’s fine, but I still think it’s got maybe a bit to do with the shortage of medical staff, qualified medical people. I guess inevitably AI is going to come in whether we like it or not, but to what degree. (DG1)

Most women accept a workforce shortage as reason to introduce AI

Many women took seriously the shortage of radiologists; those more open to this premise typically preferred either replacing one radiologist or triage. Conversely, women preferring CDS tended not to discuss workforce shortages, so may have seen this as insufficient justification to override other conditions.

AI making final decisions was a change too far for most women

Familiarity interacted with human control/decision making. The least familiar workflows were triage, and AI alone. These were described as science-fiction—allowing AI to make final decisions—thus required the greatest leap in reimagining how health status was decided. Women prioritising familiarity preferred CDS or replacing one radiologist, which represented the least change in who interpreted their mammograms.

Current evidence influences some women’s preferences, but not others

Women were invited to consider whether their preferences had changed based on the evidence on AI’s performance. Some women were unmoved, staying with their preferred workflows, typically CDS or replacing one radiologist. Here, human control and familiarity were often more important than AI performance, diminishing the impact of evidence.

Women seemed most likely to use evidence when they were already open to a workflow for other reasons. If women believed humans should read every mammogram, evidence of safety did not sway them towards triage. In contrast, if women were open to most workflows, and to the evidence, they often preferred triage, followed by replacing one radiologist.

Women debated whether false positives or false negatives were worse: fear about both scenarios influenced their decision-making. The more women wanted to avoid false positives, the more likely they were to prefer triage or replacing one specialist, based on the evidence provided.

Women were asked how accurate AI needed to be for women to have faith in it: some accepted AI performing as well as radiologists, others said it needed to perform better; some nominated 95–99% accuracy, others 100%. Women responded with the same range of views, whichever workflow they considered.

Humans mitigate risk the best, so workflows should be designed around them

All women had greater confidence in radiologists than they did in AI to read mammograms and mitigate risk. The shared view that evidence for AI was insufficient likely reinforced confidence in humans. Although researchers presented the workflows in relation to the role of AI, women centred human attributes, and gave most weight to the role of *humans* in each workflow. However, as shown, different women approached the relationship between risk and human mitigation differently. For some, every mammogram was at equally high risk of an inaccurate result: thus, humans should read every mammogram to mitigate this risk, and CDS or replacing one radiologist were preferred. Others saw only a subset of complex mammograms as high risk: thus, AI as triage was preferred to focus radiologists' attention on these high-risk mammograms.

Discussion

There is a small and growing literature on women's views on the use of AI in breast screening, which we have reviewed elsewhere [13]. To our knowledge, this is the first study of the views of women of screening age about workflow design options for AI in breast screening. Women's judgements turned on four conditions: retaining human control and decision making, high-quality evidence of excellent AI performance, familiarity, and clear reasons for change. The first two conditions are reflected in existing systematic reviews of consumer views of AI in healthcare [2–6]. The second two conditions are less prevalent in systematic reviews, so may be particularly relevant to breast screening. They may be explained in part by the cohort (older women are reportedly sometimes more apprehensive about healthcare AI [28]) and our methodology (we began from a premise for introducing AI in screen-reading). Women rejected use of AI alone in mammography screen-reading and made different judgements about three other workflows (CDS, replacing one radiologist and triage) depending on how they weighted the four conditions. These detailed judgements about the four workflows are new to the literature.

Inevitably, AI performance will improve. Women for whom accuracy is most important, if their accuracy expectations are met, are likely to welcome these advancements and be most open to expanding AI's role. But it would be wrong to assume that most women would welcome AI unconditionally. For women who value radiologists, accountability, screening services contributing to employment and the advancement of human skills and knowledge, it will be harder to accept a larger role for AI. Different women prioritise different conditions, and so make different judgements. However, it seems likely that improved accuracy will not lead all women to accept AI making final decisions, if clinicians are available to perform the same task.

Strengths and limitations

Limitations of our study include that the participants were on average somewhat younger, and more university educated and more likely to be Australian-born, than the broader population of Australian women. Despite this, we achieved age

diversity, majority non-university educated participants, and 28% overseas born, so our findings reflect diverse women's views. Another limitation is existing uncertainty about AI implementation in practice (e.g., regarding accountability, accessibility, speed of results), such that the conversations were necessarily somewhat speculative. As AI technology is evolving quickly, recent evidence about AI performance [7] is not captured in the information presented to participants. The fact that women had an opportunity to learn about breast screening and AI before the discussion groups is a strength: we note that they contain the kind of information that might be required to support informed consent to participate in breast screening using AI, so the results of this study are arguably more relevant to practice than collection of uninformed views. We have made the videos public to allow readers to make their own assessment of the information presented and its likely effect on women's views; we also hope the videos may be useful in other contexts. Although there may be some selection bias in the sample (e.g., due to recruitment of some women via social media, and the online format), we note that in Australia more than 90% of the population use social media and internet respectively [29]. As in all group-based qualitative research, group dynamics may have had some impact on the findings, although we note that the facilitator is an experienced qualitative researcher who aimed to mitigate such impacts as much as possible (for example, ensuring that all participants were able to speak).

There is some evidence that when patients consider AI in their own healthcare, they are more hesitant to its utilisation [28]. A novel strength of this study is our presentation of concrete and likely implementation scenarios, allowing women to consider these against their own experience, values and expectations.

Implications for policy and research

Breast screening is a complex public health service producing a range of benefits and harms for women, requiring complex decision making about service objectives and design. The balance of outcomes (lives lengthened by early intervention, false negatives and positives, cancer overdiagnosis) is a matter of ongoing debate [30]. Future screening modalities are also in flux, including possible future introduction of polygenic risk scoring [31]. The potential introduction of AI is one more element in this complex policy landscape [32]. However, for the women who participate, breast screening is both a trusted and a valued service [33]. This means significant changes—such as the introduction of AI screen reading—require careful attention to outcomes relevant to women and to screening policymakers [34]. There was a clear message in this study that women expected to be treated as active stakeholders in these decisions. As public knowledge about the potential risks and harms of AI grows, we suggest that services should think carefully about what is required both for public engagement and individual patient consent in the use of AI.

Conclusion

We have shown that women's judgements about workflows for AI in breast screening rely on different interpretations and combinations of four conditions: preserving human control and decision making; availability of relevant, high-quality evidence of performance; familiarity; and being convinced that the reasons for change outweigh the potential harms, risks, or costs. Although different women preferred different workflows, all rejected AI working alone in reading mammograms. Looking across the conditions, we conclude that women would ask that AI only be deployed to do tasks it can do well, only where necessary, and only to fill gaps that radiologists cannot meet.

Supporting information

S1 Appendix. Demographic and screening questions.
(DOCX)

S2 Appendix. Social Media Advertisement.
(DOCX)

S3 Appendix. Moderator guide.

(DOCX)

S4 Appendix. Standards for reporting qualitative research checklist.

(DOCX)

Acknowledgments

The authors would like to acknowledge Lucy Carolan who supported participant engagement in dialogue groups and Emma Frost who designed the videos and power point slides, and advised us on our final figures.

Author contributions

Conceptualization: M. Luke Marinovich, Nehmat Houssami, Stacy Carter.

Data curation: Diana Popic, Julie Hall, Stacy Carter.

Formal analysis: Diana Popic, Julie Hall.

Funding acquisition: M. Luke Marinovich, Nehmat Houssami.

Investigation: Julie Hall, Stacy Carter.

Methodology: Diana Popic, M. Luke Marinovich, Nehmat Houssami, Julie Hall, Stacy Carter.

Project administration: Diana Popic, Julie Hall, Stacy Carter.

Resources: Diana Popic, Julie Hall, Stacy Carter.

Supervision: Stacy Carter.

Validation: Stacy Carter.

Visualization: Diana Popic.

Writing – original draft: Diana Popic.

Writing – review & editing: M. Luke Marinovich, Nehmat Houssami, Stacy Carter.

References

1. Carter SM, Rogers W, Win KT, Frazer H, Richards B, Houssami N. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *Breast*. 2020;49:25–32. <https://doi.org/10.1016/j.breast.2019.10.001> PMID: 31677530
2. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. *Lancet Digit Health*. 2021;3(9):e599–611. [https://doi.org/10.1016/S2589-7500\(21\)00132-1](https://doi.org/10.1016/S2589-7500(21)00132-1) PMID: 34446266
3. Tang L, Li J, Fantus S. Medical artificial intelligence ethics: a systematic review of empirical studies. *Digit Health*. 2023;9:20552076231186064. <https://doi.org/10.1177/20552076231186064> PMID: 37434728
4. Yang L, Ene IC, Arabi Belaghi R, Koff D, Stein N, Santaguida PL. Stakeholders' perspectives on the future of artificial intelligence in radiology: a scoping review. *Eur Radiol*. 2022;32(3):1477–95. <https://doi.org/10.1007/s00330-021-08214-z> PMID: 34545445
5. Hogg HDJ, Al-Zubaidey M, Talks J, Denniston AK, Kelly CJ, Malawana J, et al. Stakeholder perspectives of clinical artificial intelligence implementation: systematic review of qualitative evidence. *J Med Internet Res*. 2023;25:e39742. <https://doi.org/10.2196/39742> PMID: 36626192
6. Vo V, Chen G, Aquino YSJ, Carter SM, Do QN, Woode ME. Multi-stakeholder preferences for the use of artificial intelligence in healthcare: a systematic review and thematic analysis. *Soc Sci Med*. 2023;338:116357. <https://doi.org/10.1016/j.soscimed.2023.116357> PMID: 37949020
7. Houssami N, Marinovich ML. AI for mammography screening: enter evidence from prospective trials. *Lancet Digital Health*. 2023;5(10):e641–2. [https://doi.org/10.1016/s2589-7500\(23\)00176-0](https://doi.org/10.1016/s2589-7500(23)00176-0)
8. Lennox-Chhugani N, Chen Y, Pearson V, Trzciński B, James J. Women's attitudes to the use of AI image readers: a case study from a national breast screening programme. *BMJ Health Care Inform*. 2021;28(1):e100293. <https://doi.org/10.1136/bmjhci-2020-100293> PMID: 33795236
9. Jonmarker O, Strand F, Brandberg Y, Lindholm P. The future of breast cancer screening: what do participants in a breast cancer screening program think about automation using artificial intelligence? *Acta Radiol Open*. 2019;8(12):2058460119880315. <https://doi.org/10.1177/2058460119880315> PMID: 31839989

10. Ongena YP, Yakar D, Haan M, Kwee TC. Artificial intelligence in screening mammography: a population survey of women's preferences. *J Am Coll Radiol.* 2021;18(1 Pt A):79–86. <https://doi.org/10.1016/j.jacr.2020.09.042> PMID: 33058789
11. Pesapane F, Rotili A, Valconi E, Agazzi GM, Montesano M, Penco S, et al. Women's perceptions and attitudes to the use of AI in breast cancer screening: a survey in a cancer referral centre. *Br J Radiol.* 2023;96(1141):20220569. <https://doi.org/10.1259/bjr.20220569> PMID: 36314388
12. Carter SM, Carolan L, Saint James Aquino Y, Frazer H, Rogers WA, Hall J, et al. Australian women's judgements about using artificial intelligence to read mammograms in breast cancer screening. *Digit Health.* 2023;9:20552076231191057. <https://doi.org/10.1177/20552076231191057> PMID: 37559826
13. Carter SM, Popic D, Marinovich ML, Carolan L, Houssami N. Women's views on using artificial intelligence in breast cancer screening: a review and qualitative study to guide breast screening services. *Breast.* 2024;77:103783. <https://doi.org/10.1016/j.breast.2024.103783> PMID: 39111200
14. Hersch J, Jansen J, Barratt A, Irwig L, Houssami N, Howard K, et al. Women's views on overdiagnosis in breast cancer screening: a qualitative study. *BMJ.* 2013;346:f158. <https://doi.org/10.1136/bmj.f158> PMID: 23344309
15. Absetz P, Aro AR, Sutton SR. Experience with breast cancer, pre-screening perceived susceptibility and the psychological impact of screening. *Psychooncology.* 2003;12(4):305–18. <https://doi.org/10.1002/pon.644> PMID: 12748969
16. University of Wollongong. What is AI?: YouTube; 2023 [updated 23 November 2023; cited 2023 December 5]. Available from: <https://www.youtube.com/watch?v=sr-sQgVnLD0>
17. University of Wollongong. The use of AI in breast screening. 2023 [Accessed 2023 December 5]. Available from: <https://www.youtube.com/watch?v=QsHrWsm05GY>
18. University of Wollongong. How well does AI work in breast screening. 2023 [Accessed 2025 February 20]. Available from: <https://youtu.be/k-47Zkfozno>
19. Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A, et al. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *BMJ.* 2021;374:n1872. <https://doi.org/10.1136/bmj.n1872> PMID: 34470740
20. Marinovich ML, Wylie E, Lotter W, Lund H, Waddell A, Madeley C, et al. Artificial intelligence (AI) for breast cancer screening: BreastScreen population-based cohort study of cancer detection. *EBioMedicine.* 2023;90:104498. <https://doi.org/10.1016/j.ebiom.2023.104498> PMID: 36863255
21. Australian Bureau of Statistics. Tablebuilder Canberra: ABS. 2021. Available from: <https://www.abs.gov.au/statistics/microdata-tablebuilder/tablebuilder>
22. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Qual Quant.* 2021;56(3):1391–412. <https://doi.org/10.1007/s11135-021-01182-y>
23. Braun V, Clarke V. Thematic analysis: a practical guide. London: SAGE Publications; 2022.
24. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology.* 2020;18(3):328–52. <https://doi.org/10.1080/14780887.2020.1769238>
25. Braun V, Clarke V. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health.* 2019;11(4):589–97. <https://doi.org/10.1080/2159676x.2019.1628806>
26. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? *Int J Qual Stud Health Well-being.* 2014;9:26152. <https://doi.org/10.3402/qhw.v9.26152> PMID: 25326092
27. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine.* 2014;89(9):1245–51. <https://doi.org/10.1097/acm.0000000000000388> PMID: 00001888-201409000-00021
28. Fritsch SJ, Blankenheim A, Wahl A, Hetfeld P, Maassen O, Deffge S, et al. Attitudes and perception of artificial intelligence in healthcare: a cross-sectional survey among patients. *Digit Health.* 2022;8:20552076221116772. <https://doi.org/10.1177/20552076221116772> PMID: 35983102
29. Statista. Internet in Australia - statistics & facts. 2025 [Accessed 2025 February 15]. Available from: <https://www.statista.com/topics/5261/internet-usage-in-australia/#editorsPicks>
30. Ding L, Poelhekken K, Greuter MJW, Truyen I, De Schutter H, Goossens M, et al. Overdiagnosis of invasive breast cancer in population-based breast cancer screening: A short- and long-term perspective. *Eur J Cancer.* 2022;173:1–9. <https://doi.org/10.1016/j.ejca.2022.06.027> PMID: 35839596
31. Roberts E, Howell S, Evans DG. Polygenic risk scores and breast cancer risk prediction. *Breast.* 2023;67:71–7. <https://doi.org/10.1016/j.breast.2023.01.003> PMID: 36646003
32. Houssami N, Kerlikowske K. AI as a new paradigm for risk-based screening for breast cancer. *Nat Med.* 2022;28(1):29–30. <https://doi.org/10.1038/s41591-021-01649-3> PMID: 35027759
33. Degeling C, Barratt A, Aranda S, Bell R, Doust J, Houssami N, et al. Should women aged 70–74 be invited to participate in screening mammography? A report on two Australian community juries. *BMJ Open.* 2018;8(6):e021174. <https://doi.org/10.1136/bmjopen-2017-021174> PMID: 29903796
34. Taylor-Phillips S, Seedat F, Kijauskaitė G, Marshall J, Halligan S, Hyde C, et al. UK National Screening Committee's approach to reviewing evidence on artificial intelligence in breast cancer screening. *Lancet Digit Health.* 2022;4(7):e558–65. [https://doi.org/10.1016/S2589-7500\(22\)00088-7](https://doi.org/10.1016/S2589-7500(22)00088-7) PMID: 35750402