Investigating the feasibility of using A.I. for population-level mammography image quality improvement initiatives at Leeds Teaching Hospitals NHS Trust

N Sharma*; H Heathcote-Watson; A Nielsen Moody; M Fletcher

Disclosures: *Clinical advisor, Densitas Inc.; Advisor for Hologic and BARD
Mammography Image Quality (IQ)

- The 3 core PHE standards in screening:
  1. Achieve **optimum image quality**
  2. Limit radiation dose
  3. Minimise the number of repeat examinations


- Higher quality images …
  - ↑ sensitivity[Taplin, 2002]
  - ↓ stage at detection[Rauscher 2013]
  - ↓ interval cancers[Taplin, 2002]
  - ↓ false positive rate[Guertin 2018]

- The $$$ of mammography IQ
  - **Annual** direct costs:
    - Technical recalls (2.13%[2018-2019]) = ~£1.8 million
    - QA self-reviews (1 shift/month, 40 studies) = ~£2.5 million
  - Delayed diagnosis and treatment costs
    - Breast cancer stage = strongest predictor of costs.[Hall 2015]
    - **True annual £ unknown**
Current State of IQ @ LTHT

- IQ monitoring
  - Technical recall/repeat aggregate rates monitored monthly

- Self-evaluations
  - Screening service: 40 mammograms self-reviewed monthly
  - Diagnostic service: no mandated reviews

- IQ improvement initiatives
  - No active IQ educational interventions
  - Until now, population-level IQ data inaccessible
  - Prospective trial planning

Alignment with NHS mandates

- NHS Long Term Plan for Cancer
  - Calls for improvement in national screening programmes through investment in innovative technologies;

- NHSX Strategic priorities
  - Calls for introducing technologies reducing burden on clinicians and staff, to focus on patients
IQ Assessment Challenges

- **Time-consuming**
  - Infeasible at time of image acquisition (8 min/exam)
  - Delayed feedback/corrective action
  ← Resource challenges further intensified by mammography workforce pressures

- **Visual**
  - Subjective

**UK NHS positioning evaluations**

- Nipple in profile
- MLO Pectoral muscle to nipple level
- MLO Pectoral muscle at appropriate angle
- Symmetrical images
- MLO IMF shown clearly
- CC Medial border demonstrated
- CC Some axillary tail shown
- CC Back of breast clearly shown with some medial central & lateral

Ambiguous terms

**Does the Pectoralis muscle extend within 1cm of the PNL?**

1\textsuperscript{st} read: 9/15 reviewers = YES
2\textsuperscript{nd} read: 9/15 reviewers = YES, BUT…

6/15 (40\%) reviewers flipped their assessment (3 flipped to present, 3 flipped to absent) [Sharma 2020]
Research Study Objectives

**Primary:** To investigate the current state of mammo IQ in breast imaging services at Leeds Teaching Hospitals NHS Trust (n~60 000 images)

**Secondary:** To compare the population-based AI prevalence rate with visual prevalence rates in a validation sample (n~200 images).

- The study was waived by research ethics
- The study was approved by institutional quality committee
Methodology

• Population-based image processing
  - Densitas® IntelliMammo™ was installed
  - Studies acquired over 12/2021 to 03/2022 were processed
    ○ [N=59 264 images (n_{CC}=29964 n_{MLO}=29300)]

• Manual data collection (random sample of 50 symptomatic studies)
  - A pair of lead radiographers reviewed together for a consensus
  - 196 images (98 CC & 98 MLO)

• Analysis
  - Event rate per positioning error (stratified by CC/MLO)
  - Weekly average error rate time plots
  - Agreement assessed by Cohen’s kappa (validation dataset [n=198])

*Kappa Classifications:*
  Less than chance agreement (<0);
  Slight agreement (0.01-0.20);
  Fair agreement (0.21-0.40);
  Moderate agreement (0.41-0.60);
  Substantial agreement (0.61-0.80);
  Almost perfect agreement (>0.80)
Results

- Population-based error prevalence 12/2021-03/2022
  - As low as 3% with IR placement error
  - As high as 51% with IMF missing error
- 3% to 51% by AI in the population-level data, compared to…
  - … 9% to 56% by expert assessment in the validation set
  - … 5.5% to 40.4% by AI assessment in the medical literature
- Kappa range from ‘substantial’ (＞0.60) to ‘almost perfect’ (＞0.80)

<table>
<thead>
<tr>
<th>Positioning Error</th>
<th>Leeds Teaching Hospitals NHS Trust</th>
<th>Medical Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population (N=59,264)</td>
<td>Validation set (N=198)</td>
</tr>
<tr>
<td></td>
<td>AI Error Rate</td>
<td>Expert Error Rate</td>
</tr>
<tr>
<td>Pectoralis Muscle Length</td>
<td>16%</td>
<td>10%</td>
</tr>
<tr>
<td>Pectoralis Muscle Concave</td>
<td>13%</td>
<td>12%</td>
</tr>
<tr>
<td>IMF Missing</td>
<td>51%</td>
<td>56%</td>
</tr>
<tr>
<td>IR Placement</td>
<td>3%</td>
<td>9%</td>
</tr>
<tr>
<td>MLO Posterior Tissues Missing</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>CC Posterior Tissues Missing</td>
<td>23%</td>
<td>12%</td>
</tr>
<tr>
<td>CC Excessive Exaggeration</td>
<td>34%</td>
<td>17%</td>
</tr>
</tbody>
</table>

*Source: Norwegian error rate data [Waade 2021]; *Source: Canadian error rate data [Rouette 2021]
Weekly rate variation

- **Stable** weekly error rates

← Provides ‘baseline’ assessment for monitoring improvement initiatives (e.g. educational interventions)
Discussion

Key findings
- These findings support the use of AI for reliable and reproducible quantitative mammography positioning image quality assessments.
- Aligns with other studies suggesting AI may agree with expert assessments
  - Slight 0.06 (pec shape) to substantial agreement 0.69 (nipple not in profile) [Waade 2021]
- With population-based error rate information at your fingertips, it is possible to evaluate image quality improvement initiatives (such as tailored educational sessions).

Future work
- Investigate Radiographer and Assistant Practitioner error rates
  - Stratification of results by years of experience
- Investigate needs-based image quality improvement initiatives
- Implement interventions and monitor the impacts on baseline error rates
- Stratify error rates by presence of patient associated limitations

Limitations
- Validation data set sample size was small with low positioning error event rates
- Did not stratify analysis by screening and diagnostic mammograms
  - Differences with imaging requirements
Conclusion

These study findings suggest that automated A.I. mammography positioning error assessments may provide a **feasible approach** to measuring and monitoring the impact of **image quality improvement initiatives** at Leeds Teaching Hospitals NHS Trust.

References


O'Leary, D. and Rainford, L. (2011). Can radiation dose in mammography be further reduced by increasing the image quality?. Breast Cancer Research, 13(S1).


