ORIGINAL ARTICLE

Single Reading with Computer-Aided Detection for Screening Mammography

Fiona J. Gilbert, F.R.C.R., Susan M. Astley, Ph.D., Maureen G.C. Gillan, Ph.D.,
Olorunsola F. Agbaje, Ph.D., Matthew G. Wallis, F.R.C.R.,
Jonathan James, F.R.C.R., Caroline R.M. Boggis, F.R.C.R.,
and Stephen W. Duffy, M.Sc., for the CADET II Group*

ABSTRACT

BACKGROUND

The sensitivity of screening mammography for the detection of small breast cancers is higher when the mammogram is read by two readers rather than by a single reader. We conducted a trial to determine whether the performance of a single reader using a computer-aided detection system would match the performance achieved by two readers.

METHODS

The trial was designed as an equivalence trial, with matched-pair comparisons between the cancer-detection rates achieved by single reading with computer-aided detection and those achieved by double reading. We randomly assigned 31,057 women undergoing routine screening by film mammography at three centers in England to double reading, single reading with computer-aided detection, or both double reading and single reading with computer-aided detection, at a ratio of 1:1:28. The primary outcome measures were the proportion of cancers detected according to regimen and the recall rates within the group receiving both reading regimens.

RESULTS

The proportion of cancers detected was 199 of 227 (87.7%) for double reading and 198 of 227 (87.2%) for single reading with computer-aided detection (P=0.89). The overall recall rates were 3.4% for double reading and 3.9% for single reading with computer-aided detection; the difference between the rates was small but significant (P<0.001). The estimated sensitivity, specificity, and positive predictive value for single reading with computer-aided detection were 87.2%, 96.9%, and 18.0%, respectively. The corresponding values for double reading were 87.7%, 97.4%, and 21.1%. There were no significant differences between the pathological attributes of tumors detected by single reading with computer-aided detection alone and those of tumors detected by double reading alone.

CONCLUSIONS

Single reading with computer-aided detection could be an alternative to double reading and could improve the rate of detection of cancer from screening mammograms read by a single reader. (ClinicalTrials.gov number, NCT00450359.)

From the Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen (F.J.G., M.G.C.G.); the Department of Imaging Science and Biomedical Engineering, University of Manchester, Manchester (S.M.A.); the Department of Epidemiology, Mathematics, and Statistics, Wolfson Institute of Preventive Medicine, London (O.F.A., S.W.D.); the Cambridge Breast Unit, Addenbrookes Hospital, Cambridge (M.G.W.); the Nottingham Breast Institute, Nottingham City Hospital, Nottingham (J.J.); and the Nightingale Breast Screening Unit, Wythenshawe Hospital, Manchester (C.R.M.B.) — all in the United Kingdom. Address reprint requests to Dr. Gilbert at the Aberdeen Biomedical Imaging Centre, University of Aberdeen, Lilian Sutton Bldg., Foresterhill, Aberdeen AB25 2ZD, Scotland, United Kingdom, or at f.j.gilbert@abdn.ac.uk.

THE INTERPRETATION OF SCREENING mammograms by two readers (double reading) rather than by one reader increased the rate of cancer detection by 4 to 14% in several studies[1-6] and by 10% according to a meta-analysis.[7] Double reading is standard practice in at least 12 European countries[8] but has not been widely adopted in the United States. The use of computer-aided detection systems has the potential to increase cancer-detection rates. Thus, a single reader using computer-aided detection might be able to match the performance of two readers.[9,10] Single reading with computer-aided detection could be an alternative to double reading and improve the results of single reading.

Computer-aided detection systems use computer algorithms to analyze digital mammographic images. They identify and mark potentially suspicious regions to attract the reader's attention to features that might have been overlooked or dismissed as normal.[6,11-24] In the United States, where single reading is standard practice, commercial computer-aided detection systems have been widely adopted to improve reader performance[25-36] and are used in 25 to 30% of all mammogram readings.[37] However, the benefit of computer-aided detection in screening mammography was recently questioned in an observational study by Fenton et al.[35] and remains controversial.[38]

There have been calls for robust evidence from multicenter, randomized trials involving large numbers of participants and readers to inform screening programs that are considering an investment in computer-aided detection.[36,39-42] We conducted the prospective, randomized Computer-Aided Detection Evaluation Trial II (CADET II), in which single reading with computer-aided detection was compared with double reading during routine screening mammography in the United Kingdom National Health Service Breast Screening Programme (NHSBSP). Arbitration was used in cases of disagreement between the two readers, since this is United Kingdom practice and has been shown to reduce recall rates while maintaining high cancer-detection rates.[43-45]

METHODS

STUDY DESIGN

The study was designed as an equivalence trial, with matched comparisons between the sensitivity of single reading with computer-aided detection and the sensitivity of double reading. A total of 31,057 women were recruited between September 2006 and August 2007 from the static and mobile units of three NHSBSP centers in England (Manchester; Nottingham; and Warwickshire, Solihull, and Coventry). Information about the trial was posted 3 to 6 weeks before their appointments to women who were invited to undergo routine screening by two-view film mammography. Written informed consent was obtained at the mammography appointment. The study was approved by the South East Multi-Centre Research Ethics Committee.

RANDOMIZATION

Batches of films from each screening session were randomly assigned, at a ratio of 1:1:28, to one of three reading regimens. Group A, with a planned number of 1000 subjects, was assigned to receive a double reading only; group B, with a planned number of 1000 subjects, was assigned to receive a single reading with computer-aided detection only; group C, with a planned number of 28,000 subjects, was assigned to receive both a single reading with computer-aided detection and a double reading. Assignments were made with the aid of a random-number generator and provided to each center in sequentially numbered, sealed envelopes.

FILM DIGITIZATION AND COMPUTER-AIDED DETECTION OUTPUT

Trial mammograms were digitized and analyzed by the software detection algorithms of an Image-Checker DMax computer-aided detection system, version 8.1 (Hologic/R2 Technology). The prompting markers were viewed on a CheckMate flat-panel display screen superimposed on an image of the corresponding mammogram at a reduced resolution. The PeerView facility provided the reader with a high-resolution image of any region of interest that was identified by the computer-aided detection algorithm.[46] Markers were generated for masses (asterisks) and microcalcifications (solid triangles). If both a mass and microcalcification were detected, a four-pointed star was displayed. The sizes of the markers were related to the likelihood of cancer, as determined by the algorithms. In this study, the software-detection algorithm thresholds were set to operate at a detection sensitivity of approximately 88% for masses and 95%

for calcifications, with corresponding false marker rates of 1.5 and 1.0, respectively, per four-film examination.[46]

FILM READERS

The film readers were 17 radiologists, 2 breast-cancer clinicians, and 8 trained film-reading technologists (radiographers). All film readers met the requirements of the NHSBSP quality-assurance guidelines by reading at least 5000 screening mammograms per year and participating in the annual PERFORMS self-assessment test.[47] Readers acting as single readers with computer-aided detection or as the first of the two readers in a double reading had a median of 6 years of mammography experience (interquartile range, 4 to 14 years); the second readers had slightly less experience, with a median of 5 years (interquartile range, 2 to 10 years). Before the study, the readers completed a training course in the use of computer-aided detection involving 300 to 400 subjects.[42,48]

FILM READING

Films were independently double read or single read with computer-aided detection in separate reading sessions by different readers with no access to the decision of the recorded outcome from the other reading regimen. A reader who was one of the two double readers was not permitted to read the same subject's films with computer-aided detection. Mammograms from a previous round were mounted for viewing.

The readers recorded the results for each subject as "recall for further assessment" or "return to routine screening." Discordant results from double readings were arbitrated by a third reader or, in Manchester, by another pair of readers. In single reading with computer-aided detection, the readers viewed the mammogram and then accessed the computer image with prompts and reviewed any areas marked by the computer-aided detection system before recording in the National Breast Screening System their overall recommendation for "recall for further assessment" or "return to routine screening."

OUTCOME MEASURES

The primary outcome measures were the proportion of cancers detected by the two reading regimens and the recall rate. These were analyzed in a matched-pair comparison.
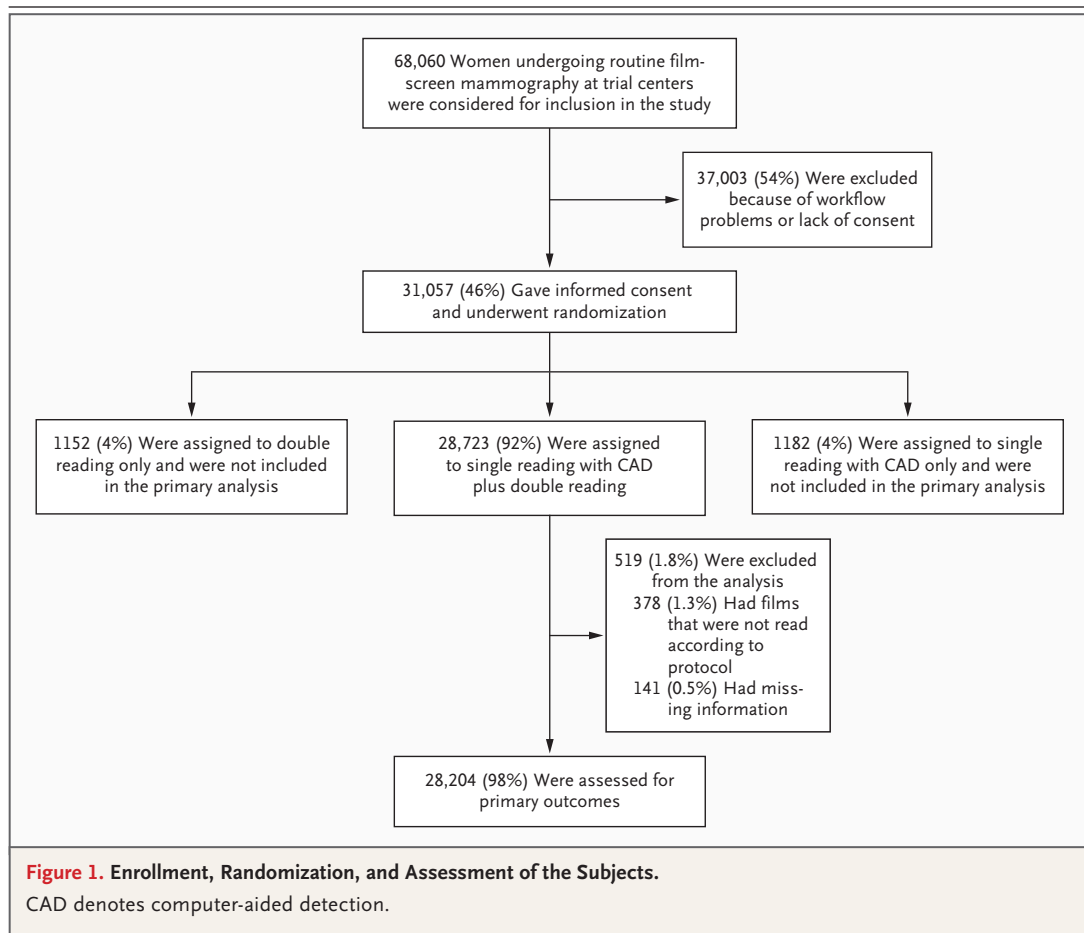
STATISTICAL ANALYSIS

The trial was designed as an equivalence study,[49,50] with a matched-pair comparison of the two regimens in the group that received both single reading with computer-aided detection and double reading. Equivalence was defined as a 95% confidence interval that ruled out a difference of more than 10% in either direction in the rate of detection of cancers. The 10% equivalence criterion is more rigorous than the 15% criterion often used in pharmaceutical equivalence studies.[51,52] Within the group receiving both regimens, the differences between recall rates for all subjects and for those with cancer were compared with the use of McNemar's method.[53] This method takes into account the fact that the decisions resulting from single reading with computer-aided detection and from double reading are made on the same subjects. For the primary end point, detection of cancers, the 95% confidence interval for the difference between detection rates was calculated. If the 95% confidence interval did not extend beyond ±10%, it was considered to be equivalence.

In 2004, the rate of detection of invasive and in situ breast cancers in the United Kingdom NHSBSP with the use of double reading was 6.91 cancers per 1000 women screened.[54] Increasing this rate by 11% because each mammogram was read according to two regimens in our study would result in a detection rate of 7.67 per 1000. Power calculations on the assumption of a 20% disagreement rate between the two regimens indicated a requirement of 215 cancers to establish equivalence according to our criterion of 10%. Thus, 28,000 mammograms would have to be read according to both regimens to achieve a power of 90%. In addition, to minimize reader bias, so that a reader using one of the regimens would not know whether the mammogram would be read according to the other regimen, a further 2000 mammograms, randomly assigned to either double reading or single reading with computer-aided detection, were included.[55]

RESULTS

Figure 1 shows a flow diagram of the trial. In the group receiving both single reading with computer-aided detection and double reading, 227 cancers were detected among 28,204 subjects, for an over-

68,060 Women undergoing routine film-screen mammography at trial centers were considered for inclusion in the study

37,003 (54%) Were excluded because of workflow problems or lack of consent

31,057 (46%) Gave informed consent and underwent randomization

1152 (4%) Were assigned to double reading only and were not included in the primary analysis

28,723 (92%) Were assigned to single reading with CAD plus double reading

1182 (4%) Were assigned to single reading with CAD only and were not included in the primary analysis

519 (1.8%) Were excluded from the analysis
378 (1.3%) Had films that were not read according to protocol
141 (0.5%) Had missing information

28,204 (98%) Were assessed for primary outcomes

Figure 1. Enrollment, Randomization, and Assessment of the Subjects.
CAD denotes computer-aided detection.

all screening-detection rate of 8.0 per 1000. The corresponding numbers for the group receiving double reading only and the group receiving single reading with computer-aided detection only were 12 cancers among 1152 subjects (10.4 per 1000) and 8 cancers among 1182 subjects (6.8 per 1000), respectively. Table 1 shows the distribution of subjects in the group receiving both single reading with computer-aided detection and double reading according to age, type of screening (prevalence or incidence screening), and center.

Table 2 shows the recall decisions after a single reading with computer-aided detection and after a double reading for all subjects and for all subjects with cancer in the group of women whose mammograms were read according to both regimens. The concordance between the two reading systems was 74.9% (170 of 227). The cancer-detection rates after single reading with computer-aided detection and after double reading were similar (7.02 and 7.06 per 1000, respectively; P=0.89), as were the proportions of cancers de-

tected (87.2% and 87.7%, respectively; P=0.89). The 95% confidence interval for the difference between 87.2% and 87.7% is −7.4% to +6.6%, which is within the 10% definition of equivalence.

The recall rates were 3.9% for single reading with computer-aided detection and 3.4% for double reading (P<0.001). The absolute difference in recall rates was 0.5 percentage point, with a 95% confidence interval of 0.3 to 0.8 percentage point corresponding to a relative difference of 15%. The estimated sensitivity, specificity, and positive predictive value for single reading with computer-aided detection were 87.2%, 96.9%, and 18.0%, respectively. The corresponding figures for double reading were 87.7%, 97.4%, and 21.1%. The sensitivity values are overestimates, because data on interval cancers (cancers detected within 1 year after a negative mammography result) were not yet available for all subjects. The same interval cancers were missed by both regimens, and hence the difference in sensitivity will remain the same for both regimens.

Table 3 shows the recall rates according to center and screening regimen for subjects participating in the trial and for women not participating in the trial who were undergoing routine contemporaneous screening, with double reading of their mammograms. The recall rates were similar for trial subjects whose mammograms were read by a single reader with computer-aided detection and for nontrial subjects at the Nottingham center and the Warwickshire, Solihull, and Coventry center. However, at Manchester the recall rate among trial subjects undergoing single reading with computer-aided detection was 1.4 percentage points higher than the recall rate among trial subjects undergoing double reading and 1.2 percentage points higher than the recall rate for nontrial subjects. Overall, the recall rate was 0.3 percentage point higher for mammograms that were given a single reading with computer-aided detection than for routine screening mammograms of nonparticipating women. The recall decision was arbitrated for 365 of the 28,723 women (1.3%) whose mammograms received a single reading with computer-aided detection; these women included 6 of the 227 with cancer (2.6%).

Table 4 lists the pathological features of the 57 cancers with discordant recall decisions according to screening regimen. The proportion of in situ tumors was higher for tumors detected only by single reading with computer-aided detection than for tumors detected only by double reading, but the difference was not significant (P=0.33). Among the invasive cancers, those detected only by single reading with computer-aided detection had similar node status (P=0.89) to those detected only by double reading; they were smaller (P=0.35) and less likely to be grade 1 and more likely to be grade 3 (P=0.07) than those detected only by double reading, but these differences were not significant. Thus, there were no significant differences in pathological attributes between tumors detected by single reading with computer-aided detection alone and those detected by double reading alone.

DISCUSSION

The trial demonstrated that the cancer-detection rates of a single reader using computer-aided detection and of two readers are similar (7.02 per 1000 and 7.06 per 1000, respectively). However, the recall rate for mammograms given a single reading with computer-aided detection was 0.5

Table 1. Characteristics of Subjects in the Group Receiving Both Single Reading with Computer-Aided Detection and Double Reading.

Characteristic	No. (%)
All subjects	28,204 (100.0)
Age	
<50 yr	93 (0.3)
50–54 yr	6,595 (23.4)
55–59 yr	7,821 (27.7)
60–64 yr	7,565 (26.8)
65–70 yr	5,855 (20.8)
>70 yr	275 (1.0)
Type of screening*	
Prevalence	4,287 (15.2)
Incidence	23,917 (84.8)
Center	
Manchester	9,355 (33.2)
Nottingham	9,312 (33.0)
Warwickshire, Solihull, and Coventry	9,537 (33.8)

* Prevalence screening refers to a woman's first mammography screening, and incidence screening refers to all subsequent screenings.

percentage point higher than that for mammograms given a double reading (3.9% and 3.4%, respectively), a relative difference of 15%.

There has been much recent debate regarding the use of computer-aided detection in clinical practice. Eight studies making the comparison between single reading with computer-aided detection and double reading were reviewed by Bennett et al.,[56] but methodologic differences precluded any definitive conclusions. Most of the studies were small, had a high proportion of subjects with cancer, and involved a limited number of readers with variable training in the use of computer-aided detection.[23] Taylor and Potts used pooled estimates of effect sizes from two meta-analyses[7]: one included 10 U.S. studies in which the comparison was made between single reading alone and single reading with computer-aided detection, and the second included 17 studies comparing single reading with double reading. The authors concluded that there was no significant difference in cancer-detection rates between single reading with computer-aided detection and double reading with arbitration in cases of disagreement, but they found that the recall rate with double reading was significantly lower than the rate with single reading with computer-aided detection. However, differences between the screening programs of the United States and the

Table 2. Recall Decisions for All Subjects and for Those with Cancer among Women Whose Mammograms Were Read According to Both Regimens.*

Single Reading with CAD	Double Reading, All Subjects			Double Reading, Subjects with Cancer		
	No Recall	Recall	Total	No Recall	Recall	Total
	number (percent)					
No recall	26,675	429	27,104 (96.1)	0	29	29 (12.8)
Recall	582	518	1,100 (3.9)	28	170	198 (87.2)
Total	27,257 (96.6)	947 (3.4)	28,204 (100)	28 (12.3)	199 (87.7)	227 (100.0)

* CAD denotes computer-aided detection.

Table 3. Recall Rates of Trial Subjects Whose Mammograms Were Read According to Both Regimens and Nontrial Subjects.*

Center	Trial Subjects			Nontrial Subjects†
	Single Reading with CAD	Double Reading	Total‡	Double Reading
	no./total no. (%)			
Warwickshire, Solihull, and Coventry	355/9537 (3.7)	342/9537 (3.6)	538/9537 (5.6)	446/12,023 (3.7)
Manchester	490/9355 (5.2)	354/9355 (3.8)	632/9355 (6.8)	590/14,732 (4.0)
Nottingham	255/9312 (2.7)	251/9312 (2.7)	359/9312 (3.9)	288/9869 (2.9)
Overall	1100/28,204 (3.9)	947/28,204 (3.4)	1529/28,204 (5.3)	1324/36,624 (3.6)

* CAD denotes computer-aided detection.
† Nontrial subjects were women not participating in the trial who were undergoing routine contemporaneous screening at the three centers, with double reading of their mammograms.
‡ This column represents the total recalled after either or both regimens were performed.

United Kingdom make these evaluations difficult to interpret.[57] Our large, prospective study helps to clarify some of these issues. The small increase in the number of women recalled for assessment in the group given a single reading with computer-aided detection as compared with the group given a double reading is consistent with the results of previously published studies.[26,30,31,58] Furthermore, our results are in line with the pooled estimates predicted by Taylor and Potts[7] in a comparison between double reading with arbitration in cases of disagreement and single reading with computer-aided detection.

The strengths of this study are that it was a multicenter, prospective trial with more than 30,000 women randomly assigned to one of three reading groups and that the readers were unaware of the reading assignments. Reflecting practice in the United Kingdom, the readers were all high-volume readers (reading more than 5000 mammograms per year) and were breast radiologists, breast-cancer clinicians, or film-reading technol-ogists. The single readers using computer-aided detection and the first readers of mammograms given double readings had almost identical experience in reading mammograms, with a median of 6 years of experience (interquartile range, 4 to 14 years) for both groups of readers. Thus, our study was not biased by a difference in the amount of experience of the two groups of readers, but the conclusion that the two reading regimens are equivalent applies only if the single reader with computer-aided detection has experience similar to that of a first reader in a double reading. We used double reading with arbitration in cases of disagreement, which is routine practice in the United Kingdom, because it has been shown to reduce the recall rate.[1-7] In the group assigned to single reading with computer-aided detection, 1.3% of the cases were arbitrated by another reader (or pair of readers in Manchester). The recall rates at all sites were within the limits acceptable to the United Kingdom NHSBSP (<7% for prevalence screenings and <5% for incidence

Table 4. Pathological Features of Tumors with Discordant Recall Decisions.*

Feature	Positive Result from Single Reading with CAD and Negative Result from Double Reading (N = 28)	Negative Result from Single Reading with CAD and Positive Result from Double Reading (N = 29)
	number (percent)	
Histologic type		
Invasive ductal	12 (43)	16 (55)
Invasive lobular	4 (14)	1 (3)
Other invasive	3 (11)	6 (21)
In situ	9 (32)	6 (21)
Pathological size†		
≤10 mm	11 (58)	10 (43)
>10 mm	8 (42)	13 (57)
Lymph-node status†		
Negative	16 (84)	19 (83)
Positive	3 (16)	4 (17)
Grade†		
1	4 (21)	9 (39)
2	11 (58)	12 (52)
3	4 (21)	1 (4)
Unknown	0	1 (4)

* CAD denotes computer-aided detection. Percentages may not sum to 100 because of rounding.
† Data for these features are given for invasive cancers only.

screenings)[47] and other European countries.[8] The recall rates for trial subjects assigned to double reading (and at two centers, for those assigned to single reading with computer-aided detection) were similar to those for nontrial subjects assigned to double reading, a result indicating that the behavior of readers participating in double reading was not substantially altered by participation in the trial.

The main study design was a matched comparison of the results of the application of the two reading regimens to the same subject. Some interval cancers were missed by both reading regimens (i.e., the number of false negatives was the same for both regimens), but the failure to detect these cancers did not alter the absolute difference in sensitivity between the two regimens. The advantage of the matched design is that it enables a valid comparison of sensitivities to be performed before the interval cancers come to light. In the mammograms in which breast cancer was detected after discordance between the recall decisions of the two reading regimens, we observed no significant differences between the pathological attributes of the tumors detected by the two regimens.

The film-reading and arbitration procedures were not standardized among the three centers but reflected current practice at each center and were within national guidelines. Similarly, each center applied a pragmatic approach to the protocol for single reading with computer-aided detection. In two centers, there was no significant difference in recall rates between mammograms given a single reading with computer-aided detection and those given a double reading; however, the Manchester center had a significant increase of 1.4 percentage points in the rate of recalls with single reading with computer-aided detection as compared with double reading. This difference was not associated with the amount of experience or the staff category of the readers or whether the mammograms were obtained at a prevalence or an incidence screening. However, of the three centers, Manchester also had both the highest recall rates for double reading and the highest cancer-detection rates in the trial. For Warwickshire, Solihull, and Coventry; Manchester;

and Nottingham, the overall cancer-detection rates within the trial were 7.1, 10.0, and 7.0 per 1000 screened, respectively (P=0.03), a result suggesting that the increased recall rate with single reading with computer-aided detection in Manchester is due to a more cautious recall policy in general in that center. During the period of the trial, the time between mammography screenings was greater than 3 years in Manchester, an interval that might have permitted more cancers to become available for detection than a shorter period might have done.

Our results suggest that single reading with computer-aided detection is an alternative to double reading; whether to adopt this technology is a question of cost-effectiveness. The additional costs of the computer-aided–detection equipment and the costs associated with an increase in recall must be balanced against the potential savings in reader time. Clearly, the cost-effectiveness of computer-aided detection requires investigation.[6,23,37] Furthermore, comparison between the performance of computer-aided detection in full-field digital mammography and its performance in film mammography (which was used in this study) will be required.[59]

Double reading, which is recognized as the best method for the detection of small invasive cancers,[60] is often difficult to achieve in practice because of costs and the need for two readers. The results of this study are applicable to programs in which double reading is standard practice.[8] Where single reading is standard practice, computer-aided detection has the potential to improve cancer-detection rates to the level achieved by double reading.

APPENDIX

The members of the CADET II group were as follows: **Analysis and Writing Committee:** F.J. Gilbert, S.M. Astley, M.G.C. Gillan, O.F. Agbaje, M.G. Wallis, J. James, C.R.M. Boggis, S.W. Duffy; **Trial Management and Site Coordination Group:** F.J. Gilbert, M.G.C. Gillan, S.W. Duffy, O.F. Agbaje, S.M. Astley, M.G. Wallis, J. James, C.R.M. Boggis, H. Flight, K. Heer, J. Cooper; **Statistical Analysis:** S.W. Duffy, O.F. Agbaje; **Film readers at participating NHSBSP centers (Warwickshire, Solihull, and Coventry; Manchester; and Nottingham):** M.G. Wallis, C.R.M. Boggis, J. James, R.R. Ramachandra, M. Wilson, H.C. Burrell, C. Lewis, U.M. Beetles, E.J. Cornford, S. Garnett, E. Hurley, L.J. Hamilton, A.A. Duncan, A.K. Jain, A.J. Evans, V. Gaur, N.B. Barr, R. Wilson, R. Walker, M. Griffiths, A. Luck, L. Chapman, J. Johnson, J. Patel, J. Hackney, R. Roberts, S. Bundred.

REFERENCES

1. Anttinen I, Pamilo M, Soiva M, Roiha M. Double reading of mammography screening films — one radiologist or two? Clin Radiol 1993;48:414-21.

2. Deans HE, Everington D, Cordiner C, et al. Scottish experience of double reading in the National Breast Screening Programme. Breast 1998;7:75-9.

3. Blanks RG, Wallis MG, Moss SM. A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: results from the UK National Health Service Breast Screening Programme. J Med Screen 1998; 5:195-201.

4. Liston JC, Dall BJ. Can the NHS Breast Screening Programme afford not to double read screening mammograms? Clin Radiol 2003;58:474-7.

5. Ciatto S, Ambrogetti D, Risso G, et al. The role of arbitration of discordant reports at double reading of screening mammograms. J Med Screen 2005;12:125-7.

6. Helvie M. Improving mammographic interpretation: double reading and computer-aided diagnosis. Radiol Clin North Am 2007;45:801-11.

7. Taylor P, Potts HWW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. Eur J Cancer 2008;44:798-807.

8. Yankaskas BC, Klabunde CN, Ancelle-Park R, et al. International comparison of performance measures for screening mammography: can it be done? J Med Screen 2004;11:187-93.

9. Working Party of the Radiologists Quality Assurance Coordinating Group. Computer aided detection in mammography. Sheffield, United Kingdom: NHSBSP, 2001. (Publication no. 48.)

10. Karssemeijer N, Otten JD, Verbeek AL, et al. Computer-aided detection versus independent double reading of masses on mammograms. Radiology 2003;227:192-200.

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

12. Garvican L, Field S. A pilot evaluation of the R2 image checker system and users' response in the detection of interval breast cancers on previous screening films. Clin Radiol 2001;56:833-7.

13. Moberg K, Bjurstam N, Wilczek B, Rostgard L, Egge E, Muren C. Computed assisted detection of interval breast cancers. Eur J Radiol 2001;39:104-10.

14. Brem RF, Schoonjans JM. Radiologist detection of microcalcifications with and without computer-aided detection: a comparative study. Clin Radiol 2001;56:150-4.

15. Birdwell RL, Ikeda DM, O'Shaughnessy KF, Sickles EA. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. Radiology 2001;219:192-202.

16. Huo Z, Giger ML, Vyborny CJ, Metz CE. Breast cancer: effectiveness of computer-aided diagnosis observer study with independent database of mammograms. Radiology 2002;224:560-8.

17. Ciatto S, Del Turco MR, Risso G, et al. Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography. Eur J Radiol 2003;45:135-8.

18. Brem RF, Baum J, Lechner M, et al. Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial. AJR Am J Roentgenol 2003;181:687-93.

19. Ikeda DM, Birdwell RL, O'Shaughnessy KF, Sickles EA, Brenner RJ. Computer-aided detection output on 172 subtle findings on normal mammograms previously obtained in women with breast cancer detected at follow-up screening mammography. Radiology 2004;230:811-9.

20. Astley SM, Gilbert FJ. Computer-aided detection in mammography. Clin Radiol 2004;59:390-9.

21. Castellino RA. Computer aided detection (CAD): an overview. Cancer Imaging 2005;5:17-9.

22. Nishikawa RM. Current status and future directions of computer-aided diagnosis in mammography. Comput Med Imaging Graph 2007;31:224-35.

23. Malich A, Fischer DR, Bottcher J. CAD for mammography: the technique, results, current role and further developments. Eur Radiol 2006;16:1449-60.

24. Computer aided detection (CAD) in mammography. Chicago: BlueCross BlueShield Association Technology Evaluation Center Assessment Program, December 2002.

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

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

27. Gur D, Sumkin JH, Rockette HE, 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.

28. Destounis SV, DiNitto P, Logan-Young W, Bonaccio E, Zuley ML, Willison KM. Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience. Radiology 2004;232:578-84.

29. Birdwell RL, Bandodkar P, Ikeda DM. Computer-aided detection with screening mammography in a university hospital setting. Radiology 2005;236:451-7.

30. Cupples TE, Cunningham JE, Reynolds JC. Impact of computer-aided detection in a regional screening mammography program. AJR Am J Roentgenol 2005;185:944-50.

31. Morton MM, Whaley DH, Brandt KR, Amrami KK. Screening mammograms: interpretation with computer-aided detection-prospective evaluation. Radiology 2006;239:375-83.

32. Dean JC, Ilvento CC. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. AJR Am J Roentgenol 2006;187:20-8.

33. Ko JM, Nicholas MJ, Mendel JB, Slanetz PJ. Prospective assessment of computer-aided detection in interpretation of screening mammography. AJR Am J Roentgenol 2006;187:1483-91.

34. Georgian-Smith D, Moore RH, Halpern E, et al. Blinded comparison of computer-aided detection with human second reading in screening mammography. AJR Am J Roentgenol 2007;189:1135-41.

35. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med 2007;356:1399-409.

36. Gromet M. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. AJR Am J Roentgenol 2008;190:854-9.

37. Lindfors KK, McGahan MC, Rosenquist CJ, Hurlock GS. Computer-aided detection of breast cancer: a cost-effectiveness study. Radiology 2006;239:710-7.

38. Ciatto S, Houssami N. Computer-aided screening mammography. N Engl J Med 2007;357:83.

39. Houssami N, Given-Wilson R. Incorporating new technologies into clinical practice without evidence of effectiveness in prospective studies: computer-aided detection (CAD) in breast screening rein-forces the need for better initial evaluation. Breast 2007;16:219-21.

40. Brem RF. Blinded comparison of computer-aided detection with human second reading in screening mammography: the importance of the question and the critical numbers game. AJR Am J Roentgenol 2007;189:1142-4.

41. Jiang Y, Miglioretti DL, Metz CE, Schmidt RA. Breast cancer detection rate: designing imaging trials to demonstrate improvements. Radiology 2007;243:360-7.

42. Gilbert FJ, Astley SM, McGee MA, et al. Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom National Breast Screening Program. Radiology 2006;241:47-53.

43. Duijm LE, Groenewoud JH, Hendriks JH, de Koning HJ. Independent double reading of screening mammograms in The Netherlands: effect of arbitration following reader disagreements. Radiology 2004;231:564-70.

44. Cornford EJ, Evans AJ, James JJ, Burrell HC, Pinder SE, Wilson AR. The pathological and radiological features of screen-detected breast cancers diagnosed following arbitration of discordant double reading opinions. Clin Radiol 2005;60:1182-7.

45. Mucci B, Athey G, Scarisbrick G. Double reading of screening mammograms: the use of a third reader to arbitrate on disagreements. Breast 1999;8:37-9.

46. Hologic mammography. Santa Clara, CA: R2 Technology, 2007. (Accessed September 22, 2008, at http://www.hologic.com/imagechecker.)

47. Quality assurance guidelines for breast cancer screening radiology. Sheffield, United Kingdom: National Health Service Breast Screening Programme, 2005. (Publication no. 59.)

48. Astley S, Quarterman C, Al Nuaimi Y, et al. Computer-aided detection in screening mammography: the impact of training on reader performance. In: Proceedings of IWDM [International Workshop of Digital Mammography] 2004, Durham, NC, June 18–21, 2004:231-6.

49. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996;313:36-9. [Erratum, BMJ 1996;313:550.]

50. Senn SJ. Statistical issues in drug development. Chichester, England: John Wiley, 1997.

51. Lara LF, Cisneros G, Gurney M, et al. One-day quadruple therapy compared with 7-day triple therapy for Helicobacter pylori infection. Arch Intern Med 2003;163:2079-84.

52. Gallant JE, Staszewski S, Pozniak AL, et al. Efficacy and safety of tenofovir DF vs stavudine in combination therapy in antiretroviral-naive patients: a 3-year randomized trial. JAMA 2004;292:191-201.

53. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947;12:153-7.

54. National Health Service Breast Screening Programme. Annual review. Sheffield, United Kingdom: NHSBSP, 2004.

55. Wald NJ, Murphy P, Major P, Parkes C, Townsend J, Frost C. UKCCCR multicentre randomised controlled trial of one and two view mammography in breast cancer screening. BMJ 1995;311:1189-93.

56. Bennett RL, Blanks RG, Moss SM. Does the accuracy of single reading with CAD (computer-aided detection) compare with that of double reading? A review of the literature. Clin Radiol 2006;61:1023-8.

57. Smith-Bindman R, Chu PW, Miglioretti DL, et al. Comparison of screening mammography in the United States and the United Kingdom. JAMA 2003;290:2129-37.

58. Khoo LAL, Taylor P, Given-Wilson RM. Computer-aided detection in the United Kingdom National Breast Screening Programme: prospective study. Radiology 2005;237:444-9.

59. Del Turco MR, Ciatto S, Lazzari B, Zatelli G. New imaging techniques: the role of digital mammography. Eur J Cancer 2002;38:Suppl:S31-S32.

60. Blanks RG, Wallis MG, Given-Wilson RM. Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography. J Med Screen 1999;6:152-8.

Copyright © 2008 Massachusetts Medical Society.