Review Article

Acta Radiologica
0(0) 1–6

DOI: 10.1177/0284185118770917
journals.sagepub.com/home/acr
SAGE

The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review

Emilie L Henriksen[1,2], Jonathan F Carlsen[1], Ilse MM Vejborg[1], Michael B Nielsen[1] and Carsten A Lauridsen[1,2]

Abstract

Background: Early detection of breast cancer (BC) is crucial in lowering the mortality.

Purpose: To present an overview of studies concerning computer-aided detection (CAD) in screening mammography for early detection of BC and compare diagnostic accuracy and recall rates (RR) of single reading (SR) with SR + CAD and double reading (DR) with SR + CAD.

Material and Methods: PRISMA guidelines were used as a review protocol. Articles on clinical trials concerning CAD for detection of BC in a screening population were included. The literature search resulted in 1522 records. A total of 1491 records were excluded by abstract and 18 were excluded by full text reading. A total of 13 articles were included.

Results: All but two studies from the SR vs. SR + CAD group showed an increased sensitivity and/or cancer detection rate (CDR) when adding CAD. The DR vs. SR + CAD group showed no significant differences in sensitivity and CDR. Adding CAD to SR increased the RR and decreased the specificity in all but one study. For the DR vs. SR + CAD group only one study reported a significant difference in RR.

Conclusion: All but two studies showed an increase in RR, sensitivity and CDR when adding CAD to SR. Compared to DR no statistically significant differences in sensitivity or CDR were reported. Additional studies based on organized population-based screening programs, with longer follow-up time, high-volume readers, and digital mammography are needed to evaluate the efficacy of CAD.

Keywords

Computer-aided detection, mammography screening, diagnostic accuracy, breast cancer, early detection

Date received: 29 October 2017; accepted: 14 March 2018

Introduction

Breast cancer (BC) is the most common cancer among women worldwide (1). In 2012, 1.67 million cases were diagnosed and approximately 522,000 deaths were reported (1). Early detection is crucial in lowering the mortality, and screening mammography is effective in the detection of BC in its early stages (2,3). Screening programs differ between countries and factors such as screening interval, age, technology, and organization vary considerably (4). The differences between the USA and Europe are major (4). In Europe, population-based screening with at least one high-volume reader is encouraged (2), and it is recommended that each mammogram is read by two radiologists; double reading (DR) is standard practice in Europe (2,5,6). In the USA, population-based screening programs are not used and single reading (SR) is common (5,7).

[1]Department of Diagnostic Radiology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[2]Department of technology, Faculty of Health and Technology, Metropolitan University College, Copenhagen, Denmark

Corresponding author:
Emilie L Henriksen, Sigurdsgade 26, 2200 Copenhagen, Denmark.
Email: emilie92@live.dk

The diagnostic accuracy of mammography depends on factors such as breast structure, density, and the radiologist's perception and level of experience (5,8). For SR, it is estimated that 70% of all missed BCs are due to misinterpretation while 30% are overlooked lesions (8,9). Computer-aided detection (CAD) has been developed to increase the sensitivity of mammographic examinations (10,11) by marking suspicious regions on mammograms such as microcalcifications and masses (10,12). In 1998, the first CAD system for mammography was approved by the U.S. Food and Drug Administration and in 2016 the use of digital facilities had expanded to 91.8% in the USA (13).

The aim of this review is to present an overview of studies on the use of CAD in screening mammography for early detection of BC. This review compares the diagnostic accuracy and recall rates (RR) of SR with SR + CAD and DR with SR + CAD.

Material and Methods

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (14) were used as a review protocol.

The literature search was performed in four databases: PubMed; Web of Science; Embase; and Cochrane Library on 8 June 2017. We combined the Medical Subject Heading (MeSH) terms in PubMed with a free text search to include studies without MeSH terms resulting in the following search string: mammography AND (screening OR mass screening OR breast screening OR X-ray screening OR early detection) AND (cad OR computer assisted diagnosis OR computer aided detection OR decision support systems) AND (breast neoplasm OR breast carcinoma OR breast cancer).

Only human trials, articles in English, and articles published within the last ten years were searched. Three authors screened the records by abstract. Clinical trials concerning CAD in screening mammography were included, knowing that screening populations are dissimilar between countries. The remaining articles were read in full text by the same authors. Studies that did not concern CAD in screening mammography and studies that tested CAD systems/algorithms, which were not available for screening facilities, were excluded. Studies with previously published data and studies with selective data acquisition were also excluded. Reference lists were searched for further eligible articles.

We registered demographic data and diagnostic values: cancer rate (CR); cancer detection rate (CDR); sensitivity; specificity; RR; and positive predictive value (PPV) for all included studies. CR was defined as screen-detected cancers + interval cancers (detected in the follow-up time). The CDR expressed the screen-detected cancers. If possible, at least two of the abovementioned authors calculated missing values from information in the articles.

Results

Literature search

The database searches yielded 1522 records of which 1491 were excluded by abstract. Of the remaining 31 articles, 18 were excluded after full text reading, leaving a total of 13 included studies (Fig. 1).

Suppl. Tables 1–4 show the demographic data and the results of each study.

Suppl. Tables 1 and 2 show data from studies (15–22) on SR vs. SR + CAD. In Suppl. Table 1 the reported diagnostic accuracies are calculated from the number of examinations; in Suppl. Table 2 the reported diagnostic accuracies are calculated from the number of mammograms (two views). Suppl. Tables 3 and 4 show data from studies (23–25) on DR vs. SR + CAD. In Suppl. Table 3 the reported diagnostic accuracies are calculated from the number of examinations; in Table 4 the reported diagnostic accuracies are calculated from the number of mammograms (two views). Two studies (26,27) are presented in both Suppl. Tables 1 and 3.

$$SR vs. SR + CAD$$

Suppl. Table 1 presents data from nine studies concerning SR vs. SR + CAD (15–21,26,27). Suppl. Table 2 presents data from one study (22). Nine studies were conducted in the USA (15–19,21,22,26,27) and one in Spain (20).

The follow-up time and inclusion criteria varied between studies. CR was in the range of 2.8–9.6 per 1000 examinations in seven studies (15,16,18–20,26,27) and 4.9–5.1 per 1000 mammograms in two cohorts of one study (22). Two studies did not report CR (17,21).

Eight studies reported sensitivities (15,16,18–20, 22,26,27). Five studies (15,16,18,20,27) showed an increased sensitivity using SR + CAD compared to SR, two studies showed a decrease (19,22), and one showed no difference (26).

All ten studies reported CDRs. Seven studies (15–18,20, 21,27) showed an increased CDR using SR + CAD compared to SR, two studies showed a decrease (19,22), and one study revealed no difference (26).

Three studies reported separate results for malignant mammographic masses and malignant microcalcifications (16,17,20). Two of these studies showed that CAD marked 65–67% of masses or asymmetries and 100% of microcalcifications (16,17), while radiologists detected 90–97% and 89–84%, respectively (16,17).

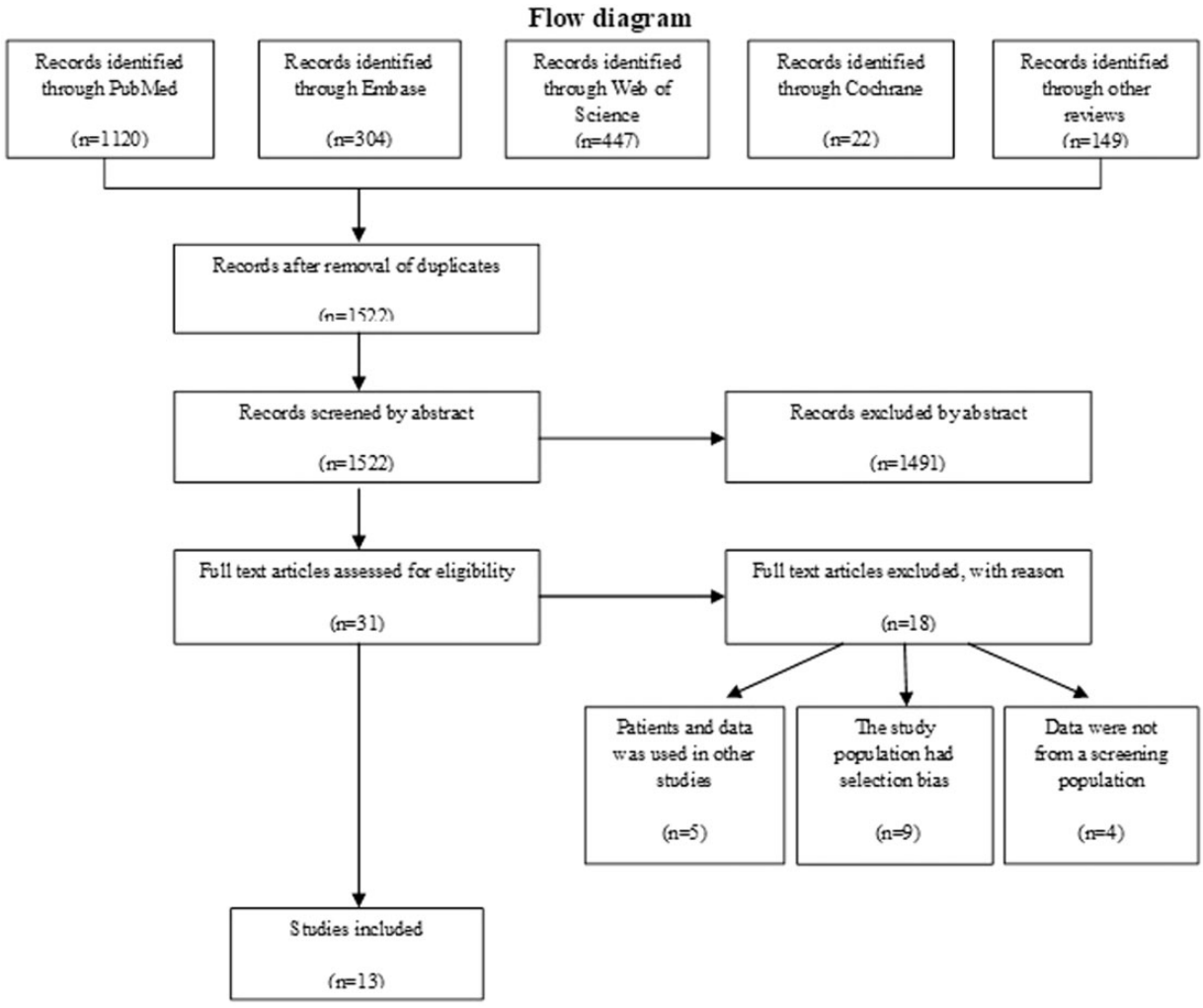

Fig. 1. Flow diagram according to Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA).

The third study showed that CAD marked 77.6% of masses and 91.3% of microcalcifications (20).

Five studies reported separate results for carcinomas in situ and invasive cancers (15,18,19,21,27). In two of the studies the sensitivity was reported for the radiologist, the CAD system, and both the radiologist and CAD (15,27). CAD had higher sensitivity for carcinomas in situ than the radiologist whereas the radiologist had higher sensitivity for invasive cancers (15,27). The overall sensitivity increased when using SR + CAD compared to SR (15,27). In the last three studies the CDR was reported separately for invasive cancer and for ductal carcinoma in situ (DCIS) (18,19,21). Two showed an increase in CDR for DCIS and no difference for invasive cancer when using CAD (18,21). One study showed no difference in CDR for DCIS and a decreased CDR for invasive cancers when using CAD (19).

Nine studies reported data for RRs (15–20,22,26,27). RRs increased when using CAD in seven studies (15–18, 20,26,27), decreased in one (22), and showed no difference in the last study (19). The study in Suppl. Table 2 reported a RR of 8.7% for SR + CAD and 9.1% for SR (22).

Seven studies stated the specificity for SR and for SR + CAD (16,18–20,22,26,27). The specificity decreased in four studies (18,20,26,27), increased in one study (22), and showed no difference in two studies (16,19) when adding CAD. The specificity was in the range of 90.2–99% for SR and 87.2–99% for SR + CAD.

In two studies (16,20) the rate of false-positive CAD marks was 97% and 77.4%, respectively.

DR vs. SR + CAD

Suppl. Table 3 presents data from four studies concerning DR vs. SR + CAD (23,24,26,27). Suppl. Table 4 presents data from one (25). Two were conducted in the USA (26,27), one in England (23), and two in Spain (24,25).

Three studies reported CRs showing a variation in the range of 2.8–8 per 1000 examinations (23,26,27).

Three studies reported sensitivities (23,26,27) and two of these showed a decreased sensitivity (23,26), while one showed an increased sensitivity using SR + CAD compared to DR (27). The sensitivity of DR was in the range of 83.3–88% and the sensitivity of SR + CAD was in the range of 72.2–90.4%.

All five studies reported CDRs. Three studies showed a lower CDR (23,24,26), while two studies showed a higher CDR when using SR + CAD compared to DR (25,27).

Two studies reported results for carcinomas in situ and/or invasive cancers (24,27). One study showed a sensitivity for invasive cancer and carcinomas in situ of 85.3% and 95% for DR and 87.2% and 98.1% for SR + CAD (27). The other study showed the fraction of carcinomas in situ and reported 3.8% fewer cases of carcinomas in situ in the cohort using SR + CAD compared to the cohort using DR (24).

Four studies reported the RR (23,25–27). One study showed no difference between SR + CAD compared to DR (26), two showed an increased RR (23,25), and the final study revealed a decreased RR when using SR + CAD compared to DR (27). Three studies reported data on specificity (23,26,27), of which two studies showed a decrease (23,26) and one showed an increase (27) when using SR + CAD compared to DR. The specificity was in the range of 88.4–97.4% for DR and 90–96.9% for SR + CAD.

Risk of bias

Due to differences in screening programs between countries, the origin of studies encompasses a risk of bias. Nine out of the 13 studies were conducted in the USA (15–19,21,22,26,27). Reader experience and volume vary between countries and screening facilities. This is essential when assessing the efficacy of CAD. Only four studies (18,23–25) reported reader volumes which reveal a risk of bias. Earlier detection with CAD does not only affect the sensitivity but could also increase the lead time. Also, the fact that the interval cancers are used to compute the sensitivity presents a risk of bias. The follow-up time was incompletely described or not reported in six studies (15,17,23–25,27) and presents a major shortcoming because the CR is highly dependent on follow-up time. No studies had a follow-up time > 12 months. In itself, this presents a high risk of bias. To be able to evaluate interval CR, at least a two-year period is necessary.

Discussion

To evaluate the efficacy of CAD in mammography screening we included 13 studies comparing either SR and SR + CAD or DR and SR + CAD. All but two studies on SR vs. SR + CAD showed an increased sensitivity and/or CDR when adding CAD whereas no difference was seen for the DR vs. SR + CAD group. Adding CAD to SR increased the RR and decreased the specificity in all but one study. Only one study showed a significant difference in RR between DR and SR + CAD.

The results for SR vs. SR + CAD indicate that the addition of CAD increases sensitivity and CDR; however, two studies showed a decreased sensitivity when adding CAD to SR (19,22). For DR vs. SR + CAD, no significant differences in sensitivity and CDR were reported. The results demonstrated that CAD had a higher sensitivity or CDR for carcinomas in situ compared with invasive cancer. The studies including results concerning different mammographic tumor characteristics showed that CAD was better at detecting microcalcifications than mammographic masses.

The variation in CR of the 13 studies may be caused by differences in study set-up and/or differences in screening programs. Differences in patient inclusion depends mainly on whether the screening program is population-based or based on self-reference. CR is also influenced by follow-up time and definitions of subsequent cancers. Subsequent cancers increase with the follow-up time and result in a higher CR. All except one of the studies that reported follow-up time had a follow-up time of one year, which is a short period for assessing a sufficient detection rate compared to the interval cancers.

All CAD prompts needs assessments by a radiologist who decides whether a CAD-mark should be dismissed or not. This means that CAD depends on readers. Several studies noted the potential of CAD and showed that CAD marked cancers which were dismissed by the radiologist (9,20,28,29). Furthermore, it has been shown that the effect of CAD differs with reader experience and volume. Less experienced and low-volume readers benefit more from CAD than experienced, high-volume readers (30–33).

The sensitivity of CAD depends on mammographic findings. Three articles state that CAD detected malignant microcalcifications with sensitivities > 90% whereas the detection of other findings was less accurate (9,32,34). Our study supports this and shows that CAD detected 90–100% of microcalcifications but only 65–67% of masses. In parallel, our results indicate that CAD detects carcinomas in situ better than invasive cancers.

Our study shows that RRs generally increased and the specificity decreased when using SR + CAD compared to SR and DR. Nine studies showed an increase in RR and five studies showed a decreased specificity. CAD is associated with a large number of false-positive marks (32–35) and two studies showed a false-positive rate of CAD marks of 97% and 77.4% (16,20). The high rate of false-positive marks may lead to false-positive findings and affect the specificity and RRs. The RRs reported in the studies were generally high and

in seven studies the RR were > 7% (17–20,22,26,27). The European Guidelines for Quality Assurance in BC screening and diagnosis recommend that the RR for initial screening should be < 7% and preferably < 5% (6). For subsequent examinations, the RR should be < 5% and preferably < 3% (6).

CAD marks suspicious areas on mammograms for further evaluation by radiologists. The most common algorithms to identify regions of interest are pixel-based and region-based methods (13). Although the techniques used in different systems vary, the CAD evaluation encompasses some similar steps (10). The first step is segmentation of the breast region from the surrounding regions, this is followed by pre-processing which enhances contrast between suspicious and normal breast tissue (10). Subsequently, an extraction of features such as lesion shape and size is used as an input to differentiate suspicious from non-suspicious findings (10). Because of the high rate of false-positive marks vendors have developed CAD systems to provide the ability to adjust the thresholds for sensitivity and specificity and thereby regulate the number of false-positive marks. A high sensitivity setting results in more false-positive prompts and lower specificity. One study reported a study period during 2010–2012 for SR + CAD and thereby had the newest data.

One limitation of this study is that the inclusion of women as screening patients varied in the included studies. This is a result of differences in screening programs between countries and facilities. In Europe there are organized population-based screening-programs whereas the USA employs screening for self-referred women. Furthermore, differences in reader experience and volume imply a limitation (19). Another limitation is the small number of studies using full-field digital mammography. In recent years, the technology of image acquisition in mammography has undergone a transition from analogue to digital. A concurrent transition in CAD technique has followed. Only four studies (21,22,24,25) reported results from digital mammography which is currently standard practice (7). When using screen-film mammography for CAD evaluation, the mammograms need to be digitalized which induces increased image noise. This may affect the CAD evaluation (19). The sensitivity is computed using interval cancers which propose a limitation. Cancers are often visible on priors and could be detected earlier. Earlier detection with CAD will not only change the sensitivity but may also lead to increased lead time. The short follow-up time presents a limitation. One year is insufficient to determine the CR which is used to calculate the different measures of diagnostic accuracy. The number of patients in the studies was in the range of 5631 to 1,621,206. Due to the low incidence of cancer in a screening population, the evaluation of CAD efficacy can be difficult because a large number of examinations is necessary to achieve significant differences (34). One more limitation of the present study was the variation in reporting of findings and diagnostic values. Also, two of the studies used the number of mammograms consisting of two views to calculate the diagnostic values whereas the remaining 11 studies used the number of examinations.

In conclusion, all but two studies showed that SR + CAD increases RRs, sensitivity, and CDR in mammography screening compared to SR alone. Compared to DR no statistically significant differences in sensitivity or CDR were reported. Additional studies are needed to evaluate the effect of CAD in an organized population-based screening program with high-volume readers. Furthermore, studies with a longer follow-up time are needed for a sufficient assessment of cancer rates. Finally, studies based on digital mammography are needed to evaluate the efficacy of CAD in the technology that is presently standard practice. Due to recent breakthroughs in artificial intelligence we may expect improvements in CAD systems.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Emilie L Henriksen http://orcid.org/0000-0003-1411-0596

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:359–386.
2. Sardanelli F, Hildegunn SA, Álvarez M, et al. Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Israel, Lithuania, Moldova, The Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and Turkey. Eur Radiol 2016;27:2737–2743.
3. Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening–viewpoint of the IARC Working Group. N Engl J Med 2015;372:2353–2358.
4. Jacobsen KK, Abraha L, Buist DSM, et al. Comparison of cumulative false-positive risk of screening mammography in the United States and Denmark. Cancer Epidemiol 2015;39:656–663.

5. Bazzocchi M, Mazzarella F, Del Frate C, et al. CAD systems for mammography: a real opportunity? A review of the literature. Radiol Med 2007;112:329–353.
6. Perry N, Broeders M, de Wolf C, et al. European Commission. European guidelines for quality assurance in breast cancer screening and diagnosis. 4th edn. Luxembourg, Office for Official Publications of the European Communities, 2006.
7. Azavedo E, Zackrisson S, Mejáre I, et al. Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. BMC Med Imaging 2012;12:12.
8. Mello-Thoms C. Perception of breast cancer: eye-position analysis of mammogram interpretation. Acad Radiol 2003;10:4–12.
9. Birdwell RL. The preponderance of evidence supports computer-aided detection for screening mammography. Radiology 2009;253:9–16.
10. Giger ML, Chan HP, Boone J, et al. Anniversary paper: History and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. Med Phys 2008;35:5799–5820.
11. Dromain C, Bover B, Canale S, et al. Computed-aided diagnosis (CAD) in the detection of breast cancer. Eur J Radiol 2013;82:417–423.
12. Taylor P, Champness J, Given-Wilson R, et al. Impact of computer-aided detection prompts on the sensitivity an specificity of screening mammography. Health Technol Assess 2005;9:1–58.
13. Keen JD, Keen JM, Keen JE. Utilization of computer-aided detection for digital screening mammography in the United States, 2008 to 2016. J Am Coll Radiol 2017;15: 44–48.
14. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009;6:e1000100.
15. Dean JC, Ilvento CC. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: Prospective study of 104 cancers. Am J Roentgenol 2006;187:20–28.
16. Ko JM, Nicholas MJ, Mendel JM, et al. Prospective assessment of computer-aided detection in interpretation of screening mammography. Am J Roentgenol 2006;187: 1483–1491.
17. Morton MJ, Whaley DH, Brandt KR, et al. Screening mammograms: Interpretation with computer-aided detection-prospective evaluation. Radiology 2006;239: 375–383.
18. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med 2007;356:1399–1409.
19. Fenton JJ, Abraham L, Taplin SH, et al. Effectiveness of computer-aided detection in community mammography practice. J Natl Cancer Inst 2011;103:1152–1161.
20. Gomez SS, Tabanera MT, Bolivar AV, et al. Impact of a CAD system in a screen-film mammography screening program: A prospective study. Eur J Radiol 2011;80:317–321.
21. Fenton JJ, Xing G, Elmore JG, et al. Short-term outcomes of screening mammography using computer-aided detection: a population-based study of medicare enrollees. Ann Intern Med 2013;158:580–587.
22. Lehman CD, Wellman RD, Buist DS, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med 2015;175:1828–1837.
23. Gilbert FJ, Astley SM, Gillan MG, et al. Single reading with computer-aided detection for screening mammography. N Engl J Med 2008;359:1675–1684.
24. Romero C, Almenar A, Pinto JM, et al. Impact on breast cancer diagnosis in a multidisciplinary unit after the incorporation of mammography digitalization and computer-aided detection systems. Am J Roentgenol 2011; 197:1492–1497.
25. Bargallo X, Santamaría G, del Amo M, et al. Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program. Eur J Radiol 2014;83:2019–2023.
26. Georgian-Smith D, Moore RH, Halpern E, et al. Blinded comparison of computer-aided detection with human second reading in screening mammography. Am J Roentgenol 2007;189:1135–1141.
27. Gromet M. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. Am J Roentgenol 2008;190: 854–859.
28. Skaane P, Kshirsagar A, Stapleton S, et al. Effect of computer-aided detection on independent double reading of paired screen-film and full-field digital screening mammograms. Am J Roentgenol 2007;188:377–384.
29. Skaane P, Kshirsagar A, Hofvind S, et al. Mammography screening using independent double reading with consensus: is there a potential benefit for computer-aided detection? Acta Radiol 2012;53:241–248.
30. Feig SA, Sickles EA, Evans WP, et al. Re: Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. J Natl Cancer Inst 2004;96:1260–1261.
31. Alberdi E, Povyakalo AA, Strigini L, et al. CAD in mammography: lesion-level versus case-level analysis of the effects of prompts on human decisions. Int J Comput Assist Radiol Surg 2008;3:115–122.
32. Houssami N, Given-Wilson R, Ciatto S. Early detection of breast cancer: Overview of the evidence on computer-aided detection in mammography screening: Radiology - Review article. J Med Imaging Radiat Oncol 2009;53:171–176.
33. Noble M, Bruening W, Uhl S, et al. Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis (Structured abstract). Arch Gynecol Obstet 2009;279:881–890.
34. Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation? Radiology 2009; 253:17–22.
35. Brem RF, Hoffmeister JW, Zisan G, et al. A computer-aided detection system for the evaluation of breast cancer by mammographic appearance and lesion size. Am J Roentgenol 2005;184:893–896.