A Medical Imaging Perspective on Effectiveness

Robert M. Nishikawa, Ph.D.

Department of Radiology

University of Pittsburgh
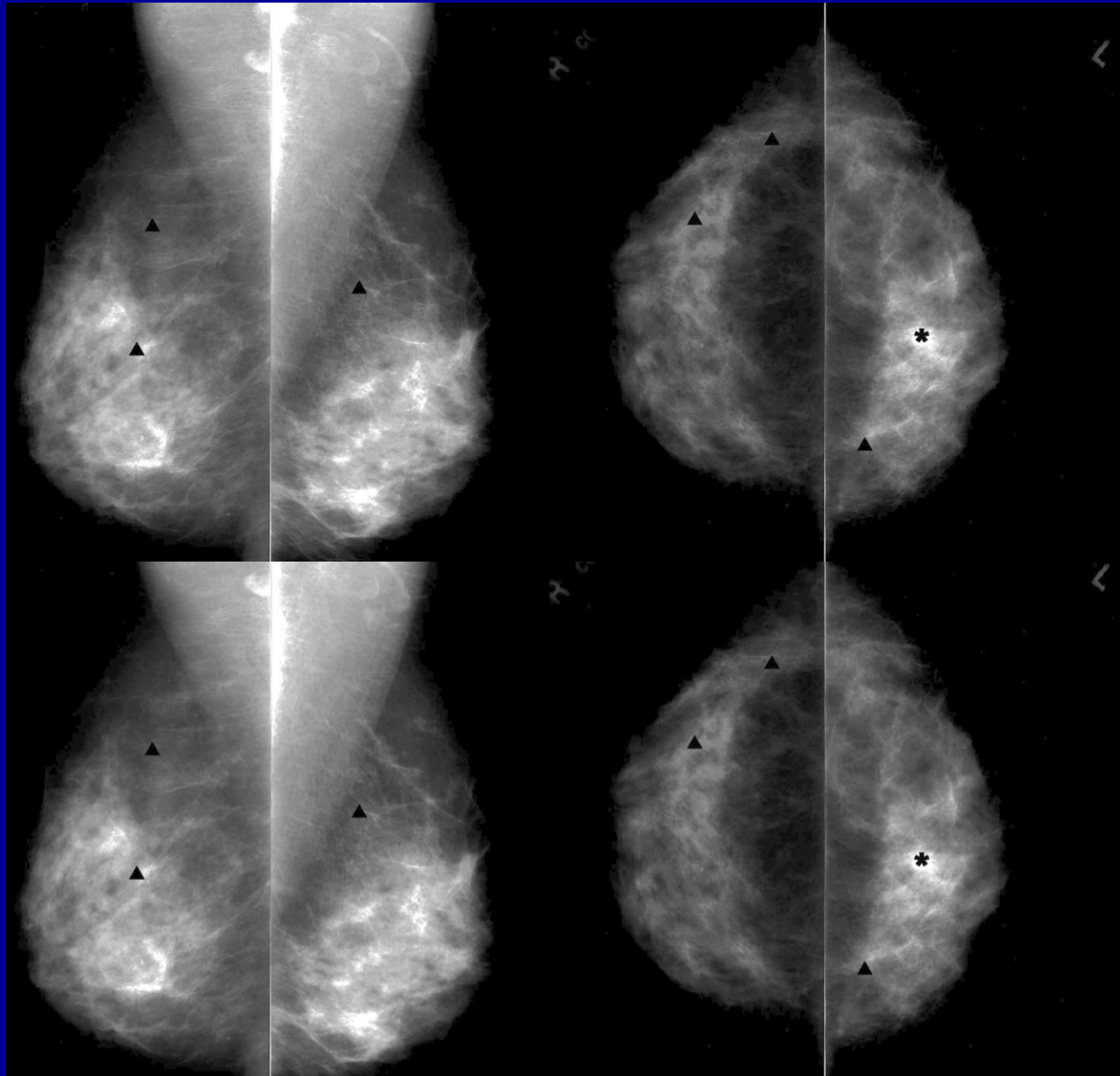
rmn29@pitt.edu

ADSA20 May 15, 2019

User Interface

- **Space:** *computer-aided detection (CAD)*
- **Problem:** *CAD is suboptimal when implemented clinically*
- **Solution:** *Need interactive user interface*
- **Results: Radiologists gain more benefit from using interactive interface**
- **TRL:** *10 (medical)*
- **Contact me: rmn29@pitt.edu**

Second Reader CAD

Interactive CAD

- **No CAD marks are shown**

- **Radiologist clicks on location(s) that they want help with**

- **CAD score is given, if CAD detected a lesion**

0.06
1
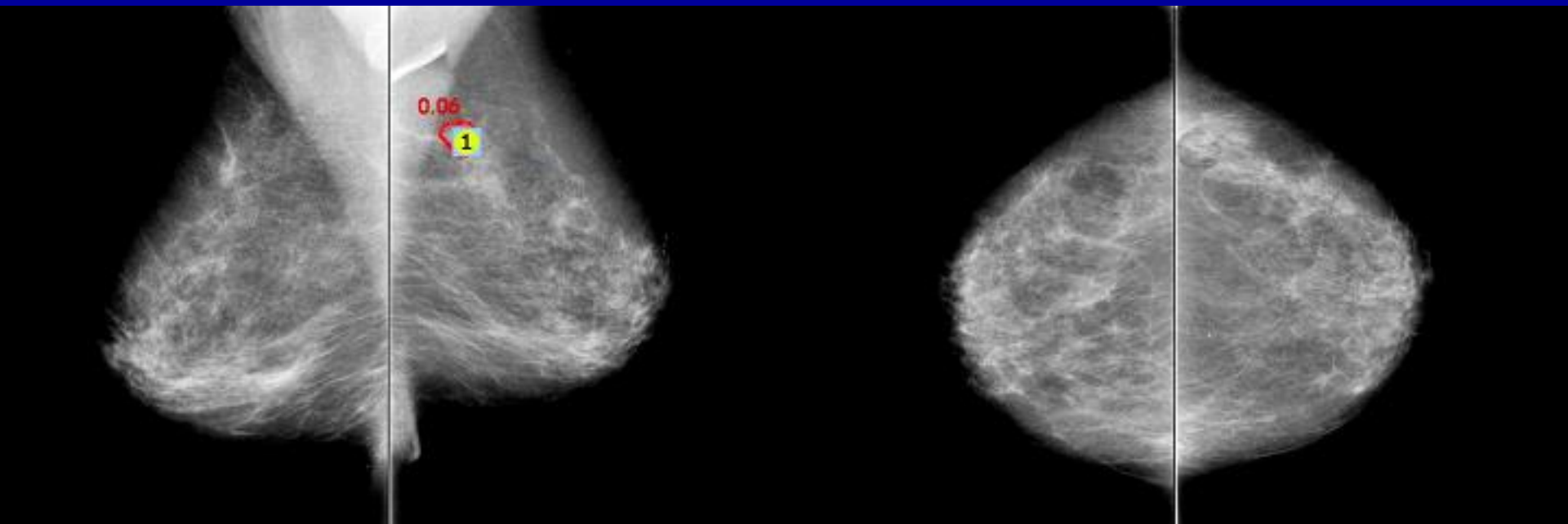
Interactive CADe

- **Takes advantage of high negative predictive value**
  - **Avoids the negative effects of low positive predictive value**

PD – PFA Curves

- **Space:** *Optimizing PD-PFA Curves*
- **Problem:** *How to optimize curve to maximize gain to user (both area under the curve and operating point)*
- **Solution:** *Determine what targets are most beneficial to user to detect and which false targets are most detrimental to the user to detect*
- **Results: Maximize benefit to user**
- **TRL:** *5 (medical)*
- **Contact me: rmn29@pitt.edu**

In Medical Imaging

- **Higher sensitivity (PD) is at the cost of higher false detection rate (PFA)**
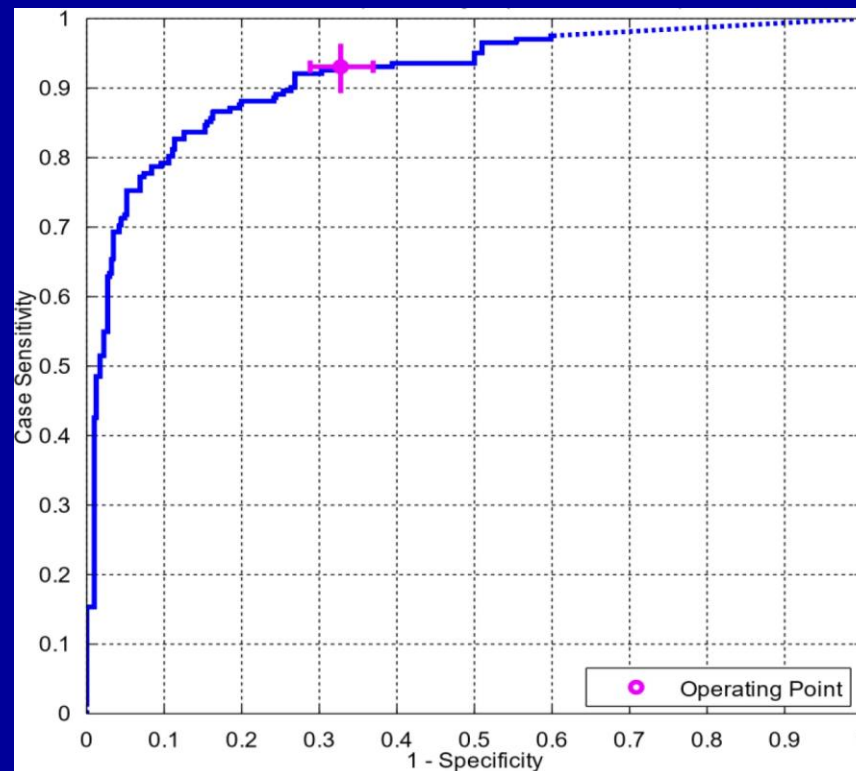
In Medical Imaging

- **At very high sensitivity, the computer will be marking cancers that radiologists will not call a cancer (because the radiologist will also be calling many non-cancers a cancer)**

- **Maybe possible that lower sensitivity and lower false detection rate will provide radiologists with better improvement in performance**

Extra slides

Radiologists Ignore Correct CAD Marks

- **Nishikawa et al. AJR 2012**

- **Observer study with 8 radiologists reading 300 screening cases (with a prior exam)**

- **105 cancers were initially missed by a radiologist and marked by CAD**
  - 30 times the radiologist recognized that they missed a cancer
  - 75 times they ignored the correct CAD mark

- **Radiologists ignored 70% of correct CAD marks**

Missed by 8 observers reading without CAD

3 picked up the cancer using CAD

Pathology: IDC
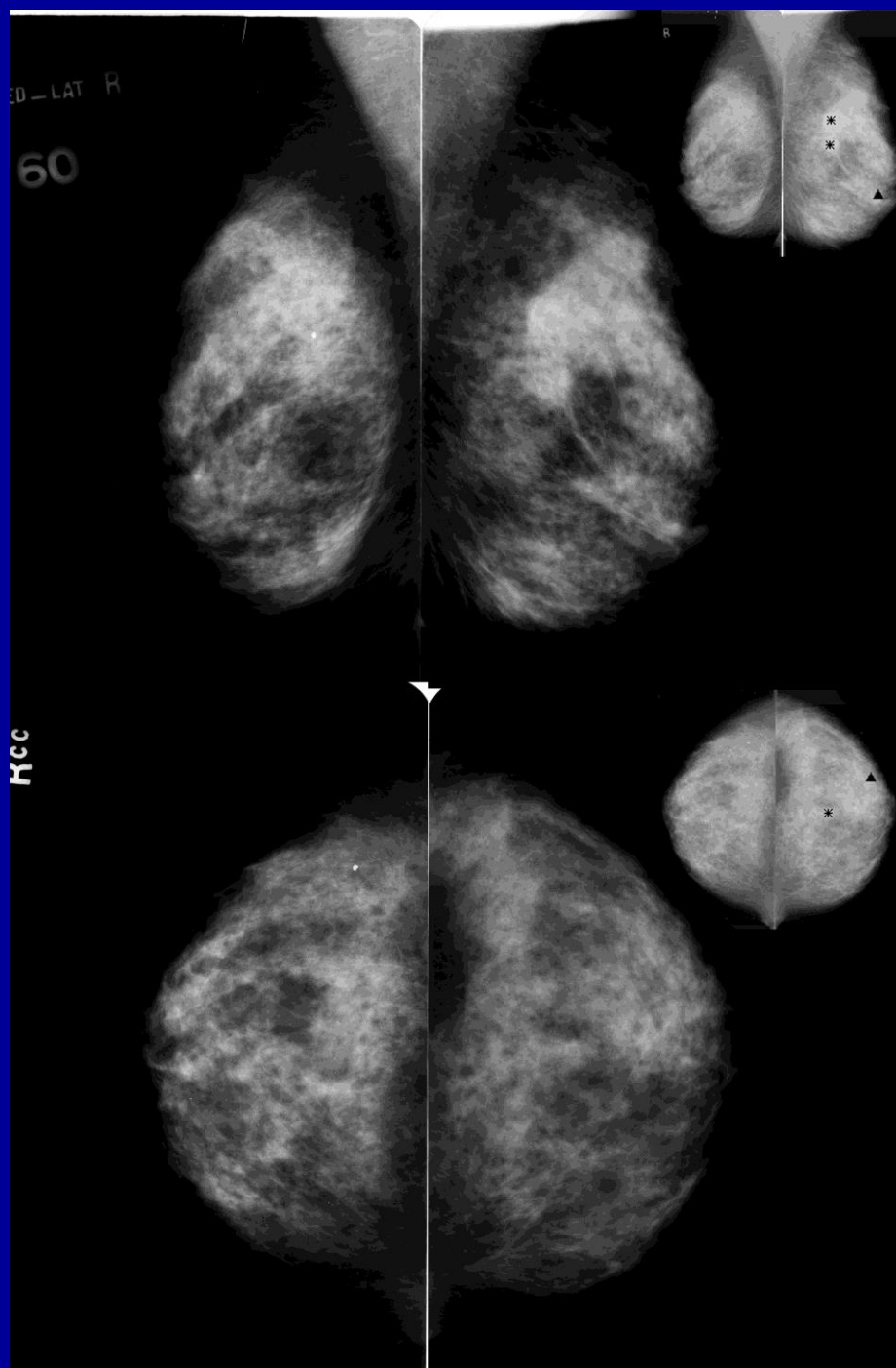
Miss Rate is Higher at Lower Prevalence

If You Don't Find It Often, You Often Don't Find It: Why Some Cancers Are Missed in Breast Cancer Screening

Karla K. Evans[1]*, Robyn L. Birdwell[2], Jeremy M. Wolfe[1]

1 Visual Attention Lab, Brigham and Women's Hospital, Harvard Medical School, Cambridge, Massachusetts, United States of America, 2 Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America

Abstract

Mammography is an important tool in the early detection of breast cancer. However, the perceptual task is difficult and a significant proportion of cancers are missed. Visual search experiments show that miss (false negative) errors are elevated when targets are rare (low prevalence) but it is unknown if low prevalence is a significant factor under real world, clinical conditions. Here we show that expert mammographers in a real, low-prevalence, clinical setting, miss a much higher percentage of cancers than are missed when the mammographers search for the same cancers under high prevalence conditions. We inserted 50 positive and 50 negative cases into the normal workflow of the breast cancer screening service of an urban hospital over the course of nine months. This rate was slow enough not to markedly raise disease prevalence in the radiologists' daily practice. Six radiologists subsequently reviewed all 100 cases in a session where the prevalence of disease was 50%. In the clinical setting, participants missed 30% of the cancers. In the high prevalence setting, participants missed just 12% of the same cancers. Under most circumstances, this low prevalence effect is probably adaptive. It is usually wise to be conservative about reporting events with very low base rates (Was that a flying saucer? Probably not.). However, while this response to low prevalence appears to be strongly engrained in human visual search mechanisms, it may not be as adaptive in socially important, low prevalence tasks like medical screening. While the results of any one study must be interpreted cautiously, these data are consistent with the conclusion that this behavioral response to low prevalence could be a substantial contributor to miss errors in breast cancer screening.
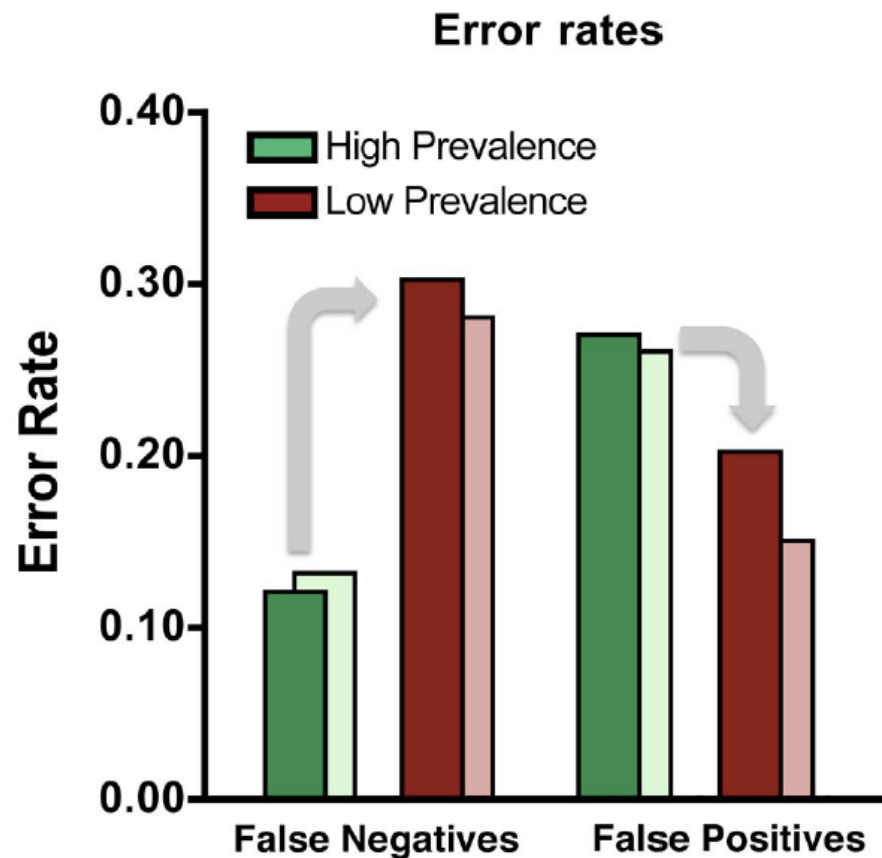
Figure 1. Error rates for rare targets (red bars, ~1% prevalence) and common targets (green bars, 50% prevalence) for two types of errors, false negatives and false positives. The dark colored bars represent data average over all 14 observers. The light red bars represent low prevalence average errors (false negatives and false positives) for the six observers who participated in both arms of the study (low and high prevalence). The light green bars represent high prevalence average errors (false negatives and false positives) restricted to the cases that the six high prevalence observers did not also see during the low prevalence arm of the study. Regardless of these filtering of the data, low prevalence, false negative errors are markedly higher than high prevalence false negative errors.
doi:10.1371/journal.pone.0064366.g001

Prevalence of TP CADe Prompts

- **Cancer prevalence in a screening population: 5 per 1000 women screened**

- **True positive CADe mark in a screening population: 2.5 per 1000 computer marks**
  (assuming 2 false detections per case and 90% sensitivity)

Lessons to Learn from Current CAD Implementation: Lesson #2

- **Radiologists don't use CAD like they are supposed to do**

FDA: Labeling of CAD Device

1. **Radiologists must review mammograms in the conventional manner prior to reviewing the CAD results**

    – **Reviewing the CAD results before reviewing the films could cause the radiologist to fail to examine the unmarked areas with adequate care**

    – *If not, might miss a cancer*

FDA: Labeling of CAD Device

2. **The CAD results assist only in the detection of suspicious regions of the mammogram**

    – **the absence of a mark should not dissuade a radiologist from investigating suspicious findings**

    – *i.e., recall rate should only increase*

Trend of Changes in Sensitivity and Recall Rate With the Use of CAD

Study	Sensitivity			Recall Rate (per 100 women screened)		
	Prior to CAD	With CAD	% Change	Prior to CAD	With CAD	% Change
Fenton (2007)	0.80	0.84	4.5%	10.1	13.2	30.7%
Fenton (2011)	0.80	0.81	1.8%	8.4	8.9	5.6%
Lehman (2015)	0.87	0.85	-2.3%	9.1	8.7	-4.4%

Trend in recall rate was statistically significant, $p<0.0001$
Trend in sensitivity was not statistically significant, $p=0.45$

Computer Human Interface

- DL algorithms will not necessarily be used in the way that they were intended

- More research is needed on how to implement AI so as to maximize added value to the radiologist

- *Without significant advances in how to best implement these tools clinically, the impact of more accurate DL algorithms are destined to be limited*