National Mammography Database, and Practice Audits - Overview

Society of Breast Imaging
2016- Austin, Texas
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
- Additional Audit Uses and updates
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
- Ed Sickles
  - Discrete mammography results
- ACR NCI ACS - BI-RADS® Committee
  - Standardization!!
  - National Mammography Data base
- BCSC –Implementation and Methods
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.
Mammography Audit Basics-MQSA

  - 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
Basic Audit - ACR BI-RADS®

- 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
- Calculate for Individuals and for Facility/Group
Audit Basics II

- Calculations for individuals
- Calculations for Facility
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 (can’t link to a 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
What do I Do with the Results?

Next Step: Compare Internally and to Standards

Key Questions:
- 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 Feb 2016 statistics
  - 8,733 Facilities, 15,564 Machines
  - 15,228 Digital machines in 8,471 Facilities
  - 39,213,944 mammograms
- NMD 2015 Data
  - 177- facilities
  - Data from 149
  - 1,425,000 exams/yr
  - ~5000 cancers/yr
## NMD Facilities

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Number of Facilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
<td>16</td>
</tr>
<tr>
<td>Community</td>
<td>78</td>
</tr>
<tr>
<td>Multi-specialty</td>
<td>31</td>
</tr>
<tr>
<td>Freestanding</td>
<td>77</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>205</strong></td>
</tr>
</tbody>
</table>
## NMD Facilities

<table>
<thead>
<tr>
<th>Annual Procedure Volume</th>
<th>Number of Mammograms</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 5000</td>
<td>75</td>
</tr>
<tr>
<td>5- 10,000</td>
<td>41</td>
</tr>
<tr>
<td>10-30,000</td>
<td>51</td>
</tr>
<tr>
<td>30,000+</td>
<td>7</td>
</tr>
</tbody>
</table>
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 %
## ALL FACILITIES Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Number</th>
<th>Result</th>
<th>BCSC Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>All exams</td>
<td>1,747,500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall Rate</td>
<td>174,000</td>
<td>9.99%</td>
<td>10.95%</td>
</tr>
<tr>
<td>Biopsy Recommendation</td>
<td>26,000</td>
<td>1.5%</td>
<td>1.38%</td>
</tr>
<tr>
<td>Cancers and PPV</td>
<td>5,324/26,175</td>
<td>20.3%</td>
<td>22.9%</td>
</tr>
<tr>
<td>Cancer Detection Rate</td>
<td>6645/29,237</td>
<td>3.8</td>
<td>4.18</td>
</tr>
<tr>
<td>DCIS %</td>
<td>13,100/5,513</td>
<td>23.8%</td>
<td>23.6%</td>
</tr>
<tr>
<td>Nodal Negative Status</td>
<td>2,086/2,503</td>
<td>83.3%</td>
<td>76%</td>
</tr>
<tr>
<td>Tumor size 11-15 mm</td>
<td>658/2,506</td>
<td>26.3%</td>
<td>————</td>
</tr>
</tbody>
</table>
Peer 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
Sample NMD Summary Data – Recall Rate
CDR 2012-13

Median: 3.44
Mean: 3.42
Std Dev: 1.88
Comparison across Facility Types

Recall Rate
Radiologist Level Comparison

![Graph showing recall rate and cancers and PPV2 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
Breast Cancer Surveillance Consortium (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 – Similarity to NMD is Validation
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
Minimal Cancers

BCSC  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?**
- Someone is below average?
- What is really an outlier!!
- How to be sure of the Statistics?
  - What is real, what is Random?
  - 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

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.
Diagnostic Mammography: Carney et al
Breast Lump

- Cut Points:
- sensitivity, less than 85%;
- specificity, < 83% or > 95%;
- Abnl Interpretation 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.
Performance of 187 U.S. radiologists who interpreted images from screening mammographic examinations.

Elmore J G et al. Radiology 2009;253:641-651

Sensitivity Specificity TOGETHER
Sensitivity Specificity Combined

AJR April 2015 - 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, PPVs are not Robust
  - May need to use several years of data, and whole facilities
    - This limits recent and individual’s contribution
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
    - Check cancer Detection High?, on Target?
    - Low?
Options

- Outliers within the Radiologists?
- Review of Cancer cases - Review of Statistics - Randomness?
  - Is Additional education Needed?
    - Last resort??: Who reads mammography
- Also – Review Technical Quality?
  - 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

<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>SFM Examinations (n = 889,347)</th>
<th>FFDM Examinations (n = 113,929)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall Rate</td>
<td>&lt;.0001</td>
<td>.001</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>&lt;.0001</td>
<td>.019</td>
</tr>
<tr>
<td>Specificity</td>
<td>&lt;.0001</td>
<td>.003</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>&lt;.0001</td>
<td>.088</td>
</tr>
<tr>
<td>Cancer detection rate</td>
<td>&lt;.0001</td>
<td>.0001</td>
</tr>
</tbody>
</table>

FFDM, full-field digital mammography; SFM, screen-film mammography.
Influence of Technologists
Digital Exams

CDR

Recall Rate
Diagnostic Performance

Effects seen for Film screen, but not Digital

Film Screen performance Measures

Do Mammographic Technologists Affect Radiologists’ Diagnostic Mammography Interpretative Performance? AJR April 2014
Do you get Enough Deep Lateral Tissue?
Get behind the IMF?
Enough Medial Tissue

CC
Enlarged CC
Spot CC
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 suggestions

- Review False Negative cases
- Cancers may be apparent on review of prior mammograms
  - May show the earliest possible findings.
  - USUALLY no real finding!! Just some tissue
- 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
- How are New Modalities 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%) and <~10%
- 4b Average biopsy recommendation ~30-35%
- 4c likely cancer, but not a BI-RADS®5 (>95%)
BI-RADS® 4 a,b,c Uses

- Refinement of skills: BiRads 3 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%)
Feedback improves performance!

- 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

Tomosynthesis

- May/Should 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?
    - What does tomosynthesis miss that is on standard images or Ultrasound
Sample Tomosynthesis Audit

<table>
<thead>
<tr>
<th>Exam Type</th>
<th># Cases</th>
<th>Recall Rate</th>
<th>CDR</th>
<th>PPV1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammogram</td>
<td>25000</td>
<td>8.3</td>
<td>4.1</td>
<td>5.0</td>
</tr>
<tr>
<td>Tomosynthesis</td>
<td>6000</td>
<td>5%</td>
<td>5.8</td>
<td>11.4</td>
</tr>
</tbody>
</table>

Caveat: Cases for tomosynthesis are not randomly selected!! Tomosynthesis cases selected by paying for it, and or Patients with increased Density. Risk 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 Expected?
- Find cancers not seen on conventional mammography?
- What does it do to overall recall rate?
- How do you compare to literature?

MRI and the Audit - BiRads 5th and Accreditation

- **Breast MRI Accreditation:** 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 is Here

- 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
Not so New BiRads 5th Edition -

- Vendors should have implemented by now
- Audit procedures incorporate Ultrasound and MRI
- Audit Parallels 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 - Real soon now.
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
  – Standard images are one in each quadrant and one retroareolar
  – So two views of a cyst makes it a diagnostic!!
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”

• Viewed as increasingly important
• Anxiety is viewed as a decrement to mammography utility
  – Delays in interpretation
  – High recall rates
  – Delays in diagnostic evaluations and biopsies
• Competition
• Contracts may specify targets?
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!!
Facility Variation of Follow up of abnormal mammograms

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

Shaded area = 50% of facilities  ○ = outliers
Way out there? : Seeding Cases and Audit

- Insert known positive cases into routine screening list
  - Change exam date, name, and don’t bill for it.
- Adds interest and experience to screening
- Can assess actual sensitivity for same cases
  - Could also insert “problem” cases into mix to see frequency of recall.

- J Wolfe et al – RSNA Nov 30 2011
Justifications for screening mammography – and the Audit

• How do you compare to the Average?
  – Recall Rate
  – False Negatives
  – Timeliness of screening process
  – Availability of Studies
  – Cancer Detection Rate?
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
European Guidelines - Complete System Measures and Recommendations

- European Guidelines for Quality Assurance in Breast Cancer Screening And Diagnosis (European Commission) - Primarily Expert opinion: Summary

- Mammography
  - Radiologist
  - Technologist

- Pathology

- Surgery

- What and Why
References on performance


• Diagnostic Mammography: Identifying Minimally Acceptable Interpretive Performance Criteria Edward A. Sickles, MD, Diana L. Miglioretti, PhD et al.