EQUITABLE IMPACT OF AN AI-DRIVEN BREAST CANCER SCREENING WORKFLOW IN REAL-WORLD US-WIDE DEPLOYMENT

Leeann D. Louis, Edgar A. Wakelin M, Matthew P. McCabe, Annie Y. Ng, Jiye G. Kim, Christoph I. Lee, Diana S. M. Buist, A. Gregory Sorensen & Bryan Haslam

Mahidol, CEB, Journal Club

23.01.2026

original paper published in nature health on 17th November 2025

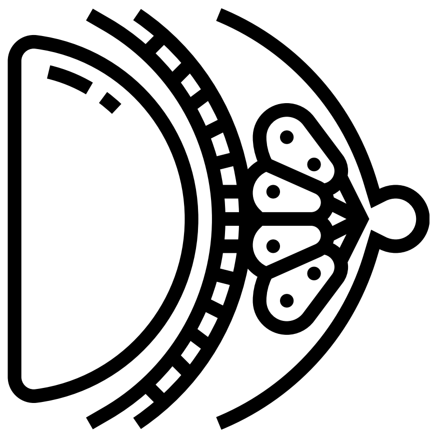
presented by Unt Ko

1. Breast density and risk of breast cancer
2. Health and Racial Disparity in Breast Cancer

Introduction

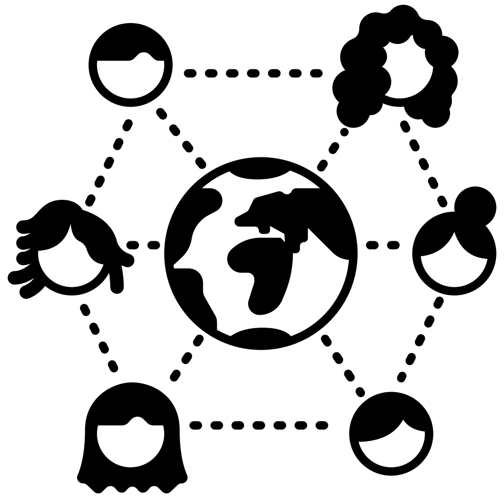
Persistent disparities create an urgent need for better screening

challenge of breast density



- Women with dense breast tissue face a **higher risk** of cancer.
- Dense tissue can mask cancer lesions in mammograms, leading to **missed diagnosis**[1].

challenge of racial inequity



- Black women in the U.S. experience significantly **higher** breast cancer **mortality** despite lower incidence rates compared to white women[2].
- This disparity is linked to systemic barriers, reduced access, and delayed diagnoses.

Research Problem

- Breast cancer screening in USA is unique (compared to EU) with a diverse range of population including increased risk groups*, **requiring scalable and equitable impact approach** in adapting AI integrated workflows.
- Artificial intelligence (AI) shows promising results for improving early breast cancer detection and overall screening outcomes, particularly in European studies.

*Black women in USA experience significantly higher breast cancer mortality.

Objectives

- Integrating AI workflows in breast cancer screening program tailored to USA clinical practice and evaluate the approach for **scalability** and **equity**.

Unique U.S. Screening paradigm

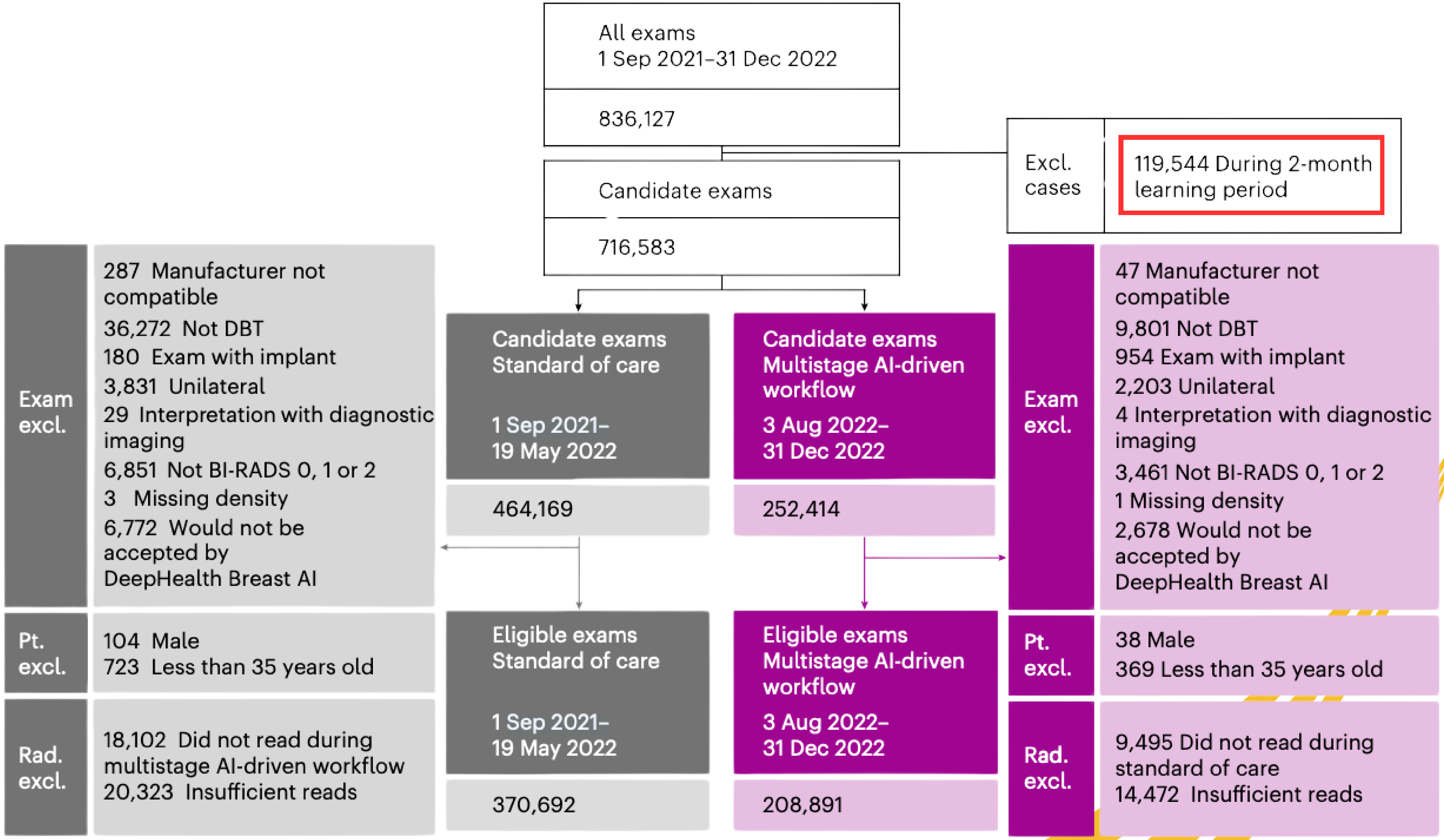
European Screening	USA Screening
Bi- or tri-ennial screening	Annual, opportunistic screening
Double reading (two radiologists)	Single-reading workflow
Primarily full-field digital mammography	Primarily Digital Breast Tomosynthesis (DBT)
	Highly diverse population

Selection criteria

Exam criteria: *bilateral screening DBT without implants or additional diagnostic imaging, BI-RADS* (0,1,2), valid breast density, compatibility with DeepHealth Breast*

Patient Criteria: *≥35 years old and self-reported as female*

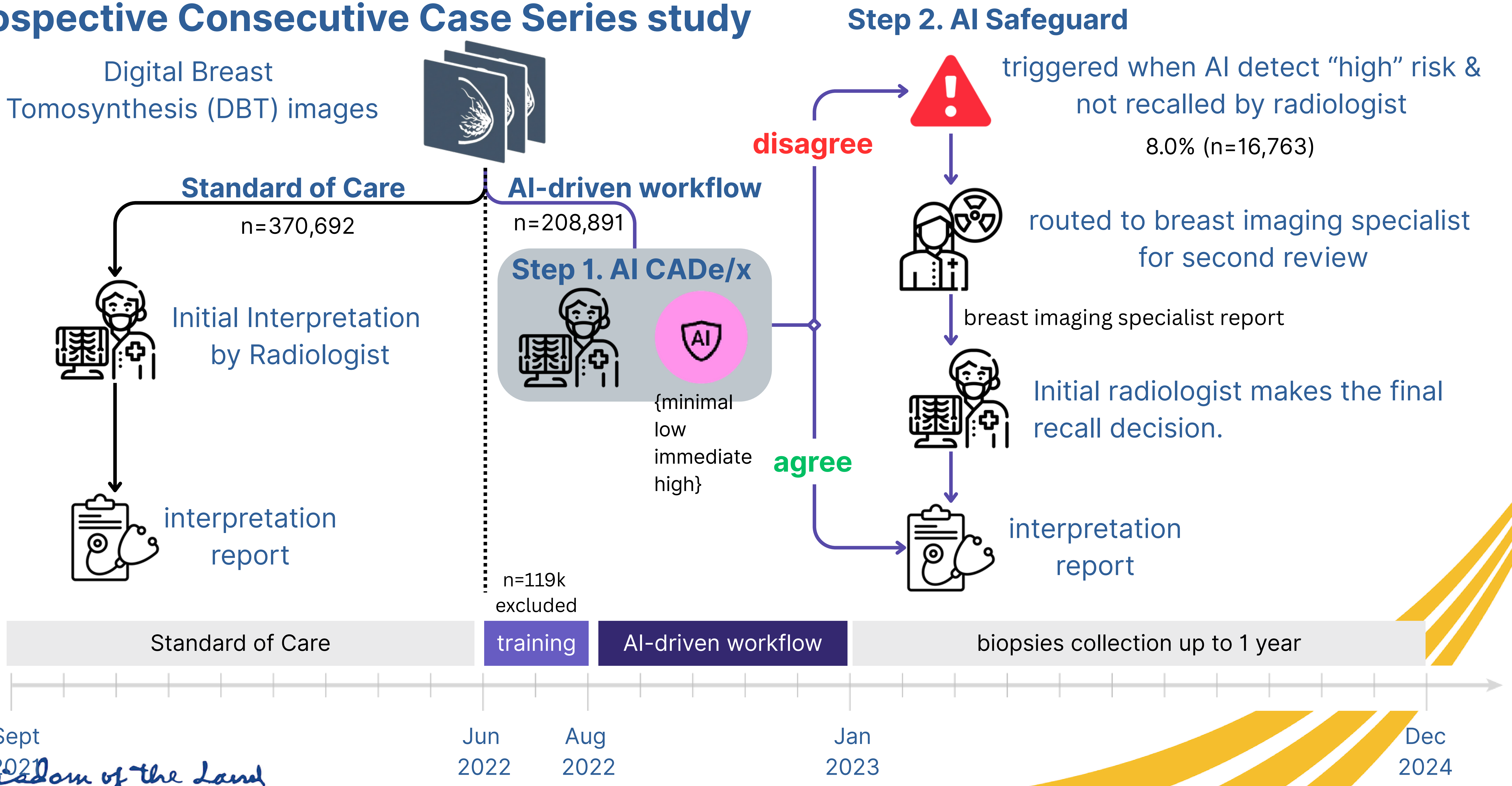
Radiologist Criteria: *interpreted screening mammograms during both study periods based on the MQSA required minimum of 960 every 2 years*



*American College of Radiology Breast Imaging Reporting and Data System (BI-RADS)

Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

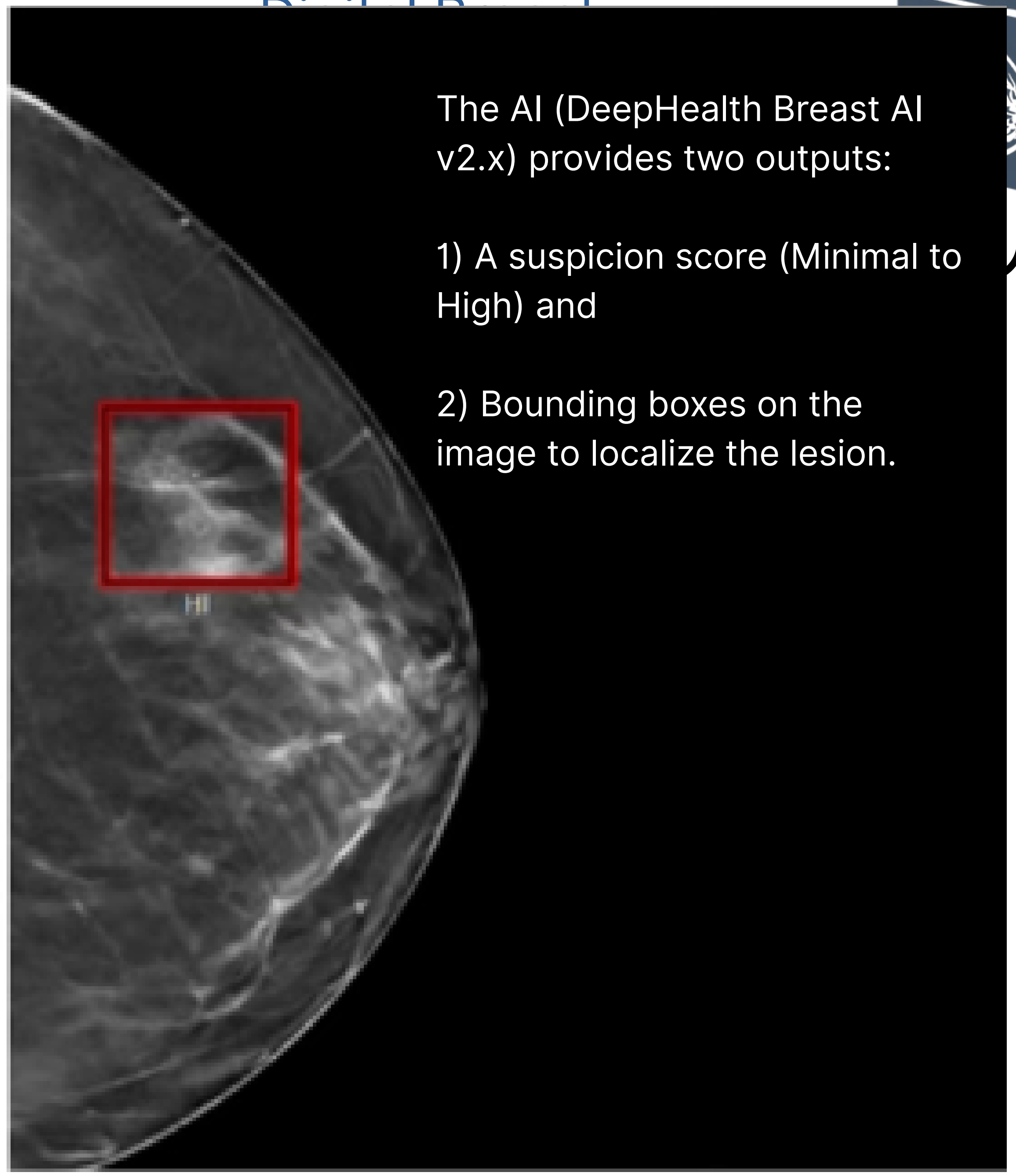
Prospective Consecutive Case Series study



Wisdom of the Land

Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

Prospective Consecutive Case Series study



AI-driven workflow
n=208,891

n=119k excluded

Step 1. AI CADe/x

{minimal
low
immediate
high}

disagree

agree

Step 2. AI Safeguard

triggered when AI detect "high" risk & not recalled by radiologist

8.0% (n=16,763)

routed to breast imaging specialist for second review

breast imaging specialist report

Initial radiologist makes the final recall decision.

interpretation report

training	AI-driven workflow	biopsies collection up to 1 year

Sept 2021 — Jun 2022 — Aug 2022 — Jan 2023 — Dec 2024

Wisdom of the Land

Statistical methods

Primary Analysis

- **Chi-squared test** for unadjusted rates between SoC and AI-driven cohort

Adjustments

- used **Generalized Estimation Equations (GEE)** to account for correlations among interpretations performed by the same radiologist.

Covariates

- Models adjusted for patient age, race/ethnicity, and breast density.

The study did not apply corrections for multiple comparisons.
The primary outcomes (CDR, Recall Rate, PPV) are highly correlated biologically and operationally in screening mammography.

Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

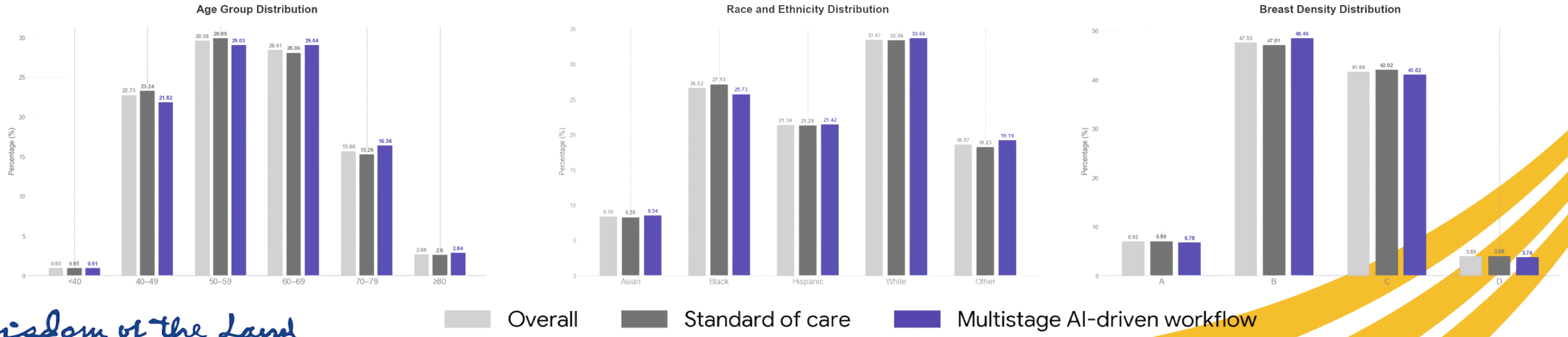
Demographic Comparison

- **Racial Diversity**: Both groups maintained high diversity with consistent distributions of White (~33%), Black (~26–27%), and Hispanic (~21%) patients.

- **Breast Density**: Tissue density was comparable across cohorts, with approximately 45% of women in both groups classified as having dense breasts (BI-RADS C or D)

- **Age Distribution**: Patient age remained matched between study periods, with the majority of patients (~62%) aged 55 years or older.

No statistically significant differences in age, race, or density between the two cohorts.



Wisdom of the Land

Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

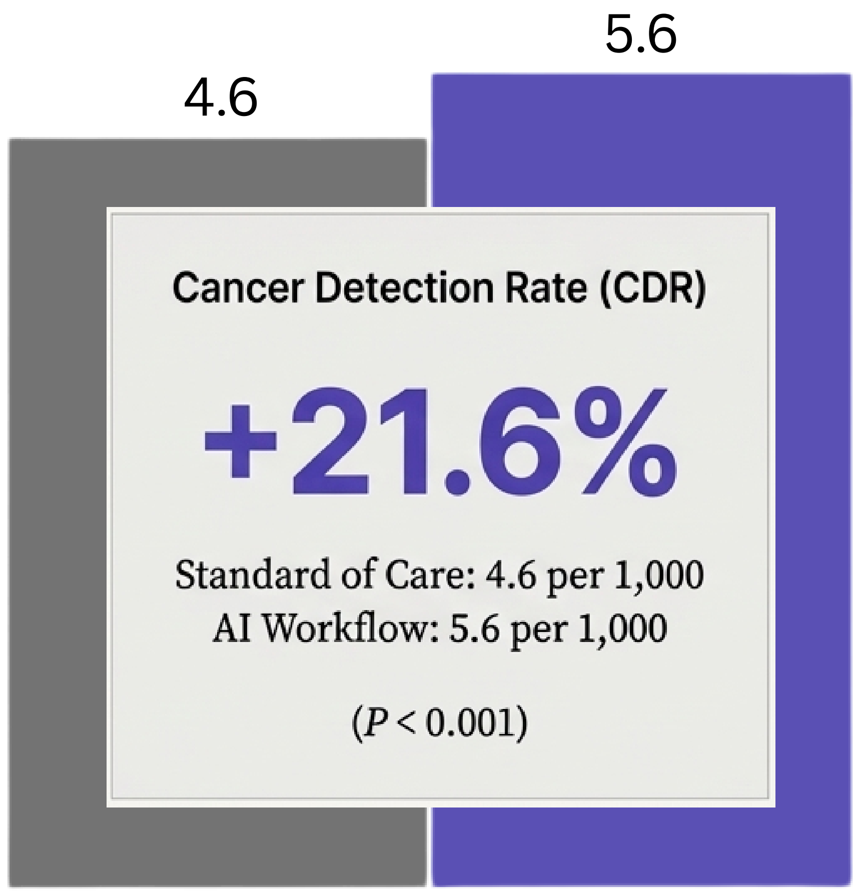
579,583
Total Screening

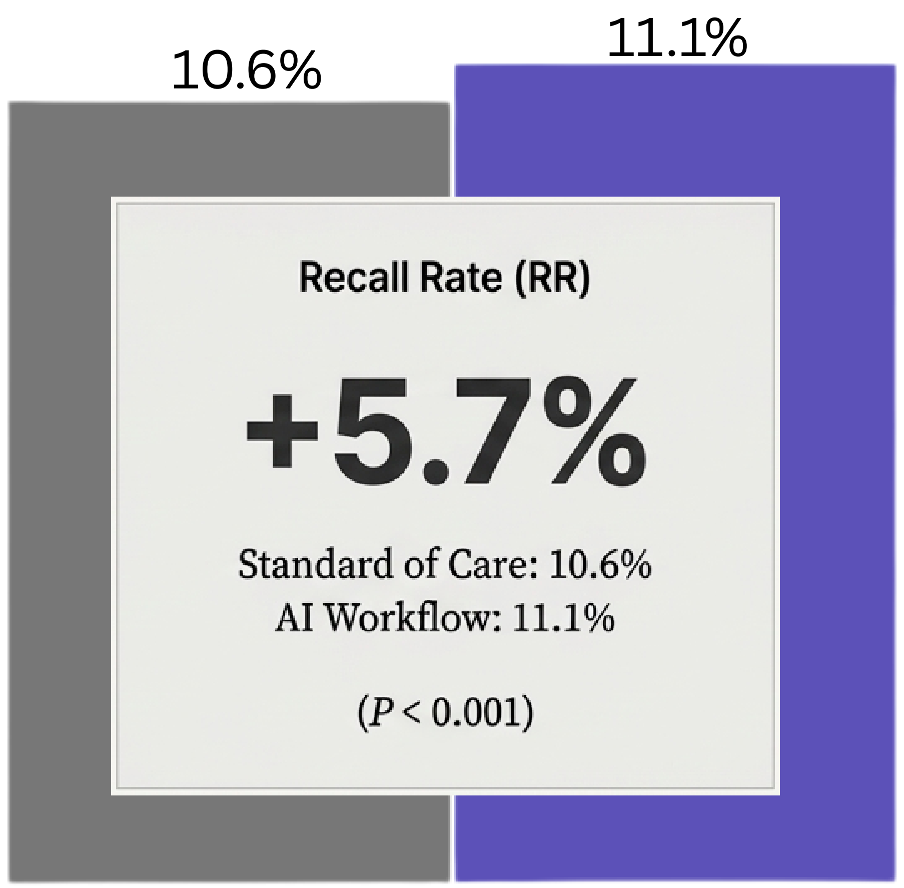
579,583 total screening exams, divided between the Standard of Care (64%) and AI-driven (36%) cohorts.

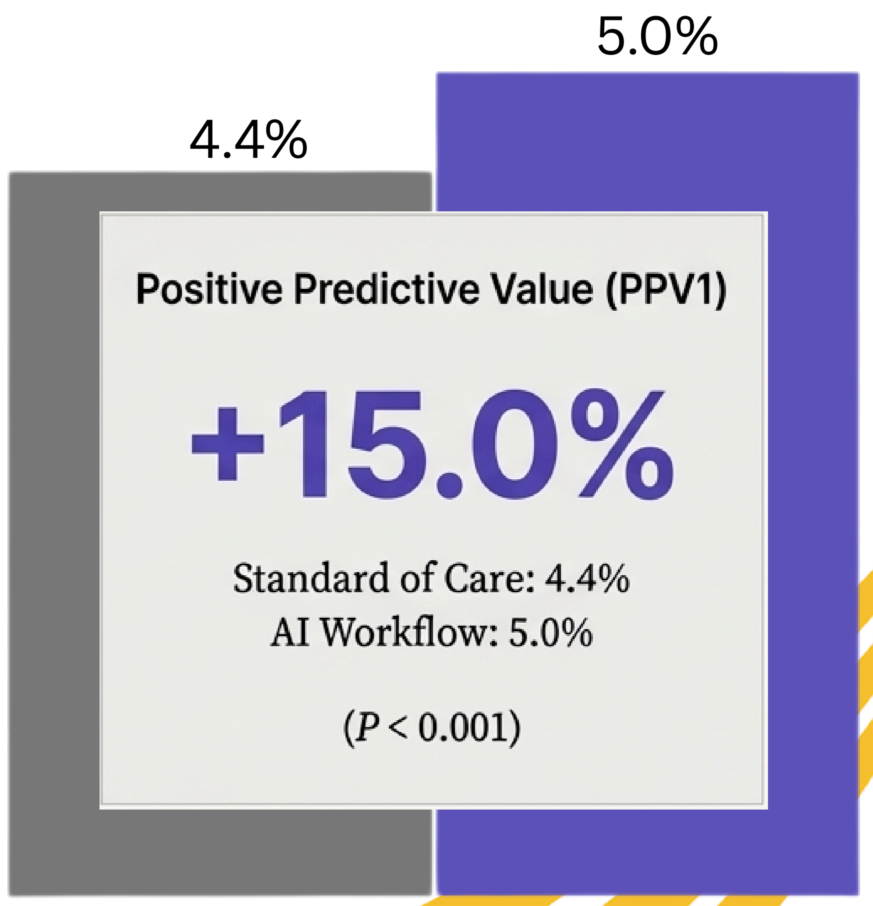
109
imaging sites

4
U.S states (CA, DE, MD, NY)

96
radiologists



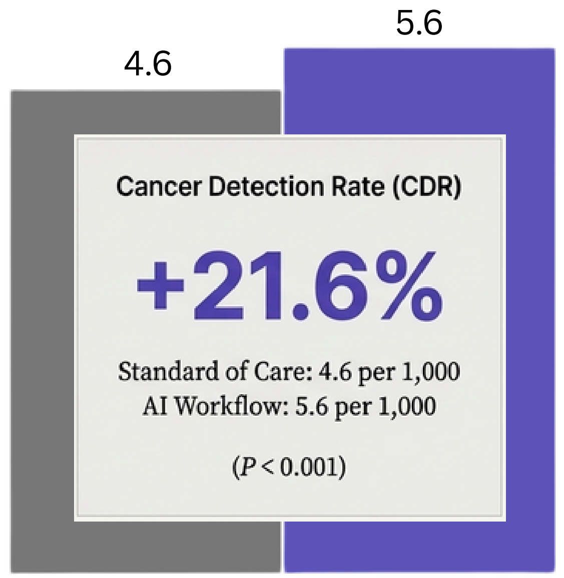




Standard of care | Multistage AI-driven workflow

Wisdom of the Land

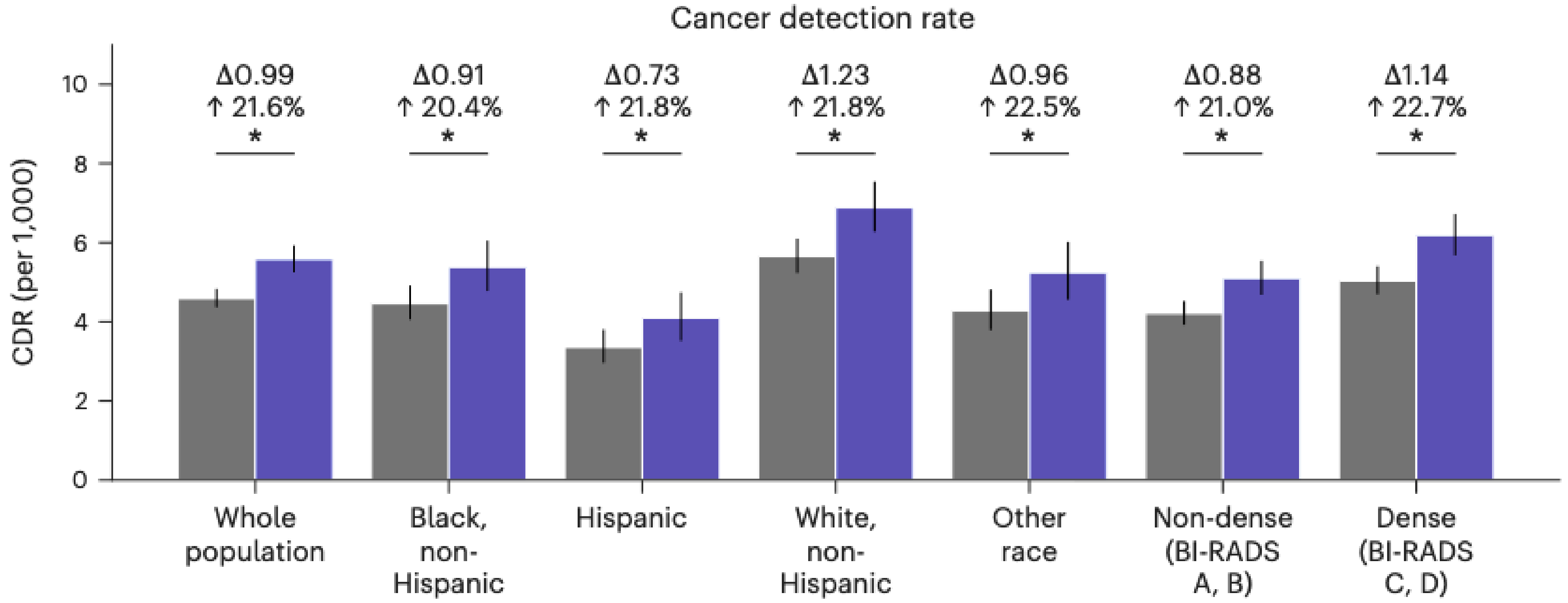
Results

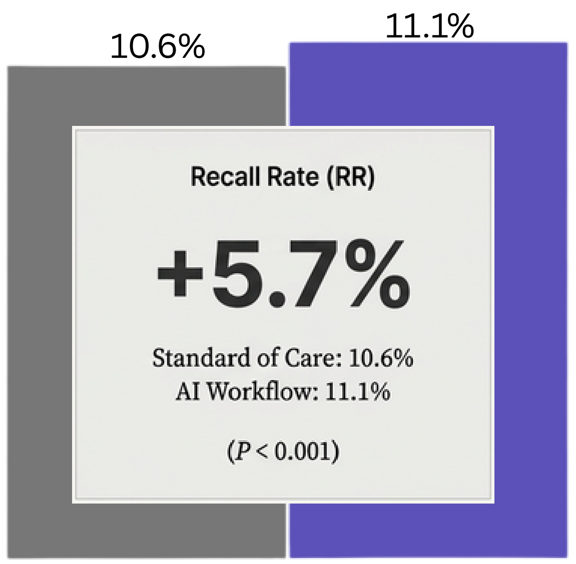


Cancer Detection Rate (CDR)

defined as the number of BI-RADS 0 (positive) exams with a malignant biopsy divided by the total number of exams multiplied by 1,000.

CDR (Δ0.99 cancers per 1,000 exams = 21.6%, 95% confidence interval (CI) 12.9–31.0%, $P < 0.001$)

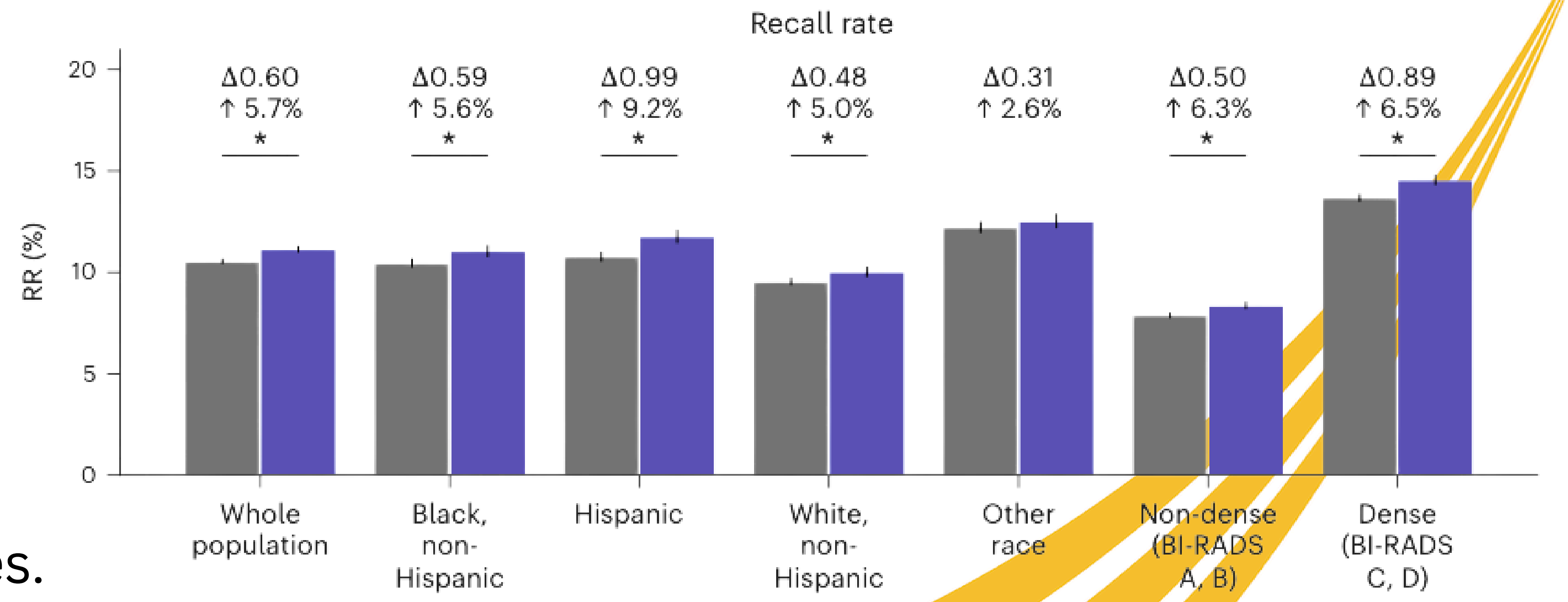




Recall Rate (RR)

defined as the percentage of screening exams that were positive

RR (Δ0.60 recalls per 100 exams = 5.7%, 95% CI 4.1–7.3%, $P < 0.001$)

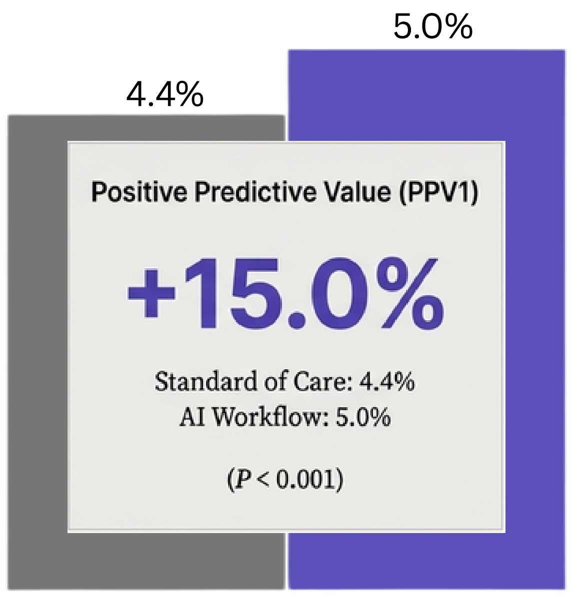


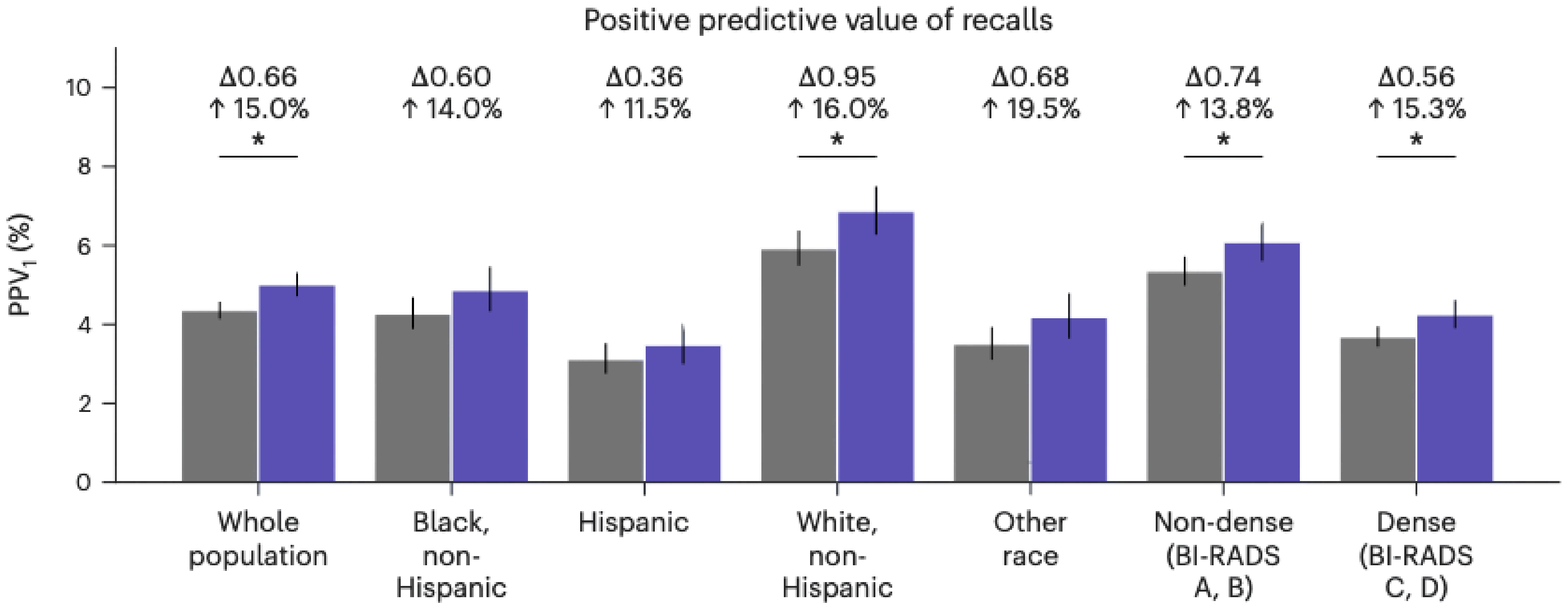
CDR increases were greater than RR increases in all cases.

Positive Predictive Value (PPV1)

defined as the percentage of positive exams that resulted in a malignant biopsy

PPV1 (Δ0.66 cancers per 100 recalls = 15.0%, 95% CI 7.0–23.7%, P < 0.001)





Discussion

Widespread Adoption Could Find Over 34,000 Additional Cancers Annually in the U.S.

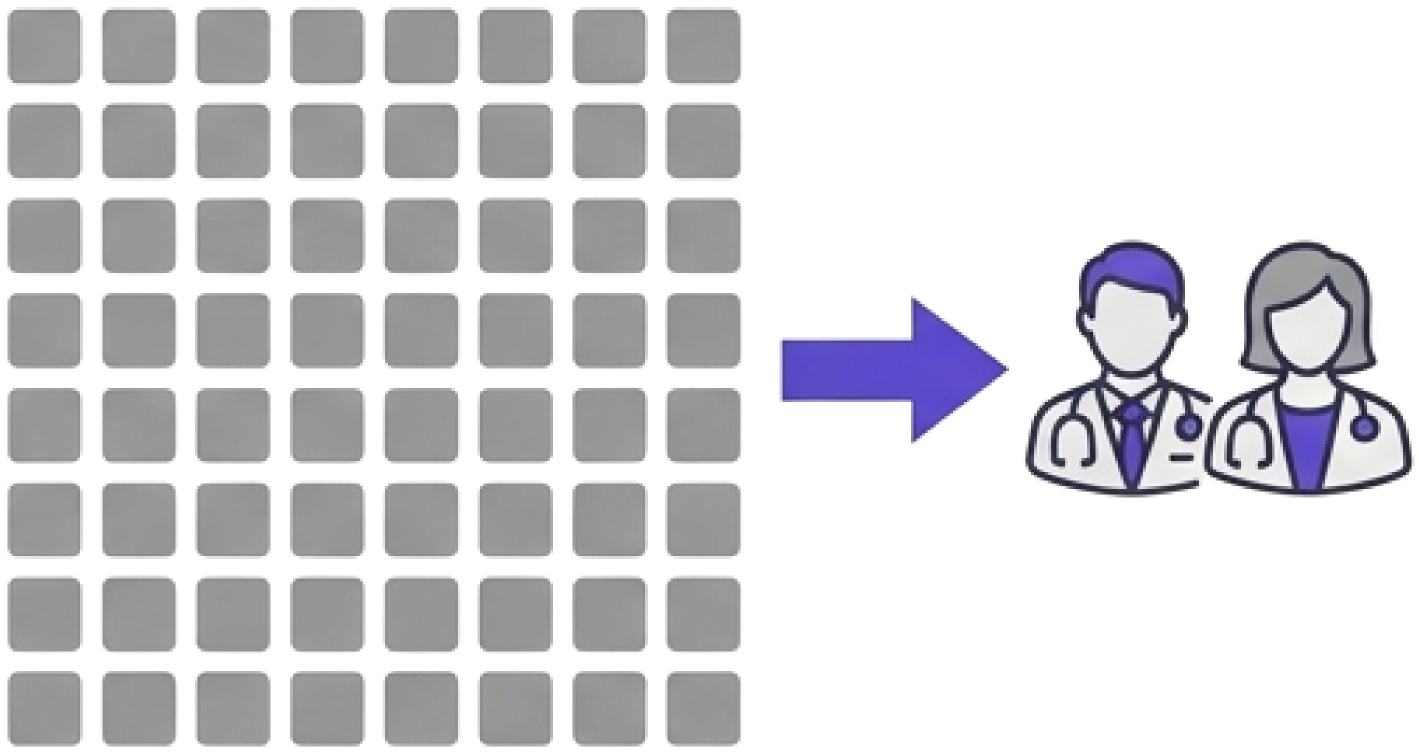
34,097 Potential additional cancers found each year through early screening.

Based on the study's observed increase in CDR applied to the **43 million mammograms performed annually** in the U.S. (assuming 80% are for screening).

By increasing the CDR, this workflow may facilitate the detection of cancers in earlier screening exams, especially for minority populations who have historically faced diagnosis at later stages.

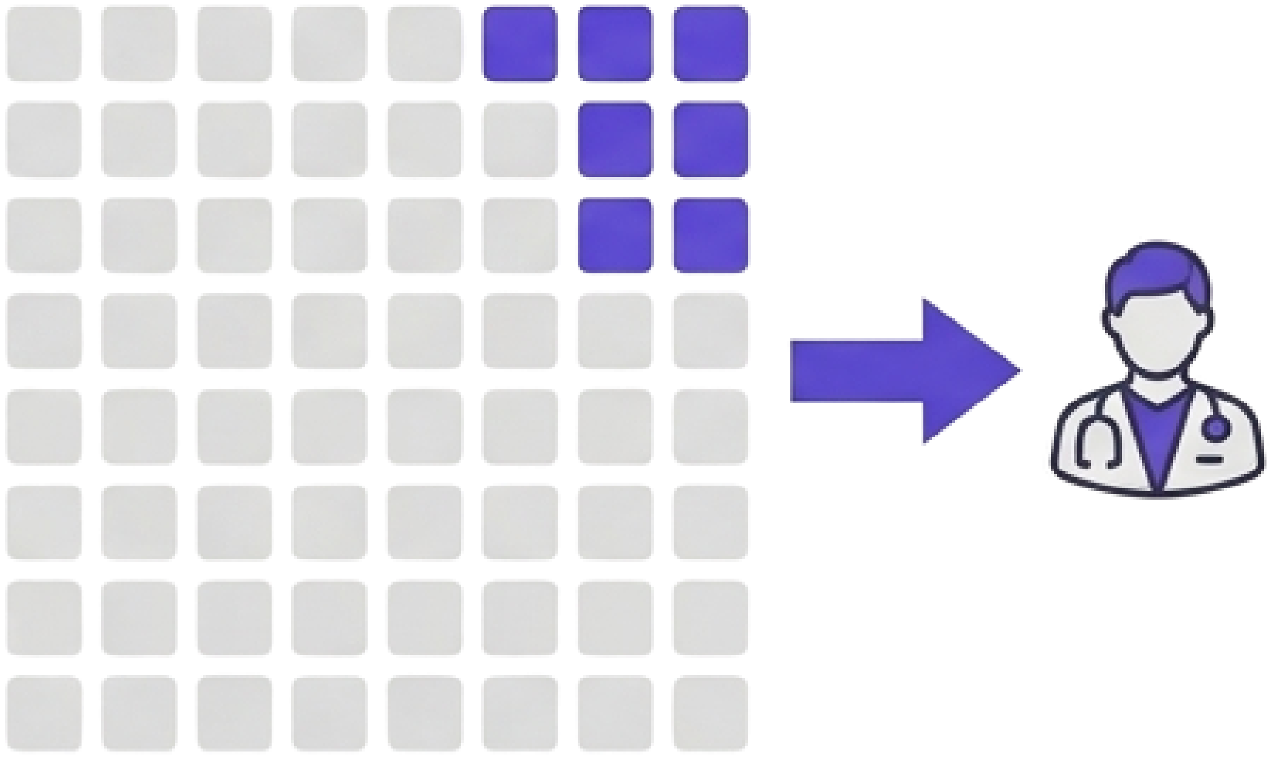
Discussion

Traditional Double Reading (EU)



100% of exams go to a second radiologist.

AI SafeGuard Review



Only 8.0% of high-risk cases are routed to a second specialist for review.

The observed 21.6% increase in CDR is greater than the estimated 11% increase associated with double reading 100% of exams in the U.S.

Challenges and Limitations

Design Limitation

- **Sequential Design**: The sequential "before-and-after" cohorts cannot fully control for unknown biases or confounders like a randomized trial would.
- **History Bias:** External factors differing between the 2021 (historical) and 2022 (concurrent) periods other than AI implementation. (e.g., post-COVID healthcare recovery).

Selection and Implementation limitations

- Potential **Selection Bias:** excluded 83 low-volume radiologists, leaving 96 high-volume radiologists. Excluding low-volume readers systematically enriches the study pool with "better" performing radiologists, can potentially inflating baseline metrics.
- **Learning Period Bias**: excluded data from the first 2 months of AI use. This might overestimate the benefit because they removed the period where radiologists were struggling to learn the system.

Analytical & ‘Black Box’ Issues

- lack long-term follow-up on interval cancers or false-negative rates.
- did not apply standard corrections for multiple comparisons despite testing across numerous subpopulations.
- underlying AI model is “blackbox” nature and study data are inaccessible intellectual property, preventing independent verification.

Conclusion

SIGNIFICANTLY IMPROVED PERFORMANCE

The multistage AI workflow is associated with a 21.6% increase in cancer detection and a 15.0% increase in positive predictive value in a real-world U.S. setting.

DEMONSTRATED EQUITY

These significant benefits are delivered equitably across all key racial, ethnic, and breast density subpopulations, helping to address longstanding disparities in care.

EFFICIENT & SCALABLE

The unique workflow provides the cancer detection benefits of double reading for a fraction of the cost and effort, making it a practical solution for the U.S. single-reader paradigm.

Thank You

Q&A

Q. Did the increase in Recall Rate (RR) lead to over-diagnosis?

- A. No. The PPV (Positive Predictive Value) increased by 15%. This means the extra recalls were effective.

Q. Why not Randomized Controlled Trial (RCT)?

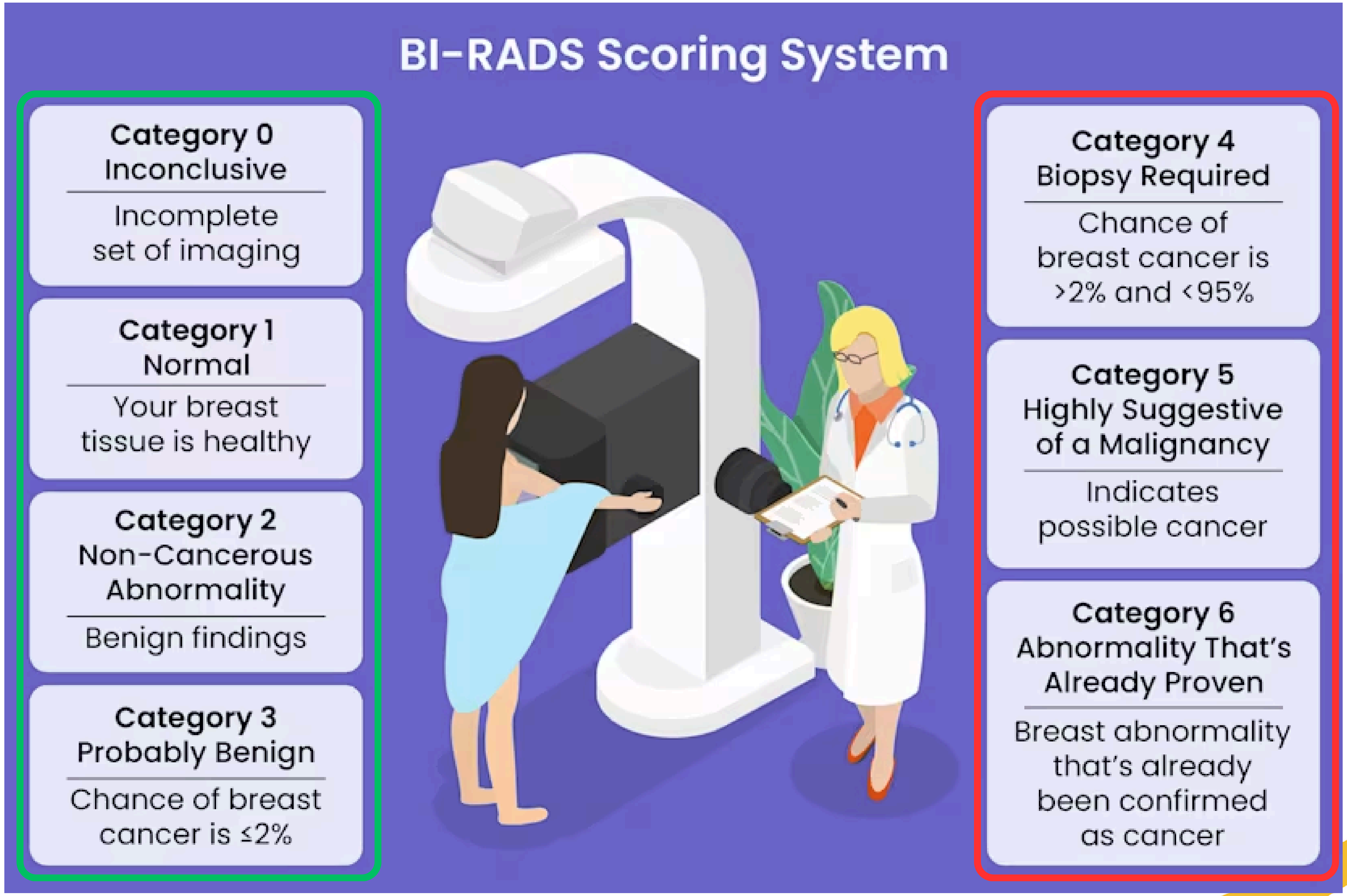
- It wasn't feasible to blind the radiologists to the AI tool while they were reading. Also, they wanted to test "real-world" implementation across 109 sites.

Q. How did they handle the clustering of data?

- They used Generalized Estimating Equations (GEE) to account for the correlation of exams read by the same radiologist

BI-RADS Scoring System

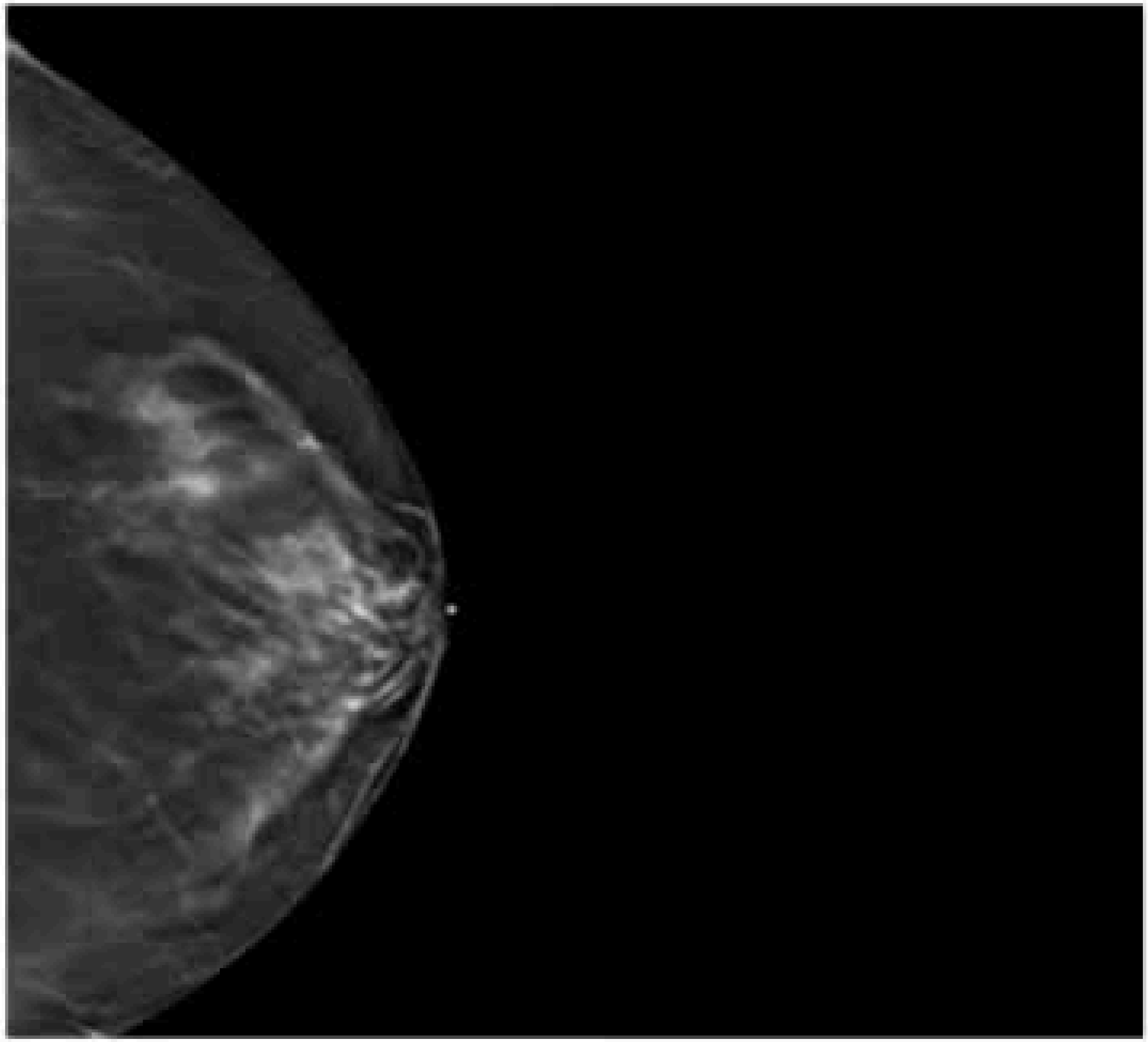
included in cohort



Getty Images/TarikVision (illustration)

excluded from cohort

Mammography



DBT