Computer Assisted Detection (CAD) - Negative Recalls in Screening Mammography: What are the Characteristics of Cancers Missed by CAD?

Pasteur Rasuli[1]*, Gerard Collin[2], Allan Gitterman[3], Stavroula Karmanos[3] and Jean Seely[1]

[1]Departments of Radiology, The Ottawa Hospital, Canada

[2]Departments of Radiology, Montfort Hospital, Canada

[3]Departments of Radiology, Queensway-Carleton Hospital, Canada

Abstract

Purpose: The purpose of this prospective study was to determine the characteristics of cancers missed by CAD and to determine if the missed cancers were significant.

Methods: The study included 37,752 consecutive women aged 50 or older undergoing screening mammography through an organized screening program in Ontario Canada. All mammograms obtained with Hologic units read using R2 CAD Image Checker with intermediate sensitivity. All CAD-negative recalls were investigated with appropriate diagnostic workup and biopsies as needed.

Results: 37,752 women were screened in the study period, with an invasive cancer detection rate of 5.9/1000, and the abnormal call rate was 4.7%. Of the 1,789 recalled cases, 109 were true positive cancers. The interval cancer rate was 15 per 12,000 in 12 months. 108 (6%) of the recalled cases were CAD-negative. Of these, 7 (6%) were found to be malignant representing a Negative Predictive Value (NPV) of 93.5% (101/108) for CAD-negative recalls. Comparatively, among the 1,681 CAD positive-recalls, 368 (21%) were found to represent breast cancer. Of the 7 missed cancers by CAD, four lesions were 7 mm or less in diameter, of which three were invasive and three were in situ ductal carcinomas. The seventh lesion was an invasive lobular carcinoma measuring 9 mm × 15 mm × 17 mm in diameter. All missed lesions were masses without any mammographically perceptible calcifications. One lesion was discovered on the first screen and six were on rescreens.

Conclusion: CAD has a high NPV, and CAD negative recalls are likely to yield small cancers at early stages.

Introduction

CAD is a computer software program designed to assist radiologists by marking abnormal breast findings on digital mammograms. Early studies of CAD confirmed retrospectively it could detect cancers missed on screening mammography [1]. These data propelled CAD to FDA approval in 1998, and it received the reimbursement nod from Medicare in 2001.

In 2001 a prospective study based on 13,000 screening exams performed in a community clinic bolstered CAD's credibility by showing 20% increased cancer detection during the one-year duration of the study [2]. But other prospective studies have found no significant difference in breast cancer detection or patient recall rates with the use of CAD [3,4].

A study by Georgian-Smith in 2007 investigated use of CAD as a second independent read and its ability to improve detection rate [5]. Similarly, a large-scale study by Gilbert et al. [6] compared a radiologist plus CAD reading versus double reading by two radiologists in a screening mammography setting. More recently Lehman et al. [7] in a large-scale study compared performance outcome of 107 radiologists using CAD with same radiologists when did not have access to CAD. All three studies reported absence of any significant increase in the cancer detection rate by using CAD. It is estimated that with CAD there is one cancer in every 2,000 CAD markings that have to be ignored, and many radiologists have found CAD more of a burden than help [8]. Use of CAD was also found to be associated with an increased discovery of DCIS but not of invasive breast cancer [9].

An increased abnormal interpretation rate has been identified as increasing the "harms "of

Table 1: Age and recall rate.

Age group	Screened cohort	Recall total Average =62.6	CAD-Positive recall Average =63.0	CAD- Negative recall Average =61.0
50-54	5897	442 (7.4%)*	410 (6.9%)*	32(7%)*
54-60	7522	328 (4.3%)	311 (4.1%)	17 (5%)
60-64	8605	348 (4.0%)	322(3.7%)	26 (7%)
65-69	8485	356 (4.1%)	341(4%)	15 (4%)
70-74	5756	251 (4.3%)	237(4.1%)	14 (5%)
75+	1487	64 (4.3%)	60(4%)	4 (6%)
Total	37752	1789 (4.7%)	1681 (4.4%)	108 (6%)

*Higher recall rate in- 50-54 age group is due to higher number of first screen patients

screening. Methods to reduce these false positives have been sought, and digital breast tomosynthesis has been helpful to reduce the false positive rate. However other methods such as CAD to reduce the false positives have not been well evaluated and this study aimed to investigate the outcome of CAD-negative abnormalities, to identify whether the recall rates could be reduced without adversely affecting the diagnostic accuracy of the mammogram if only CAD-positive lesions were to be considered for recall. To assess this, we determined number of breast cancers found among the CAD-negative calls and the grade/histological make of these cancers and their mammographic appearance.

Materials and Methods

This prospective IRB-approved, HIPPA compliant study included 37,752 consecutive, average and high-risk consenting women aged 50 to 82 with median age of 62.6 years (Table 1) who took part in screening mammography through a dedicated screening program in province of Ontario, Canada.

In Ontario, organized breast screening begins at age 50 and excludes average-risk women with history of prior breast cancer or women with breast implants. Of the cohort, 3,199 women (8.4%) were undergoing their first screen and the remaining 34,553 women were on their 2nd to 21st rescreens. Of the women who were being rescreened 11,337 (32%) had their mammogram on annual basis due to high density, family history, and/or previous high-risk breast lesion and the remaining 23,2169 (68%) had their mammogram on a biennial basis, as per Canadian guidelines. There were no additional exclusions. A minimum follow-up of two years for participants was established. The mammograms were obtained by Full-Field Digital Mammography (FFDM) units (Hologic Selenia, Marlborough, MA), and were read in batches of 60 by 8 radiologists with 10 to 35 years of experience in mammography, using R2 Image Checker CAD (Sunnyvale, CA) with intermediate sensitivity on Hologic screens. A prospective sequential read study design was used to assess the impact of CAD on the radiologist's performance in a real-time clinical practice. The radiologist first interrogated the mammogram in a standard fashion and made a decision as to whether to recall the patient. At this point, the CAD marks were deployed. In those cases where the radiologist had perceived a lesion and determined that the case should be recalled, the ultimate category of that recall was labeled as "CAD-positive" if the lesion had a CAD marking on at least one of the two views, and "CAD-negative" if there was no CAD marking on the lesion on either of the two views. Due to temporary nature of storage of CAD markings, all CAD-negative recalls were saved permanently on hard copy screen-shot images (Figure 1A) showing the discordance of CAD and radiologist's call. These hardcopy images were then saved for future reference. All recalled abnormalities (CAD
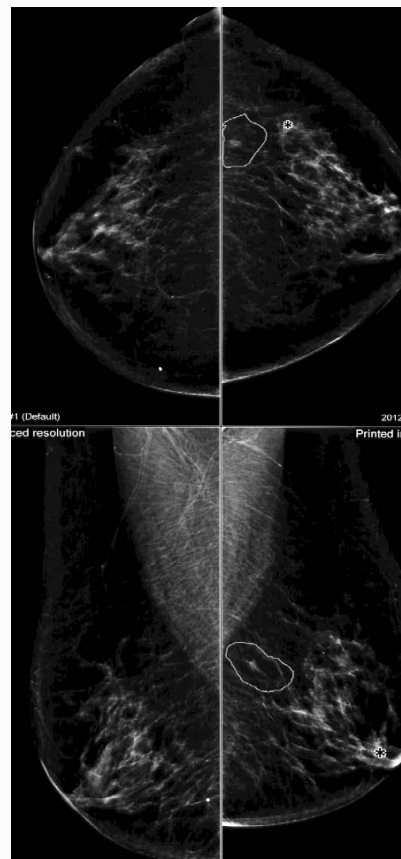

Figure 1A: The image is reproduction of hardcopy screen-shot image from patient 1 in the cancer positive CAD negative cases. Note the discordance between the Radiologist's drawing and CAD markings (*) on CC and MLO view in the left breast.

positive and CAD negative) underwent appropriate diagnostic workup including additional mammographic views including magnification, spot compression, rolled or 90degree lateral views, tomosynthesis, breast ultrasound and ultrasound-guided or stereotactic-guided breast biopsy as needed. The outcome of the diagnostic work up and any biopsies for CAD-negative lesions were tabulated on a database and formed the basis of this analysis. All women with negative workup were further followed up mammographically for two years. If no cancers were detected in that period, the findings were considered as "benign findings". Statistical analysis was performed with SPSS software (version 19.0.0; IBM, Armonk, New York).

Results

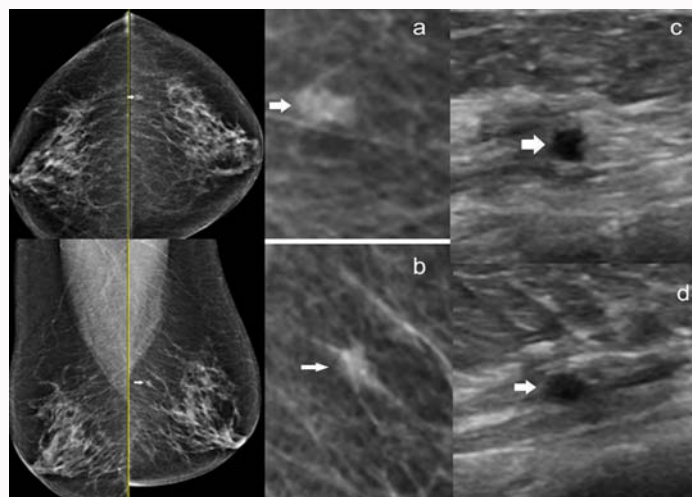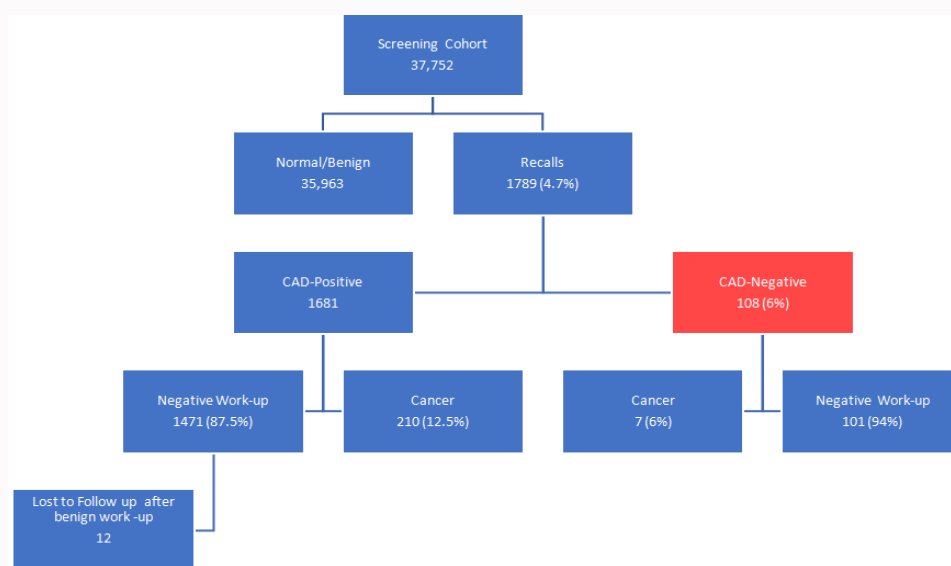During the study period, 223 screen-detected breast cancers were

Figure 1B: Same 75-year-old woman as in Figure 1A (9th screen) showing a new 4 mm irregular very dense mass with no CAD markings in the lower outer quadrant. Corresponding US shows an irregular hypoechoic mass which was biopsied as DCIS.

Flow Chart:

found in 37,752 women, yielding a screen cancer detection rate of 5.9 in 1000. As shown in the Flow Chart, there were 1,789 recalls for further diagnostic work-ups, producing a recall rate of 4.7% and a positive predictive value of 20.84% (Table 1). Of the 1,789 recalls, 108(6%) were CAD-negative. This subgroup had an average age of 61 which was statistically indifferent from the CAD positive cohort which was 63. Of the CAD-negative recalls 21 (19.4%) were on their first screen while in the recalled CAD-positive cohort only 143 in 1681 (8.5%) were in their first screen (p<0.01). Among the 108 CAD-negative recalls, there were 7 (6%) lesions that on follow up biopsies were found to be breast cancer, representing a CAD miss rate of 1.8% (95% CI: 0.57% to 3.1%) and a Negative Predictive Value (NPV) of 93.5% (101/108) for CAD-negative recalls. Comparatively, among the 1681 (1789-108) CAD positive-recalls, 366 (21.7%) were found to have breast cancer (p>0.001). Of the 7 CAD-missed cancers (Table 2), only one was discovered on first screen and all others were on rescreens. All CAD-negative cancers were masses without any perceptible mammographic calcifications, representing a NPV of 100% for

calcific lesions by CAD. All tumors except one in a 52-year-old were found in women over age 70 (Median: 71, range: 52 Y to 82 Y.). Five of seven lesions were visible on both CC and MLO projections. Four lesions were in fatty breasts and six were 7 mm or less in diameter. Four were invasive carcinomas (one high grade, one intermediate and two low grade, but all four were node-negative). Three were Ductal Carcinoma *in situ* (DCIS), one mucinous, one papillary and the other cribriform type (Figures 2-6). The mucinous DCIS lesion had been recalled one year earlier and ultrasound study had shown a simple cyst in the same area without any intracystic component (Figure 3h). The only CAD-negative cancerous lesion larger than 7 mm in this study was a node-negative invasive lobular carcinoma, measuring 9 mm × 15 mm × 17 mm in diameter (Figure 7). Complete attributes of the seven CAD-Negative lesions that were found to be cancerous are provided on Table 2.

Discussion

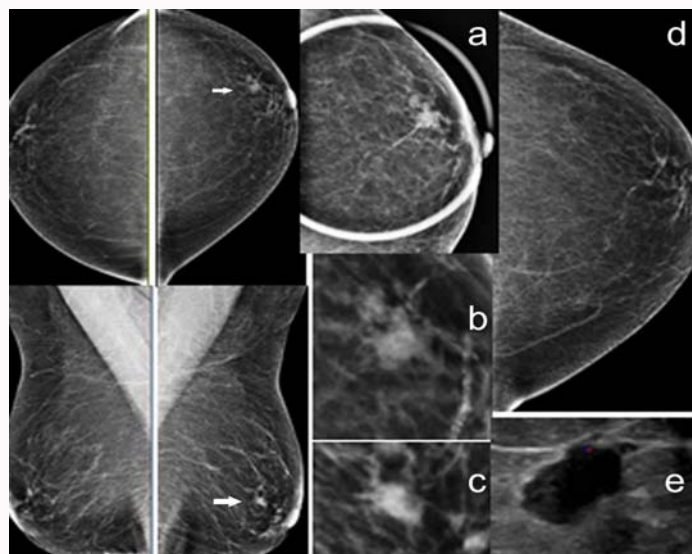Previous studies have evaluated mammographic CAD as an

Figure 2: 70-year-old woman (9th screen) with entirely fatty breasts showing a new 5 mm irregular high-density mass with some spiculation in the retroareolar region unmarked by CAD. Tomosynthesis was not contributory (5d). Corresponding US (5e) showed an irregular hypoechoic mass which was biopsied showing DCIS. The lesion is better outlined on spot compression view (2a). Figures 2b and 2c represent photographic magnification of the lesion in the original mammogram.
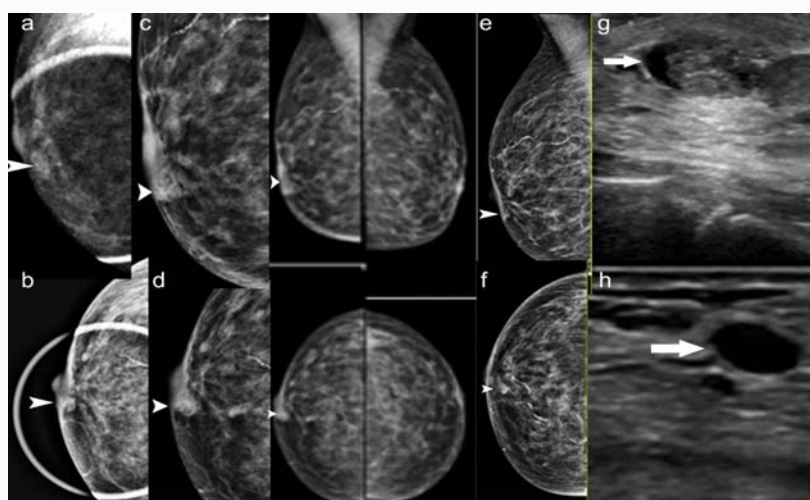

Figure 3: 82-year-old woman (2nd screen) with heterogenous breasts showing a 5 mm × 5 mm × 5 mm lobulated and irregular subdermal density in the retroareolar region of the right breast which increased in size since the previous screen (3e and 3f). This at the time was shown on ultrasound (3h) to be a cyst. One year later US showed an intracystic mass (3g) which on biopsy was found to be a mucinous intracystic high grade DCIS. The lesion is better outlined on spot compression views (3a and 3b).

"independent interpreter" and have compared CAD in the setting of radiologist reading *vs.* radiologist+CAD reading [1-4], or double radiologist reading *vs.* single radiologist+CAD reading [5,6].

Mammographic interpretation consists of two parts: (a) detection (finding of the abnormal tissue), which is a function of sensitivity and (b) analysis (assessment of likelihood that the findings are real), which is a function of specificity. CAD is known to have 99% sensitivity [1] but suffers from very low specificity which is reported to be as low as 22% to 30% [10]. These attributes make CAD inferior in "analysis" but superior in "detection" compared to a radiologist with average sensitivity and specificity of 85% and 91% respectively [7]. A radiologist is prone to missing fine microcalcifications and lesions due to fatigue, low target prevalence and environmental distraction [11]. CAD on the other hand is unable to use comparison with the previous studies and is blind to temporal stability in establishing

malignancy. It also lacks the ability to correlate the findings in CC to MLO views or *vice versa* and is blind to three-dimensional presence of a mammographic lesion. Furthermore, CAD, like all Expert systems, is unable to learn from experience and lacks intuition and common sense.

Given the complementary nature of the sets of strengths and weaknesses between CAD and radiologists in mammographic screening, it was interesting to examine the outcome of a collaborative approach between the two. It is for this reason that we evaluated the outcome of CAD-negative lesions in the screening cohort. As a secondary outcome, we retrospectively evaluated the possible consequences of the "what if" scenario if all CAD-negative lesions were to be ignored or reevaluated with more scrutiny, and only CAD-negative lesions with a higher degree of suspicion were permitted to proceed to the recall phase. This would have meant
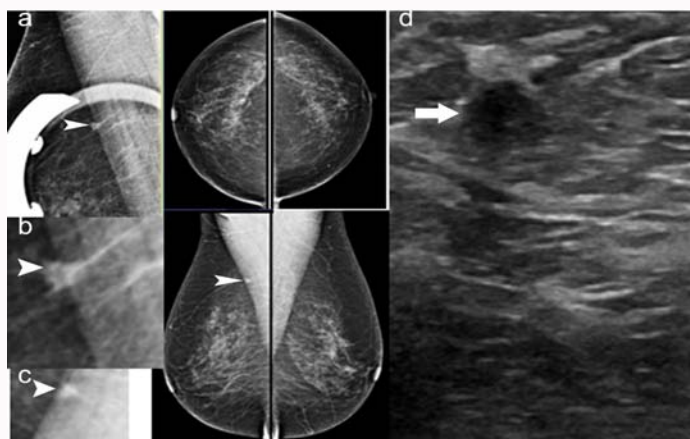
Figure 4: 52-year-old woman (3rd screen) with heterogeneous breasts showing a new 2 mm × 2 mm × 2 mm spiculated and highly dense lesion seen only on MLO view of the right breast consistent with a developing asymmetry. A corresponding irregular hypoechoic mass on ultrasound was biopsied and was discovered to be an invasive low grade ductal carcinoma. There was no nodal involvement. Figures 4b and 4c represent photographic magnification views of the density on the spot compression views and the original mammogram respectively.
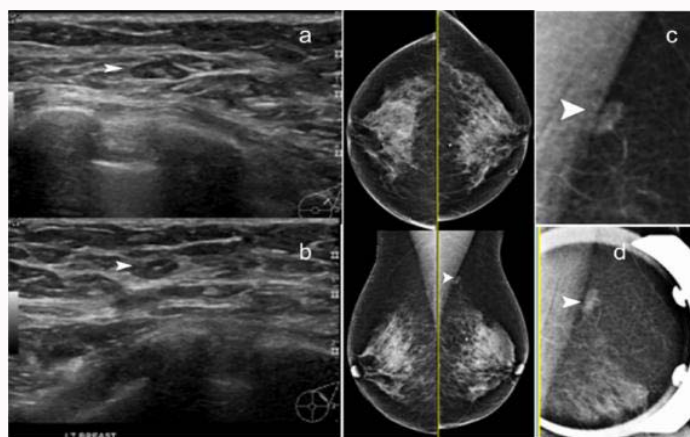

Figure 5: 72-year-old woman (9th screen) with heterogeneously dense breasts showing a 4 mm new asymmetry seen only on MLO projection of the left breast consistent with a developing asymmetry. Corresponding US was not contributory (5a and 5b), and on stereotactic core biopsy this was discovered to be an invasive ductal carcinoma. There was no nodal involvement. The lesion is better outlined on spot compression view (5d). Figure 5c represents photographic magnification of the lesion in the original mammogram.
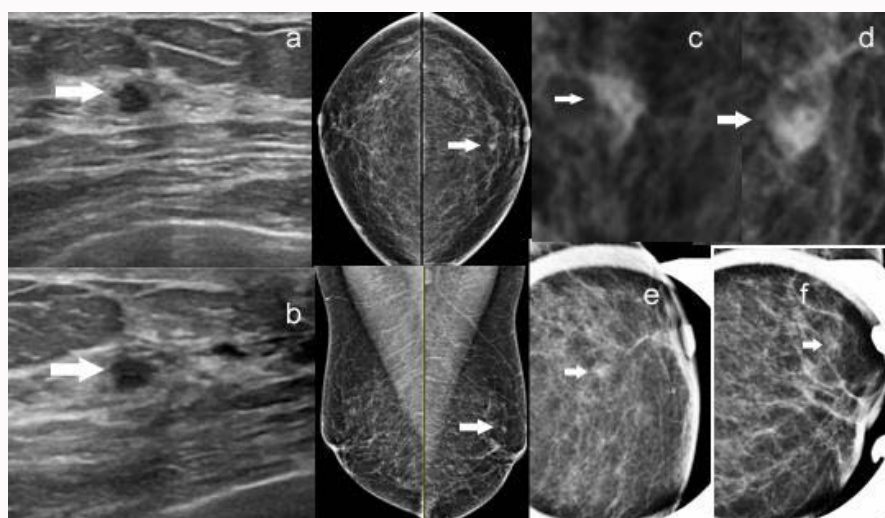
Figure 6: 73-year-old woman (1st screen) with entirely fatty breasts showing a 3 mm very dense irregular mass CAD-negative lesion in the retroareolar region. The lesion was poorly seen on spot compression views (6e and 6f) Corresponding US showed an irregular hypoechoic mass (6a and 6b) which on biopsy was discovered to be an invasive ductal carcinoma. Figures 6c and 7d represent photographic magnification of the lesion in the original mammogram.
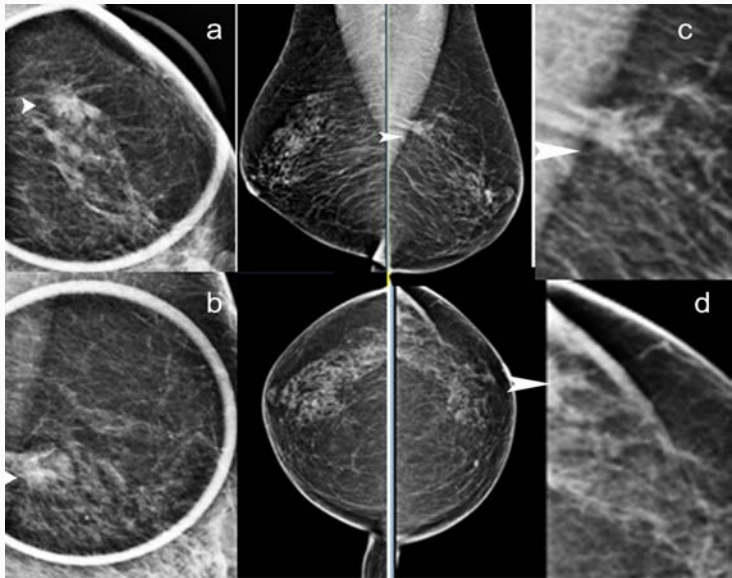
Figure 7: 73-year-old woman (3rd screen) with entirely fatty breasts showing a 17 mm × 15 mm × 9 mm new asymmetry in the posterior upper outer quadrant of the left breast only partially visualized on the CC view (7d) consistent with a developing asymmetry (arrowheads). On biopsy this was discovered to be an invasive lobular carcinoma. The lesion is better outlined on spot compression views (7a and 7b). Figures 7c and 7d represent photographic magnification of the lesion in the original mammogram.

Table 2: Attributes of the CAD-negative lesions.

	AGE (yrs)	SCRN #	MAMMOGRAPHIC		TUMOR Size (mm)	TUMOR Histology	TUMOR Nodal Stage	TUMOR Laterality	TUMOR Quadrant	TUMOR Location
			Appearance	Density						
1	75	9	Irregular density	Entirely fatty	4 × 4 × 4	Cribriform DCIS, intermediate Grade	N/A	Left	LOQ	Post.1/3
2	70	9	Irregular density	Entirely fatty	5 × 4 × 6	Papillary DCIS, intermediate Grade	N/A	Left	Retro-areolar	Ant.1/3
3	82	2	Irregular lobulated subdermal density	Hetero-geneous	5 × 5 × 5	Mucinous intracystic DCIS, high Grade	N/A	Right	Retro-areolar	Ant 1/3
4	52	3	Spiculated density	Hetero-geneous	2 × 2 × 2	Invasive ductal carcinoma, low Grade	Negative	Right	UOQ	Axillary Tail
5	72	9	Irregular density	Hetero-geneous	5 × 3 × 3	Invasive ductal carcinoma, intermediate Grade	Negative	Left	UOQ	Post.1/3
6	73	1	Irregular density	Entirely fatty	2 × 3 × 3	Invasive ductal carcinoma, high Grade	Negative	Left	Retro-areolar	Ant. 1/3
7	73	3	Irregulard Density	Entirely fatty	9 × 15 × 17	Invasive lobular carcinoma, low Grade	Negative	Left	UOQ	Post.1/3

that CAD markings are deployed at the start of the mammographic interrogation rather than the end, and CAD be used by the radiologist to produce a joint reading, instead of each producing an independent reading as in other studies.

In this study, 6 out of 7 CAD-negative calls that yielded cancer were due to new or enlarging lesions in patients who had previous studies for comparison. It is therefore logical to utilize findings arising from comparison with previous studies as a condition to overrule CAD when dealing with CAD-negative lesions. Hypothetically, if we had applied this proviso to the entire CAD negative cohort, the only cancer missed by ignoring all CAD-negative lesions would have been a 2 mm × 3 mm, node negative, invasive cancer found in the first screen.

The high cancer detection rate (5.9 cancers per 1,000 screens) and low recall rate (4.7%) in this study are indicators of high expertise among the screening radiologists in the center. It is possible that in centers with higher recall rates, the number of CAD-negative recalls would have been even higher than the 108 patients found in this study. Besides misuse of resources, the negative recalls have also been found to be a source of anxiety among women undergoing screening mammography and a major factor in poor reattendance in future screenings [12,13]. The results of this study are generalizable to programs screening women 50 years and older when using R2 CAD with intermediate sensitivity and Hologic units. There are other CAD units in the market which may not function as well, bringing disrepute to all CAD programs. Unfortunately, many publications on CAD fail to indicate the CAD unit used in the study when assessing

their performance.

Strength of this prospective study is that it was conducted through a dedicated provincial screening program, and all work ups were done at a dedicated tertiary referral breast center. The authors had access to a centralized provincial database for calculation of various stats on breast cancer diagnoses and other performance outcome indicators. The population-based screening program allowed for many breast cancers to be diagnosed.

A limitation to our study is that the results are only applicable to women 50 years or older. In accordance with the provincial screening policy, women aged 40 to 50 are considered average risk and are not screened in Ontario. But since 5 of 7 CAD-negative cancers were in fatty breasts, it is likely that the inclusion of denser breasts in women aged 40 to 50 would not have changed the results since the CAD-negative cancers in this study would have been likely masked by dense breast tissue. We also relied on the stability of lesions for two years as evidence of benignity, rather than having a biopsy and tissue sample in every case, although this is an accepted practice defined in the ACR lexicon [14]. Another possible limitation is that because cases were considered as CAD-positive based on a single concordant marking, some CAD-positive cases may not have recalled the same area as the breast cancer and could have been recalled serendipitously for another abnormality in the same breast region on single view while being incorrectly attributed as detecting the cancer.

In conclusion, the results of this study suggest that given the high negative predictive value of CAD, breast radiologists may consider using extra scrutiny when calling a lesion that is CAD-negative, unless it falls in CAD-blind category which are lesions that are new in comparison to the previous studies as well as lesions that can be seen in two views. Such an approach could help reduce false positive rates without affecting clinical outcomes given the small number and early stages of cancers missed by CAD. Further research is required to confirm the generalizability of our results to other screening centers. In particular, additional research involving women between the ages of 40 and 50 is needed to determine whether the results of this study would also apply to this age group with denser breast tissues.

References

1. Warren-Burhenne LJ, Wood SA, D'Orsi CJ, Feig SA, Kopans DB, O'Shaughnessy KF, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology 2000;215:554-62.

2. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center. Radiology. 2001;220(3):781-6.

3. Gur D, Sumkin JH, Rockette HE, Ganott M, Hakim C, Hardesty L, et al. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. J Natl Cancer Inst. 2004;96:185-90.

4. Cole EB, Zhang Z, Marques HS, Edward Hendrick R, Yaffe MJ, Pisano ED. Impact of computer-aided detection systems on radiologist accuracy with digital mammography. AJR Am J Roentgenol. 2014;203(4):909-16.

5. Georgian-Smith D, Moore RH, Halpern E, Yeh ED, Rafferty EA, D'Alessandro HA, et al. Blinded comparison of computer-aided detection with human second reading in screening mammography. AJR Am J Roentgenol. 2007;189(5):1135-41.

6. Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. N Engl J Med. 2008;359(16):1675-84.

7. Lehman CD, Wellman RD, Buist DSM, Kerlikowske K, Tosteson ANA, Miglioretti DL; Breast Cancer Surveillance Consortium. Diagnostic accuracy of digital screening mammography with and without computer-Aided Detection. JAMA. 2015;175(11):1828-37.

8. Kohli A, Jha S. Why CAD failed in mammography? JACR 2018;15(3):535-37.

9. Fenton JJ, Xing G, Elmore JG, Bang H, Chen SL, Lindfors KK, et al. Short-term outcomes of screening mammography using computer-aided detection: A population-based study of Medicare enrollees. Ann Intern Med. 2013;158(8):580-7.

10. Thurfjell E, Thurfjell G, Egge E, Bjurstam N. Sensitivity and specificity of computer-assisted breast cancer detection in mammography screening. Acta Radiol 1998;39(4):384-8.

11. Birdwell RL. The preponderance of evidence supports computer-aided detection for screening mammography. Radiology. 2009;253(1):9-16.

12. Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. Psychooncology 2005;14(11):917-38.

13. Brewer NT, Salz T, Lillie SE. Systematic review: The longterm effects of false-positive mammograms. Ann Intern Med. 2007;146(7):502-10.

14. American College of Radiology. Breast imaging reporting and data system (BI-RADS) Atlas. 5th ed. Reston, VA: American College of Radiology; 2013.