Reduction in technical repeat and recall rate after implementation of artificial intelligence driven quality improvement software

2022 RSNA Annual Meeting, Chicago, USA

Peter R. Eby[1], Linda Martis[2], Jeremy Paluch[1], Jiyeon Jang[1], Ariane Chan[2]

Corresponding Author:
Peter R. Eby - peterebymd@proton.me

1 Virginia Mason Franciscan Health™

2 Volpara HEALTH

Disclosures

Research and data analysis performed in partnership with Volpara Health. Volpara Health did not provide any equipment, funding or other resources in exchange for this research.

The authors have no other relevant financial, business or other relationships to disclose.

Virginia Mason Franciscan Health™

Background & Purpose

- Technical repeats/recalls (TR) due to suboptimal mammographic image quality (IQ) are undesirable
- Breast positioning is a major factor, contributing to ~47%-81% of avoidable TR
- Current methods for evaluating breast positioning are subjective and time-consuming, limiting scalability of quality improvement studies[1,2,3]

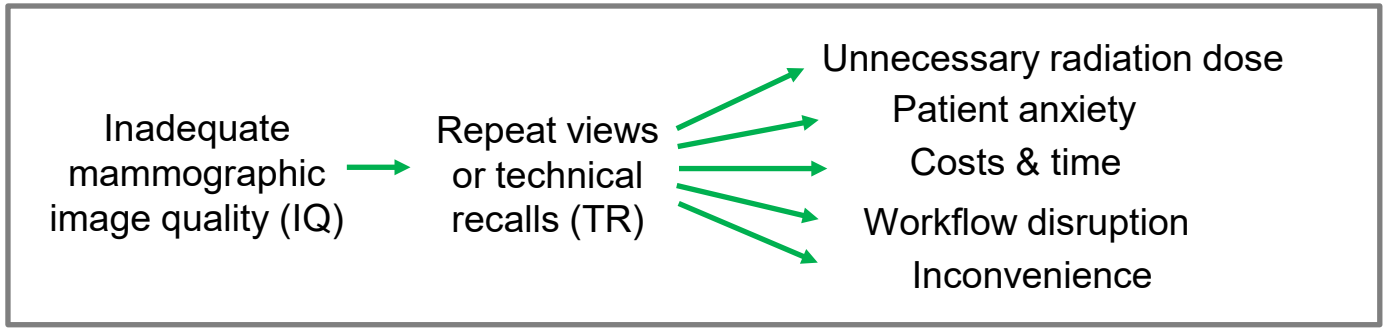


Table 1 Indications for technical recall based on modality.

Indication for technical recall	Number of cases overall	% FFDM indications	% DBT[a] indications
Motion	98	31.1	10.4
Positioning issues	192	47.1	80.6

Salkowski LR et al... J Med Imaging. 2019;6(3):031403.

We sought to evaluate whether implementation of Artificial Intelligence (AI) software was associated with improved objective Image Quality (IQ) indicators and reduced TR across Virginia Mason Franciscan Health

Virginia Mason Franciscan Health™

1. Pal S. et al. AJR Am J Roentgenol. 2018;210(4):807-815
2. Rauscher GH, et al. J Am Coll Radiol. 2020;17(11):1420-8.
3. Rouette J, et al. CMAJ Open. 2021;9(2):E607-E12.

Methods

- In 2019, AI IQ software (Volpara Health) was installed in 11 machines across 9 clinics at Virginia Mason Franciscan Health

  - AI software objectively measured breast positioning and compression **IQ indicators** for ALL images and ALL Techs

IQ indicators evaluated

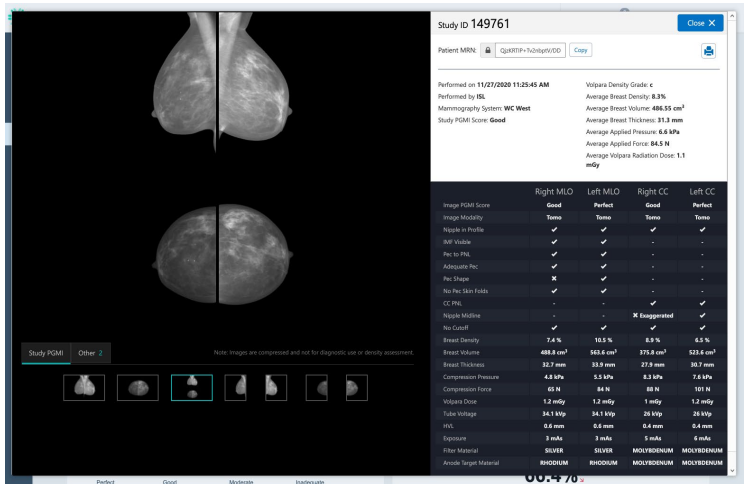
Positioning score
Perfect, Good, Moderate, Inadequate (PGMI)

Compression pressure score
Target (7-15 kPa), Low (<7 kPa), High (>15 kPa)

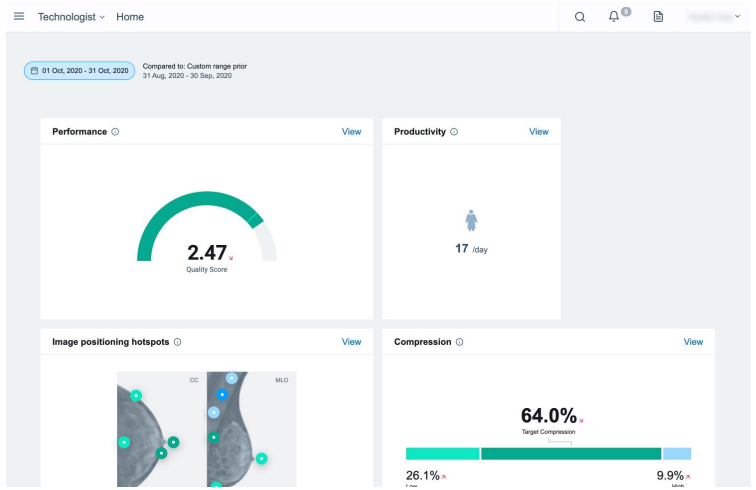
Quality score
Weighted breast positioning and compression score ranging from 0 to 4

- Interactive dashboards provided **continuous feedback** and **benchmarking** for technologists, which enabled them to identify **focus areas for improvement** and facilitated **goal setting**

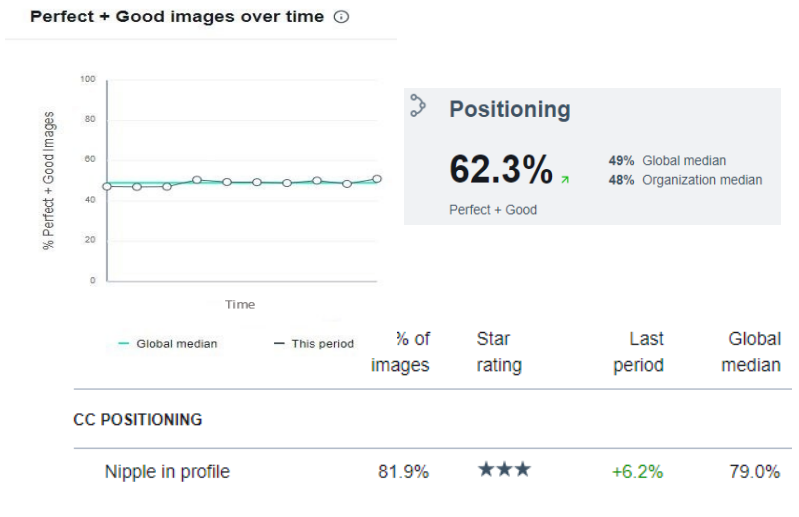
Vendor-neutral display images with image-, metric-, and study-level metrics



Interactive dashboards and automated reports



Trend analyses and global/organizational benchmarking



Methods

- Radiologists selected patients for TR during standard clinical review of images without knowledge of IQ scoring
- TR rates were extracted from Centricity (GE Healthcare)
- Aggregated IQ data and patient demographics were extracted from Analytics (Volpara Health)
- Analysis was restricted to non-implant, mammography (2D) exams acquired in the first ("Baseline") and most recent ("Current") 12 months following AI software installation

Data was aggregated per tech:

- TR rates
- IQ indicators (PGMI scores, target compression frequencies, overall Quality Scores)
- Patient demographics (age, breast volume, volumetric breast density)

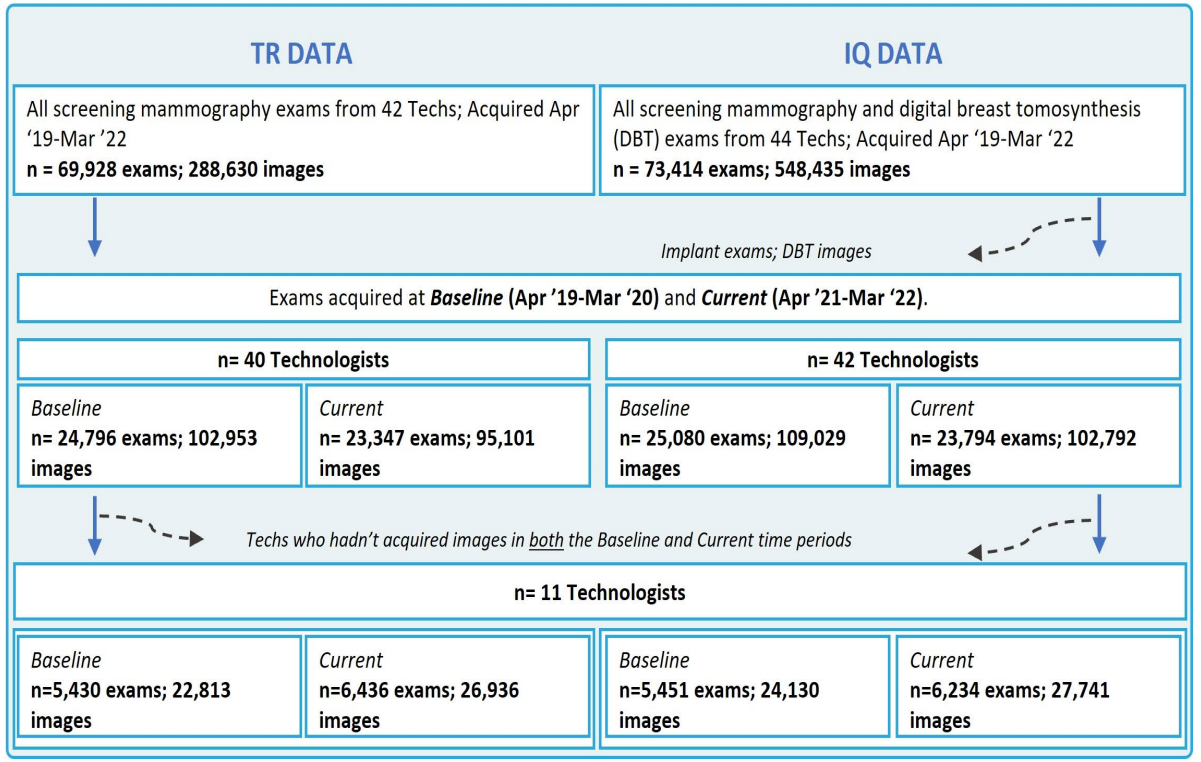
Changes between Baseline and Current periods were evaluated using Chi2, Kruskal Wallis and paired t-tests

A subset of Techs (n=11) who acquired images in both Baseline and Current periods were categorized based on the percentage of images scored Inadequate in Baseline vs Current periods:

- "Improvers" (>3% to <3%*)
- "Non-improvers" (>3% to >3%*)

Using Chi2 test, the Current TR was compared between Improvers and Non-improvers.

Data workflow to select exams and Techs for analysis



* 3% based on target TR recommended in the United Kingdom. Public Health England. Guidance - Breast Screening: repeat mammograms. 2017

Results

Patient demographics, TR and IQ indicators compared between Baseline and Current periods

		Baseline[c]		Current[c]		P-value
		Median or count	Mean ± SD or count (%)	Median or count	Mean ± SD or count (%)	
Age (y)[a]		63	62.18±10.996	62	61.23±11.069	<0.001
Volumetric breast density (%)[a]		5.60	7.60±5.707	5.34	7.62±6.124	<0.001
Breast volume (cm^3) [a]		804	926.87±575.815	790.60	912.90±569.092	<0.001
Overall Quality Score [a,d]		2.27	2.28±0.218	2.45	2.42±0.307	0.001
Image- level breast positioning score categories [a]	**Perfect + Good**	59862	56.36%	58712	59.78%	<0.001
	Moderate + Inadequate	46351	43.64%	39509	40.22%	
Image-level compression pressure categories [a]	**Target (7-15 kPa)**	63201	59.06%	62846	63.57%	<0.001
	High or Low (<7 or >15 kPa)	43807	40.94%	36020	36.43%	
Technical repeat/recall (TR) rate[b]	**Good images**	102165	99%	94941	100%	<0.001
	Repeated/Recalled images	788	1%	160	0%	

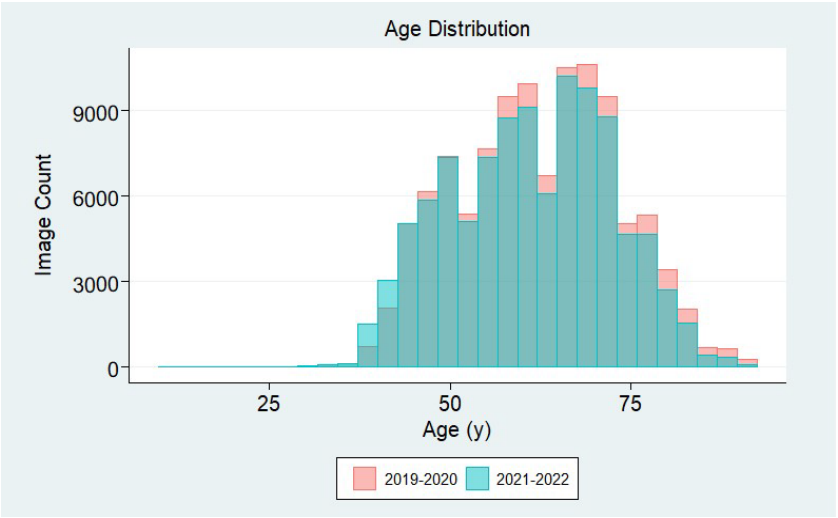
[a] Data from Volpara Analytics (n = 198,054 images; 40 techs); [b] Data from Centricity (n= 211,821 images; 42 techs). Total numbers for each variable differs due to missing data.
[c] Data shown is for the all technologists who had acquired images in the Baseline and Current periods.
[d] Except for Quality Score, which used paired t-test, Kruskal-Wallis and Chi2 tests were used for continuous and categorical variables, respectively.

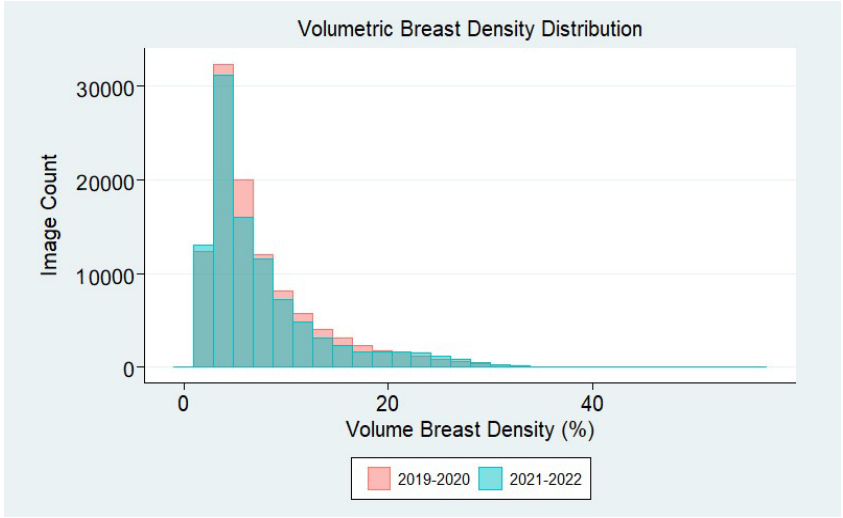
Virginia Mason Franciscan Health™

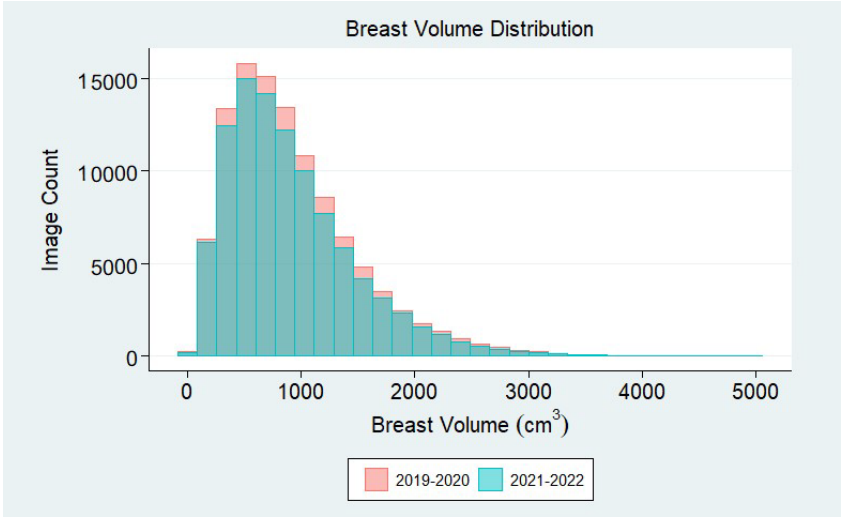
Results

Patient demographics significantly differed from *Baseline* to *Current*, although distributions very similar

- Median age (63 vs 62 y)
- Breast volume (804 vs 791 cm^3)
- Volumetric breast density (5.6% vs 5.3%)





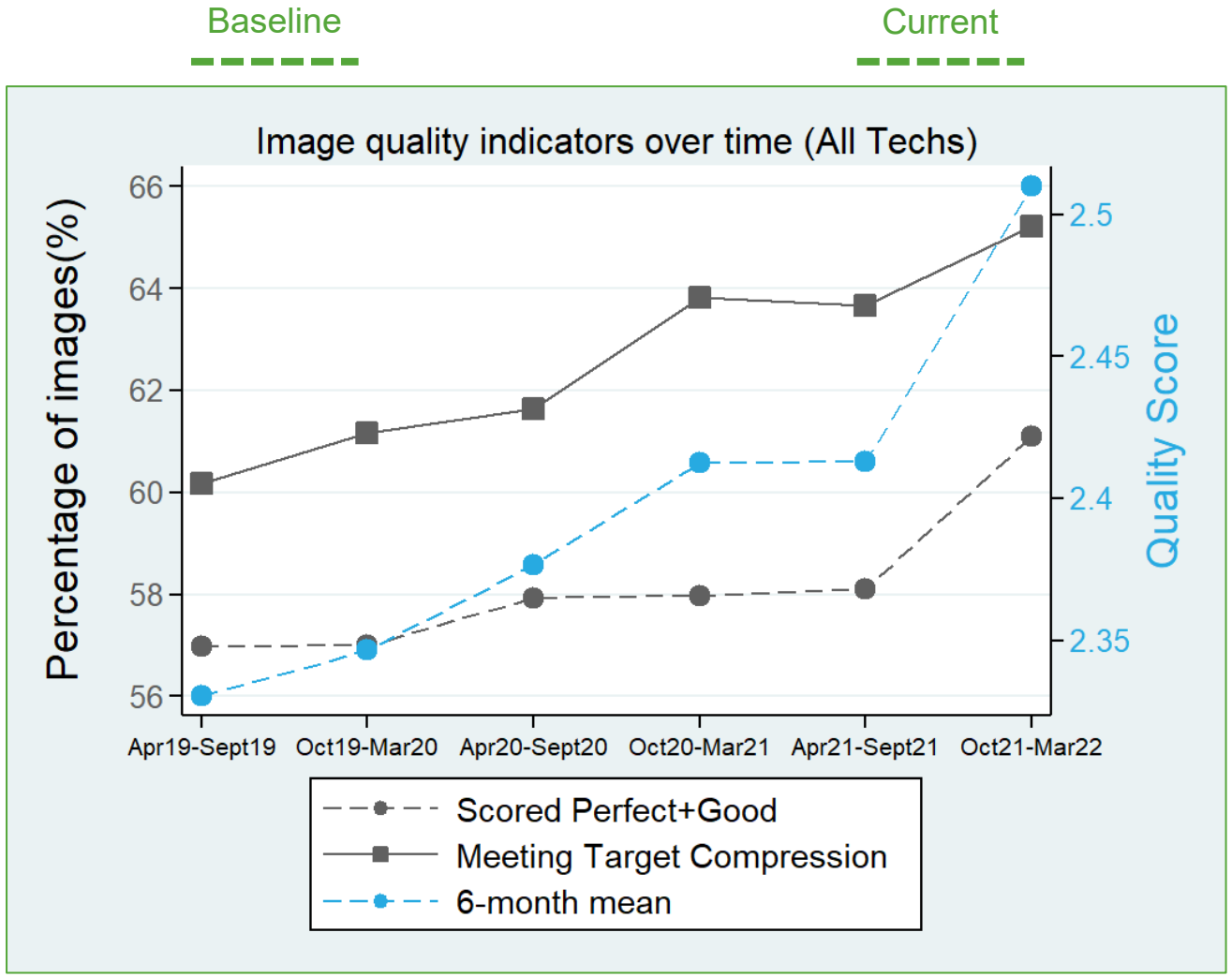


Virginia Mason Franciscan Health™

Results

Comparing Baseline (first 12-months) vs Current (recent 12-months), significant improvements in IQ indicators were observed

- 6% increase in the proportion of images scored Perfect or Good (56.36% vs 59.78%)
- 8% increase in the proportion of images meeting Target Compression (59.06% vs 63.57%)
- 6% increase in mean Quality Score (2.28 vs 2.42)

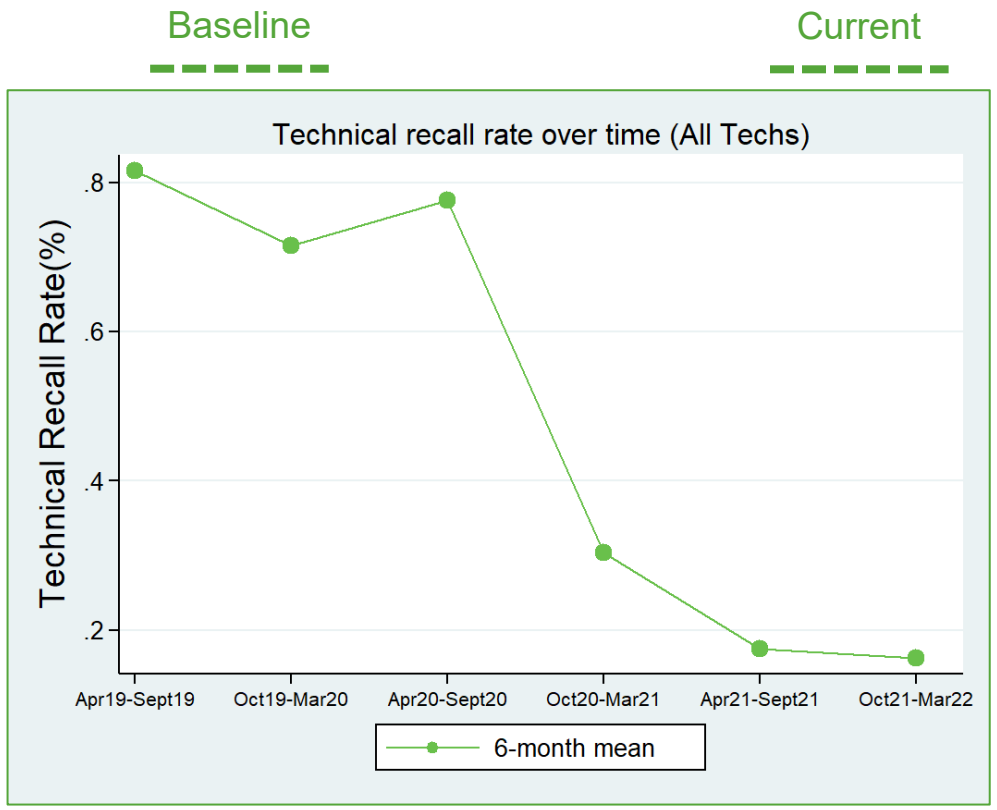


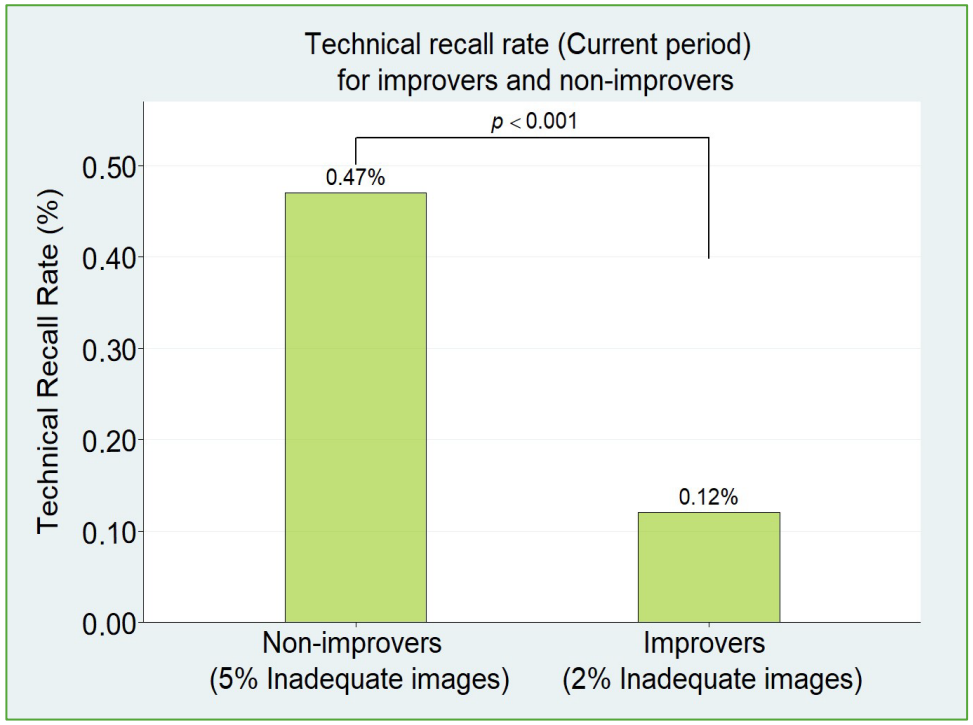
Virginia Mason Franciscan Health™

Results

Comparing Baseline (first 12-months) vs Current (recent 12-months), significant improvements in TR indicators were also observed

- 78% reduction in TR rate (0.77% vs 0.17%)
- Current TR was significantly lower for improvers vs non-improvers (0.12% vs 0.47%) suggesting a correlation between IQ improvement and TR rates (Fig. 4).





Conclusion & Clinical Impact

- To our knowledge, this study represents the **largest North America—based mammographic IQ evaluation to date** and only the second to have looked at impacts of IQ improvement initiatives and TR rates[1].
- Over a 2.5-year period following installation of AI IQ software, we **observed significant improvement** (6% and 8%) in objectively measured breast positioning and compression IQ respectively, as well as a concomitant 78% **reduction in TR rates.**
- Future evaluations **matching TR and IQ data at the patient-level and extension to DBT images**, would facilitate more direct measures of patient and provider outcomes (e.g. cancer detection rates, costs, radiation dose) and allow for analyses to more definitively associate IQ improvements with clinical outcomes.

Compared to conventional, manual assessment, AI software has the potential to:

- Revolutionize mammography IQ by facilitating mammographic IQ evaluation on an unprecedented scale
- Provide objective, continuous feedback and benchmarking for Techs
- Improve IQ and improve outcomes for both providers and consumers of mammography screening

Virginia Mason Franciscan Health™

1. Pal S. et al. AJR Am J Roentgenol. 2018;210(4):807-815