# Issues in the Use of AI for Breast Cancer Screening

Etta D. Pisano, MD FACR FSBI
The American College of Radiology and the University of Pennsylvania
Beebe Symposium at the National Academy of Medicine
March 14, 2025

# Conflicts

- I am an employee of the American College of Radiology (ACR) AND of ARPA-H, a federal government agency under HHS. At ARPA-H, I lead the Advancing Clinical Trials Readiness (ACTR) Initiative.
- The ACR is working with many AI vendors to help them prove the safety and efficacy of their products, sometimes for FDA premarket evaluation.
- The ACR is also working to help radiologists understand how AI products may be useful in their unique practice settings.
- All views expressed in this talk are mine and do not represent official positions of the ACR OR ARPA-H.

# Fundamentals of US Breast Cancer Screening

- **For those without signs or symptoms of breast cancer**
- **Annual or biennial starting at 40, ending depends on woman**
- **2 views per breast for digital mammography, or multiple slices for tomosynthesis**
- **Generally read "off line" (Not while the woman waits)**
- **7-15% "called back" for additional work-up which includes extra mammographic images, US, MRI. Some of these are technical repeats.**
- **Approximately 8-10% of those called back get breast biopsies (usually imaging-guided, but some need surgical biopsy)**
- **Of those who get biopsy, ~20% have cancer (or 5-10/1000 people screened).**
- **And, some cancers are still missed (~40% of breast cancers are visible on prior mammograms).**

# Fundamentals of US Breast Cancer Screening

- **For those without signs or symptoms of breast cancer**
- **Annual or biennial starting at 40**
- **2 views per breast for digital mammography, or multiple slices for tomosynthesis**
- **Generally read "off line" (Not while the woman waits)**
- **7-15% "called back" for additional work-up which includes extra mammographic images, US, MRI. Some of these are technical repeats. We can reduce False Positives at call back!**
- **Approximately 8-10% of those called back get breast biopsies (usually imaging-guided, but some need surgical biopsy).**
- **Of those who get biopsy, ~20% have cancer (or 5-10/1000 people screened). We can reduce False Positives at BX!**
- **And, some cancers are still missed (~40% of breast cancers are visible on prior mammograms). We can improve True Positives!**

# A Potential Schema for Evaluating AI and the Relative Importance of Each Factor

- **What is effect on patient outcomes (cancer dx/death/etc.)?**
  **FEWER DEATHS from BREAST CANCER and LESS MORBIDITY FROM DX/RX**
- **How SAFE (Accuracy/Sens/Spec) is the technology for patients?**
  **MUCH SAFER…NO CHANGE IN SAFETY…MUCH LESS SAFE**
- **How does this technology effect costs?**
  **MUCH MORE EXPENSIVE…NO CHANGE…MUCH LESS**
- **Does the technology increase convenience for patients?**
  **MUCH MORE…NO CHANGE…MUCH LESS**
- **How much would this help radiologists do their jobs better?**
  **EXTREMELY HELPFUL…SLIGHTLY HELPFUL…NOT HELPFUL**

“The Food and Drug Administration (FDA) is responsible for protecting the public health by assuring the **safety, efficacy**, and security of human and veterinary drugs, biological products, medical devices, our nation's food supply, cosmetics, and products that emit radiation. The FDA also provides accurate, science-based health information to the public.”

~46 AI products applied to breast imaging have been authorized for sale in the US as of 1/27/2025.

Breast AI Software = Medical Device

Is Autonomous Breast AI allowed under MQSA for the interpretation of screening mammograms?

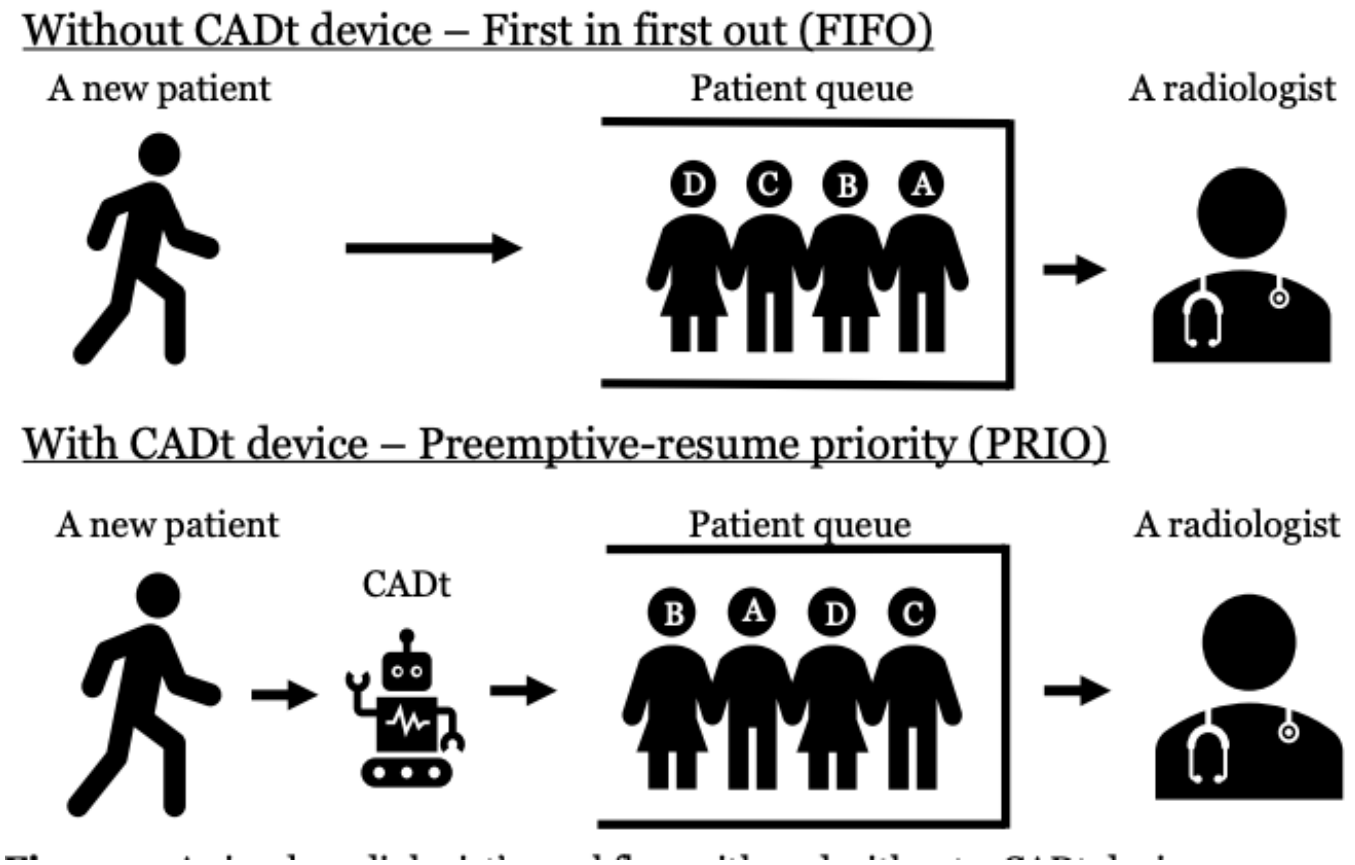
Breast AI Software = Medical Device

Is Autonomous Breast AI allowed under MQSA for the interpretation of screening mammograms?  NO!
But there are companies lobbying to change MQSA.

# Current Regulation of AI/ML Software – Intended Uses

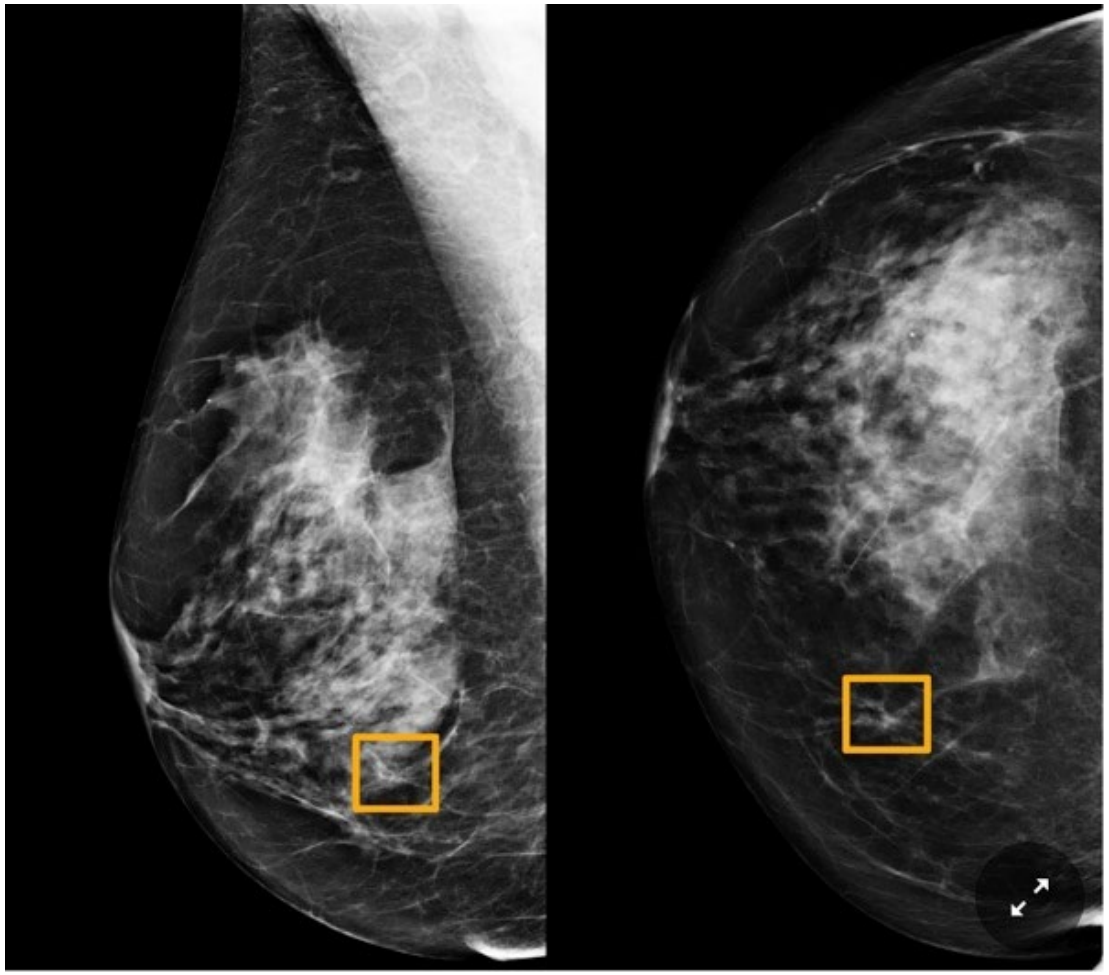
- CADt- Triage
- CADe-Detection/Localization
- CADx-Diagnosis/Characterization
- CADe/x-Both Detection and Diagnosis/Classification
- CADa/o-Acquisition/Optimization

# How CADt software is intended to be used*



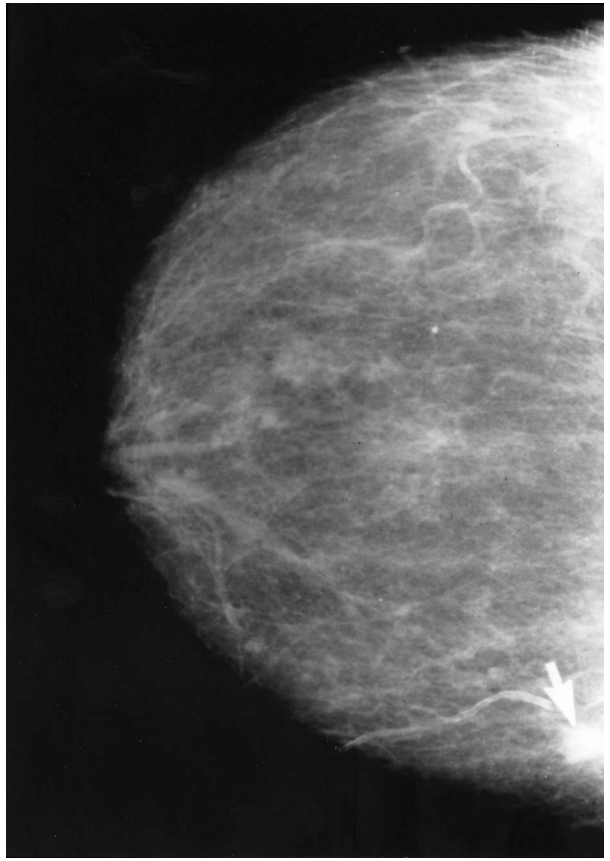

*From poster by E Thompson, et al.of FDA CDRH
https://www.fda.gov/media/148986/download

# How CADe software is intended to be used*



*From NYTimes. “AI is learning to read mammograms” by Denise Grady. 1/1/2020. Image from Northwestern University.

# How CADe/x software is intended to be used*



AI SYSTEM OUTPUT

BIRADS 5- HIGHLY SUGGESTIVE of MALIGNANCY

*https://upload.wikimedia.org/wikipedia/commons/3/35/Mammogram_with_obvious_cancer.jpg

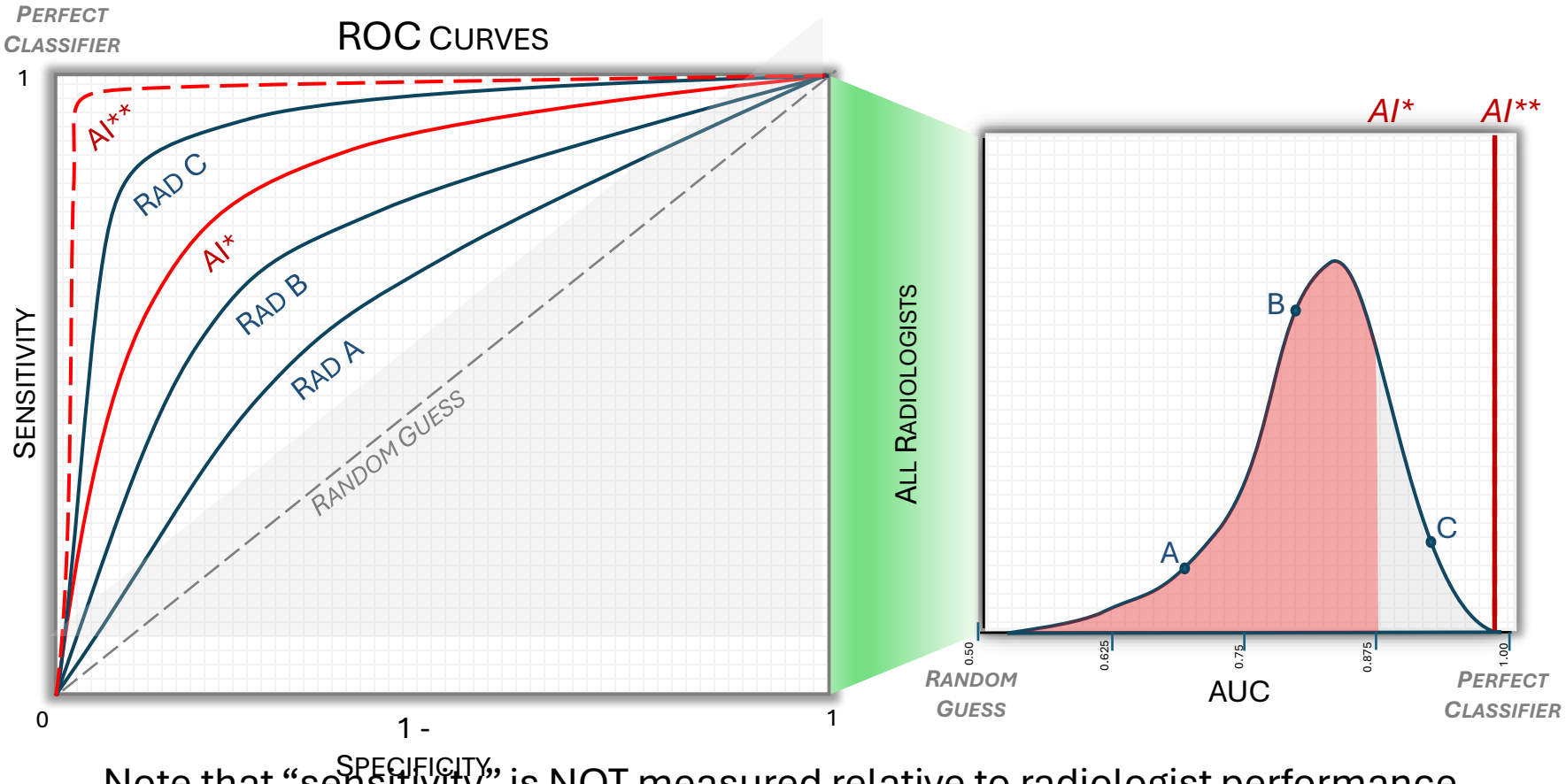
# Measuring Patient Outcomes

- Requires years of study to assess accurately (breast cancer deaths happen 5-25 years after initial diagnosis)
- **Impractical in the regulatory environment**
- Requires randomized clinical trials to reduce confounding by factors not related to the AI (SES, race, ethnicity, comorbid conditions, etc).
- Can be studied through registries with excellent ascertainment of outcomes (e.g. BCSC)

# Potential Methods to Assess Safety and Efficacy (Diagnostic Accuracy/Sensitivity/Specificity) of Breast AI Software

- Standalone performance testing – practical in the regulatory environment
- Reader Studies – practical in the regulatory environment
- Clinical Trials – too expensive in the regulatory environment, so impractical

# An ideal AI system will improve AUC and increase both sensitivity and specificity.




Note that "sensitivity" is NOT measured relative to radiologist performance, but to the actual presence of cancer in the breast up to a set time* after the examination, whether visible on the mammogram or not.

# AI Software Standalone Performance Testing

→

Diagnostic Accuracy, Sensitivity, Specificity, PPV, NPV, Call back rates

# Limitations of Standalone Performance Testing

- Were cases **properly categorized** as negative or positive and what was that categorization based on?
- Were training and testing sets **representative** of the population the algorithm will be applied to?
- Will results **generalize** to other datasets?
- What sorts of data were **missing** from the training and testing?
    - breast density
    - lesion types/sizes/features
    - Subtlety of cancers
    - Rare conditions

# Reader Studies

→

Individual Reader Data with and without the AI

Diagnostic Accuracy
Sensitivity
Specificity
PPV
NPV
Call back rates

AND

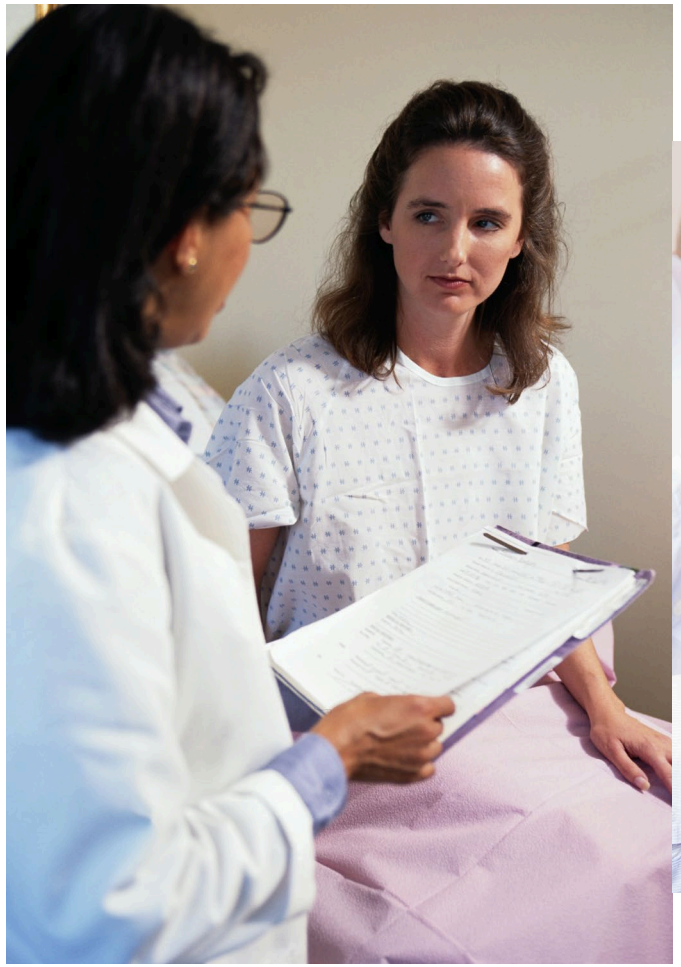
GROUP PERFORMANCE DATA

# Limitations of Reader Studies –Same as Standalone

+

- Were readers representative of the radiologists who will use the product?
- Since no patient is impacted by the readings, do readers use the software differently than in real life?
- In order to obtain data on accuracy, etc, radiologists must use a different scale than they use in usual practice – not just Call Back or not, but also likelihood of malignancy.
- Cases must be enriched for cancers compared to the usual screening population (4-7/1000 versus e.g. ~30/100 in a sample of 300 cases). How does this impact the data collected?

# Clinical Trials



1000s of volunteers

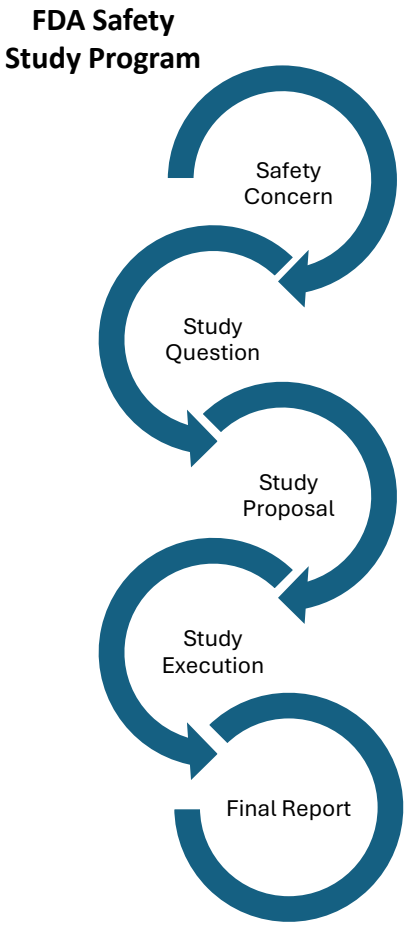
Extensive Data Entry and Cleaning

Central QC, Management and Analysis

# Limitations of Clinical Trials – THE GOLD STANDARD

- They take a long time to do.
- They are very expensive.
- Sample sizes are gigantic when testing a screening tool (3-7/1000 will have cancer, so roughly 80,000+ needed for power).
- Are patient populations enrolled representative of the US population?
- Are the sites/physicians where the trials are run representative of all US clinical sites?
- Do clinicians participating in trials behave the same way they do in clinical practice?

# Real-world Evidence Safety Study: Radiological AI for Large Vessel Occlusion (LVO) – slides courtesy of Robert Ochs of FDA CDRH

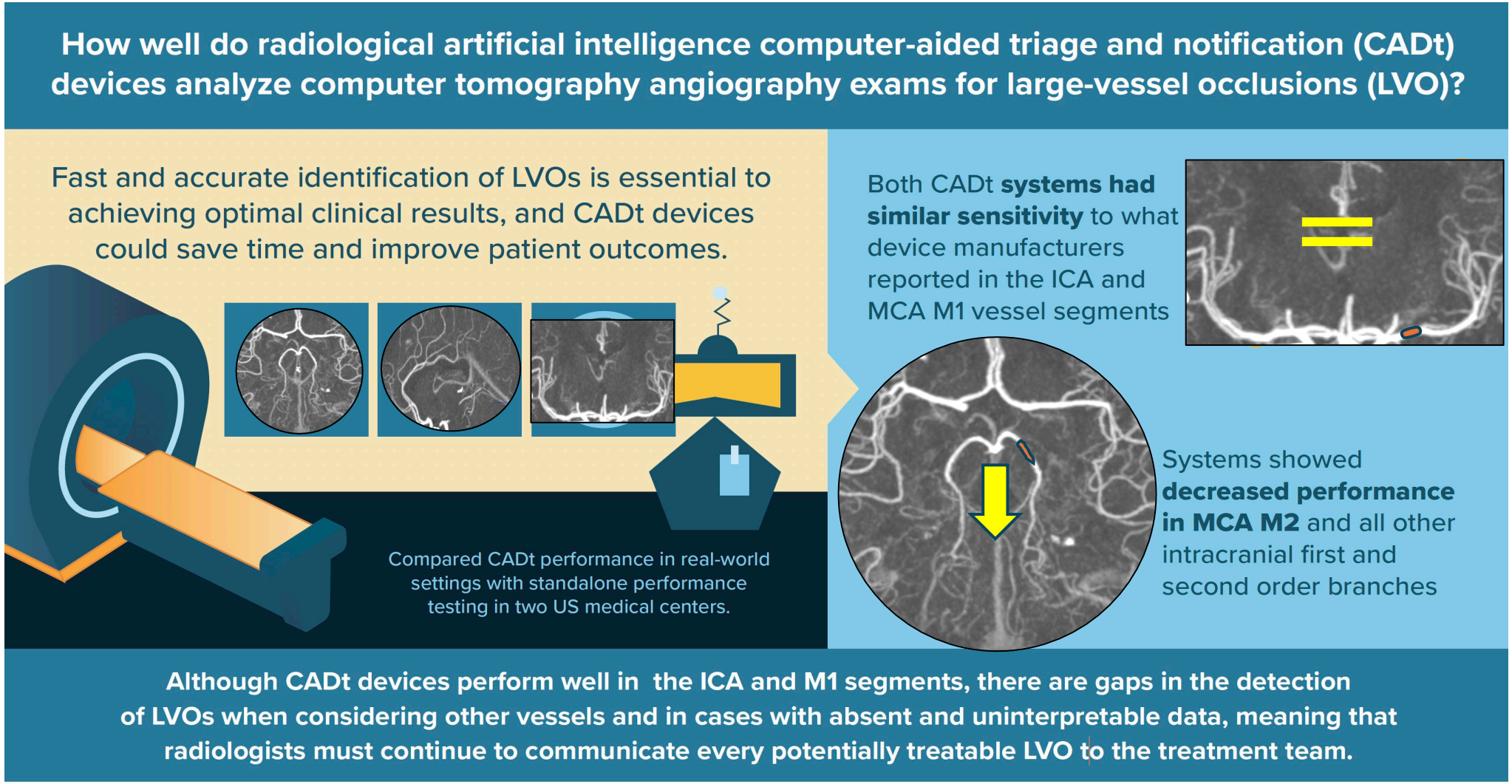



- Purpose: to evaluated AI computer-aided triage and notification (CADt) devices indicated for **prioritization** and **notification** of *suspected* LVO
- FDA initiated; protocol developed with ACR; conducted by MGH and Lahey
- Retrospective evaluation to compare **real world performance** vs. **labeling**
- Results: CADt performance consistent with the manufacturer-stated performance for ICA and M1 vessels, but not for other potentially treatable intracranial vessel occlusion in 2nd and 3rd order segments
- The instructions for use for the CADt devices in the study did not provide a sub-analysis of performance by vessel location
- **Informed** discussions with manufacturers about labeling revisions and an FDA safety communication

Intended Use of Imaging Software for Intracranial Large Vessel Occlusion - Letter to Health Care Providers

Kunst et al, Real-World Performance of Large Vessel Occlusion Artificial Intelligence–Based Computer-Aided Triage and Notification Algorithms—What the Stroke Team Needs to Know, -J Am Coll Radiol. 2023 May 16:S1546-1440(23)00335-6.

# Real-world Evidence Safety Study: Radiological AI for Large Vessel Occlusion (LVO)




Kunst et al, Real-World Performance of Large Vessel Occlusion Artificial Intelligence–Based Computer-Aided Triage and Notification Algorithms—What the Stroke Team Needs to Know, -J Am Coll Radiol. 2023 May 16:S1546-1440(23)00335-6.

# Evidence Reader Studies Don’t Necessarily Predict ACTUAL Performance in the Real World

Research

Original Investigation | LESS IS MORE

## Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection

Constance D. Lehman, MD, PhD; Robert D. Wellman, MS; Diana S. M. Buist, PhD; Karla Kerlikowske, MD; Anna N. A. Tosteson, ScD; Diana L. Miglioretti, PhD; for the Breast Cancer Surveillance Consortium

JAMA Intern Med . 2015 Nov;175(11):1828-37.
doi: 10.1001/jamainternmed.2015.5231.

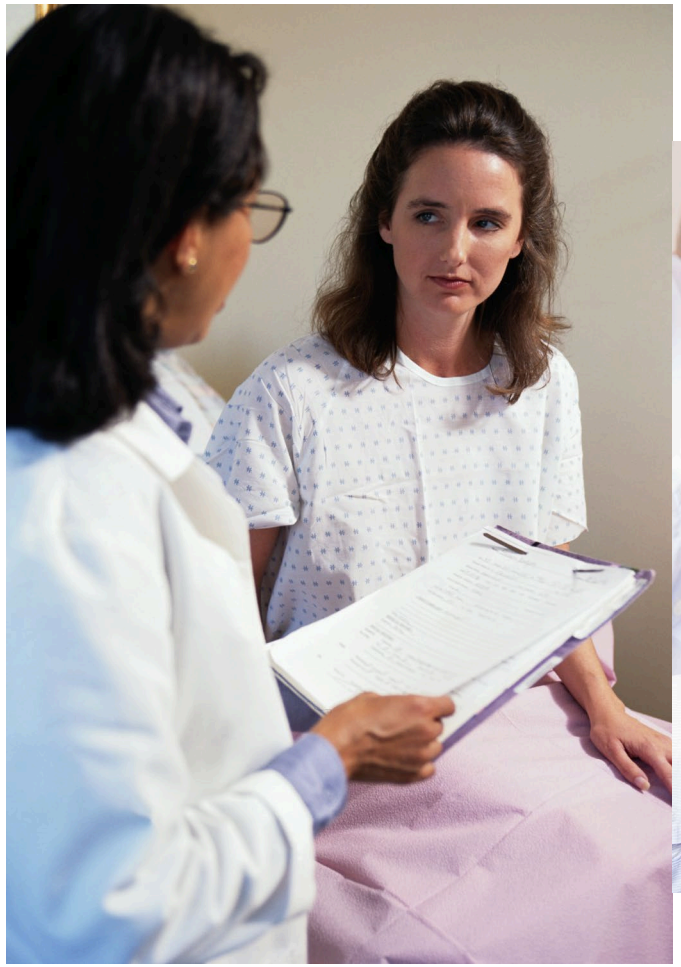
# Ethical Issues with Using AI in 2025

- **Dismantling the current regulatory framework** likely will expose patients to uncertain risks. Radiologists will have to develop new methods to assess products on the market, especially if marketing hype drives patient demand.
- Using software **“off label”** may increase risk to patients.
- AI tools may have been proven effective on different patient populations or with readers who are not typical of individual practices, causing **uncertainty of risks and benefits for different groups of patients and readers.**
- AI tools **drift** over time and need correction for that when it happens. (ARPA-H Precise-AI seeks to address this issue.)

# Could Real World Evidence be a pathway to the use of Autonomous AI in breast ca screening?

- Install AI algorithms that have reached certain performance levels through standalone performance tests in clinics before FDA authorization to learn how radiologists actually perform when they are used.
- Allow patients to opt out of the use of the AI at clinic check-in.
- For those who opt in, provide the standard of care interpretation by a single reader.
- Provide a second radiologist reading that uses the AI- one read for cases that AI determines as needing the radiologist to review, and an AI read alone (autonomous AI) plus a “quick read” by a radiologist for those cases that the AI determines as “almost certainly normal”.
- If there are disagreements between the two radiologists, decide upon patient pathway – call-back or not – through consensus conversation.

# How RWE Differs from a clinical trial?



1000s of volunteers

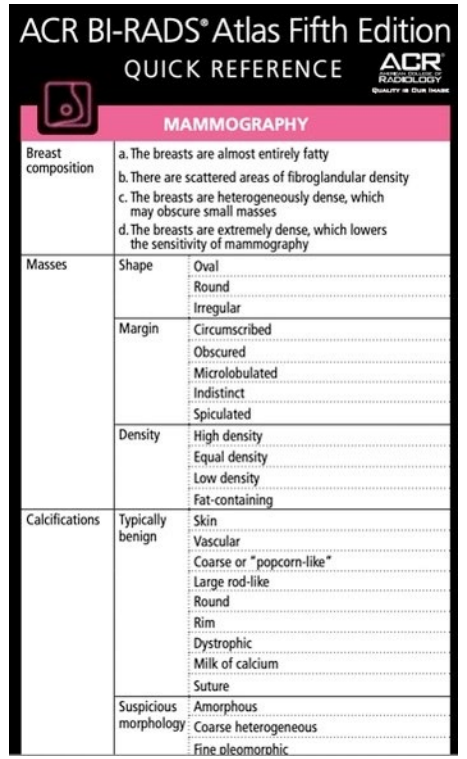
Extensive Data Entry and Cleaning

Central QC, Management and Analysis

# How RWE might work

Regular check-in with opt-out

Usual Clinical Workflow Utilized Double-reading to prove safety of autonomous AI?



ACR BI-RADS® Atlas Fifth Edition
QUICK REFERENCE

MAMMOGRAPHY

| | | |
|---|---|---|
| Breast composition | a. The breasts are almost entirely fatty<br>b. There are scattered areas of fibroglandular density<br>c. The breasts are heterogeneously dense, which may obscure small masses<br>d. The breasts are extremely dense, which lowers the sensitivity of mammography | |
| Masses | Shape | Oval |
| | | Round |
| | | Irregular |
| | Margin | Circumscribed |
| | | Obscured |
| | | Microlobulated |
| | | Indistinct |
| | | Spiculated |
| | Density | High density |
| | | Equal density |
| | | Low density |
| | | Fat-containing |
| Calcifications | Typically benign | Skin |
| | | Vascular |
| | | Coarse or "popcorn-like" |
| | | Large rod-like |
| | | Round |
| | | Rim |
| | | Dystrophic |
| | | Milk of calcium |
| | | Suture |
| | Suspicious morphology | Amorphous |
| | | Coarse heterogeneous |
| | | Fine pleomorphic |

Reporting and Path Data Auto-Extracted

# Other ways AI may be useful in breast imaging

- What if we could predict short-term risk of breast cancer?
  - Maybe we could use imaging and other factors (genetics, risk factors, etc) to predict 3-5 year risk.
  - That would allow individualized Screening Paradigms- MRI, US, contrast mammo for a few years rather than every year for life.
  - Maybe we could reduce risk through drug therapy in a wider group of women
  - Clairity and iCAD are working to get approval of such products. Scientific group at Karolinska also working on this idea.

# Have all technical repeats done routinely before the patient leaves the exam room!

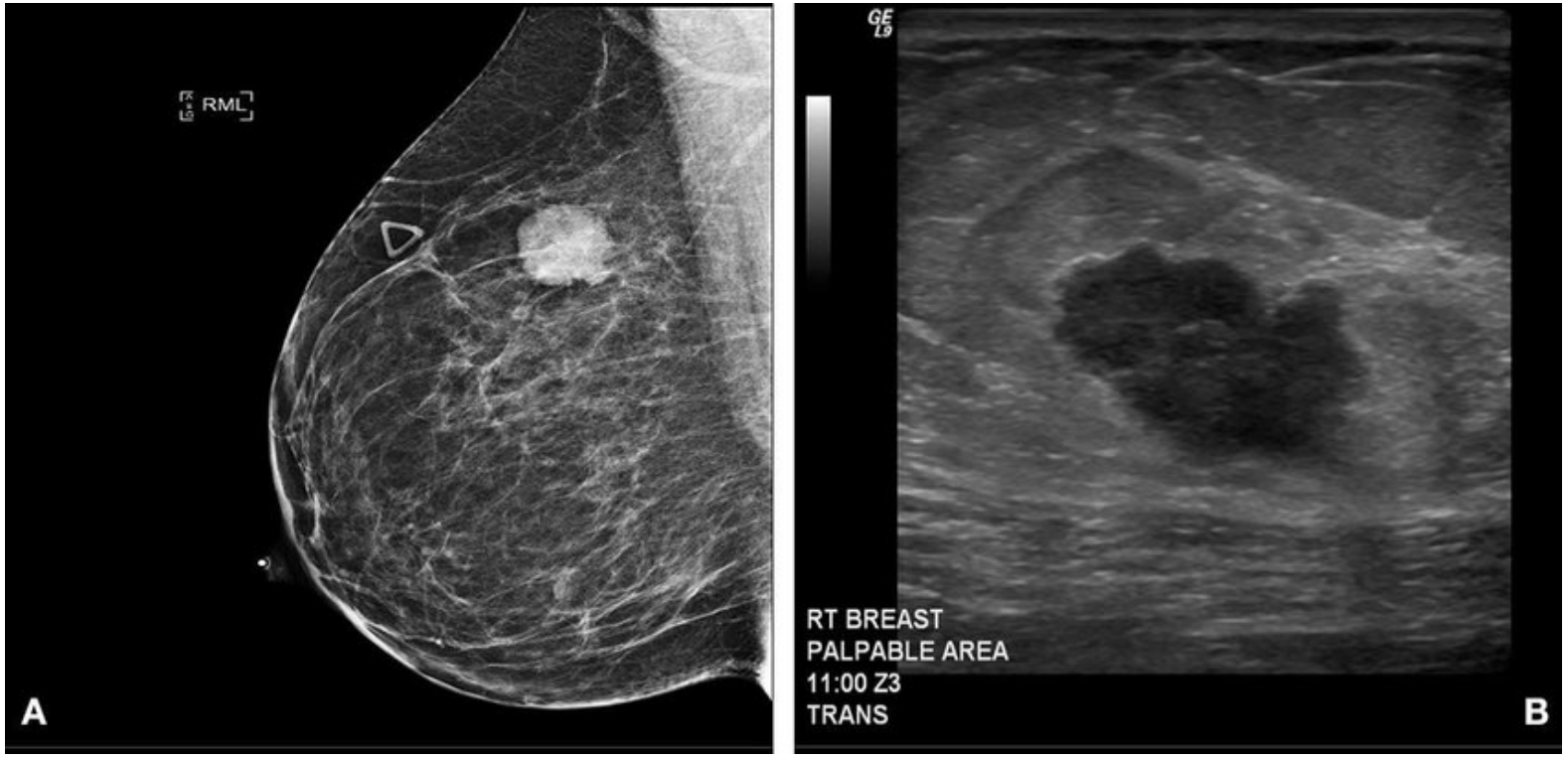
AI could alert tech of need to repeat one or more images before patient leaves the room!

NO CHANGE IN PATIENT OUTCOMES?
NO CHANGE in RISK TO PATIENT.
LIKELY SAVES MONEY for Patient and Facility.
MORE PATIENT CONVENIENCE.
REDUCES RADIOLOGIST TIME.

# AI could recommend typical work-up for usual findings


CHANGE IN PATIENT OUTCOMES UNCERTAIN
UNCLEAR EFFECT ON PATIENT SAFETY
COSTS REDUCED
PATIENT CONVENIENCE SIGNIFICANTLY IMPROVED.
RADIOLOGIST PRACTICE SIGNIFICANTLY IMPROVED.

# AI could eliminate work-up, go straight to biopsy.




Likely no Effect on patient outcomes.
Patient Safety likely unchanged.
Costs reduced.
Patient Convenience likely improved.
Radiologist work reduced.

# TWO FINAL POINTS

- The EU and UK are ahead of the US in implementing and testing AI for clinical practice.
- The UK National Health Service **may** fund a Breast Cancer Screening AI trial in the near future that will likely cluster randomize women to-
  - AI for triage
  - AI as the second reader
  - Control group - 2 radiologist readers (the Standard of Care in the UK).*

*Fiona Gilbert of Cambridge University


Image courtesy of NIST