Cureus
Part of SPRINGER NATURE

Open Access Review Article

Artificial Intelligence (AI) Support in Mammography Screening: Evidence From the MASAI (Mammography Screening With Artificial Intelligence) Trial and Gaps in Device Regulation

Sohum Shah [1], Ripal Gandhi [2]

1. Radiology, Sage Hill School, Newport Coast, USA 2. Miami Cardiac and Vascular Institute, Baptist Health South Florida, Miami, USA

Corresponding author: Sohum Shah, sohum323@gmail.com

Review began 11/09/2026
Review ended 11/18/2026
Published 11/24/2026

© **Copyright** 2026
Shah et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.120202

Abstract

Artificial intelligence (AI) tools for reading mammograms are already being sold to hospitals, yet few have been evaluated in randomized trials. This narrative review examines the Mammography Screening with Artificial Intelligence (MASAI) trial in Sweden, a randomized controlled trial of 105,915 women, and sets it alongside earlier retrospective studies and newer prospective and real-world studies from Sweden, Germany, Denmark, and Hungary. AI-supported screening detected 29% more cancers than standard double reading without any increase in false positives, and it reduced radiologists' reading workload by 44%. Interval cancers, which are those diagnosed between screenings, were 12% lower in the AI group (proportion ratio: 0.88, 95% confidence interval (CI): 0.65 to 1.18), but the difference was not statistically significant. A sample size estimate in this review shows that confirming a reduction of that size would require about 1.14 million women, more than 10 times MASAI's enrollment. The review also shows that the workload savings depend on double reading, which is less common in the United States, and that the AI system tested in MASAI was first cleared by the U.S. Food and Drug Administration (FDA) in 2018 on the basis of a retrospective reader study of 240 cases. In single-reader U.S. screening, AI's main value would likely be the time saved on each read, a benefit that no large screening trial has reported, and one small prospective study found no time saved. Of the 1,614 AI-enabled devices on the FDA's list as of September 2026, 1,553 (96.2%) were cleared through the 510(k) pathway, which does not require a clinical trial. This review proposes two potential solutions: a registry-linked, cluster-randomized trial in single-reader U.S. breast cancer screening with interval cancer as its primary endpoint, and a tiered evidence rule that requires patient outcome data for AI devices that determine how many radiologists read an exam.

Categories: Radiology
Keywords: artificial intelligence, breast cancer screening, computer-aided detection, double reading, fda 510(k), interval cancer, mammography, masai, medical device regulation, randomized controlled trial

Introduction And Background

In Sweden, two radiologists read every screening mammogram. The second reader is there because the first, however skilled, will sometimes miss a tumor. Many European countries use double reading, whereas in the United States, a single reader working with computer-aided detection software is more common [1]. Double reading may catch more cancers, but it also doubles the work, and it still cannot catch every cancer. Some cancers are found between screenings, after a scan has come back normal. These are called interval cancers. Their rate is one of the main ways a screening program is judged, because each one is a cancer that screening did not catch in time.

Artificial intelligence (AI) software is meant to help with both problems, and it can be used in two main ways [2]. As a triage tool, it scores each exam and decides how many radiologists read it, so that clearly normal exams get one reader instead of two. As a concurrent reading aid, it marks areas worth a second look while a radiologist reads, and the radiologist still makes every decision. A triage tool changes who looks at a scan, whereas a reading aid only changes what the reader sees. Studies on archived images can show how accurate either tool is, but demonstrating how it works in a real screening program requires prospective studies, and randomized trials provide the strongest evidence.

The Mammography Screening with Artificial Intelligence (MASAI) trial in Sweden is one of the few randomized trials of medical AI that reports results for patients rather than accuracy on a test set [3]. It tested an AI system that performed both triage and reading support within a national screening program. This narrative review uses MASAI as a test case. It sets the trial's results alongside earlier retrospective studies and later screening programs that tested AI, then examines what MASAI cannot settle on its own. Its most important endpoint has a sample size problem, and its workload benefit depends on the Swedish reading system. The device it tested was also cleared in the United States for a different use, which raises the review's regulatory question: should software that decides how many radiologists read an exam be held to

the same evidence standard as software that only marks images? Radiologists who use tools like it can also be pulled toward the AI's answer even when that answer is wrong. The review ends with the authors' recommendations for a trial design and a regulatory rule to address these gaps.

Review

2. Methods

This is a narrative review, not a systematic review, so it does not follow the full Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) process. The search and selection steps are described below to show readers how studies were chosen. PubMed and Google Scholar were searched for the terms "artificial intelligence" and "mammography screening" in studies published between January 2018 and November 18, 2026, the date of the last search. The FDA's list of AI-enabled medical devices and its 510(k) database were searched for mammography devices and for Transpara by name. ClinicalTrials.gov and the ISRCTN registry were checked for ongoing trials, and the reference lists of included studies and recent reviews were checked for studies the search had missed.

Studies were included if they tested an AI system for reading mammograms and reported its accuracy or its effect on screening, such as cancer detection, recall, false positives, interval cancers, reading workload, or reading time. Studies were selected on the basis of their size and whether they tested AI prospectively in real screening rather than on archived images. Randomized trials were given the most weight. PRAIM, which was observational, and other prospective and real-world studies from Sweden, Denmark, and Hungary were included as supporting evidence. Earlier retrospective studies, a 2021 systematic review, and two 2026 meta-analyses were included to show where the evidence stood before and after MASAI. A reader study was included because it examined how AI suggestions change radiologists' decisions, and a prospective study of reading time was included because it bears on the gap in Section 7.2, even though it also included diagnostic exams. FDA records and an analysis of cleared AI devices were used to describe how these tools reach the U.S. market. EDITH and PRISM are the largest ongoing trials, but their results are not yet available.

No formal risk-of-bias tool was used. Instead, each study's design determined how much weight its results carried: randomized trials first, then prospective paired-reader studies, then observational and before-and-after studies, and finally retrospective studies on archived images. The main weakness of each study, such as self-selection in PRAIM, the before-and-after design in Denmark, and industry funding in Hungary, is noted where its results are discussed. Results were summarized in words and tables rather than pooled, because the studies used different AI systems, reading workflows, and outcomes.

3. Design of the MASAI trial

MASAI was conducted within Sweden's national screening program. Between April 2021 and December 2022, 105,915 women were randomly assigned to two groups: 53,043 to AI-supported screening and 52,872 to standard double reading without AI [3]. In the AI group, a commercial system (Transpara version 1.7.0) scored each exam, routed lower-risk exams to one radiologist and higher-risk exams to two, and flagged suspicious findings for the readers [4]. Even so, every exam in the AI group was read by at least one radiologist [3]. The primary endpoint was the interval cancer rate. The trial used a non-inferiority design, meaning that it aimed to show that AI-supported screening was no worse than double reading within a prespecified margin of 20% [3].

4. Results of the MASAI trial

4.1. Safety Analysis (2023)

Finding more cancer only helps if the tool does not also flag healthy women and send them for scans and biopsies they never needed. In 2023, the researchers published an early safety analysis of the first 80,020 women, split almost evenly into two groups of about 40,000 [5]. The AI group had 244 cancers detected, and the standard group had 203, roughly 6 per 1,000 women compared with 5. The share of women called back for extra tests barely moved, from 2.0% to 2.2%. The false positive rate stayed at 1.5% in both groups, and reading workload fell by 44% [5].

4.2. Screening Performance (2025)

With the full group of about 106,000 women, the AI-supported group had 29% more cancers detected than standard double reading, with no rise in false positives [4]. The drop in reading workload seen in the early analysis held at 44%, and it came from the triage step, which sent lower-risk exams to one reader instead of two [4].

4.3. Interval Cancers (2026)

The interval cancer results were published in The Lancet in January 2026 [3]. In the standard group, 93 women had an interval cancer. In the AI group, 82 did, which corresponds to a 12% reduction (1.55 compared

with 1.76 per 1,000 women; proportion ratio 0.88, 95% CI: 0.65 to 1.18) [3]. The AI group also had 16% fewer invasive interval cancers and 27% fewer cases of the aggressive subtypes [3]. Sensitivity rose from 73.8% (95% CI: 68.9 to 78.3) to 80.5% (95% CI: 76.4 to 84.2; p = 0.031), while specificity remained at 98.5% in both groups [3]. These numbers must be read carefully. MASAI was a non-inferiority trial, so it was designed to show that AI-supported screening was no worse than double reading, and it achieved that goal. It was not designed to show that AI is better. The 12% reduction was not statistically significant (p = 0.41), and its confidence interval ranges from a 35% reduction to an 18% increase, so it could be due to chance [3]. Table *1* summarizes the three main MASAI publications.

Publication	Women analyzed	Main outcome	Key result (AI vs. standard)	Limitation noted
Lång et al. [5]. The Lancet Oncology, 2023	80,020	Safety: detection, recall, false positives	244 vs. 203 cancers; recall 2.2% vs. 2.0%; false positives 1.5% in both; 44% less reading workload	Interval cancer data not yet available
Hernström et al. [4]. The Lancet Digital Health, 2025	About 106,000	Screening performance	29% more cancers detected; no rise in false positives; 44% less reading workload	Swedish population only; long-term outcomes unknown
Gommers et al. [3]. The Lancet, 2026	105,915	Interval cancer rate	82 vs. 93 interval cancers (12% fewer); sensitivity 80.5% vs. 73.8%; specificity 98.5% in both	Drop in interval cancers not statistically significant

TABLE 1: Main published results from the MASAI trial

MASAI: Mammography Screening With Artificial Intelligence; AI: artificial intelligence

5. Evidence beyond MASAI

5.1. Retrospective Studies

Before MASAI, most of the evidence for AI in mammography came from archived images. In 2019, Rodriguez-Ruiz et al. compared an AI system with 101 radiologists on 2,652 exams, 653 of which showed cancer, drawn from nine datasets in seven countries [6]. The AI system had an area under the curve of 0.840, compared with an average of 0.814 for the radiologists, and it scored higher than 61.4% of them [6]. In 2020, McKinney et al. tested a different system on screening data from the United States and the United Kingdom and reported absolute reductions in false positives of 5.7% and 1.2% and in false negatives of 9.4% and 2.7% [7].

Test sets like these are chosen in advance, and the 2019 set was about one-quarter cancers, a much higher proportion than a screening population would have. A 2021 systematic review in The BMJ looked at AI in screening programs and found a less favorable picture. It included 12 studies covering 131,822 screened women. In the three studies that compared AI with the radiologist's original decision, 34 of 36 AI systems (94%) were less accurate than a single radiologist, and all of them were less accurate than the consensus of two or more radiologists [8]. Five smaller studies, with 1,086 women and 520 cancers, found AI to be more accurate than a single radiologist, but the review judged them to be at high risk of bias [8].

5.2. Prospective and Real-World Studies

In Germany, the PRAIM study followed 463,094 women who were screened at 12 sites between July 2021 and February 2023, and 119 radiologists decided for themselves whether to use an AI system [9]. The AI-supported group had a cancer detection rate of 6.7 per 1,000 compared with 5.7 per 1,000 in the group without AI, a 17.6% increase, and the recall rate stayed level (37.4 compared with 38.3 per 1,000) [9]. Because PRAIM was observational, radiologists who chose AI may differ from those who did not. Its detection results therefore support MASAI but do not carry the same weight as a randomized trial.

In Stockholm, the ScreenTrustCAD study used AI as an independent reader in a paired design, so each reading strategy could be compared on the same women [10]. Of the 58,344 women aged 40 to 74 screened between April 2021 and June 2022, 55,581 were included. One radiologist plus AI detected 261 cancers, compared with 250 for two radiologists, a 4% higher detection rate that met the study's non-inferiority goal [10]. AI alone detected 246 [10].

A Danish study examined what changed after the Capital Region of Denmark started using AI routinely in November 2021. Before that, two radiologists read every exam. Afterward, exams the AI judged likely

normal, about two-thirds of the total (66.9%), went to one senior breast radiologist, and the rest went to two radiologists with AI support [11]. This resembles MASAI's triage step, but it took place in routine care rather than in a trial. Comparing 60,751 women screened before the change with 58,246 screened after it, Lauritzen et al. found that cancer detection rose from 0.70% to 0.82%, the recall rate fell from 3.09% to 2.46%, and the false positive rate fell from 2.39% to 1.63% [11]. After the change, about one in three recalls led to a cancer diagnosis (33.6%), up from fewer than one in four (22.6%), and the reading workload dropped by 33.5% [11]. Unlike MASAI, the Danish program lowered false positives instead of only holding them steady. However, a before-and-after design cannot rule out other changes over the same period. Women in the AI group also had a longer interval since their previous screening (a median of 993 days, compared with 845), which could raise detection on its own, and the authors note that longer follow-up is needed to check for overdiagnosis [11].

In Hungary, Ng et al. used AI the other way around, to add reads instead of cutting them [12]. Two radiologists still read every exam. When both cleared an exam but the AI system, Mia, flagged it, another radiologist reviewed the case and decided whether to recall the woman [12]. In routine use at four sites from July 2022 to January 2023, after two pilot phases, this added 4% to 6% to total reading workload and found 0.7 to 1.6 more cancers per 1,000 women, with 0.16% to 0.30% more recalls [12]. Most of the extra cancers were invasive (83.3%), and 47.0% were 10 mm or smaller [12]. Kheiron Medical Technologies, the company that makes Mia, funded the study, and six of the authors were Kheiron employees who held stock options, which matters when this study is compared with independent ones. Follow-up lasted only two to nine months, which was too short to count interval cancers [12].

5.3. Recent Meta-Analyses

Two meta-analyses published in 2026 pooled much of this evidence. Sossavi et al. combined 13 prospective and retrospective studies, covering 1.03 million screens, that compared AI-supported reading with standard double reading [2]. Overall, cancer detection was about the same (risk ratio 1.01, 95% CI: 0.96 to 1.07), and so was the recall rate (1.00, 95% CI: 0.88 to 1.15) [2]. The results depended on how the AI was used. When it was used for triage, as in MASAI, cancer detection held steady, and recalls fell by 11% (0.89, 95% CI: 0.82 to 0.96). When it replaced the second reader, first reads were roughly halved, but the number of disagreements that a third radiologist had to settle rose sharply, so some of the saved work moved to that step [2].

Wulandari et al. looked only at AI systems cleared by the FDA or carrying a European CE mark and pooled 42 studies [13]. The AI systems reading on their own reached an area under the curve of 0.89 (95% CI: 0.85 to 0.92) on digital mammograms, which the authors judged comparable to radiologists [13]. But 39 of the 42 studies were retrospective, and the authors found that the evidence was strongest for triage and grew weaker as the AI was given more independence [13].

5.4. Ongoing Trials

Two newer trials are underway. In England, the EDITH trial was announced with plans to invite nearly 700,000 women at 30 screening sites and test whether one specialist working with AI can do work that now takes two [14]. Its registered protocol describes a randomized trial of 660,000 women at about 30 screening centers that will test two types of AI-assisted screening against standard double reading, with women followed through cancer registries for four years [15]. In the United States, the PRISM trial, funded with $16 million from the Patient-Centered Outcomes Research Institute [16], randomly assigns individual screening exams to a radiologist alone or a radiologist with Transpara version 2.1 on 3D mammograms. It plans to include about 400,000 exams read by about 108 radiologists at six health systems [17]. Its co-primary outcomes are cancer detection and recall rates, and it lists interval cancers within 365 days of a negative screen as a secondary outcome [17]. Table *2* compares these studies with MASAI.

Study	Country	Design	Size	Main endpoints
MASAI: Gommers et al. [3]. The Lancet, 2026	Sweden	Randomized controlled trial	105,915 women	Interval cancer rate (primary); cancer detection; workload
PRAIM: Eisemann et al. [9]. Nature Medicine, 2025	Germany	Observational: radiologists chose whether to use AI	463,094 women	Cancer detection rate; recall rate
ScreenTrustCAD: Dembrower et al. [10]. The Lancet Digital Health, 2023	Sweden	Prospective, paired-reader, non-inferiority study	55,581 women	Cancer detection with one radiologist plus AI vs. two radiologists
Lauritzen et al. [11]. Radiology, 2024	Denmark	Before-and-after comparison of routine AI use	60,751 women before AI; 58,246 after	Cancer detection; recall rate; false positive rate; reading workload
Ng et al. [12]. Nature Medicine, 2023	Hungary	Prospective: AI flags cases for an additional human read	28,811 women across three phases	Additional cancers detected; recalls; added workload
EDITH: UK Department of Health and Social Care [14], 2025; ISRCTN81384017 [15]	England	Randomized trial (registered protocol); AI-assisted screening vs. standard double reading	660,000 women (planned)	Whether one reader with AI can do the work of two
PRISM: UCLA Fielding School of Public Health [16], 2025; NCT06934239 [17]	United States	Randomized by individual screening exam	About 400,000 screening exams (planned)	Cancer detection rate; recall rate (co-primary); interval cancers within 365 days (secondary)

TABLE 2: Large studies of AI-supported mammography screening

AI: artificial intelligence

6. How AI devices reach the U.S. market

Most medical AI products have far less evidence behind them than Transpara has now. As of September 2026, the FDA's list of AI-enabled medical devices included 1,614 entries, and 1,230 (76.2%) were in radiology [18]. These counts come from the version of the list that was accessed on October 14, 2026. Each device's pathway can be determined from its submission number, since 510(k) numbers start with K, De Novo numbers with DEN, and premarket approval numbers with P. By that count, 1,553 of the 1,614 devices (96.2%) were cleared through the 510(k) pathway, 40 were authorized through De Novo classification, and 21 received premarket approval [18]. Under 510(k), a company clears a device by showing that it is substantially equivalent to one already on the market. Clinical trial data are not required.

Transpara's own history shows how this works. The FDA cleared its first version in November 2018 under 510(k), about two and a half years before MASAI enrolled its first woman [19]. The predicate device named in that submission was OsteoDetect, software that detects wrist fractures on X-rays [19]. The main clinical evidence was a retrospective reader study in which 14 radiologists read 240 archived cases with and without the software, and their average area under the receiver operating characteristic curve (AUC) rose from 0.866 to 0.886 [19]. The cleared use was a concurrent reading aid that marks suspicious regions while a radiologist reads [19]. In MASAI, the software performed a task with higher stakes: it decided whether an exam got one reader or two [4]. A 2026 analysis in PLOS Digital Health found the same pattern across the field. Of the 1,357 cleared AI devices, only 34 (2.5%) were linked to a registered prospective trial, and only 3 (0.2%) had been tested on patient outcomes such as mortality or readmissions [20].

7. Gaps in the current evidence

7.1. The Interval Cancer Result Is Too Small to Confirm in One Trial

MASAI was sized to show non-inferiority, and a non-inferiority trial is not built to prove a benefit. Showing that AI actually lowers interval cancers would take a superiority test, which needs a much larger sample. A rough estimate comes from the standard formula for comparing two proportions with a two-sided test, assuming equal group sizes and that each woman’s outcome is independent of the others. The approximate number of women needed in each group is given by Equation 1:

$$n = \left(z_1 - \frac{\alpha}{2} + z_1 - \beta\right)^2 \times [p_1(1-p_1) + p_2(1-p_2)]/(p_1 - p_2)^2$$

Using MASAI's observed interval cancer rates, p_1 = 93/52,872 (1.76 per 1,000) for standard reading and p_2 = 82/53,043 (1.55 per 1,000) for AI-supported reading, with α = 0.05 and 80% power (z values of 1.960 and 0.842), Equation 1 gives about 570,500 women per group, or about 1.14 million in total. This estimate treats MASAI's observed 12% drop as the true effect, and the answer changes a lot if that is wrong. At 90% power, the total rises to about 1.53 million. If the true drop were 20%, the total would fall to about 400,000. MASAI's confidence interval (0.65 to 1.18) is wide enough that a 20% drop and no drop at all are both possible [3]. The estimate also ignores dropout and clustering by screening site or radiologist, and both would raise it. Even EDITH, with 660,000 women registered [15], would fall short of 1.14 million if it were split into two equal groups to test the same question.

7.2. The Workload Savings Depend on Double Reading

MASAI's workload savings came from its triage step, which dropped the second reader on lower-risk exams [4]. That saving exists only where a second reader exists. In U.S. practice, where a single reader with computer-aided detection is more common [1], there is no second read to remove, and AI is unlikely to replace that one reader. Its value there would have to come from saving time on each read. Most screening exams are normal, and when the AI scores an exam as normal, and the radiologist reads it as normal too, the radiologist could be more confident and finish faster. The time saved on normal exams could then go to complex cases. MASAI counted how many reads each exam got but did not time them.

The early evidence on timing does not yet show a saving. In a prospective study at one Korean hospital, Lee et al. turned AI on and off in alternating months for four radiologists reading 4,577 screening and diagnostic mammograms [21]. Average reading time was almost the same with and without AI (65.0 compared with 64.3 seconds, p = 0.723), while the cancer detection rate was higher in the AI months (22.3 compared with 11.5 per 1,000) [21]. That study was small, mixed screening exams with diagnostic ones, and reported average times rather than times for normal exams alone, so it cannot settle the question. PRISM plans to measure reading time, but only at its UCLA site [17].

7.3. The Tested Use Differs From the Cleared Use

MASAI tested Transpara version 1.7.0 as a triage tool [4]. The 2018 U.S. clearance covered an earlier version as a concurrent reading aid [19]. Software is updated more often than large trials are run, so by the time a trial result is published, the product on the market may differ from the one that was tested. The FDA has started to respond to this problem. Under final guidance issued in 2024, a manufacturer can submit a predetermined change control plan that describes, in advance, the changes it plans to make to an AI device and how it will test them, so that those changes do not each need a new submission [22].

7.4. The Trial Population Was Narrow

MASAI enrolled women in Sweden, and the researchers note that the results may not carry over to other health care systems [4]. The trial also did not collect data on race or ethnicity [5], so it cannot show whether the tool works equally well for all women. Follow-up across later screening rounds and studies of cost-effectiveness are still to come [3].

7.5. Radiologists Can Follow the AI When It Is Wrong

In MASAI's AI group, the software also highlighted suspicious findings for the radiologists [4]. That raises a question the trial was not designed to answer: what happens when the marks are wrong? In a 2023 reader study, Dratsch et al. had 27 radiologists read 50 mammograms with category suggestions from what they were told was an AI system, and for 12 of the mammograms, the suggestion was deliberately incorrect [23]. Inexperienced radiologists rated 79.7% of mammograms correctly when the suggestion was right, but only 19.8% when it was wrong. Even very experienced radiologists fell from 82.3% to 45.5% [23]. The AI in that study was simulated, and the image set was small, so it cannot show how often this happens in real screening. A U.S. trial could measure it by recording how often radiologists change their read to match an AI suggestion that later proves wrong.

8. Proposed directions

The two proposals below are the authors' recommendations, not findings from the studies reviewed. Each comes from the gaps in Section 7. The trial design in Section 8.1 addresses Sections 7.1, 7.2, 7.4, and 7.5, and the regulatory rule in Section 8.2 addresses Section 7.3 and the evidence gap described in Section 6.

8.1. A Registry-Linked Trial in Single-Reader Screening

PRISM's registered protocol already counts interval cancers within 365 days of a negative screen as a secondary outcome and measures reading time at one site [17]. A U.S. trial designed to address the questions in Section 7 would go further in four ways. First, its primary endpoint would be the interval cancer rate, which would be measured by linking screening records to state cancer registries over at least two screening rounds (Section 7.1). Second, it would randomize by screening site or reading day across a network large enough to approach the 1.14 million women estimated in Section 7.1, with that figure increased to account for clustering. Third, it would prespecify results by race and ethnicity, which MASAI could not report [5] (Section 7.4). Fourth, it would record how long radiologists spend on each exam at every site, with normal exams reported separately, and how often radiologists change a read to match an AI suggestion that later proves wrong (Sections 7.2 and 7.5). It would also report sensitivity, specificity, and positive and negative predictive values.

A lower-cost first step is possible. PRISM already randomizes exams between a radiologist alone and a radiologist with AI [17]. Linking PRISM's participants to state cancer registries after the trial would extend its interval cancer follow-up beyond 365 days and into later screening rounds without enrolling anyone new, and the results could then be pooled with MASAI and EDITH to reach the sample size given in Equation 1.

8.2. A Tiered Evidence Rule for AI Devices

Not every AI tool needs a trial the size of MASAI. A tool that marks images for a radiologist who still reads every exam carries different risks from those of software that decides whether a second radiologist ever sees a scan. A tiered rule would scale the evidence to that difference. Tools that only mark images for a reader could stay on the 510(k) pathway, with a requirement to report how they perform after reaching the market. Tools whose output changes how many people read an exam would need prospective patient outcome data, such as interval cancer rates, either before clearance or within a fixed period afterward, as a condition of staying on the market. Because 96.2% of listed AI devices already go through 510(k) [18], such a rule would leave most products on their current path.

Conclusions

MASAI showed that a large randomized trial of medical AI can be run inside a national screening program, and that AI-supported screening found more cancers with less work and without a rise in false positives. Its interval cancer result points in a favorable direction, but confirming it would require a sample about 10 times larger. Its workload savings rest on double reading, and the device it tested was cleared in the United States years earlier, for a narrower use, based on retrospective data. In the United States, where one radiologist usually reads each exam, the question is less about removing a reader than about how much time AI saves per read. The difference in MASAI came down to 93 women compared with 82, and further research must determine whether that gap is real.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Sohum Shah

Acquisition, analysis, or interpretation of data: Sohum Shah, Ripal Gandhi

Drafting of the manuscript: Sohum Shah

Critical review of the manuscript for important intellectual content: Ripal Gandhi

Supervision: Ripal Gandhi

Disclosures

AI use statement: The authors used Claude as follows: language polishing, structure or formatting, data cleaning or exploratory analysis in Abstract, Introduction, Results. I had to analyze the data from the AI tools that I used to determine relevance because the complexity of the topic required review with a practicing radiologist in detail. I had other mentors provide guidance and suggestions that I then incorporated into the article as well. The authors confirm that they have reviewed and verified all AI-assisted content and take full responsibility for the work. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear

to have influenced the submitted work.

References

1. Taylor-Phillips S, Stinton C: Double reading in breast cancer screening: considerations for policy-making . Br J Radiol. 2020, 93:20190610. 10.1259/bjr.20190610
2. Sossavi E, Roy C, Molière S: Artificial intelligence in breast cancer screening: a systematic review and meta-analysis of integration strategies. Eur J Radiol Open. 2026, 16:100727. 10.1016/j.ejro.2026.100727
3. Gommers J, Hernström V, Josefsson V, et al.: Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study: a randomised, controlled, non-inferiority, single-blinded, population-based, screening-accuracy trial. Lancet. 2026, 407:505-14. 10.1016/S0140-6736(25)02464-X
4. Hernström V, Josefsson V, Sartor H, et al.: Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study. Lancet Digit Health. 2025, 7:e175-83. 10.1016/S2589-7500(24)00267-X
5. Lång K, Josefsson V, Larsson AM, et al.: Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol. 2023, 24:936-44. 10.1016/S1470-2045(23)00298-X
6. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al.: Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. J Natl Cancer Inst. 2019, 111:916-22. 10.1093/jnci/djy222
7. McKinney SM, Sieniek M, Godbole V, et al.: International evaluation of an AI system for breast cancer screening. Nature. 2020, 577:89-94. 10.1038/s41586-019-1799-6
8. Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A, Taylor-Phillips S: Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. BMJ. 2021, 374:n1872. 10.1136/bmj.n1872
9. Eisemann N, Bunk S, Mukama T, et al.: Nationwide real-world implementation of AI for cancer detection in population-based mammography screening. Nat Med. 2025, 31:917-24. 10.1038/s41591-024-03408-6
10. Dembrower K, Crippa A, Colón E, Eklund M, Strand F: Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. Lancet Digit Health. 2023, 5:e703-11. 10.1016/S2589-7500(23)00153-X
11. Lauritzen AD, Lillholm M, Lynge E, Nielsen M, Karssemeijer N, Vejborg I: Early indicators of the impact of using AI in mammography screening for breast cancer. Radiology. 2024, 311:e232479. 10.1148/radiol.232479
12. Ng AY, Oberije CJ, Ambrózay É, et al.: Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. Nat Med. 2023, 29:3044-9. 10.1038/s41591-023-02625-9
13. Wulandari PI, Gandomkar Z, Rickard M, Hooshmand S, Suleiman M, Brennan P: Diagnostic performance of FDA-cleared and CE-marked AI systems for mammography screening: a systematic review and meta-analysis. Eur J Radiol. 2026, 204:113136. 10.1016/j.ejrad.2026.113136
14. World-leading AI trial to tackle breast cancer launched . (2025). Accessed: November 1, 2026: https://www.gov.uk/government/news/world-leading-ai-trial-to-tackle-breast-cancer-launched .
15. Early Detection using Information Technology in Health (EDITH): ISRCTN81384017 . (2025). Accessed: November 18, 2026: https://www.isrctn.com/ISRCTN81384017.
16. UCLA to lead $16 million national study on artificial intelligence in breast cancer screening . (2025). Accessed: November 1, 2026: https://ph.ucla.edu/news-events/news/ucla-lead-16-million-national-study-artificial-intelligence-breast-cancer.
17. Pragmatic Randomized trial of artificial Intelligence for Screening Mammography (PRISM): NCT06934239 . (2025). Accessed: November 18, 2026: https://clinicaltrials.gov/study/NCT06934239.
18. Artificial intelligence-enabled medical devices . (2026). Accessed: October 14, 2026: https://www.fda.gov/medical-devices/artificial-intelligence-enabled-medical-devices/list-artificial-intelligence-enab....
19. 510(k) summary: Transpara (K181704) . (2018). Accessed: November 1, 2026: https://www.accessdata.fda.gov/cdrh_docs/pdf18/K181704.pdf.
20. Abulibdeh R, Cajas Ordóñez SA, Celi LA, Gorijavolu R, Izath N, Markussen Lunde T: 1,357 AI medical devices cleared, 3 actually tested on patient outcomes. PLOS Digit Health. 2026, 5:e0001597. 10.1371/journal.pdig.0001597
21. Lee SE, Heo SJ, Shin HJ, Kim EK: Impact of AI assistance on reading time, cancer detection rate, and abnormal interpretation rate in screening and diagnostic mammography: a prospective alternating-month study. Eur Radiol. 2026, 4:4-6. 10.1007/s00330-026-12793-0
22. Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions: guidance for industry and Food and Drug Administration staff. (2024). Accessed: November 6, 2026: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predete....
23. Dratsch T, Chen X, Rezazade Mehrizi M, et al.: Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. Radiology. 2023, 307:e222176. 10.1148/radiol.222176