Real World Clinical Impact of Implementing Artificial Intelligence on Radiologists' Performance in High Volume Mammography Screening

Gräewingholt, A. et al. Real World Clinical Impact of Implementing Artificial Intelligence on Radiologists' Performance in High Volume Mammography Screening. Oral presentation at Radiological Society of North America (RSNA) Annual Meeting; November 27-December 1, 2022; Chicago, IL.

Purpose

The purpose of this study is to compare diagnostic accuracy of radiologists in a population-based screening program before and after installation of an artificial intelligence (AI) tool designed to detect breast cancer.

Materials and Methods

This EC-approved retrospective study evaluated statistics from quality assurance reports for a single German institution, Mammographie screening Paderborn, with five radiologists interpreting examinations from two clinical sites. The screening program invites 50- to 69-year-old women across Germany for full-field digital mammography (FFDM) every two years. All are read by two radiologists independently (double reading). If either one or both radiologists detect a suspicious lesion, the case is reviewed for recall by a consensus panel of all radiologists at the institution. Period 1 includes 10 years of screening with FFDM from July 1, 2009 to June 30, 2019 ("pre AI"), ending prior to the installation of AI (iCAD ProFound AI® V2.0, high operating point). Period 2 includes 14 months of screening with FFDM from July 1, 2019 to August 31, 2020 ("post AI"), during which two of the five radiologists had access to AI for use in their reading process and additionally consensus review included use of AI. Co-primary endpoints were cancer detection rate (CDR) per 1,000 screened and abnormal interpretation rate (AIR) of radiologists interpreting FFDM images pre and post AI. To summarize across radiologists, we took the unweighted average of their individual statistics and estimated 95% confidence intervals accounting for correlation between each pair of radiologists induced by the double reading paradigm.

Results

Period 1 had 322,793 reads (1,451 true positives) and Period 2 had 30,824 reads (174 true positives). Average CDR was 4.54/1,000 screened pre-AI and 5.78/1,000 post-AI with difference 1.24/1,000, 95% CI for difference: (-0.31, 2.79). Average AIR was 4.56% pre-AI and 4.84% post-AI (difference 0.28,95% CI: -0.14, 0.70).

Conclusions

Increases in CDR while maintaining similar AIR in clinical practice are consistent with prior reader studies. The effect of AI may be attenuated by availability to only two of the five readers. Our study will be expanded to a longer post AI period and to include cancer case characteristics.

Clinical Relevance Statement

This study of the impact of AI on clinical practice shows that one potential benefit of AI is to detect 1.2 additional cancers per 1,000 screened, a 27% increase, without appreciably increasing AIR.

98 Spit Brook Road Suite 100 · Nashua, NH 03062 · 1-866-280-2239