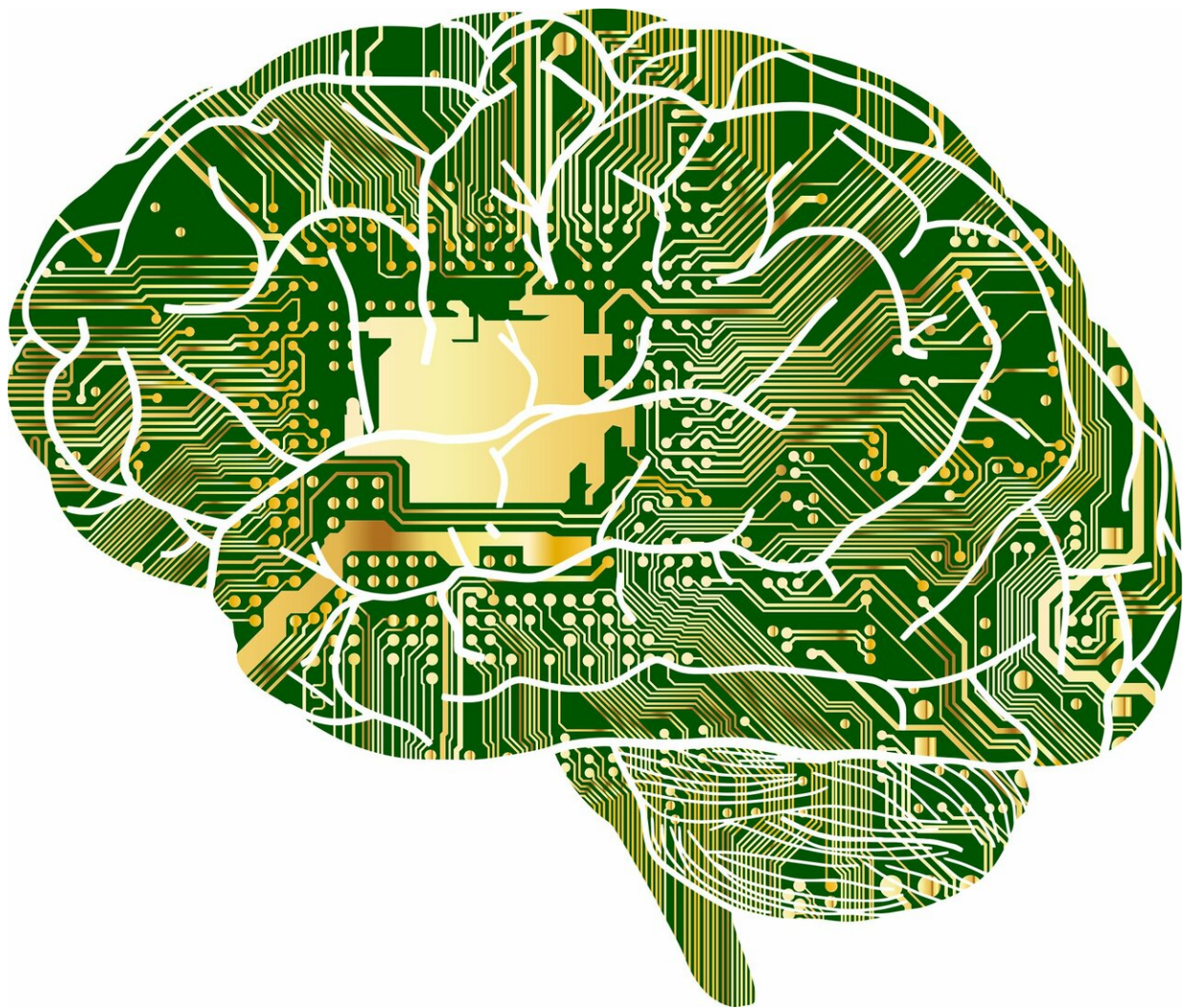


# First randomized trial finds AI-supported mammography screening is safe and almost halves radiologist workload

August 1 2023

---



Credit: CC0 Public Domain

An interim safety analysis of the first randomized controlled trial of its kind involving over 80,000 Swedish women published in *The Lancet Oncology* journal, finds artificial intelligence (AI)-supported mammography analysis is as good as two breast radiologists working together to detect breast cancer, without increasing false positives and almost halving the screen-reading workload.

However, the final trial results looking at whether the use of AI in interpreting mammography images translates into a reduction in interval cancers (cancers detected between screenings that generally have a poorer prognosis than screen-detected cancers) in 100,000 women followed over two years—and ultimately whether AI's use in mammography screening is justified—are not expected for several years.

"These promising interim safety results should be used to inform new trials and program-based evaluations to address the pronounced radiologist shortage in many countries. But they are not enough on their own to confirm that AI is ready to be implemented in mammography screening," cautions lead author Dr. Kristina Lång from Lund University, Sweden.

"We still need to understand the implications on patients' outcomes, especially whether combining radiologists' expertise with AI can help detect interval cancers that are often missed by traditional screening, as well as the cost-effectiveness of the technology."

Breast [cancer](#) screening with mammography has been shown to improve prognosis and reduce mortality by detecting breast cancer at an earlier, more treatable stage. However, estimates suggest that 20–30% of interval cancers that should have been spotted at the preceding screening mammogram are missed, and suspicious findings often turn out to be benign.

European guidelines recommend double reading of screening mammograms by two radiologists to ensure high sensitivity (to correctly identify those with disease). But there is a shortage of breast radiologists in many countries, including a shortfall of around 41 (8%) in the UK in 2020 and about 50 in Sweden, and it takes over a decade to train a radiologist capable of interpreting mammograms.

AI has been proposed as an automated second reader for mammograms that might help reduce this workload and improve screening accuracy. The technology has shown encouraging results in retrospective studies using AI to triage examinations to either single or double reading and by providing radiologists with computer-aided detection (CAD) marks highlighting suspicious features to reduce false negative results. But robust evidence from prospective randomized trials has been lacking.

Between April 2021 and July 2022, 80,033 women aged 40-80 years who had undergone mammogram screening at four sites in southwest Sweden were randomly assigned in a 1:1 ratio to either AI-supported analysis, where a commercially available AI-supported mammogram reading system analyzed the mammograms before they were also read by one or two radiologists (intervention arm), or standard analysis performed by two radiologists without AI (control arm).

This interim analysis of the Mammography Screening with Artificial Intelligence (MASAI) trial compared early screening performance (e.g., cancer detection, recalls, false positives) and screen-reading workload in the two arms. The MASAI trial will continue to establish primary outcome results of whether AI-supported [mammography screening](#) reduces interval cancers.

The lowest acceptable limit for clinical safety in the intervention group was set at a cancer detection rate above three cancers per 1,000 screened women. This was based on the premise that the cancer detection rate

might decline because the majority of screening examinations would undergo single reading instead of double reading. The baseline detection rate in the current screening program with double reading is five cancers per 1,000 screened women.

In the AI-supported analysis, the AI system first analyzed the mammography image and predicted the risk of cancer on a scale of one to 10, with one representing the lowest risk and 10 the highest. If the risk score was less than 10 the image was further analyzed by one radiologist, whereas if the AI system predicted a risk score of 10 then two radiologists analyzed the image.

The system also provided CAD marks to assist radiologists in accurately interpreting mammography images. Women were recalled for additional testing based on suspicious findings. Radiologists had the final decision to recall women and were instructed to recall cases with the highest 1% risk, except for obvious false positives.

AI failed to provide a risk score in 0.8% of cases (306/39,996) that were referred to standard care (double reading).

The recall rates averaged 2.2% (861 women) for AI-supported screening and 2.0% (817 women) for standard double reading without AI. These were similar to the average 2.1% recall rate in the clinic six months prior to the trial starting, indicating that cancer detection rates had not fallen.

In total, 244 women (28%) recalled from AI-supported screening were found to have cancer compared with 203 women (25%) recalled from standard screening—resulting in 41 more cancers detected with the support of AI (of which 19 were invasive and 22 were in situ cancers). The false-positive rate was 1.5% in both arms.

Overall, AI-supported screening resulted in a cancer detection rate of six

per 1,000 screened women compared to five per 1,000 for standard double reading without AI—equivalent to detecting one additional cancer for every 1,000 women screened.

Importantly, there were 36,886 fewer screen readings by radiologists in the AI-supported group than in the control group (46,345 vs. 83,231), resulting in a 44% reduction in the screen-reading workload of radiologists.

Although the actual time saved by using AI was not measured in the trial, the researchers calculate that if a radiologist reads on average 50 mammograms an hour, it would have taken one radiologist 4.6 months less to read the roughly 40,000 screening examinations with the help of AI compared with the roughly 40,000 in the control arm that were double read.

"The greatest potential of AI right now is that it could allow radiologists to be less burdened by the excessive amount of reading," says Lång.

"While our AI-supported screening system requires at least one [radiologist](#) in charge of detection, it could potentially do away with the need for double reading of the majority of mammograms easing the pressure on workloads and enabling radiologists to focus on more advanced diagnostics while shortening waiting times for patients."

Despite the promising findings, the authors note several limitations including that the analysis was conducted at a single center and was limited to one type of mammography device and one AI system which might limit the generalizability of the results. They also note that while technical factors will affect the performance and processing of the AI system, these will likely be less important than the experience of radiologists.

Because the AI-supported system places the final decision on whether to

recall women on radiologists, the results are dependent on their performance. In this trial, radiologists were moderately to highly experienced, which could limit the generalizability of the findings to less experienced readers. Lastly, information on race and ethnicity was not collected.

Writing in a linked Comment, Dr. Nereo Segnan, former Head of the Unit of Cancer Epidemiology and past Director of Department of Screening at CPO Piemonte in Italy (who was not involved in the study) notes that the AI risk score for [breast cancer](#) seems very accurate at being able to separate high risk from low-risk women, adding that, "In risk stratified screening protocols, the potential for appropriately modulating the criteria for recall in low-risk and high-risk groups is remarkable."

However, he cautions, "In the AI-supported [screening](#) group of the MASAI trial, the possible presence of over diagnosis (i.e., the system identifying non-cancers) or over-detection of indolent lesions, such as a relevant portion of ductal carcinomas in situ, should prompt caution in the interpretation of results that otherwise seem straightforward in favoring the use of AI...It is, therefore, important to acquire biological information on the detected lesions.

The final results of the MASAI trial are expected to do so, as the characteristics of identified cancers and the rate of interval cancers—not just the detection rate—are indicated as main outcomes. An important research question thus remains: is AI, when appropriately trained, able to capture relevant biological features—or, in other words, the natural history of the disease—such as the capacity of tumors to grow and disseminate?"

**More information:** Kristina Lång 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, *The Lancet Oncology* (2023). DOI: 10.1016/S1470-2045(23)00298-X , [www.thelancet.com/journals/lan ... \(23\)00298-X/fulltext](http://www.thelancet.com/journals/lan ... (23)00298-X/fulltext)

Provided by Lancet

Citation: First randomized trial finds AI-supported mammography screening is safe and almost halves radiologist workload (2023, August 1) retrieved 2 May 2024 from <https://medicalxpress.com/news/2023-08-randomized-trial-ai-supported-mammography-screening.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.