NHS
The Leeds Teaching Hospitals
NHS Trust

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 Leeds Teaching Hospitals NHS Trust

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

^Co-dependent: ↑ image quality = ↓ radiation dose[O'Leary, 2011] & ↓ technical recall[Salkowski, 2019]

- 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

The Leeds Teaching Hospitals NHS Trust

- 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

The Leeds Teaching Hospitals NHS Trust

- 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

Some apply literally, others say yes, only when obstructs breast tissue[Boyce 2015]

Ambiguous terms

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

1st read: **9**/15 reviewers = YES
2nd 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

The Leeds Teaching Hospitals NHS Trust

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

The Leeds Teaching Hospitals NHS Trust

- 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

The Leeds Teaching Hospitals NHS Trust

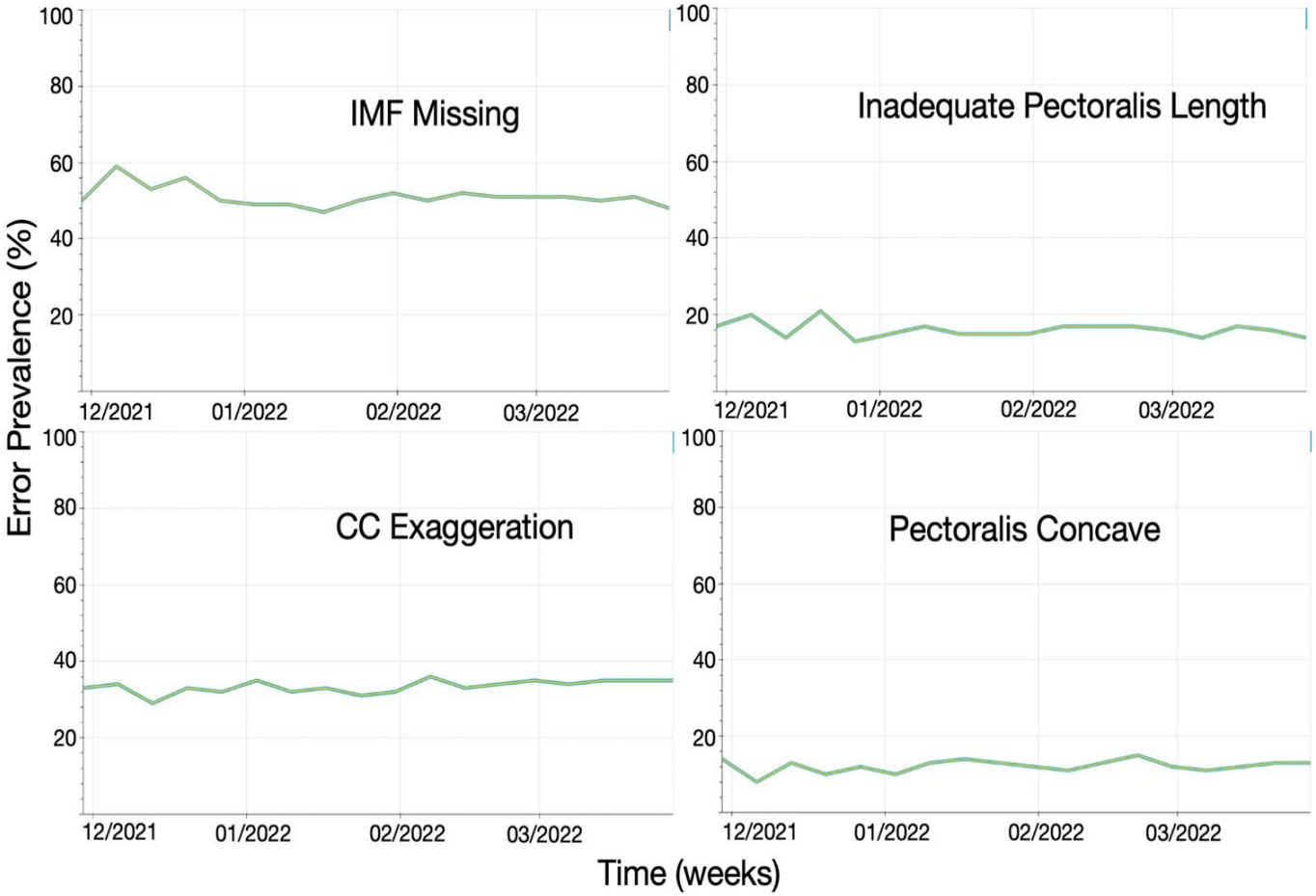
- Population-based error prevalence 12/2021-03/2022
  - As low as **3%** with IR placement error
  - As high as **51%** with IMF missing error
- 3% to 51% by AI in the population-level data, compared to…
  - … 9% to 56% by expert assessment in the validation set
  - … 5.5% to 40.4% by AI assessment in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

	Leeds Teaching Hospitals NHS Trust			Medical Literature	
	Population (N=59,264)	Validation set (N=198)			
Positioning Error	**AI Error Rate**	**Expert Error Rate**	**Kappa (95%CI)**	**AI Event Rate**	**Population***
Pectoralis Muscle Length	16%	10%	0.89 (0.69, 1.00)	5.5-37.8%	Norway; Canada
Pectoralis Muscle Concave	13%	12%	0.81 (0.61, 1.00)	16.0-19.2%	Norway
IMF Missing	51%	56%	0.71 (0.51, 0.91)	9.0-20.5%	Norway
IR Placement	3%	9%	0.78 (0.59, 0.98)	--	--
MLO Posterior Tissues Missing	11%	11%	0.85 (0.65, 1.00)	16.1%	Canada
CC Posterior Tissues Missing	23%	12%	0.95 (0.76, 1.00)	20.2%	Canada
CC Excessive Exaggeration	34%	17%	0.70 (0.50, 0.90)	24.4-40.4%	Norway

**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

The Leeds Teaching Hospitals NHS Trust

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

The Leeds Teaching Hospitals NHS Trust

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

Boyce, M., Gullien, R., Parashar, D., & Taylor, K. (2015). Comparing the use and interpretation of PGMI scoring to assess the technical quality of screening mammograms in the UK and Norway. Radiography, 21(4), 342-347.

Guertin, MH., Théberge, I., Zomahoun, HTV., Dufresne, MP., Pelletier, E., & Brisson, J. (2018) Mammography Clinical Image Quality and the False Positive Rate in a Canadian Breast Cancer Screening Program. Canadian Association of Radiologists Journal, Volume 69, Issue 2, 169 – 175

Hall, PS., Hamilton, P., Hulme, CT., Meads, DM., Jones, H., Newsham, A., Marti, J., Smith, AF., Mason, H., Velikova, G., Ashley, L., & Wright, P. (2015). Costs of cancer care for use in economic evaluation: a UK analysis of patient-level routine health system data. British journal of cancer, 112(5), 948–956.

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).

Rauscher, G., Khan, JA., Berbaum, ML., & Conant, EF. (2013). Mammogram image quality as a potential contributor to disparities in breast cancer stage at diagnosis: an observational study. *BMC Cancer*, 13(208).

Rouette, J., Elfassy, N., Bouganim, N., Yin, H., Lasry, N., & Azoulay, L. (2021). Evaluation of the quality of mammographic breast positioning: a quality improvement study. *CMAJ open*, *9*(2), E607–E612.

Salkowski, L., Elezaby, M., Fowler, A. M., Burnside, E., Woods, R. W., & Strigel, RM. (2019). Comparison of screening full-field digital mammography and digital breast tomosynthesis technical recalls. *Journal of Medical Imaging*, 6(03), p.1.

Sharma, N., Schofield, S., Fletcher, M., Hammond, RL., Larson, A. N., Bakken, AS., Handberg, E., Hopland, NM., & Iles, SE. (2019). *Mammography positioning errors: a multi-centre study.* Presented at the European Congress of Radiology, Vienna, Austria. Poster presented at the Canadian Association of Radiologists Annual Scientific Meeting, Montreal, Quebec and Radiology Research Day at Dalhousie University, Halifax, Nova Scotia.

Taplin, S., Rutter, CM., Finder, C., Mandelson, MT., Houn, F., & White, E. (2002). Screening Mammography: Clinical Image Quality and the Risk of Interval Breast Cancer. *American Journal of Roentgenology*, 178(4), pp.797-803.

Waade, GG., Danielsen, AS., Holen, ÅS., Larsen, M., Hanestad, B., Hopland, NM., Kalcheva, V., & Hofvind, S. (2021). Assessment of breast positioning criteria in mammographic screening: Agreement between artificial intelligence software and radiographers. *Journal of medical screening*, *28*(4), 448–455.