

Keywords: breast cancer; diagnostic workup; double reading; positive predictive value; referral rate; screening mammography

Comparison of the diagnostic workup of women referred at non-blinded or blinded double reading in a population-based screening mammography programme in the south of the Netherlands

Roy J P Weber^{*1}, Elisabeth G Klompenhouwer¹, Adri C Voogd^{2,3}, Luc J A Strobbe⁴, Mireille J M Broeders^{5,6} and Lucien E M Duijm⁷

¹Department of Radiology, Catharina Hospital, PO Box 1350, 5602 ZA Eindhoven, The Netherlands; ²Department of Research, Netherlands Comprehensive Cancer Organization (IKNL), PO Box 231, 5600 AE Eindhoven, The Netherlands; ³Department of Epidemiology, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands; ⁴Department of Surgery, Canisius Wilhelmina Hospital, PO Box 9015, 6500 GS Nijmegen, The Netherlands; ⁵National Expert and Training Centre for Breast Cancer Screening, PO Box 6873, 6503 GJ Nijmegen, The Netherlands; ⁶Department for Health Evidence, Radboud University Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands and ⁷Department of Radiology, Canisius Wilhelmina Hospital, PO Box 9015, 6500 GS Nijmegen, The Netherlands

Background: To determine whether referred women experience differences in diagnostic workup at non-blinded or blinded double reading of screening mammograms.

Methods: We included a consecutive series of respectively 42,996 and 44,491 screens, double read either in a non-blinded or blinded manner between 2009 and 2011. This reading strategy was alternated on a monthly basis.

Results: The overall ultrasound-guided core needle biopsy (CNB) rate and stereotactic CNB (SCNB) rate per 1000 screens were higher at blinded than at non-blinded reading (7.5 vs 6.0, $P=0.008$ and 8.1 vs 6.6, $P=0.009$). Among women with benign workup, these rates were higher at blinded reading (2.6 vs 1.4, $P<0.001$ and 5.9 vs 4.7, $P=0.013$). The benign biopsy rates were higher at blinded double reading ($P<0.001$), whereas the positive predictive value of biopsy did not differ ($P=0.103$).

Conclusions: Blinded double-reading results in higher overall CNB and SCNB rates than non-blinded double reading, as well as a higher benign biopsy rate.

Reduction in breast cancer mortality is attributed to the combination of earlier breast cancer detection through mammography screening and substantial improvements in breast cancer treatment (Kalager *et al*, 2010; van Schoor *et al*, 2011; Mandelblatt *et al*, 2013).

Screening mammograms can be assessed in several ways, for example, by single reading or double reading by certified

screening radiologists. European guidelines consider double reading by radiologists as the gold standard for the assessment of screening mammograms, as this screening strategy significantly increases the cancer detection rate when compared with single reading (Gur *et al*, 2004; Duijm *et al*, 2009). In most programmes, including the Dutch nationwide breast cancer screening programme, independent double reading rather than blinded double

*Correspondence: RJP Weber; E-mail: royweber@gmail.com

Received 14 February 2015; revised 16 June 2015; accepted 13 July 2015; published online 18 August 2015

© 2015 Cancer Research UK. All rights reserved 0007–0920/15

reading is used for the assessment of screening mammograms. We recently compared these two reading strategies and found that blinded double reading increased the sensitivity of breast cancer detection, at the expense of an increased referral rate and decreased positive predictive value (PPV) of referral (Klompshouwer *et al*, 2015). However, the influence of replacement of non-blinded by blinded double reading on the use of diagnostic resources at the workup of screen-positive women is unknown. We therefore compared the utilisation of breast imaging and biopsy techniques in women referred at blinded or non-blinded double reading in a population-based screening mammography programme in the south of the Netherlands.

MATERIALS AND METHODS

Study population. We used the information of respectively 42 996 and 44 491 full-field digital screens that had been double read, either in a non-blinded or blinded manner between July 2009 and July 2011 (Klompshouwer *et al*, 2015).

Screening procedure and referral. Details of the nationwide and local breast cancer screening programme have been described previously (Fracheboud *et al*, 2001; Nederend *et al*, 2012; Klompshouwer *et al*, 2015). In brief, free biennial screening mammography is offered to women aged 50–75 years. Before each screening examination, the woman completes a short questionnaire with questions about any previous breast surgery or breast malignancy and the use of hormone replacement therapy. Screening mammograms were double read in either a blinded (second reader was not informed about the first reader's decision) or non-blinded (second reader was informed about the first reader's decision). This reading strategy was alternated on a monthly basis. Mammograms of previous screening rounds were always available for comparison.

Women with a screening Breast Imaging Reporting and Data System (BI-RADS) category 0, 4 or 5 were referred to a dedicated breast unit (D'Orsi *et al*, 2003).

Diagnostic work-up and follow-up of referred women. A total of 15 regional and university hospitals were involved in the assessment of screen-positive women. The type of further diagnostic evaluation depended on the workup protocols of the various hospitals and could include breast tomosynthesis, (3D) breast ultrasonography, magnetic resonance imaging, ultrasound-guided fine-needle aspiration biopsy (FNAB) or core needle biopsy (CNB), stereotactic CNB (SCNB) or surgical excision biopsy. BI-RADS 0 lesions received additional imaging and biopsy if necessary. BI-RADS 4 and 5 lesions were routinely biopsied, whereas BI-RADS 3 lesions were either biopsied or followed up. Data on mammographic breast density and family history of breast cancer were extracted from the radiological and surgical records, respectively. The follow-up period of all referred women was 2 years (until the next biennial screen).

Statistical analysis. Descriptive statistics were performed using Statistical Package for Social Science 17.0 (SPSS Inc. Chicago, IL, USA). A Fisher's exact test or χ^2 -test was used to test differences in biopsy rates and workup procedures at non-blinded and blinded double reading.

RESULTS

Overall workup procedures after referral. A total of 1235/42 996 (2.9%) and 1474/44 491 (3.3%) women had been referred, respectively, at non-blinded and blinded double reading. Baseline characteristics for screened and referred women were comparable for the two reading strategies (Table 1). At both reading strategies,

workup consisted of additional breast-imaging procedures only in about half of referrals, whereas the other half underwent a combination of additional breast imaging plus biopsy ($P=0.164$; Table 2). The vast majority of biopsy procedures in both groups consisted of either CNB or SCNB, whereas <3% of referred women underwent FNAB and surgical biopsy was hardly used for diagnostic purposes. The distribution of various combinations of assessment at workup was also comparable, but the ultrasound-guided CNB rate and SCNB rate per 1000 screens were lower at non-blinded than at blinded double reading (6.0 vs 7.5, $P=0.008$ and 6.6 vs 8.1, $P=0.009$, respectively).

Work-up in false-positive referrals. A total of 951 (77.0% of referrals) and 1149 (80.0% of referrals) women experienced a false-positive referral at non-blinded and blinded double reading, respectively (Table 3). The distribution of type of assessment differed between both groups, with a larger proportion of women evaluated by additional imaging only at non-blinded double reading (65.0% vs 60.7%, $P=0.046$) and a larger proportion of CNB procedures at blinded double reading (9.9% vs 6.5%, $P=0.005$). Almost all diagnostic surgical biopsies following referral had been performed in false-positive cases, but the false-positive surgical biopsy rate was still very low at both reading strategies (0.37 and 0.34 per 1000 screens at non-blinded and at blinded double reading, respectively). The CNB rate ($P<0.001$) and SCNB rate ($P=0.013$) per 1000 screens were higher at blinded than at non-blinded reading.

Overall biopsy rates and PPV of biopsy. The two reading strategies showed comparable PPV of biopsy (46.3% vs 41.9% at non-blinded and blinded double reading, respectively, $P=0.103$; Table 4). Both the overall biopsy rate and benign biopsy rate per 1000 screens were higher at blinded double reading (17.4 vs 14.3, $P<0.001$ and 10.1 vs 7.7, $P<0.001$, respectively), as well as the proportion of women who underwent biopsy as part of a false-positive referral (39.3% vs 34.7%, $P=0.032$).

DISCUSSION

We found that blinded double reading resulted in significantly higher CNB and SCNB rates than non-blinded double reading, as well as a significantly higher benign biopsy rate. A larger proportion of women at benign work-up were evaluated merely by additional imaging at non-blinded double reading, whereas a larger fraction of referrals underwent CNB procedures at blinded double reading. The PPV of biopsy was comparable for both reading strategies.

The current study included screens during the first round of digital screening in the Netherlands. We found a biopsy rate of 14.3 per 1000 screens after non-blinded double reading and 17.4 per 1000 screens after blinded double reading. To our knowledge, data of biopsy rates after blinded double reading have not been reported previously. The biopsy rates we observed after non-blinded double reading are comparable to those of a Norwegian (61/4009) and Irish (470/35204) study during the first round of digital mammography (Skaane *et al*, 2007; Hambly *et al*, 2009).

In a Dutch study, increased detection of lesions presenting as microcalcifications during the transition from screen-film to digital screening mammography resulted in an almost fourfold increase in SCNB per 1000 screened women (Nederend *et al*, 2012). The further intensified use of SCNB at blinded double reading may have a great impact on daily practice of regional breast clinics, as it is much more time consuming than conventional breast imaging or ultrasound-guided biopsy.

Among women with benign work-up, a larger fraction was evaluated simply by additional imaging at non-blinded double reading, whereas a larger fraction at blinded double reading also underwent CNB procedures. The higher referral rate at blinded

Table 1. Baseline characteristics of women screened and referred at non-blinded and blinded double reading

	Screened population			Referred women		
	Non-blinded (n = 42 996)	Blinded (n = 44 491)	P-value	Non-blinded (n = 1 235)	Blinded (n = 1 474)	P-value
Age distribution, no (%)			0.49			0.91
50–59	22 704 (52.8)	23 469 (52.7)		653 (52.9)	778 (52.8)	
60–69	14 537 (33.8)	14 949 (33.6)		417 (33.8)	491 (33.3)	
70–75	5755 (13.4)	6073 (13.6)		165 (13.3)	205 (13.9)	
Previous breast surgery, no (%)			0.58			0.20
Yes	3354 (7.8)	3516 (7.9)		129 (10.4)	177 (12.0)	
No	39 642 (92.2)	40 975 (92.1)		1106 (89.6)	1297 (88.0)	
Hormone replacement therapy, no. (%)			0.74			0.82
Yes	2064 (4.8)	2157 (4.8)		55 (4.5)	63 (4.3)	
No	40 932 (95.2)	42 334 (95.2)		1180 (95.5)	1411 (95.7)	
Family history of breast cancer ^a , no (%)	Not available	Not available				0.59
Yes				149 (12.1)	168 (11.4)	
No				1086 (87.9)	1306 (88.6)	
Screening round, no (%)			0.85			0.15
Initial screen	4784 (11.1)	4968 (11.2)		309 (25.0)	334 (22.7)	
Subsequent screen	38 212 (88.9)	39 523 (88.8)		926 (75.0)	1140 (77.3)	
Mammographic breast density	Not available	Not available				0.83
0–50%				1014 (82.1)	1215 (82.4)	
> 50%				221 (17.9)	259 (17.6)	
Lesion abnormality at latest screen, no (%)	Not applicable	Not applicable				0.34
Suspicious mass				718 (58.1)	827 (56.1)	
Suspicious microcalcifications				334 (27.0)	445 (30.2)	
Suspicious mass with microcalcifications				85 (6.9)	95 (6.4)	
Asymmetry				22 (1.8)	28 (1.9)	
Architectural distortion				72 (5.8)	78 (5.3)	
Other mammographic abnormality				4 (0.3)	1 (0.1)	

^aAt least one first-degree relative with a diagnosis of breast cancer before the age of 50 years or at least two second-degree relatives with breast cancer.

Table 2. Non-blinded vs blinded double reading at screening mammography: assessment following referral

	Proportion					Rate (per 1000 screened women)		
	Non-blinded (n = 1 235)		Blinded (n = 1 474)		P-value	Non-blinded (n = 42 996)	Blinded (n = 44 491)	P-value
	n	(%)	n	(%)				
Assesment following referral								
No further assessment	3	(0.2)	0	(0)	—	0.1	0	—
Additional breast imaging	618	(50.0)	698	(47.4)	0.164	14.4	15.7	0.110
Additional breast imaging + FNAB	29	(2.3)	36	(2.4)	0.873	0.7	0.8	0.465
Additional breast imaging + CNB	259	(21.0)	333	(22.6)	0.310	6.0	7.5	0.008
Additional breast imaging + SCNB	285	(23.1)	362	(24.6)	0.368	6.6	8.1	0.009
Additional breast imaging + FNAC + CNB/SCNB ^a	16	(1.3)	18	(1.2)	0.862	0.4	0.4	0.808
Additional breast imaging + CNB/FNAB/SCNB ^b + SB	16	(1.3)	14	(0.9)	0.392	0.4	0.3	0.646
Additional breast imaging + CNB + SCNB	8	(0.6)	12	(0.8)	0.614	0.2	0.3	0.413
Additional breast imaging + SB	1	(0.1)	1	(0.1)	—	0	0	—

Abbreviations: CNB = core needle biopsy; FNAB = fine needle aspiration biopsy; SB = surgical (open) biopsy; SCNB = stereotactic CNB.

^aCNB and/or SCNB.

^bCNB and/or FNAB and/or SCNB.

double reading results in higher costs of the diagnostic procedures in the hospital, which will rise even further as a larger fraction of referrals is evaluated by CNB in addition to breast imaging at blinded double reading.

The higher biopsy rates at blinded double readings was accompanied by an increased cancer detection rate, which resulted in a comparable PPV of biopsy of > 40% at both reading strategies.

This PPV is within the range of those found in other European programmes and higher than those observed in the United States (Skaane *et al*, 2007; Vinnicombe *et al*, 2009; Glynn *et al*, 2011).

It remains a question of debate whether false-positive referrals, which received additional imaging or biopsy, have uniformly decreased odds of attending subsequent routine screening rounds (Fitzpatrick *et al*, 2011; Seigneurin *et al*, 2011; Maxwell *et al*, 2013).

Table 3. Non-blinded vs blinded double reading at screening mammography: assessment at false positive referral

	Proportion					Rate (per 1000 screened women)		
	Non-blinded (n = 951)		Blinded (n = 1149)		P-value	Non-blinded (n = 42 996)	Blinded (n = 44 491)	P-value
	n	(%)	n	(%)				
Assesment following referral								
No further assessment	3	(0.3)	0	(0)	—	0.1	0	—
Additional breast imaging	618	(65.0)	698	(60.7)	0.046	14.4	15.7	0.110
Additional breast imaging + FNAB	28	(2.9)	35	(3.0)	0.892	0.7	0.8	0.455
Additional breast imaging + CNB	62	(6.5)	114	(9.9)	0.005	1.4	2.6	<0.001
Additional breast imaging + SCNB	200	(21.0)	261	(22.7)	0.353	4.7	5.9	0.013
Additional breast imaging + FNAC + CNB/SCNB ^a	16	(1.7)	15	(1.3)	0.476	0.4	0.3	0.783
Additional breast imaging + CNB/FNAB/SCNB ^b + SB	16	(1.7)	14	(1.2)	0.372	0.4	0.3	0.646
Additional breast imaging + CNB + SCNB	8	(0.8)	11	(1.0)	0.780	0.2	0.2	0.539
Additional breast imaging + SB	0	(0)	1	(0.1)	—	0	0	—

Abbreviations: CNB = core needle biopsy; FNAB = fine needle aspiration biopsy; SB = surgical (open) biopsy; SCNB = stereotactic CNB.
^aCNB and/or SCNB.
^bCNB and/or FNAB and/or SCNB.

Table 4. Biopsy rates and PPVs of biopsy at non-blinded and blinded double reading of screening mammograms

	Non-blinded double reading	Blinded double reading	P-value
Assessment at referral			0.129
Biopsy, %	49.7 (614/1235)	52.6 (776/1474)	
No biopsy, %	50.3 (621/1235)	47.4 (698/1474)	
Biopsy rate per 1000 screens	14.3 (614/42 996)	17.4 (776/44 491)	<0.001
PPV of biopsy, %	46.3 (284/614)	41.9 (325/776)	0.103
False-positive referral			0.032
Biopsy, %	34.7 (330/951)	39.3 (451/1149)	
No biopsy, %	65.3 (621/951)	60.7 (698/1149)	
Benign biopsy rate per 1000 screens	7.7 (330/42 996)	10.1 (451/44 491)	<0.001

Abbreviation: PPV = positive predictive value.

Many of the women with false-positive screening results experience distress (van der Steeg *et al*, 2011), which may last for up to 3 years after a false-positive screen and is related to the invasiveness of the assessment (Bond *et al*, 2013). It is therefore important to keep the number of benign biopsies as low as possible.

Our study has certain limitations. The study population consisted of women who attended the screening programme during the transition from screen-film to digital screening mammography. Therefore, the majority of subsequent screens were compared with previously obtained screen-film mammograms. Data on screening outcome after the transition to digital screening mammography are sparse and there are no studies available that compare the use of diagnostic work-up modalities during and after the transition. All screening examinations in our study comprised two-view mammography of both breasts, including a medio-lateral oblique view and a cranio-caudal view. Recent studies have shown that the addition of breast tomosynthesis to digital mammography improves the cancer detection rate in mammography screening and may reduce false-positive referrals (Friedewald *et al*, 2014). Breast tomosynthesis, however, has not yet been implemented in the Dutch mammography screening programme.

In summary, we conclude that blinded double reading resulted in significantly higher overall and benign biopsy rate than non-blinded double reading. Although an increased screening sensitivity at blinded double reading has been reported, the

negative side effect of higher biopsy rates should be taken into account when adopting this reading strategy.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Bond M, Pavey T, Welch K, Cooper C, Garside R, Dean S, Hyde CJ (2013) Psychological consequences of false-positive screening mammograms in the UK. *Evid Based Med* **18**(2): 54–61.
 D’Orsi CJ, Mendelson EB, Ikeda DM *et al*. (2003) *Breast Imaging Reporting and Data System: ACR BI-RADS - Breast Imaging Atlas*. American College of Radiology: Reston, VA.
 Duijm LE, Louwman MW, Groenewoud JH, van de Poll-Franse LV, Fracheboud J, Coebergh JW (2009) Inter-observer variability in mammography screening and effect of type and number of readers on screening outcome. *Br J Cancer* **100**(6): 901–907.
 Fitzpatrick P, Fleming P, O’Neill S, Kiernan D, Mooney T (2011) False-positive mammographic screening: factors influencing re-attendance over a decade of screening. *J Med Screen* **18**(1): 30–33.
 Fracheboud J, de Koning HJ, Boer R, Groenewoud JH, Verbeek AL, Broeders MJ, van Ineveld BM, Hendriks JH, de Bruyn AE, Holland R,

- van der Maas PJ (2001) Nationwide breast cancer screening programme fully implemented in The Netherlands. *Breast* **10**(1): 6–11.
- Friedewald SM, Rafferty EA, Conant EF (2014) Breast cancer screening with tomosynthesis and digital mammography-reply. *JAMA* **312**(16): 1695–1696.
- Glynn CG, Farria DM, Monsees BS, Salcman JT, Wiele KN, Hildebolt CF (2011) Effect of transition to digital mammography on clinical outcomes. *Radiology* **260**(3): 664–670.
- Gur D, Sumkin JH, Hardesty LA, Clearfield RJ, Cohen CS, Ganott MA, Hakim CM, Harris KM, Poller WR, Shah R, Wallace LP, Rockette HE (2004) Recall and detection rates in screening mammography. *Cancer* **100**(8): 1590–1594.
- Hambly NM, McNicholas MM, Phelan N, Hargaden GC, O'Doherty A, Flanagan FL (2009) Comparison of digital mammography and screen-film mammography in breast cancer screening: a review in the Irish breast screening program. *AJR Am J Roentgenol* **193**(4): 1010–1018.
- Kalager M, Zelen M, Langmark F, Adami HO (2010) Effect of screening mammography on breast-cancer mortality in Norway. *N Engl J Med* **363**(13): 1203–1210.
- Klompshouwer EG, Voogd AC, den Heeten GJ, Strobbe LJ, de Haan AF, Wauters CA, Broeders MJ, Duijm LE (2015) Blinded double reading yields a higher programme sensitivity than non-blinded double reading at digital screening mammography: a prospected population based study in the south of The Netherlands. *Eur J Cancer* **51**(3): 391–399.
- Mandelblatt J, van Ravesteyn N, Schechter C, Chang Y, Huang AT, Near AM, de Koning H, Jemal A (2013) Which strategies reduce breast cancer mortality most? Collaborative modeling of optimal screening, treatment, and obesity prevention. *Cancer* **119**(14): 2541–2548.
- Maxwell AJ, Beattie C, Lavelle J, Lyburn I, Sinnatamby R, Garnett S, Herbert A (2013) The effect of false positive breast screening examinations on subsequent attendance: retrospective cohort study. *J Med Screen* **20**(2): 91–98.
- Nederend J, Duijm LE, Louwman MW, Groenewoud JH, Donkers-van Rossum AB, Voogd AC (2012) Impact of transition from analog screening mammography to digital screening mammography on screening outcome in The Netherlands: a population-based study. *Ann Oncol* **23**(12): 3098–3103.
- Seigneurin A, Exbrayat C, Labarere J, Delafosse P, Poncet F, Colonna M (2011) Association of diagnostic work-up with subsequent attendance in a breast cancer screening program for false-positive cases. *Breast Cancer Res Treat* **127**(1): 221–228.
- Skaane P, Hofvind S, Skjennald A (2007) Randomized trial of screen-film versus full-field digital mammography with soft-copy reading in population-based screening program: follow-up and final results of Oslo II study. *Radiology* **244**(3): 708–717.
- van der Steeg AF, Keyzer-Dekker CM, De Vries J, Roukema JA (2011) Effect of abnormal screening mammogram on quality of life. *Br J Surg* **98**(4): 537–542.
- van Schoor G, Moss SM, Otten JD, Donders R, Paap E, den Heeten GJ, Holland R, Broeders MJ, Verbeek AL (2011) Increasingly strong reduction in breast cancer mortality due to screening. *Br J Cancer* **104**(6): 910–914.
- Vinnicombe S, Pinto Pereira SM, McCormack VA, Shiel S, Perry N, Silva IMD (2009) Full-Field Digital versus Screen-Film Mammography: comparison within the UK Breast Screening Program and Systematic Review of Published Data. *Radiology* **251**(2): 347–358.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.