A Randomized Controlled Trial Comparing Screening Mammography With and Without Assistance from Artificial Intelligence for Breast Cancer Detection and Recall Rates in Adult Patients: Study Protocol

CLINICAL INVESTIGATION PROTOCOL

Principle Investigators:

Joann G. Elmore, MD, MPH
David Geffen School of Medicine,
University of California, Los Angeles

Diana L. Miglioretti, PhD
University of California, Davis

Sponsored by:

University of California, Los Angeles, Jonsson Comprehensive Cancer Center

Version Number: v. 1.0

Day Month Year

August 26, 2025

Protocol Approved by DSMB: August 7, 2025

SPONSOR PROTOCOL APPROVAL

Protocol Title:	A Randomized Controlled Trial Comparing Screening Mammography With and Without Assistance from Artificial Intelligence for Breast Cancer Detection and Recall Rates in Adult Patients: Study Protocol	
Protocol Number:	NCT06934239	
Version:	Version 1.0	
Name and Title	**Date**	**Signature**

PROTOCOL REVISION HISTORY

Protocol version #	Release date	Reason for change

INVESTIGATOR SIGNATURE PAGE

By signing below, the Investigators acknowledges that they have read the NCT06934239 Protocol and agree to conduct the study as outlined. The Investigators agree to maintain the confidentiality of all the information received or developed with this protocol.

_____ _____

Joann G. Elmore, MD, MPH Diana L. Miglioretti, PhD

Printed Name of Dual Principal Investigators

Table of Contents (Following SPIRIT 2025 guidelines for protocols for randomized trials) **Page**

Administrative Information

1a, 1b. Title and structured summary

1a. Title stating the trial design, population, and interventions, with identification as a protocol.

A Randomized Controlled Trial Comparing Screening Mammography With and Without Assistance from Artificial Intelligence for Breast Cancer Detection and Recall Rates in Adult Patients: Study Protocol

1b. Structured summary of trial design and methods, including items from the World Health Organization Trial Registration Data Set.

We will conduct a randomized clinical trial (RCT) comparing screening outcomes when Digital Breast Tomosynthesis (DBT or 3D) mammograms are interpreted with vs. without an FDA-cleared AI tool, Transpara.

Primary Registry and Trial Identifying Number:	clinicaltrials.gov; NCT06934239
Date of Registration in Primary Registry:	April 18, 2025
Secondary Identifying Numbers:	N/A
Source(s) of Monetary or Material Support:	Patient-Centered Outcomes Research Institute (PCORI); California Breast Cancer Research Program of the University of California
Primary Sponsor:	Jonsson Comprehensive Cancer Center, University of California, Los Angeles
Secondary Sponsor(s):	
Contact for Public Queries:	Contact person: Dr. Michelle L'Hommedieu, UCLA Email address: mlhommedieu@mednet.ucla.edu Telephone: (310) 592-9454 Postal Address: 1100 Glendon Ave., Suite 900, Los Angeles, CA 90024
Contact for Scientific Queries:	Principal Investigator: Dr. Joann Elmore, UCLA Email address: jelmore@mednet.ucla.edu Telephone: (310) 794-3763 Postal Address: 1100 Glendon Ave., Suite 900, Los Angeles, CA 90024
	Principal Investigator: Dr. Diana Miglioretti Email address: dmiglioretti@ucdavis.edu Telephone: (530) 752-7168 Postal Address: One Shields Ave., Med Sci 1C, Davis, CA 95616
	Contact person: Dr. Hannah Milch, UCLA Email address: hmilch@mednet.ucla.edu Telephone: (424) 440-3266

	Postal Address: 1250 16th Street, Suite 2340, Santa Monica, CA, 90404
Public Title:	Pragmatic Randomized trial of artificial Intelligence for Screening Mammography (PRISM)
Scientific Title:	A Randomized Controlled Trial Comparing Screening Mammography With and Without Assistance from Artificial Intelligence for Breast Cancer Detection and Recall Rates in Adult Patients: Study Protocol
Countries of Recruitment:	United States (US)
Health Condition(s) or Problem(s) Studied:	Breast Cancer
Intervention(s):	For usual care arm, exams will be reviewed and interpreted as done in many usual clinical practices (radiologist alone). For the AI arm (radiologist assisted by AI), exams will be scored using ScreenPoint Medical's trademarked and FDA-cleared AI decision support tool for 3D mammography, Transpara (referred to as "AI" in this protocol) and radiologists will be presented with the AI scoring information before and during their interpretation of the mammogram images. We will start the trial using the Transpara version 2.1 and transition to the Transpara version 2.1 with temporal comparison once available.
Key Inclusion and Exclusion Criteria:	This trial will include all radiologists interpreting screening mammography and all adult patients undergoing screening mammography at any of the participating breast imaging facilities across 6 regional health systems (UCLA, UC San Diego, University of Washington-Seattle, University of Wisconsin-Madison, Boston Medical Center, and University of Miami) during the trial period. Individuals must meet all the following inclusion criteria: 1. Be at least 18 years of age or older 2. Receive a screening mammogram at one of the participating breast imaging facilities OR be a radiologist who interprets screening mammograms at one of the participating breast imaging facilities.

	Exclusion criteria: 1. Patients who have opted out of all research at the health system.
Study Type:	Interventional, phase IV, parallel-group randomized controlled trial comparing screening outcomes when digital breast tomosynthesis (DBT) mammograms are interpreted with vs. without an FDA-cleared AI tool. Screening mammography exams will be randomized after image acquisition to be interpreted either with or without an AI decision-support tool. Radiologists cannot be blinded to study arm during screening mammography interpretation. However, interpreting radiologists and facility staff (e.g., those scheduling the exams and those acquiring the images) will not know in advance which patients will be randomized to the AI tool. Radiology department staff, including schedulers and clinical support staff, and patients will remain blinded throughout the trial. Patients will not be informed of the use of AI, because the AI program is an FDA-cleared support tool currently in use at imaging facilities throughout the US.
Date of First Enrollment:	Pending –
Sample Size:	400,000 screening mammogram exams interpreted by approximately 108 radiologists.
Recruitment Status:	Pending (participants are not yet being recruited or enrolled at any site).
Primary Outcomes:	1. Outcome name: Cancer detection rate Metric/method of measurement: Number of screening exams recommended for breast biopsy (final BI-RADS assessment of 4 or 5) resulting in detected cancer, per 1,000 screening exams Timepoint: Cancer diagnosed within 90 days of positive study entry screening mammogram 2. Outcome name: Recall rate Metric/method of measurement: Number of screening exams recalled for diagnostic work-up (initial BI-RADS assessment of 0, 3, 4, or 5), per 1,000 screening exams Timepoint: Day of interpretation

Key Secondary Outcomes:	1. Outcome name: Interval cancer rate (i.e., false-negative rate) Metric/method of measurement: Number of screening exams with a negative assessment (final BI-RADS assessment of 1 or 2), and breast cancer diagnosed within 1 year, per 1,000 screening exams Timepoint: Cancer diagnosed within 365 days of a negative study entry screening mammogram 2. Outcome name: False positive recall rate Metric/method of measurement: Proportion of screening exams recalled for additional imaging (final BI-RADS assessment of 1, 2, or 3), with no breast cancer diagnosed within 1 year Timepoint: No cancer diagnosed within 365 days of a positive study entry screening mammogram 3. Outcome name: False positive short-interval follow-up recommendation rate Metric/method of measurement: Proportion of screening exams recalled for short-interval follow-up (final BI-RADS assessment of 3) with no breast cancer diagnosed within 1 year Timepoint: No cancer diagnosed within 365 days of a positive study entry screening mammogram 4. Outcome name: False positive biopsy recommendation rate Metric/method of measurement: Proportion of screening exams recalled for breast biopsy (final BI-RADS assessment of 4 or 5) with no breast cancer diagnosed within 1 year Timepoint: No cancer diagnosed within 365 days of a positive study entry screening mammogram 5. Outcome name: Trust and confidence in AI Metric/method of measurement: Focus group and survey data Timepoint: Years 1,2 and Years 4,5

	6. Outcome name: Efficiency metrics (only for the UCLA site) Metric/method of measurement: Interpretation time required for radiologists to interpret each mammogram with versus without AI; Delivery time, using time stamp data from exam acquisition to delivery of results to patients (aka turnaround time) Timepoint: Timepoints as described above
Ethics Review Main Trial: **Date of approval:**	Approved (UCLA IRB 24-1192) 11/27/2024
Ethics Review Radiologists' Data: **Date of approval:**	Approved (UCLA IRB 24-0730) 6/6/2024
Name and contact details of ethics committee(s):	UCLA Office of Human Research Protection Program (OHRPP). (310) 825-5344 ohrpp@research.ucla.edu
Completion date:	N/A
Summary results:	N/A
IPD sharing statement:	Plan to share IPD: We will follow mandates of PCORI and individual state laws.

2. Protocol version

Version date and identifier.

Version date:	August 26, 2025
Identifier:	Version 1.0

3a, 3b, 3c, 3d. Roles and responsibilities

3a. Names, affiliations, and roles of protocol contributors.

Joann Elmore, MD, MPH	Professor of Medicine, University of California, Los Angeles	Dual-PI of PCORI award (administrative PI)
Diana L. Miglioretti, PhD	Professor and Division Chief of Biostatistics, University of California, Davis	Dual-PI of PCORI award
Hannah Milch, MD	Assistant Professor of Radiology, University of California, Los Angeles	Co-PI of PCORI award, UCLA Site Lead Radiologist
Christoph Lee, MD, MS, MBA	Professor of Radiology, University of Wisconsin-Madison	Co-PI of PCORI award
Michelle L'Hommedieu, PhD	Project Director, University of California, Los Angeles	Project Director

3b. Name and contact information for the trial sponsor.

Jonsson Comprehensive Cancer Center, University of California, Los Angeles

3c. Role of trial sponsor and funders in design, conduct, analysis, and reporting of trial; including any authority over these activities.

The principal investigators of this investigator-initiated trial have full responsibility for the design, conduct, analysis, and reporting of the trial.

The trial funders are not involved in study design, conduct, analyses, or reporting. They do not have any authority over these activities.

3d. Composition, roles, and responsibilities of the coordinating site, steering committee, end point adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable.

Administrative Coordinating Site (UCLA): Dr. Elmore will serve as the Contact PI and lead of the Administrative Coordinating Site. She will oversee all aspects of the design, conduct, analyses, and reporting of the project and provide input on interpretation and translation of findings into clinical practice. She will oversee progress on the implementation of the technical infrastructure required for the project, guide the scientific direction, and ensure scientific rigor. In addition, she will oversee successful completion of all project-related administrative tasks. Dr. Elmore will oversee the involvement of the Data and Safety Monitoring Board and the engagement with stakeholder partners and patients to ensure that their perspectives are at the forefront of the design and implementation of this study.

Data Coordinating site (UCD): Dr. Miglioretti will serve as Dual PI and the lead of the Data Coordinating Center (DCC). She will oversee data collection and management, guide the scientific direction, and ensure scientific rigor. The DCC will be responsible for maintaining data integrity, security, and protections throughout the study. They will develop a comprehensive data management plan, including manuals of operations for clinical data collection and associated data dictionaries. A REDCap database will be developed to collect aggregate data from participating facilities for monthly enrollment reports. They will also facilitate the collection of precise data on which Transpara version and functionality is in use at each site, including dates of implementation. Additionally, the DCC will design and implement a data submission process and procedures for receiving de-identified exam-level data—including patient, exam, radiologist, and facility characteristics; radiologists' interpretations; AI results; breast cancer outcomes; and radiologist survey data—from participating facilities. The DCC will review submitted data for completeness and quality, document and resolve data issues, and pool the cleaned data for analysis. The DCC will also assist participating sites with cancer registry linkages. The DCC biostatisticians will create reports for DSMB meetings and conduct analyses for reports and manuscripts.

Executive Committee: This core decision-making body will include the two Dual-PIs, two Co-PIs, the project director, and the DCC project manager. The committee will meet twice monthly to monitor progress, resolve issues, and make decisions on scientific strategy and resource allocation.

Steering Committee: The Steering committee (SC) will include the dual PIs, two Co-PIs, all site PIs, and a patient representative. The SC will provide a forum for sharing best practices across sites and advise about all matters related to implementing study procedures at the local sites. The SC will be the primary

decision-making body for all matters related to the conduct of the study at the individual local sites. The SC will meet virtually approximately monthly in year one, and then quarterly thereafter.

Endpoint Adjudication Committee: None

Data and Safety Monitoring Board (DSMB): The Data and Safety Monitoring Board (DSMB) will be chaired by Dr. Ruth Carlos, who leads the ECOG-ACRIN Cancer Care Delivery Research Committee. It will include 4 other individuals, including experts in cancer screening, RCT design, biostatistics, and one patient partner. The DSMB will evaluate interim data reports every 6 months for participant safety, study conduct, and progress, and to make recommendations concerning trial continuation, modification, or termination. An interim data analysis will be conducted once half of exams are randomized. Results by study arm will be shared with the DSMB during closed session and will not be available to the executive or steering committees until enrollment is completed, and data are cleaned and locked.

Patient Partner Advisory Board (PPAB): Patient partners (N = 8) chosen from the study data collection sites will meet approximately quarterly and help guide the study from the patient perspective. Patient partners will ensure that the study remains aligned with the needs and preferences of the populations we aim to serve. Furthermore, we will work closely with our patient partners and stakeholder groups to develop tailored dissemination materials to ensure that the findings are accessible and actionable for a wide audience.

Stakeholder Partners: Stakeholder partners (N = ~15) include advocacy groups, clinicians, researchers, industry, health system leaders, and policy makers. Stakeholder partners will meet approximately annually with the research team and patient partners to discuss study progress. They will also provide input on the study and interpretation of study results. The goals of the stakeholder meetings are to garner input from a wide range of multiple perspectives as the study progresses, to promote confidence in the utility of the AI tool if study results warrant an endorsement, and to create a wide network for disseminating results and implementing practice changes.

Participating study sites for data collection: Data collection sites include UCLA, UC San Diego, University of Washington-Seattle, University of Wisconsin-Madison, Boston Medical Center, and University of Miami. The local team at each individual data collection site will vary, but in general will include the site PI(s), one or more clinical champions, an administrative IT staff, a data analyst, a project manager, and a local patient representative. The site PI(s) at each data collection site will be responsible for overseeing implementation of the study procedures and data collection. The local teams will address unique issues that may arise at an individual site. The data managers from each data collection site will meet with the DCC approximately monthly in years 1 and 2 and as needed in years 3-5 to discuss data collection and submission protocols and issues.

Open science

4. Trial registration

Name of trial registry, identifying number (with URL), and date of registration. If not yet registered, name of intended registry.

Name of trial registry:	clinicaltrials.gov

Identifying number and URL:	NCT06934239; https://clinicaltrials.gov/study/NCT06934239
Date of Registration:	April 18, 2025

5. Protocol and statistical analysis plan

Where the trial protocol and statistical analysis plan can be accessed.

The trial protocol and statistical analysis plan can be requested from the Dual PIs.

6. Data sharing

Where and how the individual deidentified participant data (including data dictionary), statistical code, and any other materials will be accessible.

The DCC will ensure reproducibility through well-documented and maintained code and data files that will be accessible via deposit in a PCORI-designated data repository unless otherwise prohibited by law. Information related to the AI intervention, Transpara, including any restrictions in access or re-use of the intervention or its code can be accessed from ScreenPoint, Medical BV, Transpara.

7a, 7b. Funding and conflicts of interest

(7a) Sources of funding and other support (e.g., supply of drugs).

Patient-Centered Outcomes Research Institute (PCORI)	BPS-2024C2-39667 .
California Breast Cancer Research Program of the University of California	B30TR8496
National Institutes of Health/National Cancer Institute	pending

(7b) Financial and other conflicts of interest for principal investigators and steering committee members.

Financial and other conflicts of interest for Principal Investigators:

Dr. Joann Elmore (UCLA)	Serves as Editor-in-Chief for Adult General Medicine topics at UpToDate and receives textbook royalties from Elsevier unrelated to this study.
Dr. Hannah Milch (UCLA)	None
Dr. Haydee Ojeda-Fournier (UCSD)	None
Dr. Janie Lee (University of Washington)	None
Dr. Christoph Lee (University of Wisconsin-Madison)	Reports disclosures unrelated to this study. He receives textbook royalties from UpToDate, Inc., Oxford University Press, and McGraw Hill, Inc.; personal fees from American College of Radiology for journal editorial board work; and clinical advisory fees from RadNet.
Dr. Clare Poynton (Boston Medical Center)	None

Dr. Jose Net (University of Miami)	None
Dr. Diana Miglioretti (UC Davis)	Receives book royalties from Elsevier unrelated to this study and is faculty member in the Radiological Society of North America's Clinical Trial Methodology Workshop and receives an honorarium for her participation
Ms. Katherine Fultz Hollis (patient member of Steering Committee)	None

8. Dissemination policy

Plans to communicate trial results to participants, health care professionals, the public, and other relevant groups (e.g., reporting in trial registry, plain-language summary, publication).

Our Patient Partner Advisory Board and stakeholders will provide support and advice on the dissemination of aggregate study results to all the relevant stakeholders and communities. Multiple communication methods will be used to share study results from the clinical trial with patients and the methods may vary by health system site. Study sites' healthcare systems will be encouraged to publish the PCORI lay language Results Summary as part of their larger communication strategy on any patient-facing websites or newsletters.

Radiologists will not receive their aggregate performance metrics with and without AI during the trial. Each participating site has agreed to not access or view their internal performance data until the last participant is randomized. After trial completion, individual radiologists will receive their interpretive performance metrics with and without AI, shared confidentially. No data identifying individual radiologists or patients will be publicly posted.

Trial results will be posted to ClinicalTrials.gov. Manuscripts will be submitted to peer-reviewed journals to disseminate findings to the broader clinical and academic communities. We will also present findings at major national conferences (e.g., Radiological Society of North America Annual Meeting) to disseminate results to relevant professional audiences.

Introduction

9a, 9b. Background and rationale

9a. Scientific background and rationale, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention.

Computer Aided Detection (CAD) tools for mammography were FDA-cleared in 1998, based on small reader studies, and quickly adopted across the US, reaching 92% of U.S. mammography facilities by 2016. However, our study team's prior work[1] demonstrated that CAD was associated with lower diagnostic accuracy, increased false positives, and higher costs, demonstrating limitations of early adoption without robust clinical validation.[2] Our follow-up study confirmed that historical CAD for mammography was associated with lower diagnostic accuracy, more patient harm, and higher costs in clinical settings.[2] These risks were not evident in pre-market reader studies and indicate a gap in real-world evaluation of emerging technologies.

Today's FDA-cleared AI tools for mammography offer potential improvements in cancer detection and workflow efficiency. Potential harms observed with prior CAD systems warrant evaluation of current AI tools. This pragmatic randomized controlled trial adjudicates this gap in evidence for newer, FDA-cleared artificial intelligence (AI) technology that is already in use internationally and at many U.S. breast imaging facilities.

This trial is intended to inform the interpretation of screening mammography and subsequent patient outcomes, which is pertinent to both radiologists interpreting screening mammograms and patients receiving screening mammograms, by comparing the effectiveness of interpreting screening of digital breast tomosynthesis (DBT) with vs. without AI assistance in improving breast cancer detection with no more than a 1.5 percentage point increase in recall rate. The intended use of the AI intervention in the context of the clinical pathway is to provide radiologists who interpret screening mammograms with decision support at the time of interpreting these mammograms. More specifically, radiologists will be the ones using the AI intervention; the screening mammography exams are randomized after image acquisition and will be interpreted either with or without the AI decision-support tool by the radiologists (see 15a.Intervention and comparator for details). We hypothesize the following:

AI use will be associated with an increase in cancer detection and an initially higher recall rate as radiologists start using AI, followed by a recall rate comparable to that without AI (no more than 1.5 percentage-points higher) after a learning curve period. AI use will be associated with lower rates of missed breast cancers and similar rates of false positives after a learning curve period.

Improved patient outcomes with AI will be most pronounced for exams on women who are White, older, and have less dense breasts, and on baseline exams. AI will aid patient outcomes when the interpretation is by radiologists with less clinical experience, lower annual interpretive volume, and less tolerance of ambiguity, yet greater automation bias (the tendency for humans to defer to a computer algorithms' results) will be noted among these radiologists.

9b. Explanation for choice of comparator.

The comparator to interpretation of screening mammograms by radiologists alone (usual care of mammography interpretation without AI support) was selected for this trial as it is a leading AI tool designed to assist radiologists in interpreting screening mammography. We selected *Transpara*, ScreenPoint Medical's FDA-cleared AI decision support tool for DBT, because it is one of the most widely adopted AI tools in clinical practice. In the U.S., Transpara is used by over 150 institutions, including large imaging center chains, community hospitals, and academic medical centers, and it is in use at more than 100 international sites, including several national screening programs. Transpara combines traditional physics-based modeling with deep learning algorithms to detect suspicious findings on mammograms. Its goal is to assist radiologists in identifying cancers earlier and more accurately, thereby increasing the likelihood of successful treatment and reducing the need for more invasive procedures, ultimately lowering breast cancer–related morbidity.

The Transpara 2.1 version will be used at participating sites early in this trial. ScreenPoint has received FDA clearance for a temporal comparison add on (i.e., Transpara version 2.1 + temporal comparisons). This version represents the most advanced AI tool for mammography currently available in the U.S. and internationally, as it incorporates prior mammograms into the AI assessment process. It allows the AI algorithm to compare the current mammogram with up to 6.5 years of prior images (aligning with how radiologists use priors in their interpretation). Both versions share the same foundational architecture, but the temporal comparison version additionally integrates prior images when generating AI scores. We anticipate that sites will transition to the updated version with temporal comparisons when it becomes commercially available in 2026.

10. Objectives

Specific objectives related to benefits and harms.

The primary objective is to compare the effectiveness of interpreting digital breast tomosynthesis screening exams with vs. without AI assistance in improving breast cancer detection with no more than a 1.5 percentage point increase in recall rate.

Secondary objectives are to evaluate additional screening harms (false-negative rate, false positive recall rate, false positive short interval follow-up recommendation rate, false positive biopsy recommendation rate) and to assess heterogeneity of intervention effects by patient, exam, and radiologist factors. Additionally, assessing trust and confidence in AI via survey and focus group data is a secondary objective.

Methods: patient and public involvement, trial design

11. Patient and public involvement

Details of, or plans for, patient or public involvement in the design, conduct, and reporting of the trial.

Our patient partners will meet virtually with the research team quarterly, providing input on the study's progress. They will be involved in interpreting results and co-authoring papers, to incorporate patient perspectives into dissemination. We will gather input during meetings and via email, using an informal consensus process to ensure all perspectives are considered in final decisions.

Our stakeholder partners will meet virtually with the research team yearly to review study progress, contribute to interpretation of results, and discuss potential for broader dissemination and implementation of the findings. These meetings will garner input from a wide range of perspectives, build confidence in the AI tool (should results warrant endorsement), and establish a network for disseminating the study's outcomes and implementing necessary practice changes.

Patient and Stakeholder input will be sought at key decision points throughout, including design of any patient-facing materials, analysis and interpretation of results, and development of dissemination strategies.

12. Trial design

Description of trial design including type of trial (e.g., parallel group, crossover), allocation ratio, and framework (e.g., superiority, equivalence, noninferiority, exploratory).

This is an interventional, phase IV, parallel-group randomized controlled trial comparing patient-centered outcomes when DBT exams are interpreted with vs. without an FDA-cleared AI tool. Screening mammography exams will be randomized with a 1:1 allocation ratio to be interpreted either with or without an AI decision-support tool. The co-primary outcomes are cancer detection rate and recall rate. We will consider AI to be superior to usual care if the cancer detection rate is significantly higher and the recall rate is non-inferior.

Methods: participants, interventions, and outcomes

13. Trial setting

Settings (e.g., community, hospital) and locations (e.g., countries, sites) where the trial will be conducted.

This trial will be conducted at individual breast imaging facilities that are affiliated with 6 US regional health systems:	
	• University of California, Los Angeles (UCLA) • University of California, San Diego • University of Washington-Seattle • University of Wisconsin-Madison • Boston Medical Center • University of Miami

Study Setting: The onsite requirements needed to integrate the AI intervention into the trial setting include educational training for radiologists who will use the tool, the health systems' integration of the AI tool and Aidoc's randomization platform into their clinical workflows, adequate staffing, the ability to track cancer outcomes, and the ability to conduct focus groups.

We will require all radiologists, including breast imaging fellows, to complete a one-hour training session developed by the UCLA site. This session will include a review of data from the external validation study with case examples of AI false alarms and cases of AI missing cancers. The goal of this training is to encourage appropriate use of the technology and protect against automation bias (the tendency for humans to defer to a computer algorithm result). The pre-RCT training sessions will be recorded and viewed as onboarding for any new faculty hired during the RCT. In addition, ScreenPoint will provide individualized training at each site, tailored to the clinical needs and requests of that site.

All participating institutions must purchase and clinically implement the AI tool. Additionally, all participating institutions must pay for the AI tool with system-level technology use agreements with ScreenPoint and will be responsible for all necessary payments for clinical use of the AI algorithm. No patients will be charged for the use of the AI tool. Study sites will need to use Aidoc, a third-party AI delivery platform, as Aidoc developed the randomization tool that ensures AI information is overlaid on approximately 50% of exams within the existing Picture Archiving Communication Systems (PACS) at the sites. Our study is not being done within a single national chain or health system and thus there are many different internal IT and PACS systems (e.g., Fuji Synapse, Sectra), making our findings more representative of the US. Aidoc can also export all relevant AI data for analysis. These automated capabilities are critical for successful trial execution as manual randomization and data collection is not feasible for 400,000 exams across multiple health systems and imaging facilities.

Additionally, all sites have designated clinical directors that manage individual facilities or groups of imaging facilities. These clinical directors will troubleshoot and oversee day-to-day clinical trial issues on the ground and be key to smooth operations and success.

Each of our health systems has strong medical record systems to track cancer outcomes and, we have budgeted for linkage to the local and state tumor registries. Funds are included to support two waves of focus groups at all sites.

Taken together, this onsite support will allow for successful study implementation.

14a, 14b. Eligibility criteria

14a. Eligibility criteria for participants.

This trial will include all interpreting radiologists and all adult patients undergoing screening mammography at any of the participating breast imaging facilities during the trial period. To be eligible to participate in this study, an individual must meet all the following criteria:	
	1. Be at least 18 years of age or older 2. Receive a screening mammogram at one of the participating breast imaging facilities OR be a radiologist who interprets screening mammograms at one of the participating breast imaging facilities. Exclusion criteria: Patients who have requested that their data not be used for any research purposes will be excluded and their data will not be submitted to the DCC.

All patients who receive a screening mammogram at one of the participating breast imaging facilities—unless they have opted out of all research at the institution—and all radiologists who interpret screening mammograms at these facilities will be included. No screening exams will be excluded based on the nature or quality of their input data (i.e., screening mammogram images). Exams randomized to the AI arm will be included in the analysis even if the AI tool does not fully run for any reason.

14b. If applicable, eligibility criteria for sites and for individuals who will deliver the interventions (e.g., surgeons, physiotherapists).

All interpreting radiologists at any of the participating breast imaging facilities across 6 regional health systems during the trial period will participate in this trial:	
	• University of California, Los Angeles (UCLA) • University of California, San Diego • University of Washington-Seattle • University of Wisconsin-Madison • Boston Medical Center • University of Miami

15a, 15b, 15c, 15d. Intervention and comparator

15a. Intervention and comparator with sufficient details to allow replication including how, when, and by whom they will be administered. If relevant, where additional materials describing the intervention and comparator (e.g., intervention manual) can be accessed.

For usual care arm of exams not randomized to AI, exams will be reviewed and interpreted as done in usual clinical practice (i.e., single read by radiologist alone).

For the AI arm, radiologists will be assisted by Transpara AI (referred to as "AI" in this protocol). The FDA-cleared Transpara 2.1 version will be used at participating sites at study start. ScreenPoint has received FDA clearance for a temporal comparison add on (i.e., Transpara version 2.1 + temporal comparisons). This version allows the AI algorithm to compare the current mammogram with up to 6.5 years of prior images (aligning with how radiologists use priors in their interpretation). Both versions share the same foundational architecture, but the temporal comparison version additionally integrates prior images when generating AI scores. We anticipate that sites will transition to the updated version with temporal comparisons when it becomes commercially available in 2026. We will collect precise data on which Transpara version and functionality is in use at each site, including dates of implementation.

The procedure for acquiring and selecting the screening mammograms (input data) for the AI intervention is as follows: patients who visit any of the participating imaging facilities for a screening mammogram will have their screening mammograms randomly assigned to be interpreted either with or without AI assistance.

There is no human-AI interaction in the handling of the input data (i.e., screening mammogram images) at the time of acquiring the screening mammogram. Randomization occurs within minutes after the breast imaging acquisition (i.e., when the mammography technologist captures the images) by an automated system. Thus, the AI data (or lack thereof) is embedded within the mammogram before the radiologist opens the exam, preventing any option to "add AI" to an exam randomized to be interpreted without AI. Radiologists will be aware of AI availability only at the time of interpretation. The AI information will appear upon opening the exam (e.g., the AI information pops up with the exam images).

The output of the AI intervention is AI information that appears upon opening the exam (for exams randomized to AI). More specifically, for exams randomized to AI, the first image displayed to the radiologist upon opening an exam on the viewing station will be the AI report. This AI report provides the following concise, one-page, standardized information (Described below for Version 1.7):

1) Overall exam risk category: Overall exam-level risk is divided into three categories: elevated, intermediate, or low.

2) Image region scores: Image region markings and scores identify individual areas of interest demarcated with bounding boxes and state relevant slice locations for the 3D exams. The lesion score ranges from 1-100, with 100 denoting the highest level of suspicion**.** A maximum of 8 markings may be present for a given exam (2 markings per each of the 4 standard views).

After viewing the report, the radiologist will interpret the exam; any specific lesion bounding boxes flagged by AI will be visible while interpreting the exam. The radiologist can toggle these markings on and off as needed and will have complete autonomy regarding the final interpretation of the exam as

positive or negative, meaning that they can choose to ignore the AI information. The AI intervention's output can contribute to radiologists' decision making, if desired, as they will have access to the AI tool's interpretation of screening mammograms, for exams randomized to AI.

15b. Criteria for discontinuing or modifying allocated intervention/comparator for a trial participant (e.g., drug dose change in response to harms, participant request, or improving/worsening disease).

N/A

15c. Strategies to improve adherence to intervention/comparator protocols, if applicable, and any procedures for monitoring adherence (e.g., drug tablet return, sessions attended).

Each data collection site has agreed to participating in the trial at the health system level. Each site will have their own dedicated site PI(s), IT and data support staff to ensure that they are following this protocol. The site PI will meet regularly with the IT/data teams to ensure that the study protocol is being followed and that the infrastructure setup allows for seamless data integration. Data collection sites will upload monthly enrollment reports indicating how many screening exams were randomized to the AI and usual care arms and how many randomized patients opted out of having their data used for research. The DCC will monitor these reports to ensure the 1:1 randomization is being properly implemented.

Procedures for assessing and handling poor-quality or unavailable input data: The input data for this trial are the screening mammogram images and the AI report that is shown to radiologists. We do not anticipate issues with image quality or data availability. There may be extremely rare exams deemed "poor quality" and the patients are asked to return and the images are placed under the same accession number. However, in rare cases, the AI scoring may fail to process a screening mammogram. If this occurs, the exam will still be included in the analysis. The frequency of AI scoring failures will be documented and reported, if feasible.

15d. Concomitant care that is permitted or prohibited during the trial.

N/A

16. Outcomes

Primary and secondary outcomes, including the specific measurement variable (e.g., systolic blood pressure), analysis metric (e.g., change from baseline, final value, time to event), method of aggregation (e.g., median, proportion), and time point for each outcome.

Primary Outcomes:		
1.	Outcome name:	Cancer detection rate
	Metric/method of measurement:	Number of screening exams recommended for breast biopsy (final BI-RADS assessment of 4 or 5) resulting in detected cancer, per 1,000 screening exams

	Timepoint:	Cancer diagnosed within 90 days of positive study entry screening mammogram
2.	Outcome name:	Recall rate
	Metric/method of measurement:	Number of screening exams recalled for diagnostic work-up (initial BI-RADS assessment of 0, 3, 4, or 5), per 1,000 screening exams
	Timepoint:	Day of interpretation

Key Secondary Outcomes:		
1.	Outcome name:	Interval cancer rate (i.e., false-negative rate)
	Metric/method of measurement:	Number of screening exams with a negative assessment (final BI-RADS assessment of 1 or 2), and breast cancer diagnosed within 1 year, per 1,000 screening exams
	Timepoint:	Cancer diagnosed within 365 days of a negative study entry screening mammogram
2.	Outcome name:	False positive recall rate
	Metric/method of measurement:	Proportion of screening exams recalled for additional imaging (final BI-RADS assessment of 1, 2, or 3), with no breast cancer diagnosed within 1 year
	Timepoint:	No cancer diagnosed within 365 days of a positive study entry screening mammogram
3.	Outcome name:	False positive short-interval follow-up recommendation rate
	Metric/method of measurement:	Proportion of screening exams recalled for short-interval follow-up (final BI-RADS assessment of 3) with no breast cancer diagnosed within 1 year
	Timepoint:	No cancer diagnosed within 365 days of a positive study entry screening mammogram
4.	Outcome name:	False positive biopsy recommendation rate
	Metric/method of measurement:	Proportion of screening exams recalled for breast biopsy (final BI-RADS assessment of 4 or 5) with no breast cancer diagnosed within 1 year
	Timepoint:	No cancer diagnosed within 365 days of a positive study entry screening mammogram
5.	Outcome name:	Trust and confidence in AI

	Metric/method of measurement:	Focus group and survey data
	Timepoint:	Years 1,2 and Years 4,5
6.	Outcome name:	Efficiency metrics (only for the UCLA site)
	Metric/method of measurement:	Interpretation time required for radiologists to interpret each mammogram with versus without AI Delivery time, using time stamp data from exam acquisition to delivery of results to patients (aka turnaround time)
	Timepoint:	Timepoints as described above

17. Harms

How harms are defined and will be assessed (e.g., systematically, nonsystematically).

Harms are defined, consistent with UCLA's Office of Human Research Protection Program, as an adverse event as "any untoward or unfavorable medical occurrence in a human subject (physical or psychological harm) temporally associated with the subject's participation in the research (whether or not related to participation in the research".

In this study, the primary potential harm from screening mammography is being recalled for additional diagnostic work-up, which may lead to anxiety, unnecessary imaging, or invasive procedures. Therefore, recall rate will be tracked as the key adverse event. While screening mammograms will be obtained as part of routine clinical care, we will not systematically track other downstream outcomes such as biopsy complications or deaths, as doing so would require considerable additional effort.

Harms will be assessed systematically through ongoing monitoring of recall rates across trial arms. Recall rates will be reviewed by the DSMB through interim data reports every 6 months to evaluate participant safety. An interim analysis will be conducted once half of the exams have been randomized. Based on this analysis and other available information, the DSMB may recommend trial continuation, modification, or early termination. For example, if the recall rate in the AI arm is significantly inferior without a corresponding improvement in the cancer detection rate, this may warrant DSMB action.

We will also evaluate interval cancer rates and false positive rates as harms, but these outcomes cannot be estimated until after linkages are performed with state or regional cancer registries.

18. Participant timeline

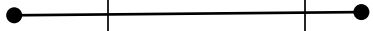
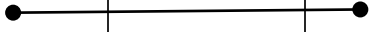
Time schedule of enrollment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure).

The RCT will continue for approximately two years, longer at the UCLA site, during which the AI will be randomized to be turned on or off (1:1) at the exam level (See Figure).

Figure. Participant timeline.

Timepoint	Day 1	Interim Days	Day 90	Day 365
Visit to Imaging Facility for Screening Mammogram	X			
Completion of Clinical Intake Survey (as part of routine clinical care)	X			
Follow Patient to Collect Breast Cancer Outcome Data for Cancer Detection Rate				
Follow Patient to Collect Breast Cancer Outcome Data for Interval Cancer Rates, False-Positive Rates, and Cancer Characteristics				

Figure. Radiologist Timeline

Timepoint	Day 1	Interim Days	Day ~730	Closeout
Radiologist Baseline Survey	X			
Interpretation of screening mammograms from both randomization arms:				
AI arm				
Usual Care arm				
Radiologist Follow-Up Survey				X

19. Sample size

How sample size was determined, including all assumptions supporting the sample size calculation.

Our target sample size of 400,000 screening exams was chosen to ensure sufficient power for evaluating our primary and secondary outcomes and for assessing heterogeneity of effects in subgroups consisting of at least 25% of the study population. Power calculations for clinical screening outcomes (Table 1) are based on baseline outcome estimates from the participating sites and national benchmarks for 3D screening mammography.[3] To account for the interim analysis, power calculations for cancer detection rate assume a two-sided alpha of 0.0492 and for the non-inferiority test for recall rate, a one-sided alpha of 0.0246. We did not adjust alpha for two primary outcomes, as we will require "success" on both primary outcomes to consider interpretation with AI to be superior to that without AI, i.e., we will conclude that interpretation with AI improves screening outcomes if both the cancer detection rate is significantly higher and the recall rate increases by no more than 1.5 percentage-points with AI-assistance. This non-inferiority margin was chosen based on expert opinion of the increase in recall rate that would be a reasonable tradeoff for an increased cancer detection rate. Power calculations assume outcomes with and without AI are independent, because this intervention varies both within and between clusters, making it impossible to predict the effect of clustering on variance estimates.[4] Detectable differences were calculated using PASS software for a Z-test of the difference between two proportions with un-pooled variance.

Table 1. Detectable differences with 80% power (and 99% power for recall rate) for overall sample of 400,000 mammograms and for subgroups comprising 50%, 25%, and 11% of the sample. Expected values without AI are based on baseline outcomes from participating sites and published 3D screening mammography benchmarks (CI Lee et al, Radiology, 2023).

	Expected value without AI	Detectable with AI in full study sample		Subgroups comprising 50% of sample		Subgroups comprising 25% of sample		Subgroups comprising 11% of sample	
Screening outcome		Value	RR	Value	RR	Value	RR	Value	RR
Cancer detection rate	0.58%	0.65%	1.12	0.68%	1.17	0.72%	1.25	0.80%	1.38
Recall rate 99% power	10.10%	9.70%	0.96	9.53%	0.94	9.30%	0.92	8.91%	0.88
Recall rate 80% power	10.10%	9.83%	0.97	9.73%	0.96	9.57%	0.95	9.32%	0.92
False negative rate	0.083%	0.059%	0.72	0.051%	0.61	0.039%	0.47	0.022%	0.27
False-positive (FP) recall	9.50%	9.76%	1.03	9.87%	1.04	10.03%	1.09	10.29%	1.07
FP short-interval follow-up	1.60%	1.71%	1.07	1.76%	1.10	1.83%	1.14	1.95%	1.22
FP biopsy recommendation	0.98%	1.07%	1.09	1.11%	1.13	1.16%	1.19	1.26%	1.28

For cancer detection rate, we expect 80% power to detect an increase from 5.8/1000 without AI to 6.5/1000 with AI (RR=1.12). With a non-inferiority margin of 1.5% for recall rate, we expect 90% power to reject the null hypothesis that AI-assisted interpretation is inferior if the recall rate with AI is increases by ≤1.2 percentage points from 10.1% to ≤11.3%. If recall rate is found to be non-inferior with AI assistance, we will test for superiority. We will have >99% power to detect a decrease in recall rate of ≥0.41 percentage-points (RR≤0.96) and 80% power to detect a decrease of 0.27 percentage-points (RR=0.97).

For our interim analysis with 200,000 exams, we will have >99% power to detect an absolute difference in recall rate of 1.5 percentage points and 80% power to detect an absolute difference in cancer detection rate of 1.3/1000 (RR=1.22) with a two-sided alpha of 0.0492.

For secondary outcomes, we expect 80% power to detect a decrease in interval cancer rate from 0.83/1000 to 0.59/1000 (RR=0.72), an increase from 9.5% to 9.76% for false positive recalls, an increase

from 1.6% to 1.7% for false positive short-interval follow-up recommendations, and an increase from 0.98% to 1.07% for false positive biopsy recommendations.

For evaluating heterogeneity of treatment effects (the), power calculations were conducted for subgroups comprising 11%, 25%, and 50% of the study sample, assuming a two-sided alpha of 0.05. We expect sufficient power to detect clinically meaningful differences for most outcomes, with detectable RRs ranging from 1.17-1.38. For recall rate, detectable RRs range from 0.92-0.96 with 80% power, and we have 99% to detect absolute differences of 0.57-1.19 percentage points. For secondary outcomes, we have 80% power to detect small increases in false positive recalls (absolute differences <1 percentage point; RRs 1.04-1.28). Power is most limited for interval cancer rate, for which we have 80% power to detect a RR of 0.27-0.61. We may have small numbers for specific woman-level or radiologist-level characteristics (e.g., extremely dense breasts occur in <10% of the population). In these cases, we will group exams into larger binary categories for the purposes of secondary analyses (e.g., women with the 2 highest vs. 2 lowest density categories). Analyses of smaller subgroups will be exploratory.

20. Recruitment

Strategies for achieving adequate participant enrollment to reach target sample size.

Our recruitment strategy will ensure a representative sample from participating health systems. We are targeting 400,000 screening exams across regional health systems in five states (WA, CA, MA, FL, WI) that provide a range of clinical settings, from academic medical centers to rural community screening facilities. Recruitment estimates were generated based on each site's annual screening volumes and prior experience in similar studies. Feasibility of conducting the study at these sites was affirmed via discussions with site leaders, administrators and clinicians. Participating radiologists expressed willingness to participate in implementation of the AI tool. Sites have agreed to participation and recruitment targets based on feasibility assessments.

Methods: assignment of interventions

Randomization

21a, 21b. Sequence generation

21a. Who will generate the random allocation sequence and the method used.

The Aidoc platform will randomize each exam at the time of image acquisition (i.e., when the mammography technologist captures the images), using a random allocation sequence.

21b. Type of randomization (simple or restricted) and details of any factors for stratification. To reduce predictability of a random sequence, other details of any planned restriction (e.g., blocking) should be provided in a separate document that is unavailable to those who enroll participants or assign interventions.

Exams will be randomized 1:1 at each participating facility using simple randomization.

22. Allocation concealment mechanism

Mechanism used to implement the random allocation sequence (e.g., central computer/telephone; sequentially numbered, opaque, sealed containers), describing any steps to conceal the sequence until interventions are assigned.

The Aidoc platform will be used to generate the random allocation sequence. Due to the automated nature of randomization that occurs immediately after imaging exam acquisition, *no one* will know in advance which patients will be randomized to the AI arm.

23. Implementation

Whether the personnel who will enroll and those who will assign participants to the interventions will have access to the random allocation sequence.

Study personnel will not have access to the random allocation sequence.

24a, 24b, 24c. Blinding

24a. Who will be blinded after assignment to interventions (e.g., participants, care providers, outcome assessors, data analysts).

Radiology department staff, including schedulers and clinical support staff, and patients will remain blinded throughout the trial. Patients will not be informed of the use of AI (per our UCLA IRB) because the AI program is an FDA-cleared support tool currently in use within the US. However, radiologists will not be blinded to study arm at the time of screening mammography interpretation because they will be shown the AI information, if present, as a support tool during their interpretation. Radiology department staff, including schedulers and clinical support staff, and patients will remain blinded throughout the trial. A strength of our randomization scheme is that we will avoid bias due to potential confounding factors (e.g., temporal trends and learning over time). Randomizing AI by exam level will best represent actual clinical practice patterns and not disrupt routine workflow.

24b. If blinded, how blinding will be achieved and description of the similarity of interventions.

Only the interpreting radiologist will know about assignment to intervention upon interpreting the screening mammogram. Whether or not AI was used will not be provided in the mammography report; thus, patients and their clinicians will not know to which arm patients were randomized.

24c. If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial.

Unblinding is not expected but may occur if the DSMB, who will evaluate interim data reports every 6 months for participant safety, study conduct and progress, make recommendations concerning trial modification.

Methods: data collection, management, and analysis

25a, 25b. Data collection methods

25a. Plans for assessment and collection of trial data, including any related processes to promote data quality (e.g., duplicate measurements, training of assessors) and a description of trial instruments (e.g., questionnaires, laboratory tests), along with their reliability and validity, if known. Reference to where data collection forms can be accessed, if not in the protocol.

We will collect data at the patient, exam, radiologist, and health system level. In addition to data collected during the trial enrollment period, a minimum of one year of baseline breast imaging data will also be obtained. A separate file will be created by site study staff that contains the exam ID, randomized arm, and AI scores. These files will be submitted to and linked by the data coordinating center (DCC).

For the collection of patient-level data, patients complete a questionnaire at most of the imaging facilities as part of routine clinical practice at each screening exam. They self-report their demographic information and complete an assessment to calculate their level of breast cancer risk. Participating sites will access the imaging facility data and EHR data to obtain clinical history and risk factors (e.g., breast procedures, breast cancer family history, menopausal status, use of hormone therapy, history of breast cancer). We will capture breast cancer diagnoses from multiple sources: local pathology labs, local health system cancer registries, and state/regional SEER and North American Association of Central Cancer Registries (NAACCR) cancer registry linkages.

Exam-level data will be collected from the EHR including radiologists' assessment, recommendations, and BI-RADS breast density recorded as part of routine clinical practice using the American College of Radiology (ACR) standardized definitions at each exam. We will also capture information on all diagnostic imaging work-up exams and biopsies following the study-entry screening mammogram. For each screening, we will collect whether it is a baseline exam (first ever) versus subsequent exam and time since last mammogram based on self-report information in combination with examinations included in the EMR. Mammography manufacturer and model will be collected at the exam level for participating facilities where this information is available; otherwise, it will be collected at the facility level.

Standardized information on radiologists will be collected via a self-report survey prior to interpreting mammograms with AI, and again at the end of the RCT. These surveys will gather information on breast imaging fellowship training, years of experience, fear of malpractice litigation, tolerance of ambiguity in clinical decision-making, and perceptions about AI and its impact on clinical workflow efficiency.[6] We will also collect radiologist-level screening volume.

Facility and health system level data will include academic vs community affiliation, imaging machine type(s), PACS system(s), payor mix and geographic location. We will also collect precise data on which Transpara version and functionality is in use at each site, including dates of implementation.

Breast cancer outcomes and characteristics will be collected through linkages with Surveillance, Epidemiology, and End Results (SEER) registries, state cancer registries, local cancer registries, and pathology laboratories. We will collect data on the following breast cancer characteristics for exploratory analyses: histology (e.g., ductal carcinoma in situ [DCIS], invasive ductal carcinoma, invasive

lobular carcinoma, inflammatory carcinoma); AJCC stage (8th edition, including prognostic and anatomic stages); SEER summary stage; lymph node status; tumor size; grade; and hormone receptor status (e.g., estrogen receptor, progesterone receptor, HER2/neu).

25b. Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols.

Screening mammography exams are one time data points; thus, we do not discuss retention or the possibility of patient participants discontinuing. Follow-up data on all patients will be linked to local and regional or state tumor registries to capture breast cancer outcomes.

26. Data management

Plans for data entry, coding, security, and storage, including any related processes to promote data quality (e.g., double data entry; range checks for data values). Reference to where details of data management procedures can be accessed, if not in the protocol.

The Data Coordinating Center (DCC) at UC Davis will serve as the central hub for data management and analysis. The DCC will collect data locally captured at each participating site and maintain it on secure computers and servers with safeguards to protect confidentiality and prevent unauthorized access, in accordance with UC Davis Health System protocols.

Data will be collected, managed, and overseen by a data scientist and a statistician from each site (e.g., via the CTSI group). Only the data scientist will have access to PHI. Data will be housed in a secure storage folder at each site (e.g., overseen by CTSI). All data will be locked until data collection is complete. All non-image data will be de-identified and sent to UC Davis for analysis. Image data will stay at each site in this secure folder for future research and quality assurance purposes.

Data collection sites will perform their own cancer registry linkages, with support from the DCC as needed. De-identified exam-level data—including patient characteristics, radiologists' interpretations, AI results, cancer outcomes, and radiologist survey data—will be submitted to the DCC for centralized processing, validation, and analysis. The DCC will provide a uniform database format, data dictionary, and study manual to ensure consistent standards and definitions across data collection sites.

Participating sites will submit to the DCC aggregate enrollment data monthly and exam-level study data at least quarterly using REDCap (Research Electronic Data Capture), a secure, web-based software platform for managing research and clinical trial data. UC Davis hosts a REDCap instance on servers that meet UC Davis Health System security requirements. REDCap offers an intuitive user interface, real-time validation rules (e.g., data type and range checks), audit trails, and automated exports to statistical software (e.g., SAS, R). The DCC will clean, verify, and pool the data and quality assurance. DCC staff will review submissions for completeness, quality, and plausibility, and resolve any data issues.

Although variables will be uniformly defined, each site will develop its own approach to populating the REDCap database based on local information systems. The preferred method will be to export data in a standard format using the national mammography database (NMD) export feature available in all mammography systems. These data will be supplemented with other study data collected from the EHR and surveys.

27a, 27b, 27c, 27d. Statistical methods

27a. Statistical methods used to compare groups for primary and secondary outcomes, including harms.

All analyses will be pre-specified in the Statistical Analysis Plan. We will fit linear probability models to estimate absolute differences and modified Poisson regression models to estimate relative risks comparing screening outcomes with vs. without AI.[7-9] We have chosen these methods because some outcomes are not rare, such as recall rates, rendering odds ratios from logistic regression models difficult to interpret. Models will be estimated using a multi-step generalized estimating equation (GEE) approach to account for non-nested clustering within patient, radiologist, and facility levels [4,10]. We will report unadjusted comparisons and, as a sensitivity analysis, comparisons adjusting for temporal trends and important patient-, exam-, and radiologist-level factors; however, we do not expect adjustment to significantly alter our results due to our randomization scheme controlling for these factors. For the non-inferiority test of recall rate, the null hypothesis is that interpretation with AI is inferior to interpretation without AI. We will reject the null hypothesis of inferiority if the upper bound of the 95% confidence interval for the absolute difference in recall rates (recall rate with AI minus recall rate without AI) does not exceed the non-inferiority margin of 1.5 percentage points.

Interim analysis: Given concerns that use of AI in mammography could decrease cancer detection rates (due to automation bias) and increase recall rates to unacceptable levels, we will conduct a formal interim analysis of our two primary outcomes once half of the screens (N=200,000) have been randomized using the O'Brian-Fleming approach with an alpha of 0.0054. This analysis will be presented to the DSMB in closed session.

27b. Definition of who will be included in each analysis (eg, all randomized participants), and in which group.

All randomized exams (i.e., exams interpreted with versus without AI assistance) fulfilling the eligibility criteria will be included in the analyses.

27c. How missing data will be handled in the analysis.

We will minimize the amount of missing data for core study variables through intensive quality assurance. The DCC will review data for potential data issues and work with sites to abstract missing data, correct data issues, and document when missing data cannot be retrieved. Despite these rigorous efforts, we may encounter some missing data, though we expect this to be rare. If <5% of data are missing, we will conduct complete case analysis. Otherwise, we will use multiple imputation methods and compare these results to complete case analysis. We will report and compare the characteristics of screening exams with and without missing data.

27d. Methods for any additional analyses (eg, subgroup and sensitivity analyses).

Learning curve: We will evaluate whether there is a "learning curve" as radiologists gain experience using the AI tool. We will evaluate changes in screening outcomes as AI use increases over time, including an interaction between time and study arm. We will assess whether screening outcomes change as a function of volume of exams interpreted with AI using modified Poisson regression including interpreting radiologist as a fixed effect and an interaction between volume and study arm to estimate the within-radiologists changes in screening outcomes with increased AI experience over the trial period.

We will also assess whether the learning curve differs for breast imaging specialists vs. general radiologists or for radiologists with high, average, or low baseline recall rates. If a learning curve is identified, we will adjust our comparative analyses by for volume.

Heterogeneity of intervention effects: We will evaluate whether the effectiveness of AI depends on patient, exam, or radiologist characteristics or on AI version. To contrast differences in effects by subgroups, we will include relevant interaction terms in the regression models and report magnitudes of effects and confidence intervals for each subgroup level along with p-values for tests of interaction. We will also evaluate whether the effectiveness of AI differs by AI version by including indicator variables for AI version in the regression models. We will compare each AI version to the control group and will compare the AI versions to each other, adjusting for temporal trends.

Exploratory analyses: Cancer characteristics will be evaluated by study arm, overall and separately for screen-detected and interval cancers. We will compare the distribution of cancer characteristics among women diagnosed with cancer (across arms [with vs. without AI]) and calculate rates of cancer characteristics per 1,000 screening mammograms among all exams. Cancer characteristics will be classified as having favorable vs. unfavorable prognosis based on definitions commonly used by the Breast Cancer Surveillance Consortium and others.[11] For example, stage will be categorized as advanced using a previously published definition based on AJCC prognostic stage II or higher.[12] Rates of cancer diagnosis with favorable prognosis and rates of cancer diagnosis with unfavorable diagnosis (among all screening mammograms) will be separately modeled and compared across study arms using logistic regression estimated via a multi-step generalized estimating equation (GEE) approach to account for clustering at the patient, radiologist, and facility levels.[13]

Efficiency Metrics: We will also assess interpretation and reporting time with AI versus without using data from UCLA. We hypothesize that radiologists reading times will be slower during the early implementation of AI, until they adapt and optimize their workflow when using the AI tool to support their interpretation of mammography. We also hypothesize that in the later steady state, radiologists will interpret mammograms 20% faster and deliver results to patients 20% faster with AI compared to without AI.

For analyses of efficiency metrics, we will analyze time required for radiologists to interpret each mammogram with versus without AI, a crucial metric in understanding any enhancement in radiologists' efficiency. We will also measure delivery time, using time stamp data from exam acquisition to delivery of results to patients (aka turnaround time), providing insights into efficiency at a health system level that are relevant for the longer-term use of AI.

Efficiency metrics power calculations**.** We expect an average interpretation time of 5 minutes with a standard deviation of 10 minutes, as more difficult mammograms can take longer to interpret. With 50% of >90,000 mammograms at just the UCLA site randomly assigned to receive AI assistance, we will have >99% power to detect a difference of 18 seconds in the average interpretation time (efficiency) of mammograms between those with and without AI. Larger observed differences in mean interpretation time will generally yield greater power and regression adjustment should reduce the variance associated with our estimates and yield even greater power.

Our large sample size will allow for well-powered subgroup analyses whereby we can explore the efficiency of this technique for cases that are traditionally harder for radiologists to interpret (i.e., dense breast tissue). We would have >99% power to detect a difference of 33 seconds in the average

interpretation time (efficiency) of mammograms between those with and without AI. Assuming a prevalence of 7.4% extremely dense breast tissue, we would have >99% power to detect a difference of 66 seconds in the average interpretation time (efficiency) between those with and without AI. These power calculations apply to the UCLA site, but these analyses may expand to the other sites.

Methods: monitoring

28a, 28b. Data monitoring committee

28a. Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and funder; conflicts of interest and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed.

Our Data and Safety Monitoring Board (DSMB) will be chaired by Dr. Ruth Carlos, who leads the ECOG-ACRIN Cancer Care Delivery Research Committee and include 4 other individuals, including experts in cancer screening, RCT design, and biostatistics, and one patient partner. The DSMB will evaluate interim data reports every 6 months for participant safety, study conduct and progress, and make recommendations concerning trial continuation, modification, or termination.

The DSMB is independent from the sponsor and funder.

Conflicts of interest for DSMB members are as follows:

Dr. Ruth Carlos reports an honorarium from Canon Medical. She also reports board membership and leadership at Academy of Radiology and Biomedical Imaging Research, ECOG ACRIN, and Association of Academic Radiologists/CERRAF. She receives travel reimbursement from the aforementioned organizations due to leadership roles (reimbursed to self). Lastly, she reports salary support paid to institution for role as editor in chief of the Journal of the American College of Radiology (JACR).

Dr. Solveig Hofvind reports no conflicts of interest.

Dr. Robert Smith reports no conflicts of interest.

Dr. Chaya Moskowitz reports that she is a faculty member in the Radiological Society of North America's Clinical Trial Methodology Workshop and receives an honorarium for her participation. She has grants from the NIH, paid to her institution, unrelated to this study.

Ms. Julie Nesbit, a patient representative, reports no conflicts of interest.

The DSM charter will be available from the study PIs upon request.

28b. Explanation of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial.

An interim data analysis will be conducted once half of exams are randomized. Based on these results and other available information, the DSMB may advise the PIs to stop the trial if the recall rate in the AI arm is significantly inferior without evidence of an improvement in the cancer detection rate. The Executive Committee will make the final decision to terminate the trial.

29. Trial monitoring

Frequency and procedures for monitoring trial conduct. If there is no monitoring, give explanation.

Data collection sites will upload monthly aggregate counts of exams randomized to the AI and usual care arm to a REDcap database developed and maintained by the DCC. Exam-level study data will be submitted to the DCC at least quarterly for data quality checks and monitoring. The DSMB will evaluate interim data reports every 6 months for participant safety, study conduct and progress, and make recommendations concerning trial continuation, modification, or termination.

Ethics

30. Research ethics approval

Plans for seeking research ethics committee/institutional review board approval.

We have received Institutional Review Board (IRB) approval from UCLA for a waiver of consent to enroll women, link datasets, and perform data analyses. The participating sites will defer to UCLA's single IRB. All procedures are Health Insurance Portability and Accountability Act (HIPAA) compliant, and we will protect identities of women, radiologists, and imaging facilities. Patients will not be informed of the use of AI, due to the lack of feasibility for this high volume of exams and because the study team will not have access to strong patient identifiers. Instead, study teams will work with their local informatics group to collect and deidentify data (at UCLA, for example, data will be collected and housed by the Clinical and Translational Science Institute/CTSI). In addition, the AI program is an FDA-cleared support tool currently in use within the US. In clinical settings, patients undergoing medical care and imaging are frequently unaware that computer-aided diagnostic tools are being utilized as part of their care, and use of these computer tools has become a standard aspect of many diagnostic imaging practices. The AI tool functions as an auxiliary support, and the final interpretation of the mammogram remains under the purview of radiologists, who may choose to consider or disregard the AI's input.

31. Protocol amendments

Plans for communicating important protocol modifications to relevant parties.

We will communicate important protocol modifications to all relevant parties (e.g., sponsor, funder, IRB, participating sites) using multiple methods including written email, zoom meetings and telephone calls as needed.

32a, 32b. Consent or assent

(32a) Who will obtain informed consent or assent from potential trial participants or authorized proxies, and how.

As noted in Section 30 above, we have received IRB approval from UCLA for a waiver of consent to enroll women, link datasets, and perform data analyses.

All eligible screening mammography exams will be included in all analyses except for the sub-analyses using radiologists' characteristics data (which are collected via radiologist surveys); radiologists will provide informed consent to have their survey data and professional characteristics used in these sub-analyses. This informed consent will be obtained from radiologists at the beginning of the study when radiologists receive the baseline survey.

(32b) Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable.

N/A

33. Confidentiality

How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial.

All research team members will abide by standards of confidentiality. No data will be available in reports or publications that could lead to the identification of individual patients, radiologists, or health systems. We will establish and maintain the appropriate administrative, technical, and physical safeguards to protect the confidentiality of the data and to prevent unauthorized use or access to it. All data management activities will take place using established data management systems with appropriate safeguards to protect the confidentiality of the data and to prevent unauthorized use or access to the data. All computers storing data will be encrypted. Secured access to computers, databases, and network domains is maintained using an isolated EtherNet Local Area Network and strong passwords. The Principal Investigators, collaborators, and research staff will not use the data for purposes other than described in our proposal. The research team will not use the data to identify any individual patient, physician, or facility.

After trial completion, individual radiologists will receive their interpretive performance metrics with and without AI, shared confidentially. No data identifying individual radiologists or patients will be publicly posted.

Multiple communication methods will be used to share study results from the clinical trial with patients and the methods may vary by health system site. This communication may be accomplished by patient portal e-messages, health system newsletters highlighting study findings, and dedicated pages on health system websites. Some breast imaging facilities and sites might want to have posters on the walls or paper handouts describing the findings. We will also use scientific dissemination approaches such as presentations/posters at national meetings and publications, including our patient partners as co-authors and co-presenters. The results that are shared will be done in accordance with institution and/or state's guidelines and laws.

34. Ancillary and post-trial care

Provisions, if any, for ancillary and posttrial care and for compensation to those who suffer harm from trial participation.

Participants will receive screening mammography as part of routine clinical care. The study involves minimal risk, as it evaluates the use of Transpara—an FDA-cleared AI decision support tool for screening mammography—without altering clinical management pathways. Because the intervention occurs within standard care and no investigational procedures are involved, no specific ancillary or posttrial care provisions are necessary. As is standard, participants who experience harm attributable to routine clinical care will be managed according to institutional policies; no additional compensation is planned for trial-related harm given the low-risk nature of the study.

References

1. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, Berns EA, Cutter G, Hendrick RE, Barlow WE, Elmore JG. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med.* 2007;356(14):1399-1409. PMC3182841.
2. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL, Breast Cancer Surveillance C. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Intern Med.* 2015;175(11):1828-1837. PMC4836172.
3. Lee CI, Abraham L, Miglioretti DL, Onega T, Kerlikowske K, Lee JM, Sprague BL, Tosteson ANA, Rauscher GH, Bowles EJA, diFlorio-Alexander RM, Henderson LM, Breast Cancer Surveillance C. National Performance Benchmarks for Screening Digital Breast Tomosynthesis: Update from the Breast Cancer Surveillance Consortium. *Radiology.* 2023;307(4):e222499. PMC10323294.
4. Miglioretti DL, Heagerty PJ. Marginal modeling of multilevel binary data with time-varying covariates. *Biostatistics.* 2004;5(3):381-398. PMID: 15208201.
5. Gerrity MS, White KP, Devellis RF, Dittus RS. Physicians' Reactions to Uncertainty: Refining the constructs and scales. *Motivation and Emotion.* 1995;19:175-191.
6. Shinners L, Aggar C, Grace S, Smith S. Exploring healthcare professionals' perceptions of artificial intelligence: Validating a questionnaire using the e-Delphi method. *Digit Health.* 2021;7:20552076211003433. PMC7995296.
7. Huang FL. Alternatives to logistic regression models in experimental studies. *The Journal of Experimental Education.* 2022;90(1):213-228.
8. Zou GY, Donner A. Extension of the modified Poisson regression model to prospective studies with correlated binary data. *Stat Methods Med Res.* 2013;22(6):661-670. PMID: 22072596.
9. Yelland LN, Salter AB, Ryan P. Performance of the modified Poisson regression approach for estimating relative risks from clustered prospective data. *Am J Epidemiol.* 2011;174(8):984-992. PMID: 21841157.
10. Miglioretti DL, Heagerty PJ. Marginal modeling of nonnested multilevel data using standard software. *Am J Epidemiol.* 2007;165(4):453-463. PMID: 17121864.
11. National Cancer Institute (NCI), Patient-Centered Outcomes Research Institute (PCORI). *Breast Cancer Surveillance Consortium (BCSC) Standard Definitions, Version 4.* February 10 2025. https://www.bcsc-research.org/data/bcsc_standard_definitions
12. Kerlikowske K, Bissell MCS, Sprague BL, Buist DSM, Henderson LM, Lee JM, Miglioretti DL. Advanced Breast Cancer Definitions by Staging System Examined in the Breast Cancer Surveillance Consortium. *J Natl Cancer Inst.* 2021;113(7):909-916. PMC8491791.
13. Henderson LM, Miglioretti DL, Kerlikowske K, Wernli KJ, Sprague BL, Lehman CD. Breast Cancer Characteristics Associated With Digital Versus Film-Screen Mammography for Screen-Detected and Interval Cancers. *AJR Am J Roentgenol.* 2015;205(3):676-684. PMC4547477.