National Mammography Database, the Audit, and OUTLIERS

Society of Breast Imaging 2015- Orlando, Florida

Updates for 2015

- BiRads - 5th final
  - Tomosynthesis
  - Screening Breast US
- NMD - Updates
- Confidence intervals for audit
- Technologist Contribution

Lecture Outline

- Background and Basics
  - What is the Audit, how to perform it
    - How to Use the Audit
- National Mammography Database (NMD)
  - How Does it Relate to NMD
- How to use audit data
  - Outliers - How to identify
- Recent Audit updates
  - BI-RADS® 5th edition, …

Audit Definition:

- Noun:
  - a systematic review or assessment of something
- Verb:
  - conduct a systematic review of

Audit purpose: Improve patient care

Background of the Audit - People

- John Thornbury and Dennis Fryback-IVP
  - individual performance varies
  - Improvement through Feedback
    - How Imaging improvement relates to patient care!
- Ed Sickles
  - Discrete mammography results
- ACR NCI ACS - BI-RADS® Committee
  - Standardization!!

Mammography Audit Basics-MQSA

- 1993 - 1999 MQSA minimal requirement 900.12(f)/900.12(f)(1):
  - Each facility shall establish and maintain a mammography medical outcomes audit program to followup positive mammographic assessments and to correlate pathology results with the interpreting physician's findings. This program shall be designed to ensure the reliability, clarity, and accuracy of the interpretation of mammograms.
- Numbers of cases per Radiologist - 960/24 Mo
- Have a mechanism to follow-up biopsy recommendations
- Correlate Pathology to Mammography Report
- Review known false negative cases

Audit - MQSA Inspection Minimum Requirements

- MQSA does not require any Statistical calculations
- Inspectors will review the audit data to ensure compliance
- Inspectors do not have standards for audit results
- Inspectors do not collect audit data.

Basic Audit - ACR BI-RADS®

- Basic Audit Recommendations beyond MQSA
- Separate Screening from Diagnostic Examinations -
- For screening examinations
  - Recall rate
  - PPV1 - Likelihood of cancer on a screen
  - Cancer Detection rate on Screens
- Diagnostic Examinations
  - PPV2,3 % cancer on a biopsy recommendation or Biopsy done

Audit Basics II

- Calculations for individuals
- Calculations for Facility

Audit Basics - Set up

- Organizing a practice to collect the data is the hard part
  - It needs to be done Prospectively!
- Follow up SYSTEM to identify abnormal cases is key
- System to acquire Pathology from all cases or at least all abnormal cases is the hardest part for many facilities

Basic Data - Prospective Set Up

- Separate Screening from Diagnostic mammograms
- Recall cases - Primarily BI-RADS® 0, but also 4, 5 (or 3 on screen) - By individual and Facility
- Diagnostic mammograms - separate by indications
  - Symptomatic - Lump, Discharge
  - Follow up from screening
  - Other - early follow up, prior cancer history, Implants?

All cases

- Did you find any Pathology?
- Core Biopsy and Needle Localizations
- Follow up of All 4’s and 5’s (and 0’s)!!!
  - categorize it - True Positive, False Positive, …
- Were there false negatives?
  - If So - Review for Findings in retrospect!
  - Must Give feedback to the Radiologist (and others)!

ACR BiRads Clinically Relevant Audit

- Positive mammogram is: 0, 4, 5, or 3 on a screen
- Assumptions to make if there is no known follow up (or can't link to local cancer registry)
  - A Negative mammogram is a True negative
  - A Positive mammogram without cancer is a False Positive
- This puts all cases into one of the four categories: TN, FN, TP, FP

ACR BiRads: Clinically Relevant Audit Calculations

- Recall Rate of Screens (# + screens/# screens)
- Positive predictive value (PPV1, PPV2, PPV3)
- Cancer detection rate on Screens
  - #TP screens/# screens

Audit Statistics for Detected Cancers

- Basic Cancer stage at diagnosis
  - DCIS or Invasive
  - Nodes + or -
- Cancer size at Pathology
  - If not available, or intervening Chemotherapy use BEST Imaging size - Usually US, but depends

Audit Calculations B

- Percentage of invasive cancers that are node-negative
- Percentage of cancers that are “minimal”
  - DCIS or ≤ 10mm invasive size
- Percentage of cancers that are stage 0 or I
  - Stage one: ≤ 20 mm and node negative (microscopic, Isolated, cells, IHC detected, those are -)

What to Do with the Results

- Next Step:
  Compare Internally
  and to Standards
- Is Improvement Necessary?
- Is Improvement Possible?

COMPARATIVE DATA – NMD and BCSC and the Literature

- What is the NMD?
  - Current Data
  - Limitations
- What is the BCSC
- Where is the literature from?

What is the NMD?

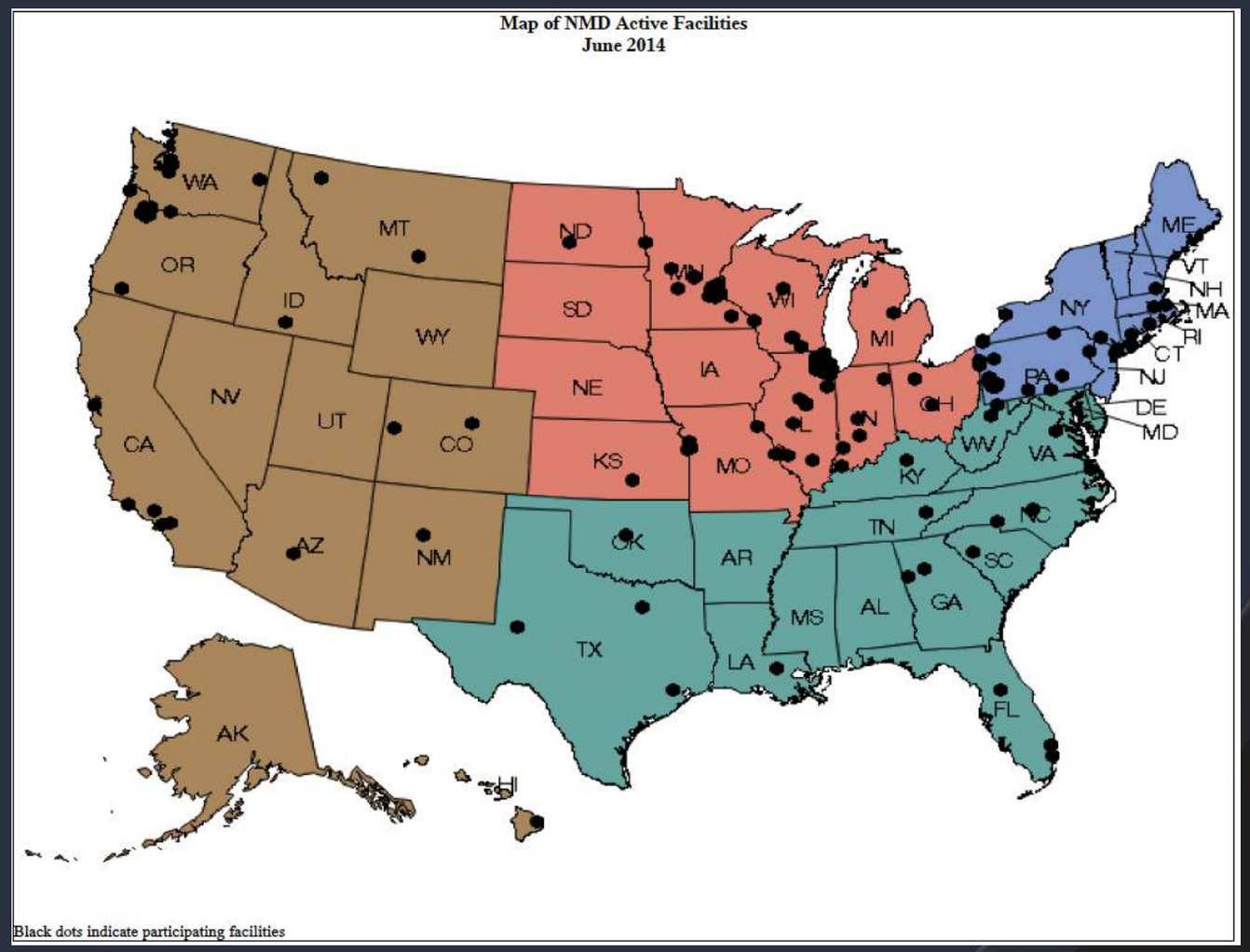
- ACR National Mammography Data Base
- Part of an ACR initiative for Comparative data
  - CTC - CT Colonography
  - GRID - General Radiology Improvement Database
  - ICE - Intravenous Contrast for CT
  - NOPR - National Oncologic PET Registry

NMD Requirements

- Voluntary Facility participation
- Some cost to facilities
  - Free to Centers of Excellence O/W $500-~$2000
- Requires Computerized systems with compatible software and export function
- HIPAA compliance Agreement (etc.)

NMD and MQSA Facility numbers

- MQSA November 2014 statistics
  - 8,718 Facilities, 13,709 Machines
  - 13,061 Digital machines in 8,215 Facilities
- NMD 2014 Data
  - 167- facilities
  - Data from 149 1,425,000 exams/yr
  - 5000 cancers/yr



NMD Composition

Facility Type	Number of Facilities
Academic	12
Community	72
Multi-specialty	29
Freestanding	53
Other	1

NMD Composition

Annual Procedure Volume	Number of Mammograms
< 5000	64
5- 10,000	32
10-30,000	45
30,000+	8

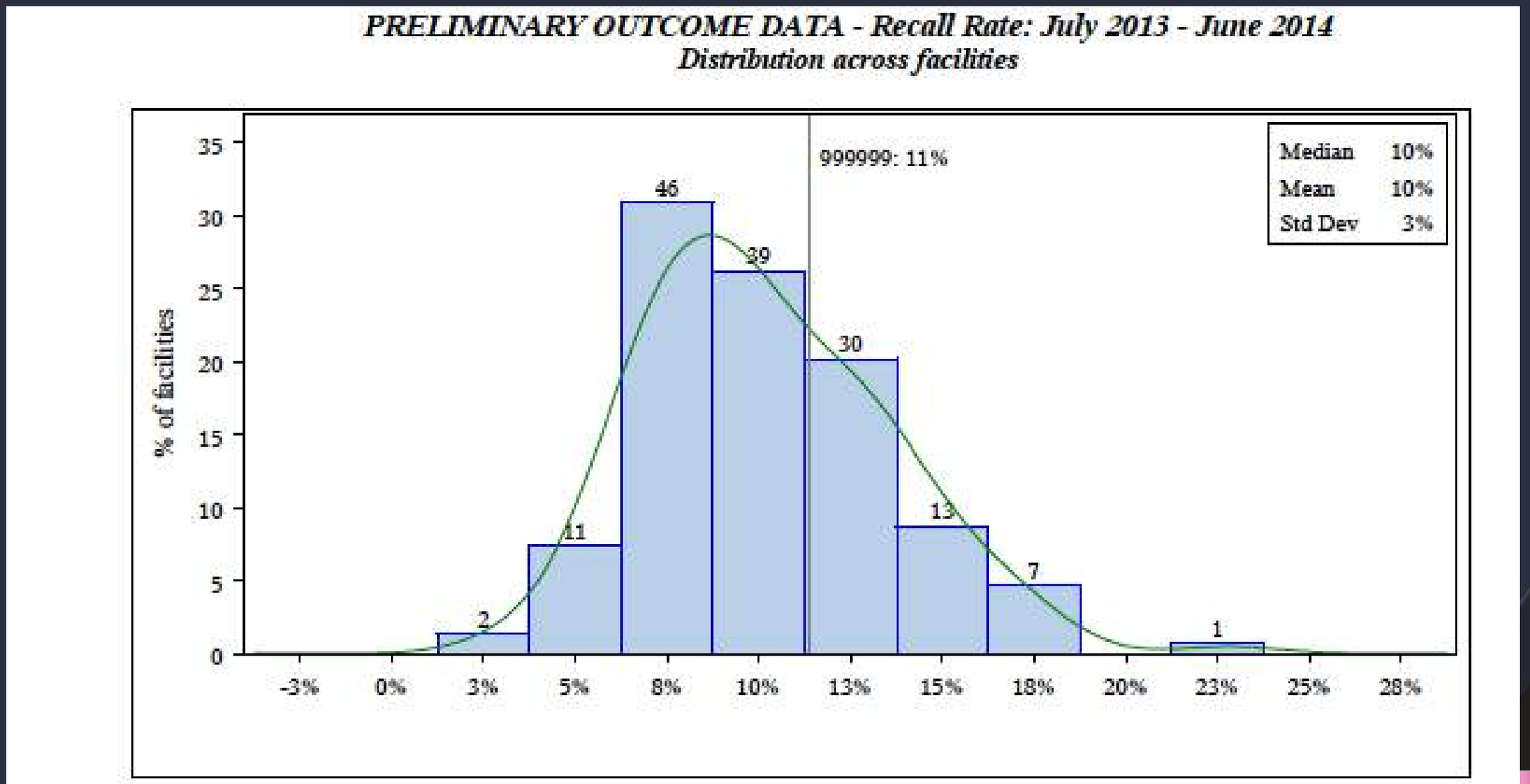
NMD provides full Audit comparison Data

- Recall rate
- PPV1(screen +)
- PPV2 (biopsy rec)
- PPV3 (Biopsy Done)
- Cancer results
  - CDR per 1000 , DCIS %, Node + Rate
    Invasive cancer Stage and size distribution
    Minimal cancer%

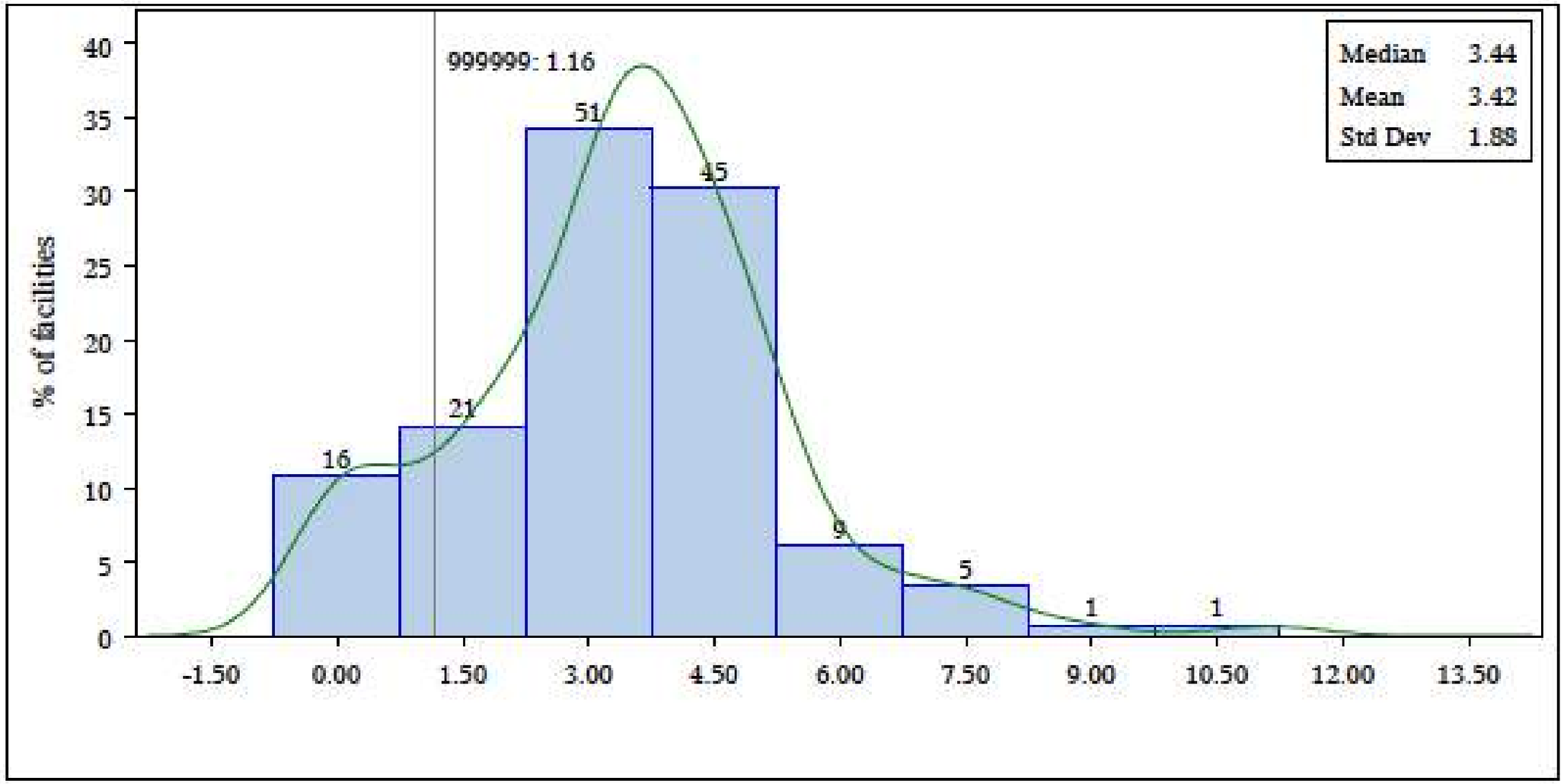
Comparisons Available

- All Facilities
- By Facility Type
- Region of the Country
- Radiologist
  - Comparison to other radiologists
- Also, includes BCSC comparative data

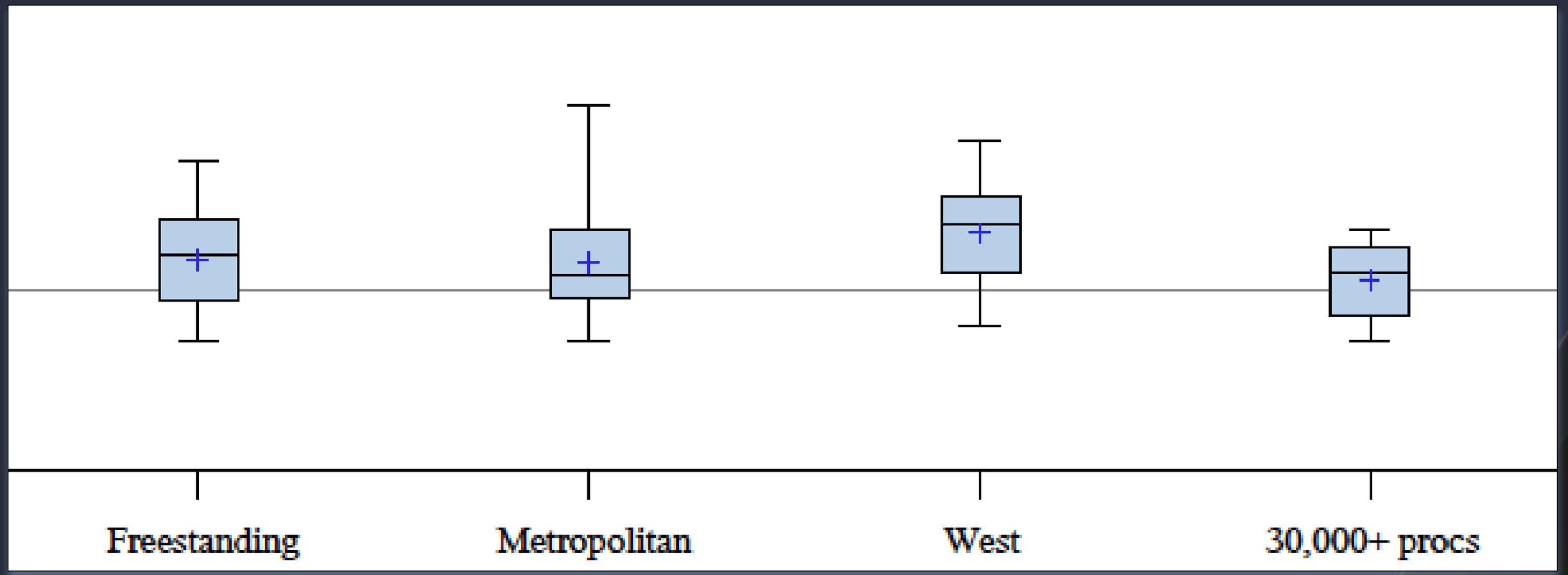
Sample NMD Summary Data – Recall Rate



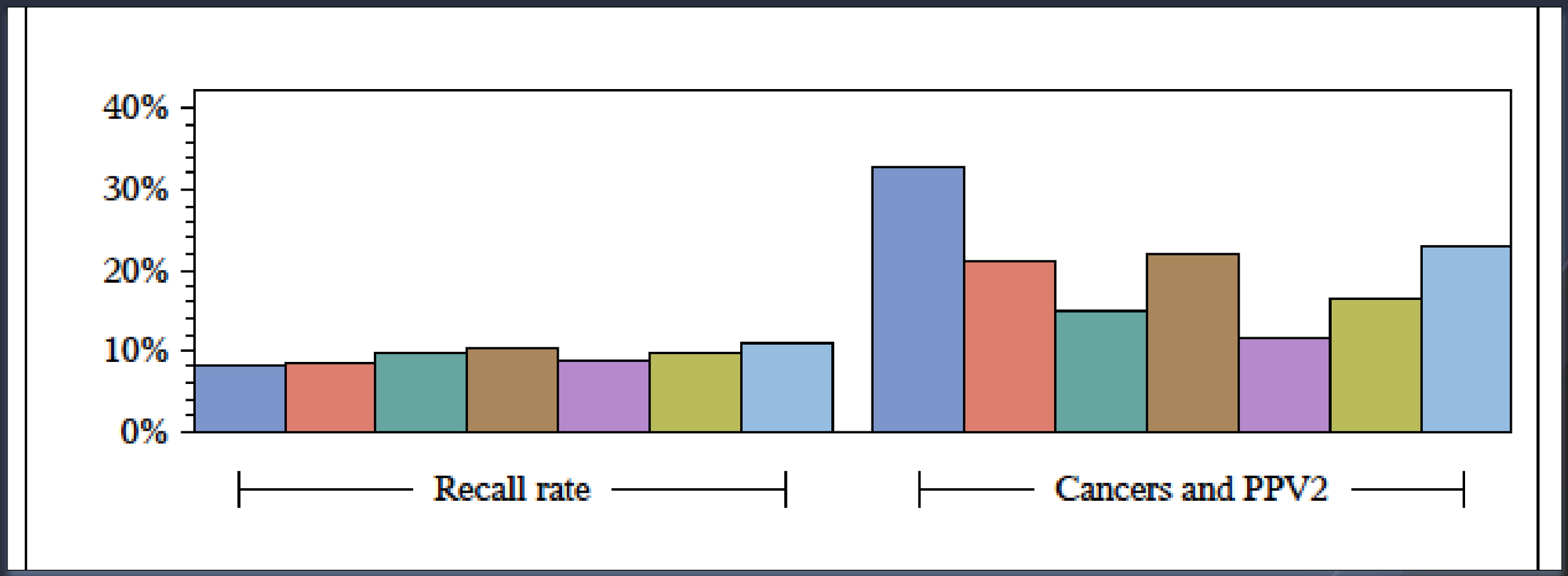
CDR 2012-13


999999: 1.16
Median 3.44
Mean 3.42
Std Dev 1.88
% of facilities
40
35
30
25
20
15
10
5
0
51
45
21
16
9
5
1
1
-1.50
0.00
1.50
3.00
4.50
6.00
7.50
9.00
10.50
12.00
13.50

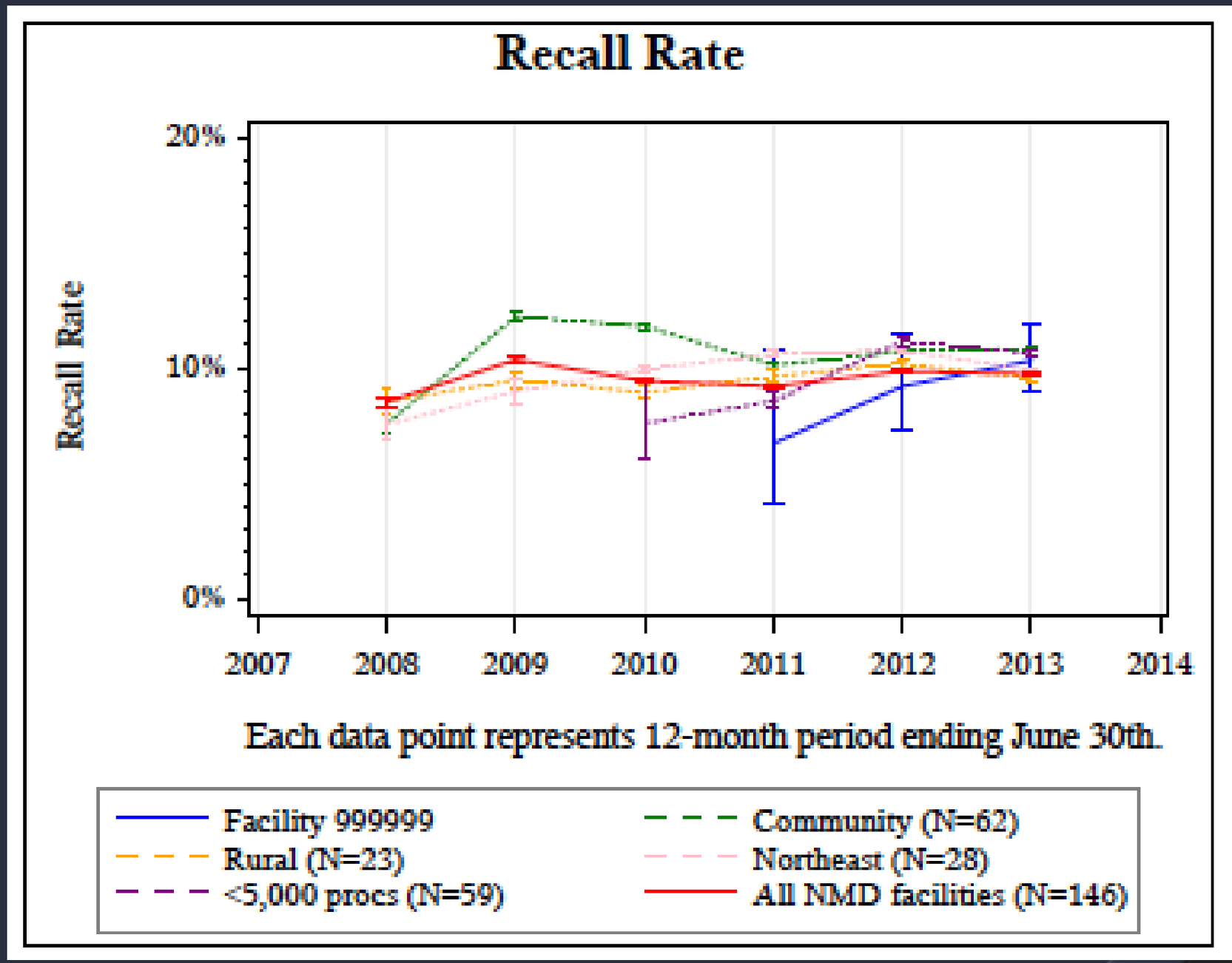
Comparison across Facility Types



Radiologist Level Comparison



Trend Info is becoming possible



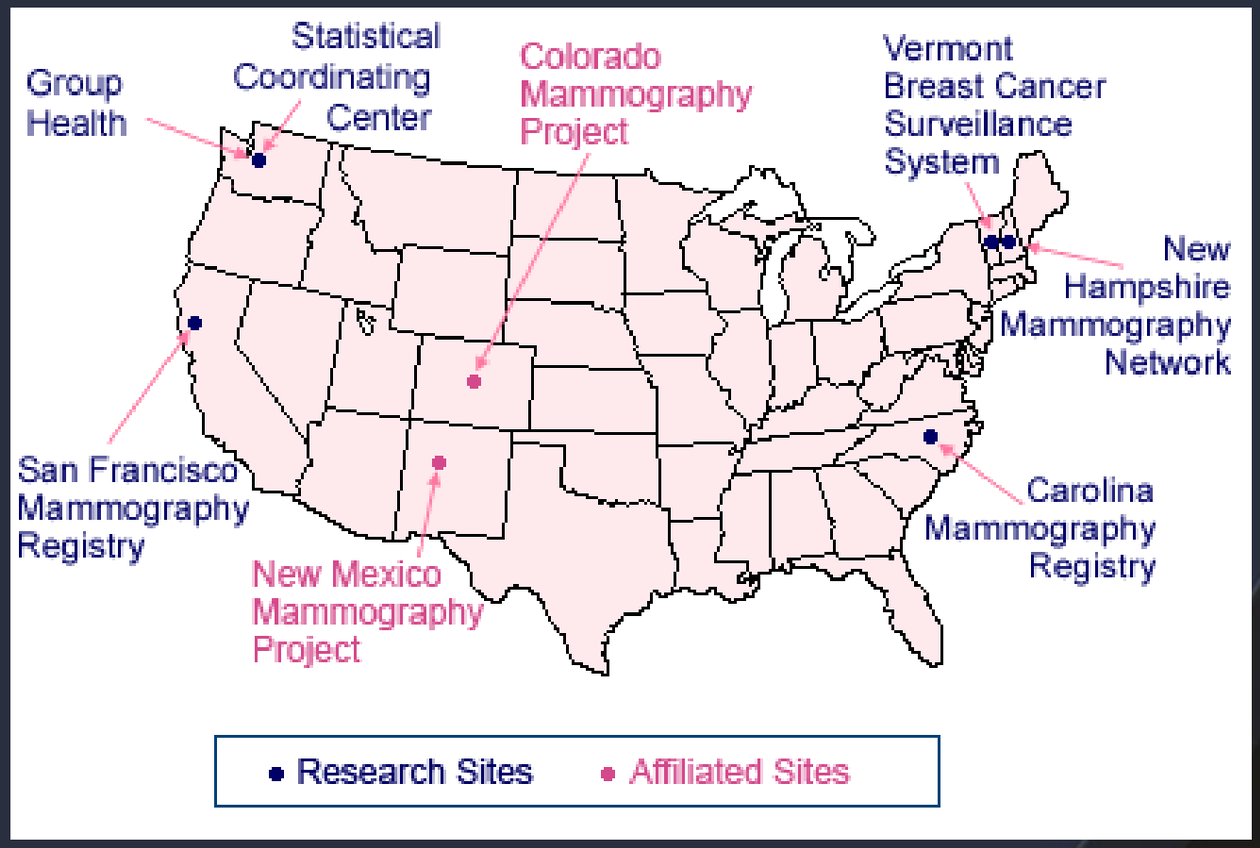
NMD and Research

- Research use of this data is beginning
- Cindy Lee of UCSF is the contact person for research questions.

- Cindy.Lee3@ucsf.edu

BCSC

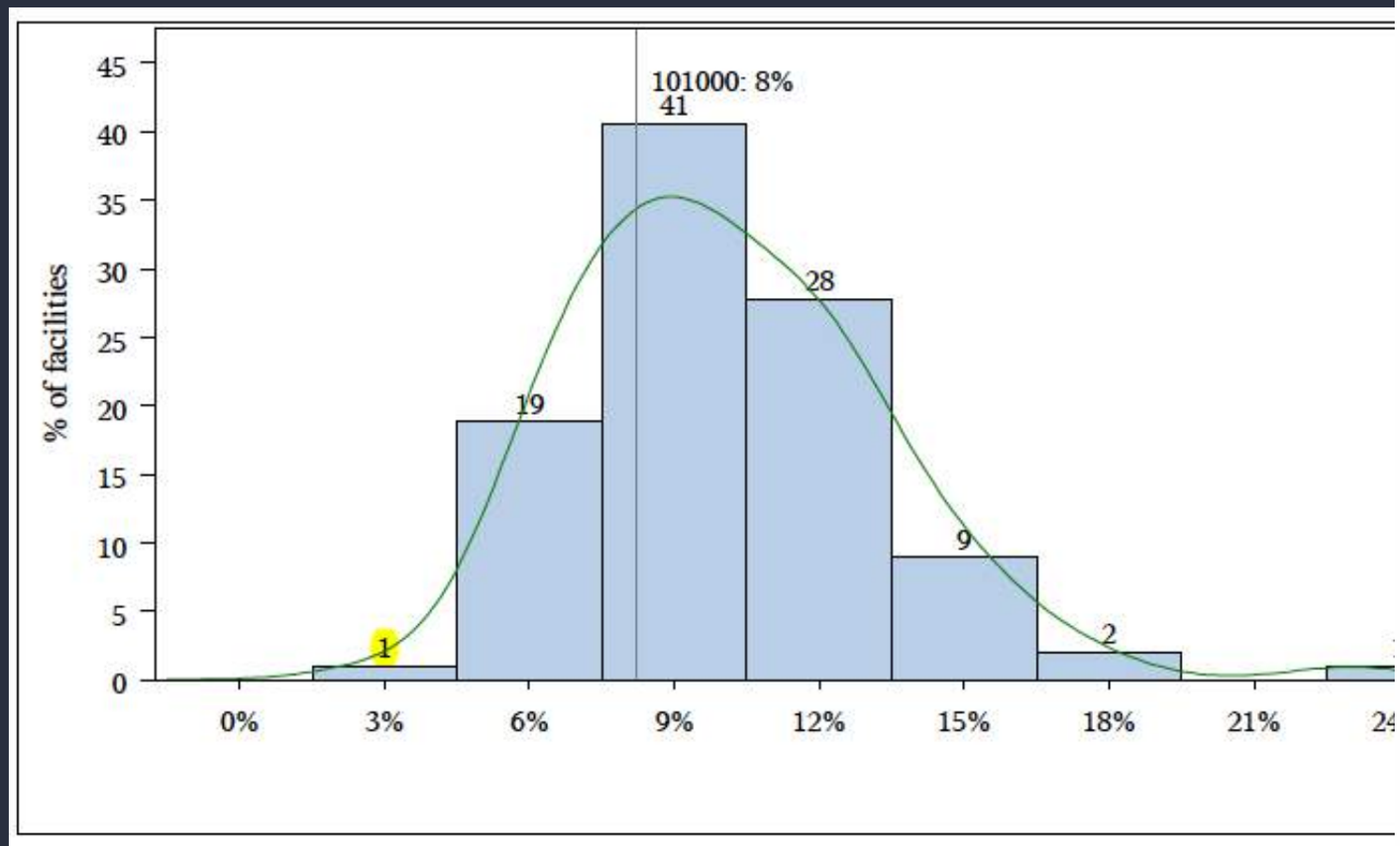
- Breast Cancer Surveillance Consortium
- Research Initiative, and Research quality data
- About 2% sample of facilities from ~1996 to 2010
- Will continue more limited data collection after 2010 - more analytic research plans
- Research Oriented Data Collection
  - Serves as reference standard



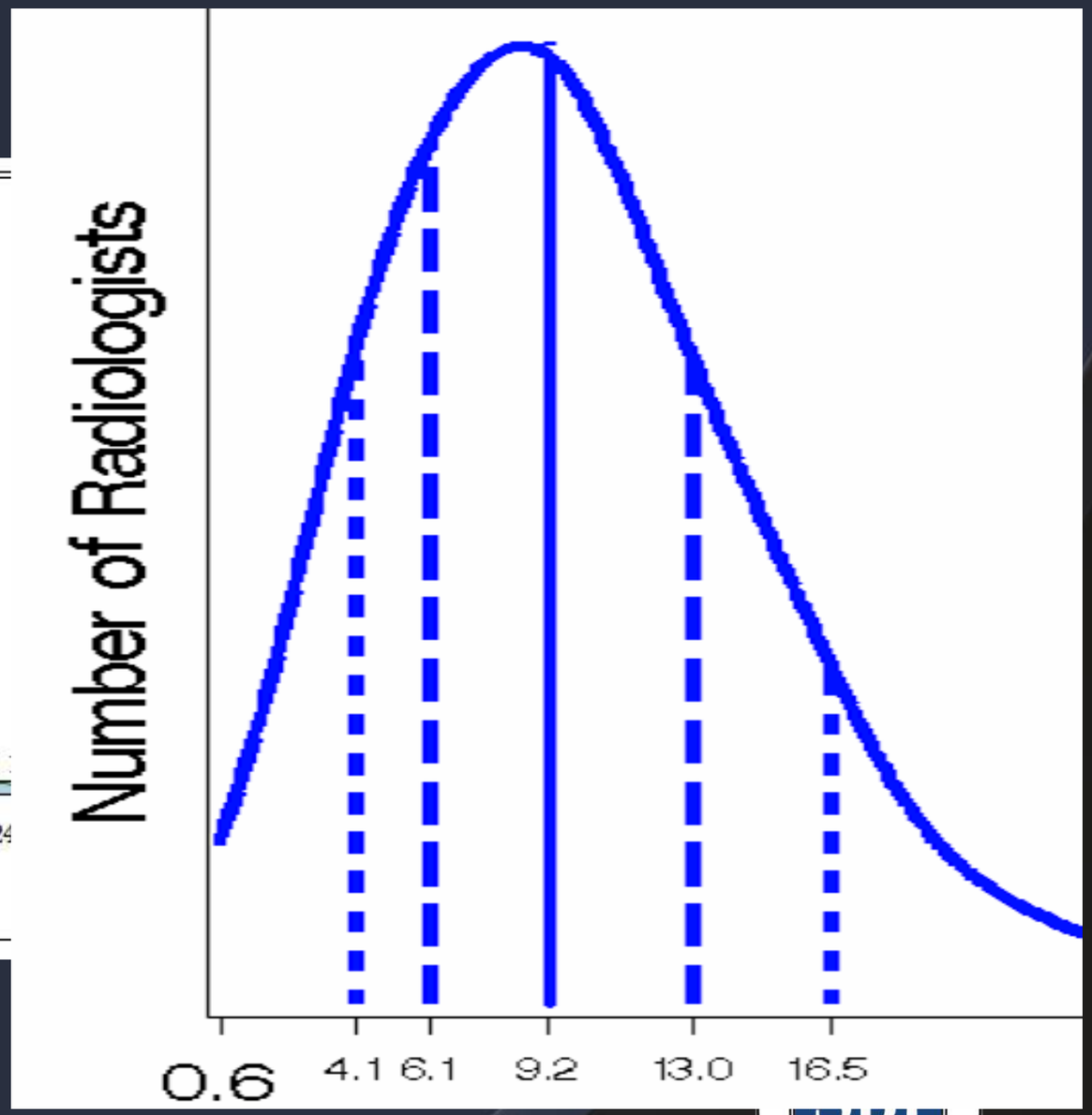
BCSC to NMD comparison

- Data at NMD: QA at the facility - NMD supplies feedback on completeness of data elements
- Data at BCSC has QA at multiple levels
- NMD facilities are voluntary
- BCSC facilities are Voluntary, but selected to be representative of US facilities
- BCSC Cancer data from registry linkage, NMD is facility report
- NMD has all measures by facility and Radiologist
- NMD is ONGOING, BCSC future data is limited

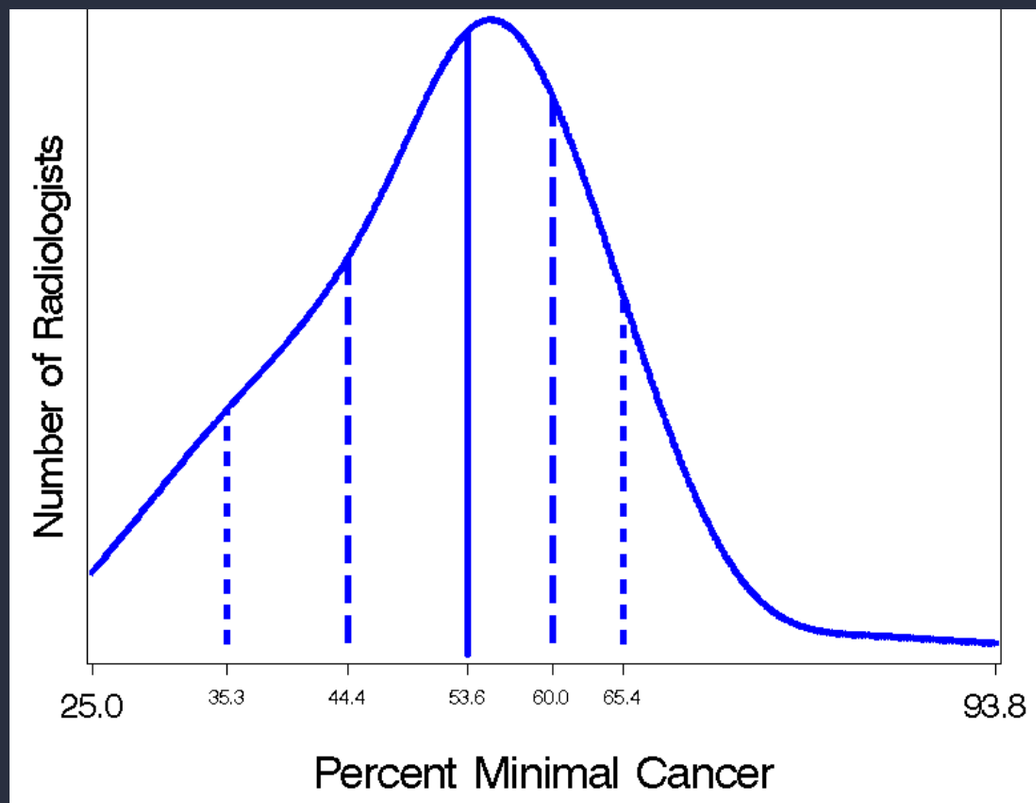
NMD Recall Rate

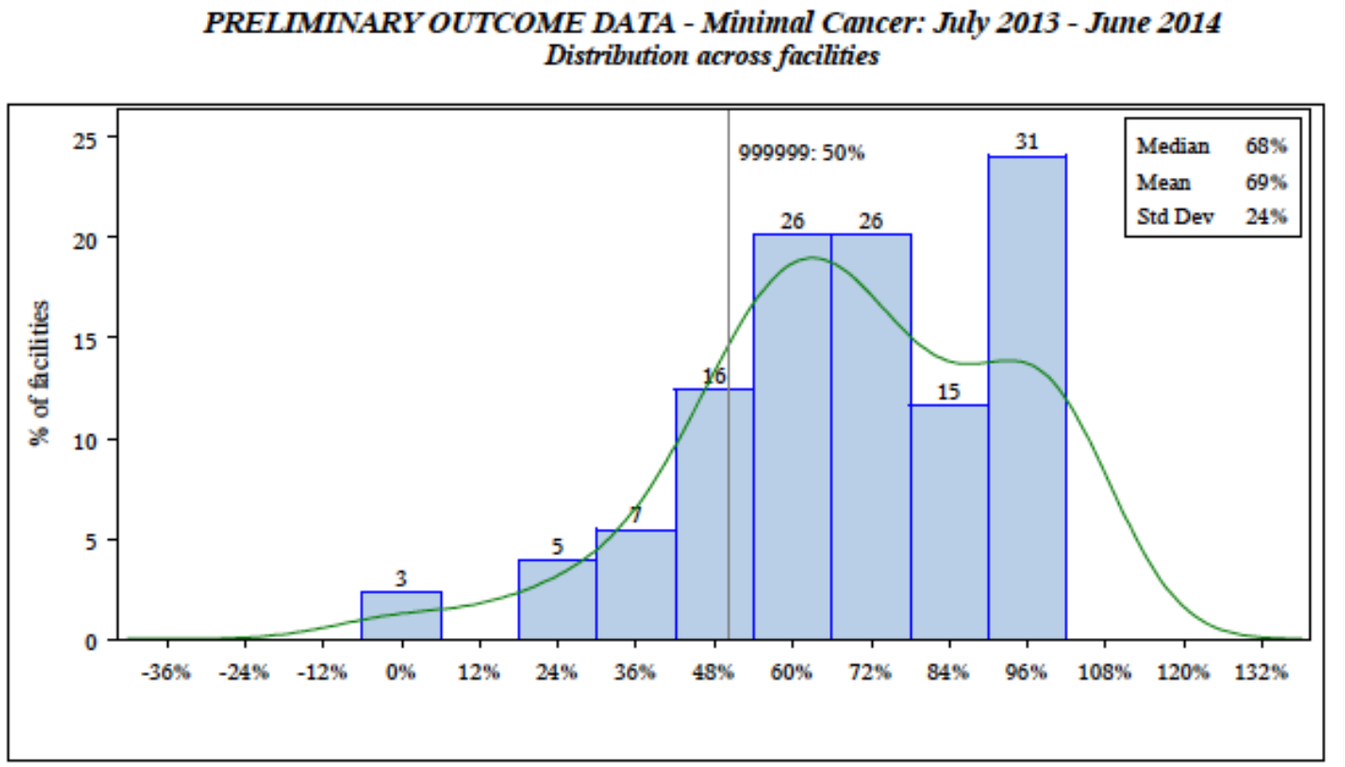


BCSC Recall Rate



Minimal Cancers BCSC and NMD





BCSC is Tumor Registry Quality Cancer data, NMD is highly Variable

Outliers

Not all measures are equal - Cancer detection is Key!

- How to Define?
  - Lake Woebegone? - How many Radiologists are self acknowledged below average?
  - How far below average is really important?
    - What is outside range matters!!
- How to be sure of the Statistics?
  - How Much Data and it's Randomness
  - Need Comparative data
  - Compare to group, and to standards where available

Carney et al. : Angoff Process - Screening and Diagnostic Exams

- 10 Experts in mammography -
  - academic and private practice
- Iterative process
  - Initial blinded recommendations
  - Look at group’s recommendations
  - Revise individual recommendations
  - Look at normative data from community - BCSC
  - Revise estimates

Carney PA, et al Identifying Minimally Acceptable Interpretive Performance Criteria for Screening Mammography Radiology May 2010 255:354-361 American Cancer Society Funded Process

Carney Angoff Process: Key Points

- Radiologists should be evaluated for multiple not just single measures.
- Outliers should be considered for additional education
- Numbers related to Cancer Measures are subject to significant statistical variation due to small numbers of cancer for most Radiologists.
- Imperfections to consider:
  - Assumes standard Screening population
  - Assumes Consistent data collection and Definitions

Screening: Carney et al

- Recall rate range 5 - 12%
- PPV1 (screen) 3-8% and PPV2 (biopsy recommendation) 20-40%
- Cancer detection rate >2.5/1000 screens
- Includes Sensitivity and specificity - not routinely available

Diagnostic Mammography: Carney et al
W/U of abnormal screening examinations

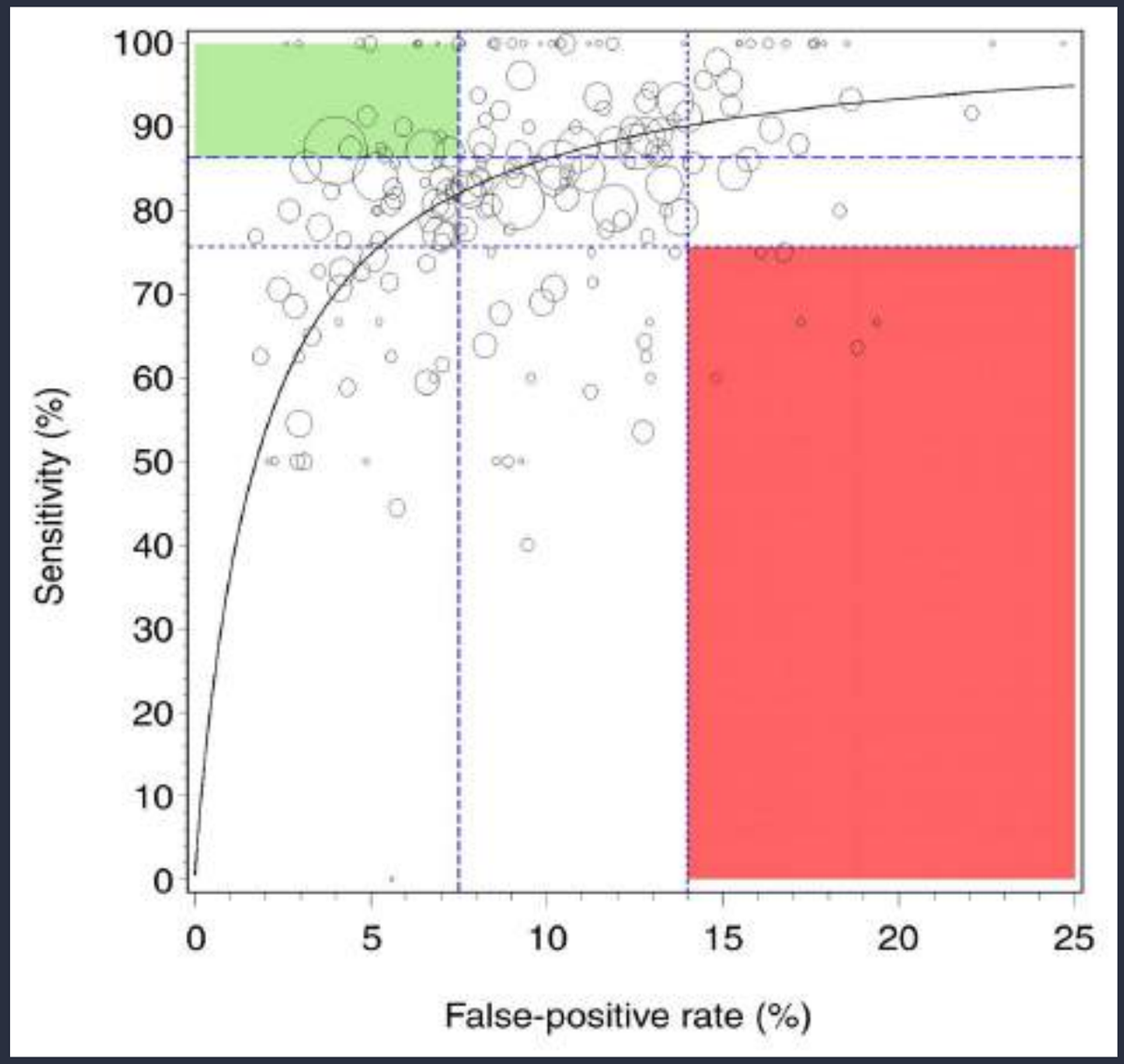
- Cut Points:
- Sensitivity, less than 80%;
- Specificity, less than 80% or greater than 95%;
- Abnormal interpretation rate, less than 8% or greater than 25%;
- PPV2 less than 15% or greater than 40%;
- PPV3 less than 20% or greater than 45%;
- cancer diagnosis rate, less than 20 per 1000 interpretations.

Radiology; Published online January 7, 2013, American Cancer Society Funded Process

Diagnostic Mammography: Carney et al Breast Lump

- **Cut Points:**
- sensitivity, less than 85%;
- specificity, < 83% or >95%;
- AbnI Intepretation rate, < 10% or > 25%;
- PPV2, less than 25% or greater than 50%;
- PPV3, less than 30% or greater than 55%;
- cancer diagnosis rate < than 40 per 1000 interpretations.

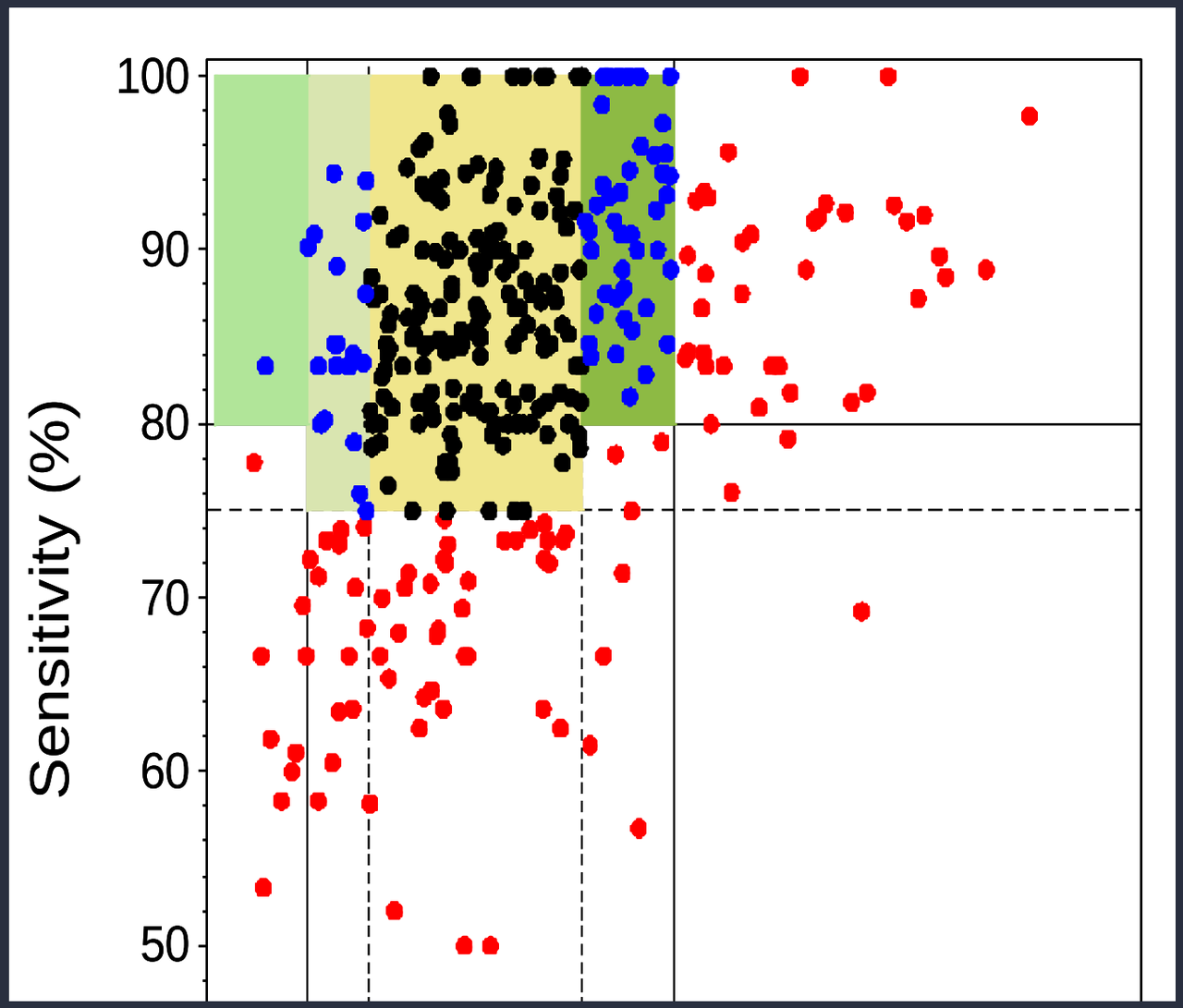
Sensitivity Specificity TOGETHER



Performance of 187 U.S. radiologists who interpreted images from screening mammographic examinations.

Elmore J G et al. Radiology 2009;253:641-65

Sensitivity Specificity Combined



In Press - AJR - Miglioretti et al

Outlier Other Considerations

- Odd Patient population?
  - Young or Old (high CDR)
  - Low SES or minorities Population (lower cancer rates)
  - More frequent screening or MRI screening (lowers mammographic CDR - Raises FN rate)

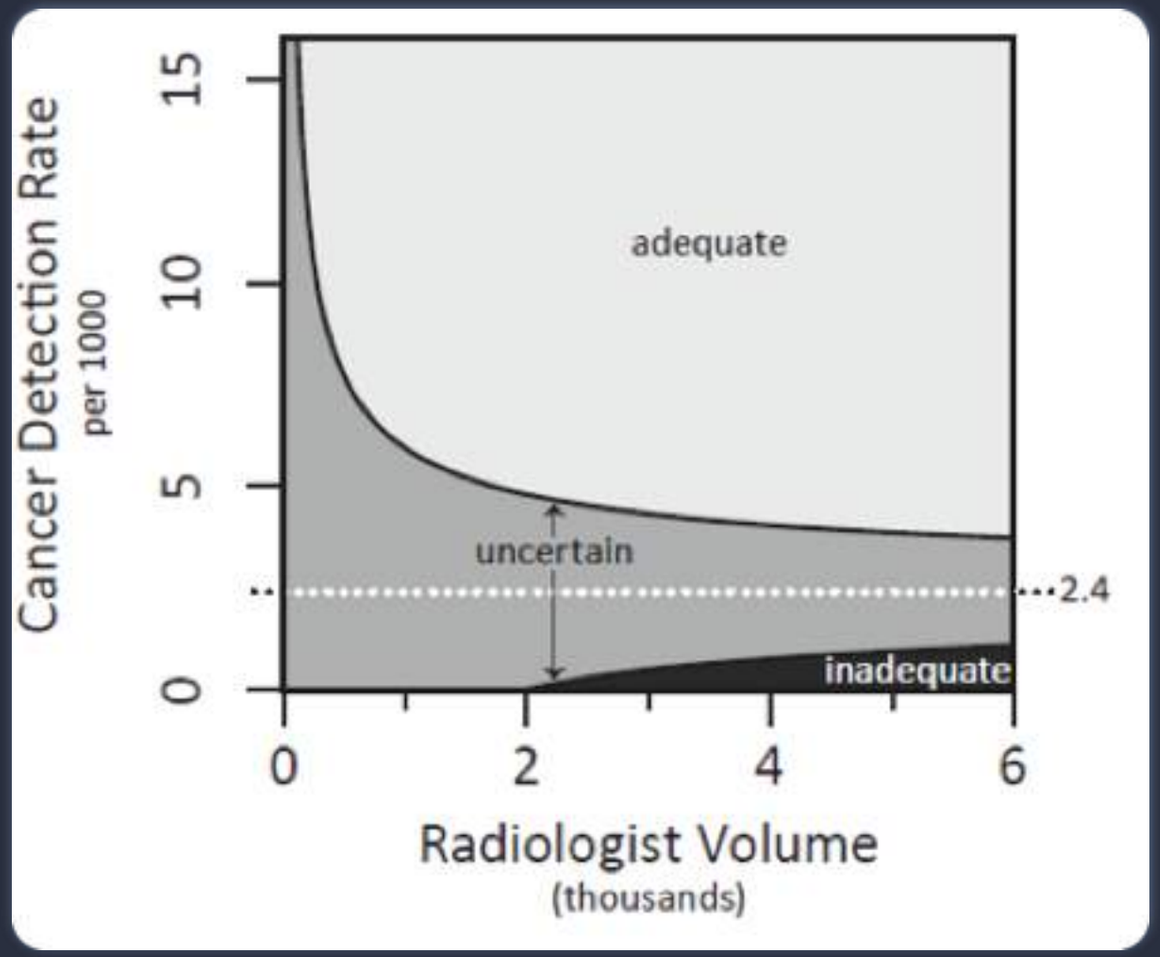
Problems in Applying “standards”

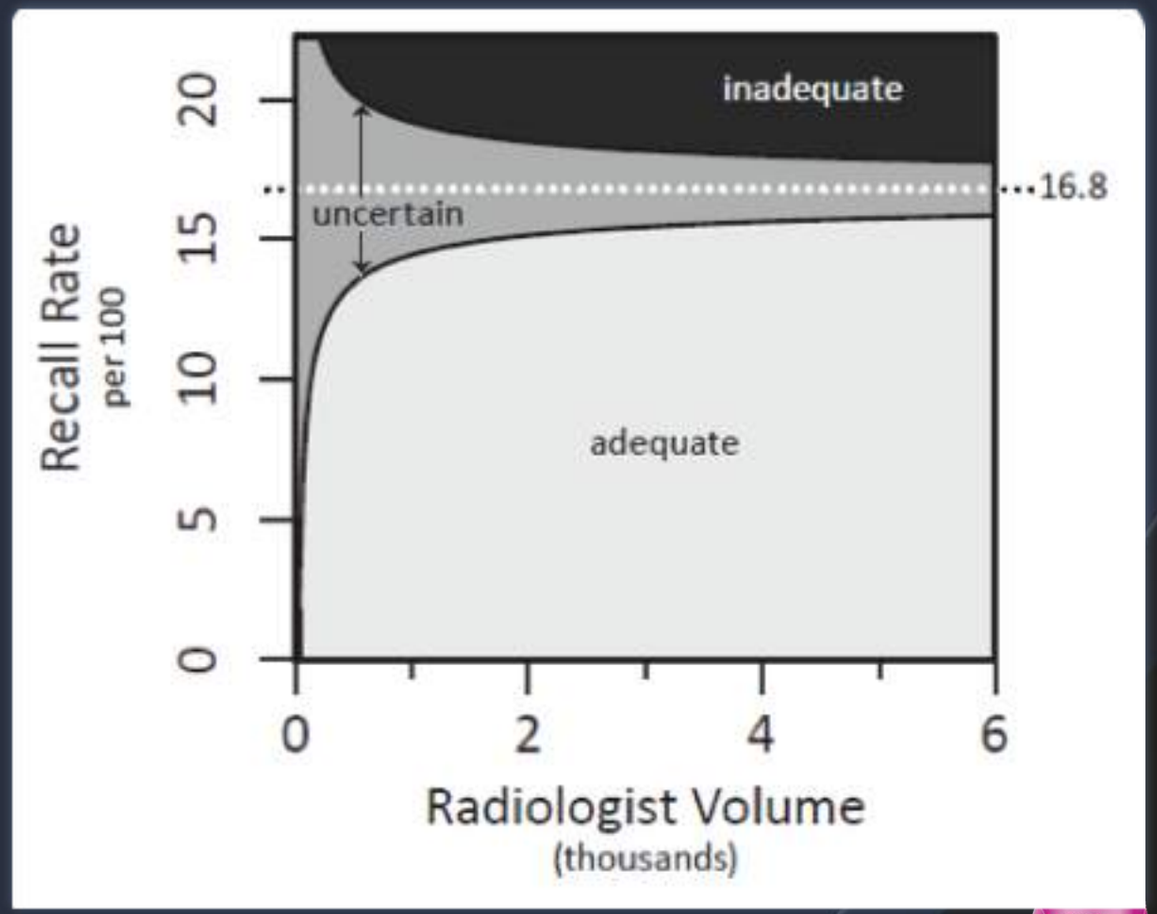
- Statistics and Randomness of small numbers
  - Will apply to anything related to Cancers - PPV’s, Cancer Detection Rate,
- Recall Rate
  - Goal is cancer detection
  - But - recall rate range is generous by most standards
    - European Guidelines is < 5 to 7% Some countries are under 4% with double reading

Small numbers Problem

- MQSA 480/yr means ~ 1 Ca/Yr, ~ 50 Recall/Yr.
- Recall Rate estimates are robust for most Readers with ~2000 mammograms
- Cancer Numbers - CDR, PPV's are not
  - May need to use several years of data, and whole facilities
    - This limits recent and individual's contribution
- Research in process dealing with the issue

Small Numbers Example





Cancer Detection Rate

Recall Rate

Courtesy of Elizabeth Burnside, University of Wisconsin, Madison

PLoS ONE 9(2): e89418. doi:10.1371/journal.pone.0089418

Underlying Problem

- Low Volume Radiologists will never really know how they are doing!!!!
- If aggregated for many years -
  - How relevant is it to now?
- If not aggregated
  - Variability of all measures - including recall rate - makes numbers meaningless

Possible Audit Conclusions:

- things are fine
- Was something found?
  - Low Cancer Detection
    - Check Recall?, FN? , PPV3 high?
- High Recall
  - One or all?
  - Check cancer Detection high?

Options

- Outliers within the Radiologists?
- Review of Cancer cases - Review of Statistics - Randomness?
  - Is Additional education Needed?
  - Last resort? Who should read mammography
- Review also Technical Issues?
  - Positioning, Prior Mammograms, etc
  - Patient population unusual?

Technologists are Important!

- Louise M. Henderson et al –
- The Influence of Mammographic Technologists on Radiologists' …
- Academic Radiology 2014 (ePub)

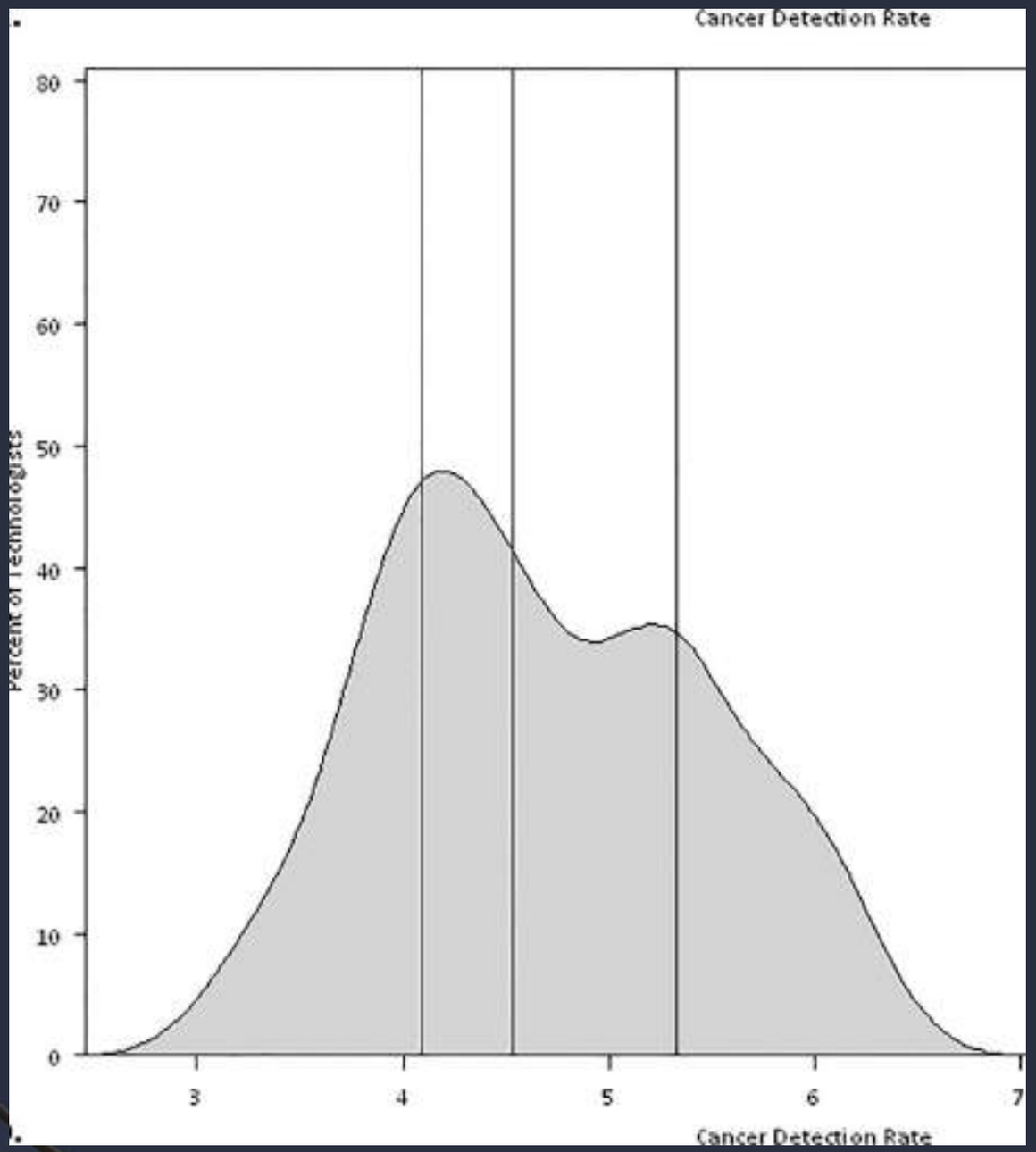
TABLE 3. Resulting *P* Values From Mixed-Effects Logistic Regression Models Assessing Variability of Performance Measures by Technologist

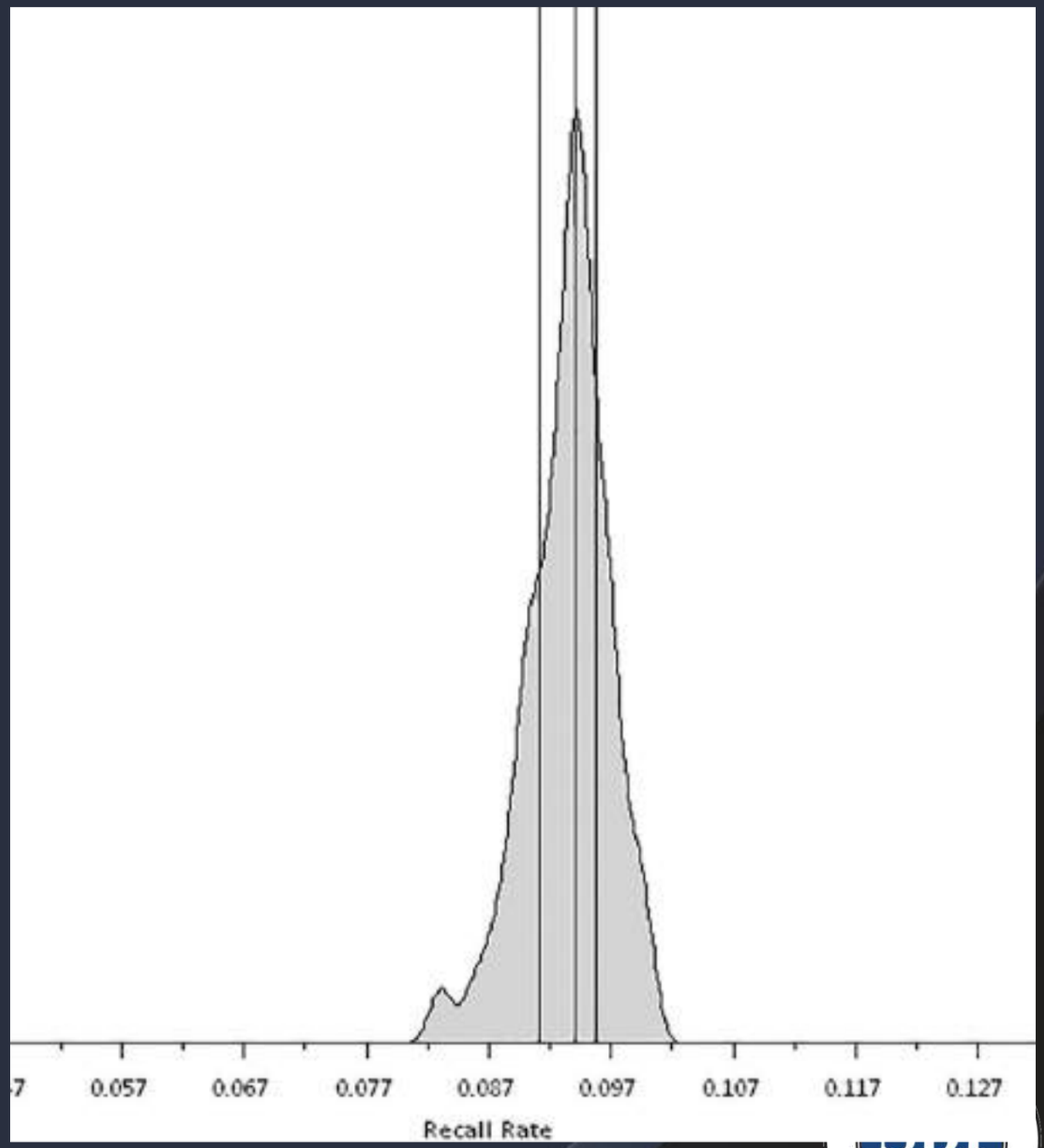
Performance Measure	SFM Examinations (n = 889,347)	FFDM Examinations (n = 113,929)
Recall Rate	<.0001	.001
Sensitivity	<.0001	.019
Specificity	<.0001	.003
Positive predictive value	<.0001	.088
Cancer detection rate	<.0001	.0001

FFDM, full-field digital mammography; SFM, screen-film mammography.

CDR Recall Rate

by Technologist





Do you get Enough Tissue?

Other uses of the results

- Consistency between Radiologists
  - It is confusing to Clinicians and Patients when there is excessive variability
  - Can something can be learned by exploring the differences?

Other important suggestions

- Review False Negative cases
- Cancers may be apparent on review of prior mammograms
  - Sometimes as subtle sub-detection findings
  - USUALLY not!!
- Therefore: review the prior mammogram whenever cancer is detected!!

Low Volume Radiologists

- Audit results are an Unknown Unknown
  - There is little information that is above the Noise to base many judgments
  - It will take a long time to accumulate any information

Other Audit Ideas

- BI-RADS® 4a,b,c
- Performance improvement by feedback
- Breast MRI Accreditation
- New Modalities
  - How is it changing outcomes
    - Tomosynthesis
    - Screening Ultrasound
- P4P - and BiRads 3
- Timeliness of Follow up

Deeper Information: ABC's of BI-RADS® 4 a,b,c

- 4a for likely benign - Example: Growing fibroadenoma like lesion - Should be > Bi-RADS 3 (< 2%)
- 4b Average biopsy recommendation
- 4c likely cancer, but not a BI-RADS®5 (> 95%)

BI-RADS® 4 a,b,c Uses

- Refinement of skills : BiRads3 and 4a's
  - Are all 4a's negative, or are many +
  - BI-RADS® 3 – are all negative? 1% or so positive?
- Use of BiRads 4a helps interpret PPV2 and 3
- 4a,b,c use **Helps colleagues** - radiologists and Pathologists
  - What biopsy recommendations are most likely to be cancer
  - Avoid Misuse of BI-RADS® 5 (>95% cancer) or 3 (<2%)

Recall Feedback

- … hypothesis that radiologists may improve their screening performance by performing the diagnostic work-up for their own recalled screening mammograms …
- Effect of Radiologists' Diagnostic Work-up Volume on Interpretive Performance
- Diana S. M. Buist, PhD, MPH, Melissa L. Anderson, MS, Robert A. Smith, PhD et al Radiology 2014

Tomosynthesis

- May consider tracking how well it works
  - Flag cases where Tomosynthesis used
    - Did Recall Rate Go Down?
  - Flag Cancers only seen on Tomosynthesis
    - Improved Cancer Detection?

Sample Tomosynthesis Audit

Exam Type	# Cases	Recall Rate	CDR	PPV1
Mammogram	25000	8.3	4.1	5.0
Tomosynthesis	6000	5%	5.8	11.4

Caveat: Cases for tomosynthesis are not randomly selected!!
Tomosynthesis cases selected by paying for it,
and or Patients with increased Density.
sk factors like DENSITY, age, family history, prior cancer history
will impact.
In our practice, prior biopsy seems
to increase likelihood of paying for Tomosynthesis

Screening Breast US

- Does it work as advertised?
- Find cancers not seen on conventional mammography?
- What does it do to overall recall rate?
- How does it compare to literature?

Brem et al – Radiology Nov 2014 epub:
Assessing Improvement in Detection of Breast Cancer with Three-dimensional Automated Breast US in Women with Dense Breast Tissue: The SomoInsight Study

MRI and the Audit - BiRads 5th and Accreditation

- **Breast MRI Accreditation requires a basic Audit**
  - Must use BI-RADS® (0-6)
  - Must establish/maintain medical outcomes audit program to follow up positive assessments and correlate pathology results with findings
  - "The audit must include evaluation of the accuracy of interpretation as well as appropriate clinical indications for the examination."
  - "Summary statistics and comparisons generated for each physician and for each facility should be reviewed annually by the lead interpreting physician."

MRI and US Audit Methods

- Just like Mammography - separation into indication is key
- Different indications have different goals
  - Screening
  - Extent of disease - and screening for multi centricity
  - Others
    - Chemotherapy monitoring
    - Additional evaluation of palpable/imaging abnormality

P4P

- Not using BiRads 3 - Probably Benign finding on screening images
- This is identified from CMS data
  - It may be useful to use audit data to confirm compliance
- Other measures have been proposed related to the Audit
  - Recall Rate range

New BiRads 5th Edition -

- Vendors have about one year from January 2014 to implement
- Audit procedures for Ultrasound and MRI
- In general they are Parallel to mammography audit process
- Similar Indications for examinations
  - Screening
  - Additional imaging following a screen
  - Symptomatic patient
  - Early follow up

NMD and BiRads 5th

- Export update by Vendors and QA by ACR is in progress
- Expect updated NMD data soon.

US Audit and BiRads 5th

- Diagnostic studies - Generally read US with the mammogram if done at the same time – One result includes both.
  - Single Report Avoids confusion - negative mammogram, positive US- which one counts? Do both get read by Clinician? Does patient get negative letter for the mammogram?
- Collaborative information - Information of the two studies are combined into one result – Oil cyst or Fat Necrosis for example
- Little benefit of reading separately - (my view)
- Audit of Diagnostic studies would include US information

Other Clarifications in new BiRads

- BiRads 3 on Screening examination is a positive assessment for audit purposes
- Double Reading
- Many FAQ answers especially examples of how to code cases

Screening Ultrasound and BiRads 5th

- Any additional images beyond routine is considered a Recall – same definition as for Mammography
- Simple for Automated US
- Less clear/complex for Hand Held studies

MRI and US Screening

- New Methods – Limited experiences
- Differences of opinion on how to integrate multiple screening methods into audit
- Differences of opinion of what is a positive screening US!!
- Variability in how it's done also contributes to this issue
- To Be continued ....

Patient Centered Medicine

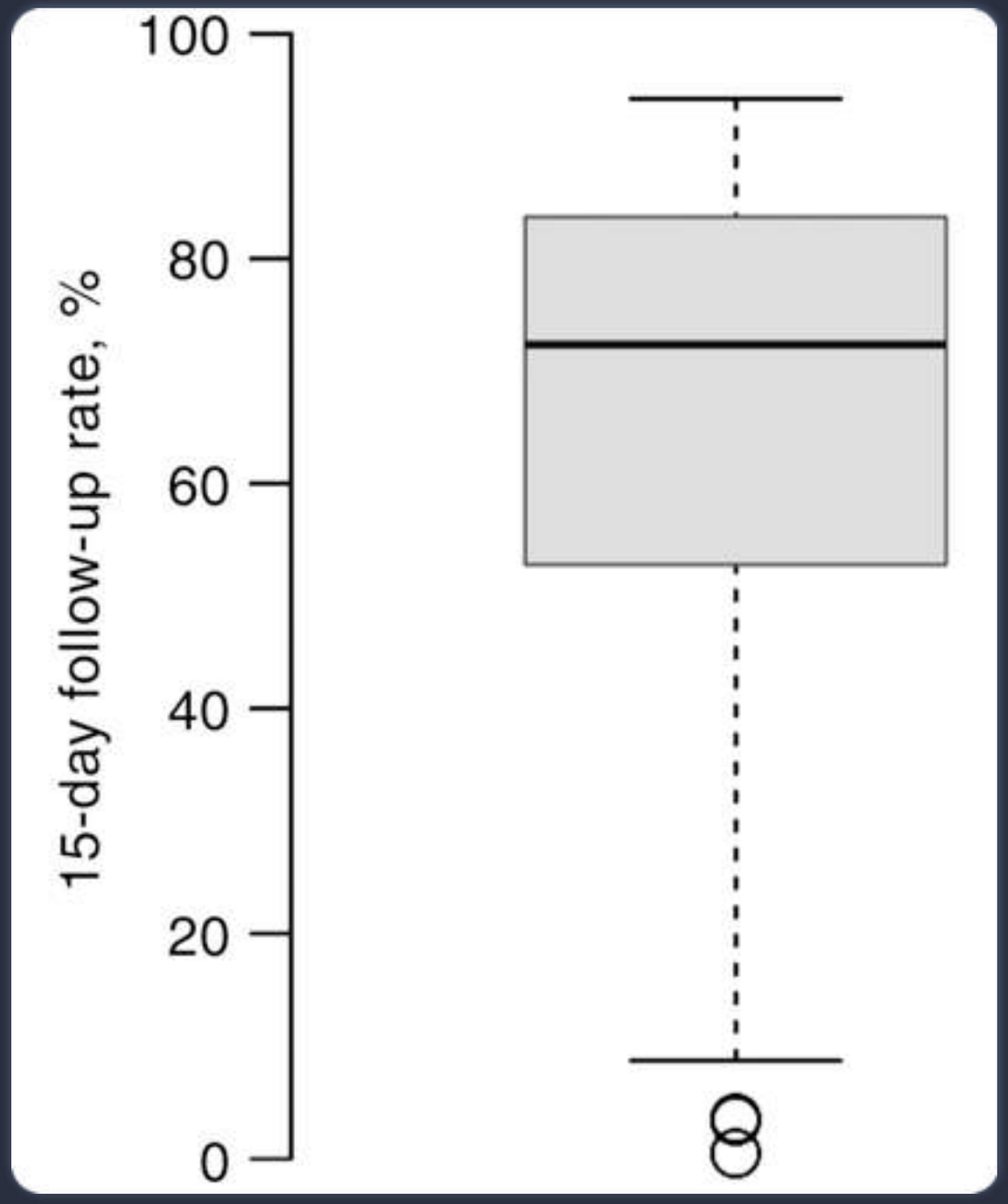
- Other measures for patient comfort and satisfaction with the process
- Time from screen to follow up and time from follow up to biopsy
- Wait times
  - for screens
  - symptomatic patients
  - Reporting delays

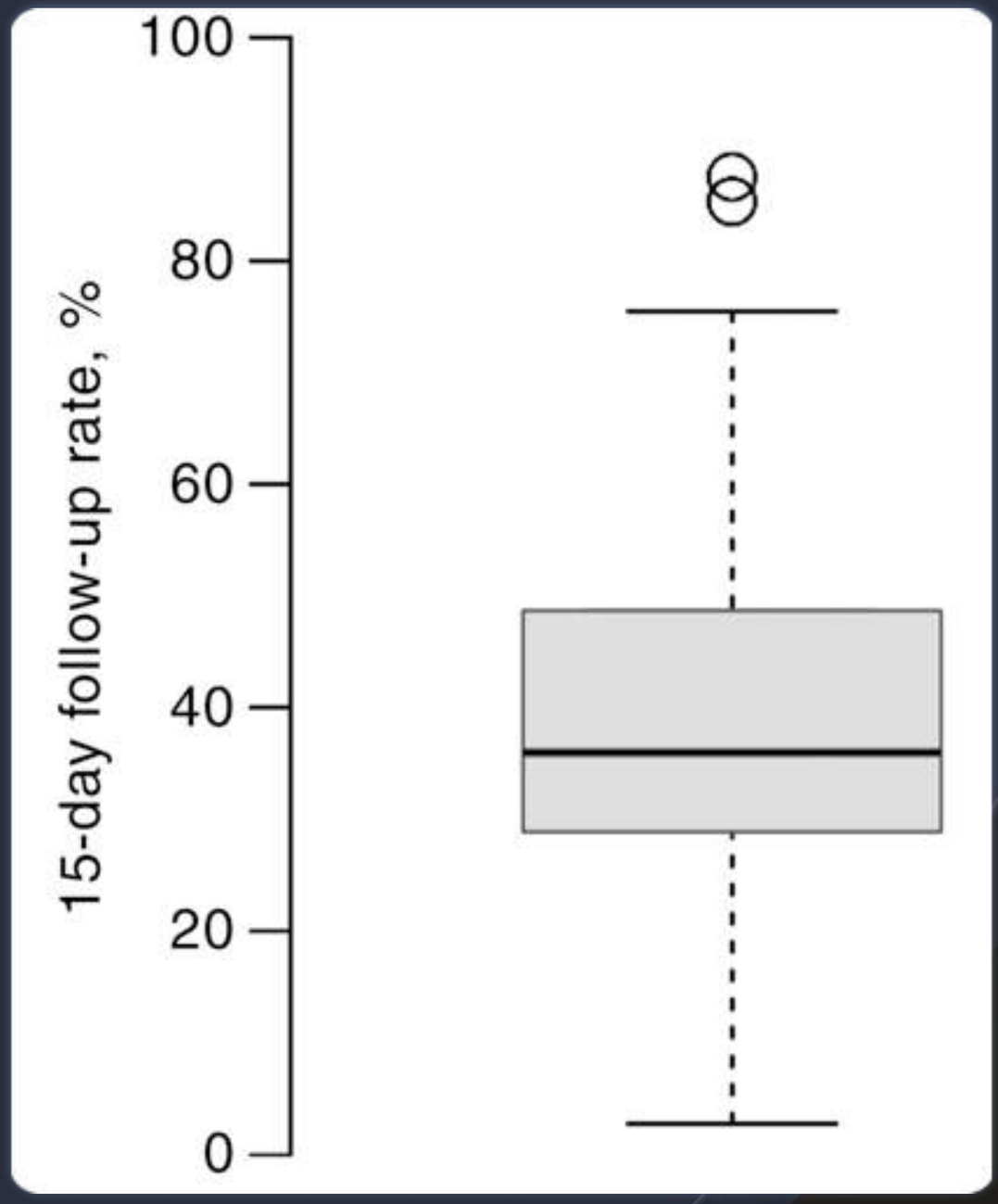
How long from abnormal screening study to the next Step?

- Patient satisfaction issue
- Comparative Data Now available
  - NCBC measurement
- Completeness is much more important - and should be followed!!

Additional Imaging

Surgical Follow up





Shaded area = 50% of facilities ○ = outliers

Facility Variation of Follow up of abnormal mammograms

Radiology

Rosenberg R D et al. Radiology 2011;261:404-413

SOCIETY OF BREAST IMAGING SBI Since 1985

Conclusions - Medical Audit

- Goal is improvement in quality and consistency
- One part of the QA process for Mammography
- The MQSA mandated audit is a minimal one
  - Performing the BiRads suggested minimal or complete audits is valuable
  - Review of images identified from the audit is valuable
- Low Volume is unknown Performance

National Mammography Database

- The NMD is a key resource for Mammography and Mammographers
- The larger the set of data submitted the more valuable the resource
- Data updated yearly
- Radiologists should participate if they can

Thank You!

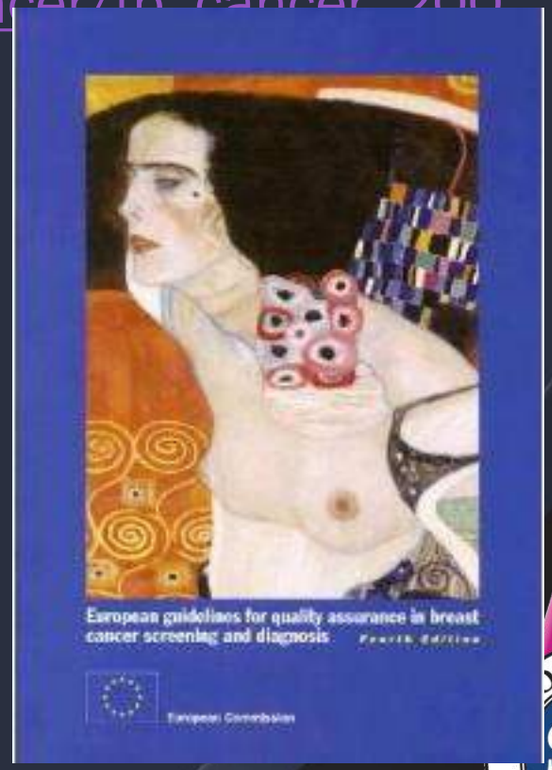
- Questions Please

References

- BCSC Website - Updated benchmarks and publications
  - http://breastscreening.cancer.gov/
- ACR National mammography database
  - http://www.acr.org/Quality-Safety/National-Radiology-Data-Registry
- ACR BI-RADS® 5th Edition
- - https://nrdr.acr.org/Portal/NMD/Main/page.aspx
- European Guidelines 4th edition
  - Executive summary available online pdf

European Guidelines - Complete System Measures and Recommendations

- European Guidelines for Quality Assurance in Breast Cancer Screening And Diagnosis (European Commission) - Primarily Expert opinion: Summary
  http://ec.europa.eu/health/ph_projects/2002/cancer/fp_cancer_2002_ext_guid_01.pdf
- Mammography
  - Radiologist
  - Technologist
- Pathology
- Surgery
- What and Why



Current Guidelines Available for Mammography Outcomes

- Identifying Minimally Acceptable Interpretive Performance Criteria for Screening Mammography. *Patricia A. Carney, et al* ***Radiology****, 2010, Vol.255: 354-361, 10.1148/radiol.10091636*
- Diagnostic Mammography: Identifying Minimally Acceptable Interpretive Performance Criteria. *Patricia A. Carney, et al.*
  - ***Radiology****, 2013, Vol.267: 359-367, 10.1148/radiol.12121216*

- National Mammography Database - ACR
- BI-RADS® Manual 5th
- European Guidelines for quality assurance in breast cancer screening and diagnosis - Fourth edition. (2006)
- ACR NMD /Updated BCSC Benchmarks data

References on performance

- Identifying Minimally Acceptable Interpretive Performance Criteria for Screening Mammography Radiology. Patricia A . Carney ,Edward A. Sickles, Barbara S. Monsees, et al: Volume 255: Number 2—May 2010
- Diagnostic Mammography: Identifying Minimally Acceptable Interpretive Performance Criteria **Edward A. Sickles, MD, , Diana L. Miglioretti, PhD, , Rachel Ballard-Barbash, MD, , Berta M. Geller, EdD, , Jessica W. T. Leung, MD, , Robert D. Rosenberg, MD, , Rebecca Smith-Bindman, MD, and , Bonnie C. Yankaskas, PhD**

Acknowledgements

- ACR and NMD Committee
- Breast Cancer Surveillance Consortium and NCI
- Participating Radiologists in the BCSC
- ACR and the BI-RADS® Committee
- Michael Linver
- American Cancer Society
- International Cancer Screening Network