# Correlation between BI-RADS Assessment Categories and Artificial Intelligence Case Scores

Mark Traill, MD; Theresa McGoff, BSN; Colleen Neal, MD; Christina Jacobs, MD; William Corser, PhD; Jeffrey Hoffmeister, MD

## BACKGROUND

An artificial intelligence (AI) algorithm (ProFound AI V2, iCAD, Nashua NH) was trained to detect breast cancer on Digital Breast Tomosynthesis (DBT). The AI system generates a 0 – 100 % case score as a relative metric for chance of underlying cancer. To evaluate the clinical utility of AI case scores, we compared AI scores to BI-RADS assessment categories determined by a radiologist without AI.

## METHODS

AI retrospectively evaluated a total sample of 890 consecutive screening DBT studies and 50 consecutive cases with biopsy-proven breast cancer detected with DBT. AI case scores were compared to a single radiologist's assessments of BI-RADS category without AI for each exam. After a series of simple Chi Square bivariate correlation tests, a series of "stepwise" regression models were conducted to examine the predictive significance of patient characteristics on outcomes.

## RESULTS

In the 890 screening exams, most AI scores and BI-RADS categories were low. Based on AI, 84.8% (755/890) had a case score <60 %, and 90.6 % (806/890) of exams were initial BI-RADS Category 1 or 2 as determined by a radiologist without AI.

Additional crosstabulation analyses between screening DBT exams with AI case score <60% and those with BI-RADS Category 1 or 2 demonstrated a statistically significant association ($x2 = 9.144$, df 1, $p=0.002$). The correlation between these two measures was also highly significant (Pearson r= + 0.101, n=890, p=0.002).

An initial BI-RADS 0 category was assigned to 84/890 (9.4 %) screening DBT exams by a radiologist. AI case scores for this group averaged 54.6 % compared to an average score of 41.0% across all exams. Recall and diagnostic work-up of 84 (9.4 %) BI-RADS 0 cases demonstrated biopsy-proven cancer in three (3.5 %) patients, each of whom had AI case scores > 60%.

Of 50 cancer patients, 38 (76.0 %) had AI case scores >60 % with an average case score of 88.3 %. Twelve (24.0 %) cancer patients had 99 % AI case scores. In the three cancer cases not detected by AI, case scores were 16, 20 and 65 % respectively. Overall cancer detection sensitivity was therefore 94.0 % (47/50).
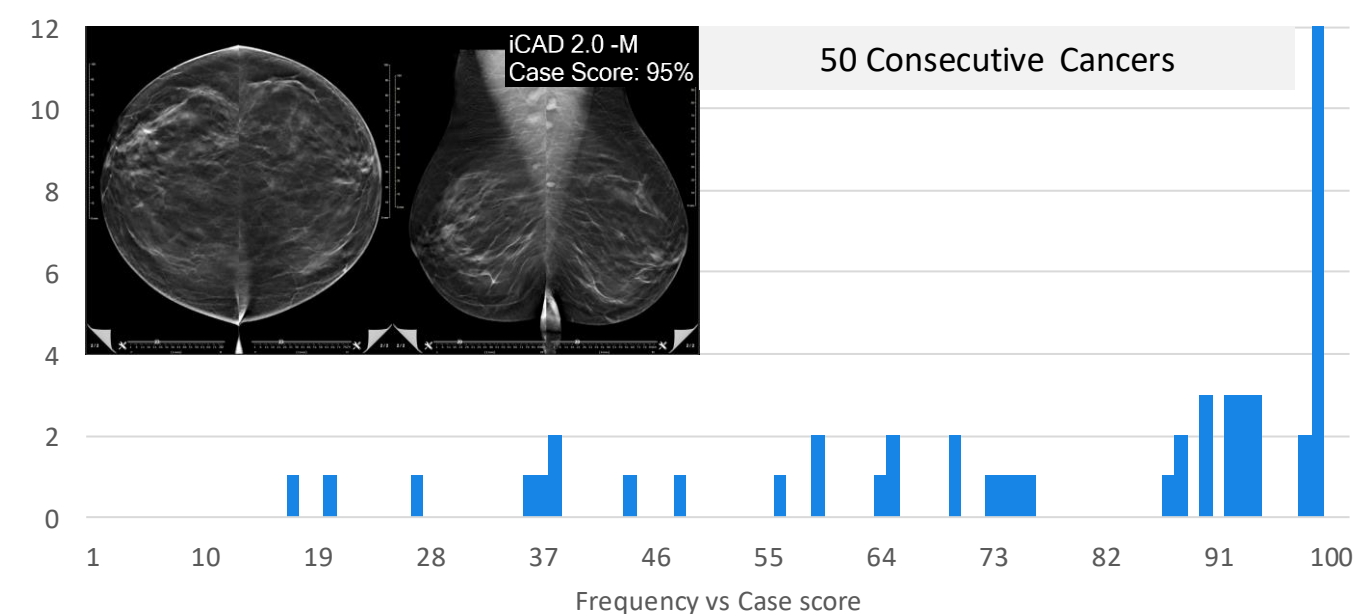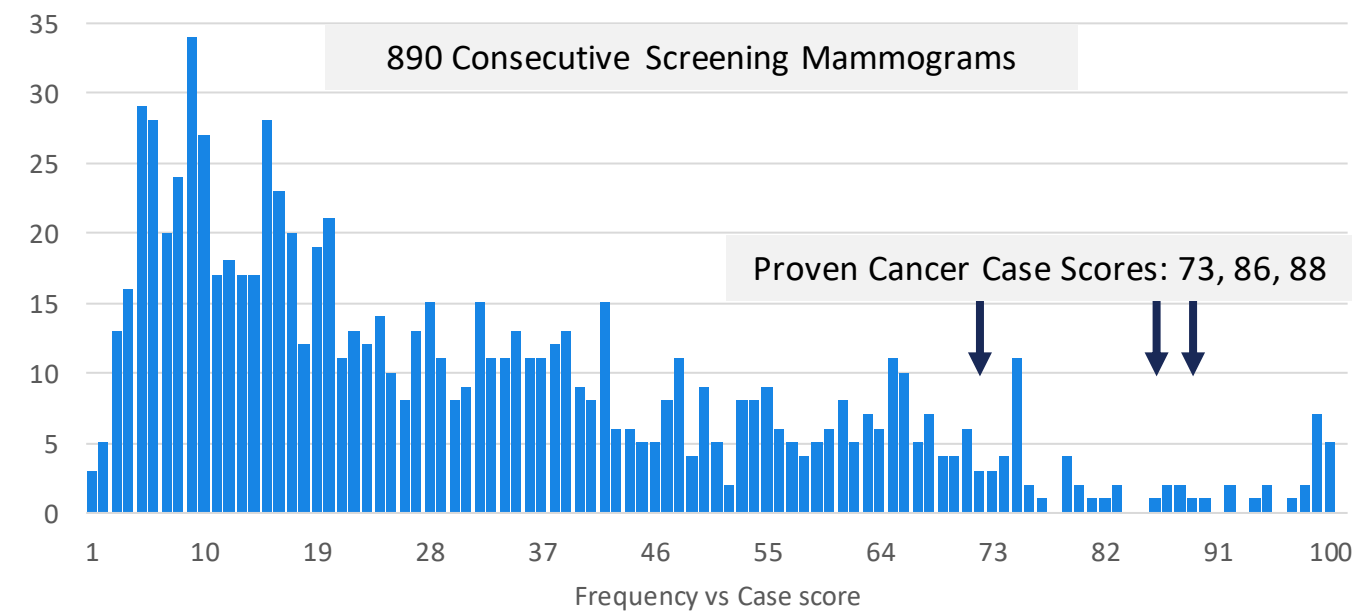




## CONCLUSIONS

Our data showed a strong positive correlation of AI case score of <60 % for patients assessed as BI-RADS 1 or 2. Most biopsy-proven cancer cases had a case score >60 % and well above the average case score of 41% for a screening population. A ProFound AI case score >60% is an indicator of increased chance of malignancy on screening DBT.