Real-World Breast Cancer Detection Before and After Implementation of an Artificial Intelligence Detection System in a Digital Breast Tomosynthesis Screening Program

Purpose

To compare radiologists' breast cancer screening performance before and after the implementation of an artificial intelligence (AI) detection system with digital breast tomosynthesis (DBT).

Materials and Methods

An IRB-approved retrospective study of Mammography Quality Standards Act (MQSA) report statistics was conducted with four radiologists reading DBT across three clinical sites and two distinct time periods. Data were collected from September 1, 2018 to August 31, 2019 with concurrent use of CAD-enhanced synthetic views (PowerLook Tomo Detection, iCAD, Nashua, NH) ("pre-PFAI") and January 1, 2020 to December 31, 2020 with concurrent use of deep learning AI detection system (ProFound AI V2.1, iCAD, Nashua, NH) ("post-PFAI"), allowing for a four-month window between time periods, including two months for ProFound AI system installation and training. The AI system provides lesion outlines, lesion scores, and an overall case score. Co-primary endpoints were cancer detection rate (CDR) per 1,000 screened and abnormal interpretation rate (AIR), both post-PFAI versus pre-PFAI. Secondary endpoints included positive predictive values (PPVs) for cancer among screenings with abnormal interpretations (PPV1) and for biopsies performed (PPV3), both post-PFAI versus pre-PFAI. Endpoints were calculated for each radiologist in each time period. Estimates of performance pre-PFAI, post-PFAI, and the difference post-PFAI – pre-PFAI were obtained as the average across radiologists. Bootstrap resampling was used to provide 95% confidence intervals (CIs) for estimates within and between time periods.

Results

Performance rates were compared for women screened pre-PFAI (n=7,627; 34 cancers) and post-PFAI (n= 4,703; 27 cancers). CDR per 1,000 screened improved from 3.8 (95% CI: 2.5, 5.3) pre-PFAI to 6.2 (95% CI: 3.9, 9.0) post-PFAI, an increase of 2.4 (95% CI: -0.4, 5.4). AIR decreased from 9.6% (95% CI: 8.9, 10.2) pre-PFAI to 7.3% (95% CI: 6.5, 8.2) post-PFAI, a reduction of 2.2% (95% CI: -3.3, -1.2). Overall, cancer detection increased, and the rate of recalled exams declined.

PPV1 doubled from 4.1% (95% CI: 2.7, 5.7) pre-PFAI to 8.8% (95% CI: 5.8, 12.1) post-PFAI, an increase of 4.7% (95% CI: 1.2, 8.2). PPV3 almost doubled from 29% (95% CI: 19, 39) pre-PFAI to 57% (95% CI: 38, 72) post-PFAI, an increase of 28% (95% CI: 7, 47).

Conclusions

Interpretation of DBT after implementation of an AI detection system resulted in increased CDR, reduced AIR, and significantly improved PPV1 and PPV3.

Clinical Relevance Statement

Real-world interpretation of DBT after implementation of an AI detection system in clinical practice resulted in improved radiologists' screening performance with a clinically relevant increase in cancer detection with fewer recalls.

MQSA Outcomes

	Pre-AI		Post-AI		Difference	
	Observed	95% CI	Observed	95% CI	Observed	95% CI
CDR per 100	3.8	(2.5, 5.3)	6.2	(3.9, 9.0)	2.4	(-0.4, 5.4)
AIR%	9.6	(8.9, 10.2)	7.3	(6.5, 8.2)	-2.2	(-3.3, -1.2)
PPV1%	4.1	(2.7, 5.7)	8.8	(5.8, 12.1)	4.7	(1.2, 8.2)
PPV3%	29	(19, 39)	57	(38, 72)	28	(7, 47)
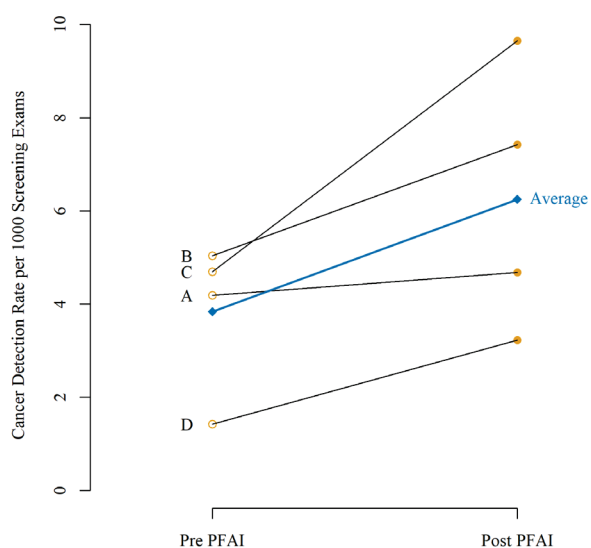
Change in cancer detection rate after the implementation of an artificial intelligence detection system
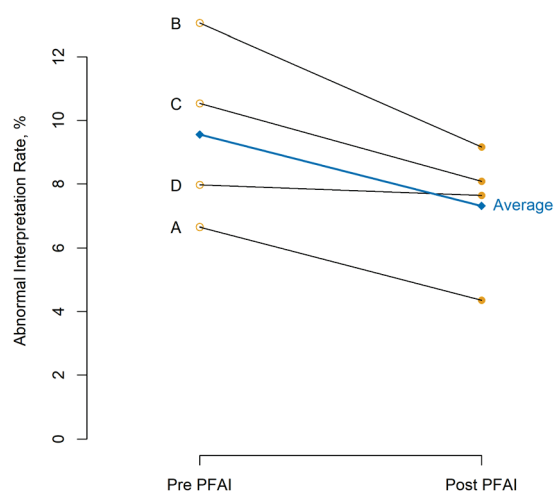
Change in abnormal interpretation rate after the implementation of an artificial intelligence detection system

iCAD®