European Radiology (2019) 29:337–344
https://doi.org/10.1007/s00330-018-5586-9

BREAST

Characteristics of screen-detected cancers following concordant or discordant recalls at blinded double reading in biennial digital screening mammography

Angela M. P. Coolen[1] · Joost R. C. Lameijer[2] · Adri C. Voogd[3,4,5] · Marieke W. J. Louwman[4] · Luc J. Strobbe[6] · Vivianne C. G. Tjan-Heijnen[5] · Lucien E. M. Duijm[7,8]

Received: 26 April 2018 / Accepted: 4 June 2018 / Published online: 25 June 2018

Abstract

Objectives To analyse which mammographic and tumour characteristics led to concordant versus discordant recalls at blinded double reading to further optimise our breast cancer screening programme.

Methods We included a consecutive series of 99,013 screening mammograms obtained between July 2013 and January 2015. All mammograms were double read in a blinded fashion. Discordant readings were routinely recalled without consensus or arbitration. During the 2-year follow-up, relevant data of the recalled women were collected. We compared mammographic characteristics, screening outcome and tumour characteristics between concordant and discordant recalls.

Results There were 2,543 concordant recalls (71.4%) and 997 discordant recalls (28.0%). The positive predictive value of a concordant recall was significantly higher (23.5% vs. 10.0%, $p < 0.001$). The proportion of BI-RADS 0 was significantly higher in the discordant recall group (75.7% vs. 56.3%, $p < 0.001$). Discordant recalls were more often an asymmetry or architectural distortion (21.8% vs. 13.2% and 9.3% vs. 6.5%, respectively, $p < 0.001$). There were no differences in the distribution of DCIS and invasive cancers and tumour characteristics were comparable for the two groups, except for a more favourable tumour grade in the discordant recall group (54.7% vs. 39.9% grade I tumours, $p = 0.022$).

Conclusions Screen-detected cancers detected by a discordant reading show a more favourable tumour grade than cancers diagnosed after a concordant recall. The higher proportion of asymmetries and architectural distortions in this group provide a possible target for improving screening programmes by additional training of screening radiologists and the implementation of digital breast tomosynthesis.

Key Points

- *With blinded double reading of screening mammograms, screen-detected cancers detected by a discordant reading show a more favourable tumour grade than cancers diagnosed after a concordant recall.*
- *The proportions of asymmetries and architectural distortions are higher in case of a discordant reading.*
- *Possible improvement strategies could target additional training of screening radiologists and the implementation of digital breast tomosynthesis in breast cancer screening programmes.*

✉ Angela M. P. Coolen
a.coolen88@hotmail.com

[1] Department of Radiology, Elisabeth-Tweesteden Hospital (ETZ), 90151, 5000 LC Tilburg, The Netherlands

[2] Department of Radiology, Catharina Hospital, Michelangelolaan 2, 5623 EJ Eindhoven, The Netherlands

[3] Department of Epidemiology, Maastricht University, GROW, P Debyelaan 1, 6229 HA Maastricht, The Netherlands

[4] Department of Research, Netherlands Comprehensive Cancer Organization (IKNL), 19079, 3501 DB Utrecht, The Netherlands

[5] Department of Internal Medicine, Division of Medical Oncology, GROW, Maastricht University Medical Centre, P Debyelaan 1, 6229 HA Maastricht, The Netherlands

[6] Department of Surgery, Canisius-Wilhelmina Hospital, PO Box 9015, 6500 GS Nijmegen, The Netherlands

[7] Department of Radiology, Canisius Wilhelmina Hospital, Weg door Jonkerbos 100, 6532 SZ Nijmegen, The Netherlands

[8] Dutch Expert Centre for Screening, Wijchenseweg 101, 6538 SW Nijmegen, The Netherlands

Keywords Mass screening · Early detection of cancer · Breast neoplasms · Mammography · Follow-up studies

Abbreviations

CAD	Computer-aided detection
CDR	Cancer detection rate
DCIS	Ductal carcinoma in-situ
FFDM	Full-field digital mammography
FNAC	Fine needle aspiration cytology
FPR	False positive rate
LCIS	Lobular carcinoma in-situ
PACS	Picture-archiving and communication system
SFM	Screen-film mammography

Introduction

Reading strategies used in breast cancer screening programmes can greatly influence the main performance indicators and have therefore been a subject of research for many years [1–6]. Breast cancer screening programmes aim to reduce patient morbidity and mortality through detection of early stage breast cancer. Obtaining an optimal balance between the main performance indicators [recall rate, cancer detection rate (CDR) and false-positive recall rate (FPR)] is important to minimise anxiety among the screened population and avoid unnecessary costs. Reading strategies to assess screening mammograms are a potential target to improve this balance.

Possible reading strategies include single reading, with or without computer-aided detection (CAD), and double reading, which can be performed in either a non-blinded (independent) or blinded fashion. At blinded double reading the second reader is not informed about the first reader's opinion. European guidelines consider radiologist double reading as the standard of reference for the assessment of screening mammograms [7]. With the implementation of full-field digital mammography (FFDM), blinded double reading became technically possible in the Dutch nationwide screening mammography programme. Previous studies have shown that non-blinded double reading significantly increases the cancer detection rate compared with single reading at screen-film mammography (SFM) [6, 8] and that blinded double reading further increases programme sensitivity compared with non-blinded double reading [4].

The purpose of the current study was to analyse which mammographic and tumour characteristics led to concordant versus discordant recalls to find strategies to further optimise the screening programme.

Materials and methods

Study population

In this prospective study we included 99,013 consecutive screening examinations (9,860 initial screens and 89,143 subsequent screens, respectively). These mammograms were performed between July 1, 2013, and January 1, 2015, in a southern screening region of The Netherlands. In this screening region, the transition from screen-film mammography (SFM) to full-field digital mammography (FFDM) was completed in 2010. Full-field digital mammograms (FFDM) were obtained at four specialised screening units. All women entering the Dutch nationwide screening programme are routinely asked to give permission to use their data for evaluation of the screening programme and for scientific purposes. One woman refused this permission and was therefore excluded from this study. This study was performed under the national permit for breast cancer screening, which is issued by the Ministry of Health, Welfare and Sports with permission of the Dutch Health Council and did not require additional ethical approval.

Screening procedure and recall

Details of the Dutch nationwide biennial breast cancer screening programme, which targets asymptomatic women aged 50-75 years, have previously been described [9]. In brief, all mammographic examinations were performed by certified technologists using a Lorad Selenia FFDM system (Hologic Inc., Danbury, CT) with a 70-μm pixel size and 232 × 286-mm field of view. After each mammographic examination, the radiographer annotated whether or not she would recall the woman. This is routine practice in our nationwide screening programme. All mammograms were then double read in a blinded fashion by a team of 13 certified screening radiologists. The radiologists were not blinded to the radiographer's opinion. In case of a subsequent screening, previously obtained mammograms were always available for comparison. Mammographic abnormalities were classified as a suspicious mass, suspicious microcalcifications, suspicious mass with microcalcifications, architectural distortion, asymmetry or another abnormality. Mammograms were classified according to the Breast Imaging Reporting and Data System (BI-RADS) [10]; the BI-RADS 3 classification is not used in the Dutch screening programme. BI-RADS classification at recall was defined as the highest BI-RADS of two readers. A discordant reading was defined as a difference in classification by two radiologists, where one classified the mammogram as negative (BI-RADS 1 or 2, i.e. no recall) and the other classified it as positive (BI-RADS 0, 4 or 5, i.e. recall). All other cases were classified as concordant readings. In addition to all concordant positive screening examinations, all discordant readings were recalled without a consensus meeting between the two radiologists or arbitration by a third reader. For the purpose of quality assurance, every 6 weeks, a supervising breast radiologist discussed all recall decisions with the radiographers. All cases that only the radiographers would have recalled were also reviewed. At this stage, a woman was recalled if the supervising radiologist considered work-up necessary.

Diagnostic work-up and follow-up after recall

In case of a positive screening result, the woman was referred to a hospital breast unit by her general physician. After physical examination by a surgical oncologist or dedicated breast nurse, additional mammographic and/or tomosynthesis views were obtained at the clinical radiologist's discretion and classified according to BI-RADS. Previous screening mammograms were routinely available for comparison via the hospitals Picture-Archiving and Communication System (PACS). Dependent on the outcome of the physical examination and clinical mammography, further work-up could consist of one or a combination of the following modalities: breast ultrasonography (US), magnetic resonance imaging (MRI) and/or biopsy [fine-needle aspiration (FNAC), core biopsy, stereotactic biopsy, open surgical biopsy]. During the 2-year follow-up (until the next biennial screening examination), screening mammography findings, clinical data as well as imaging, pathology and surgery reports of all the recalled women were collected. Screen-detected cancers were divided into ductal carcinoma in situ (DCIS) and invasive cancers. Lobular carcinoma in situ (LCIS) was considered a benign lesion.

Statistical analysis

Recalls were classified as either concordant or discordant. Chi-square and Fisher's exact tests were used to compare these two groups regarding: positive predictive value (PPV) of recall, type of screening examination (initial vs. subsequent), type of mammographic abnormality, BI-RADS classification at recall and diagnosis after recall (true positive vs. false positive). Differences in the proportion of invasive and in-situ cancers, tumour grade (using the Nottingham grading system) [9] and other tumour characteristics of screen-detected cancers were also compared for concordant and discordant recalls using chi-square and Fisher's exact tests. In case of bilateral disease, the tumour with the most advanced tumour stage was included in the analysis. In case of multiple foci of cancer, only the largest tumour was taken into account. The two-sided significance level was set at 5%. Statistical analysis was performed using IBM SPSS Statistics 23.0 (IBM SPSS Statistics for Windows, version 23.0, IBM Corp., Armonk, NY).

Results

Overall screening outcome

Out of 99,013 screened women, 3,562 were recalled for further evaluation of a mammographic abnormality (recall rate 3.6%) resulting in 704 screen-detected cancers (CDR of 7.1 per 1,000 screens and FPR of 28.9 per 1,000 screens) (Fig. 1). The majority of recalls were based on a concordant reading. Most screens were classified as BI-RADS 0 (61.8%, 2,186/3,540). The PPV of BI-RADS 0 recalls was 5.9%, of BI-RADS 4 recalls 34.1% and of BI-RADS 5 recalls 95.5% ($p < 0.001$). Twenty-two women (0.6%, 22/3,562) were recalled after discussion with the supervising breast radiologist as part of quality assurance (Fig. 1). Since the current study focuses on radiologist blinded double reading, these 22 recalls are hereafter excluded.

Mammographic characteristics of concordant versus discordant recalls

Compared with a concordant reading, a discordant reading was significantly more often an asymmetry or architectural distortion (Table 1). BI-RADS classification at recall also differed significantly between the two groups. In the discordant reading group most screens were classified as BI-RADS 0; conversely the percentage of BI-RADS 4 and 5 was higher in the concordant reading group (Table 1). The positive predictive value of recall was significantly higher in case of a concordant reading compared with a discordant reading (23.5% vs. 10.0%, $p < 0.001$, Table 1).

Tumour characteristics of screen-detected cancers detected after concordant versus discordant recalls

Proportions of DCIS and invasive cancers did not differ significantly and the DCIS grade was comparable for both groups. In particular there was no significant increase in low-grade DCIS with discordant readings compared with concordant readings (20.0% vs. 16.3%, p = 0.800, Table 2). Tumour characteristics of invasive cancers were comparable for tumours detected by concordant and discordant readings, except for a more favourable tumour grade of invasive cancers detected by discordant reading (54.7% vs. 39.9% grade I tumours, p = 0.022, Table 2). Tumour size of invasive cancers tended to be smaller for cancers detected following a discordant reading, with fewer invasive cancers showing lymph node metastases, but these differences were not statistically significant.

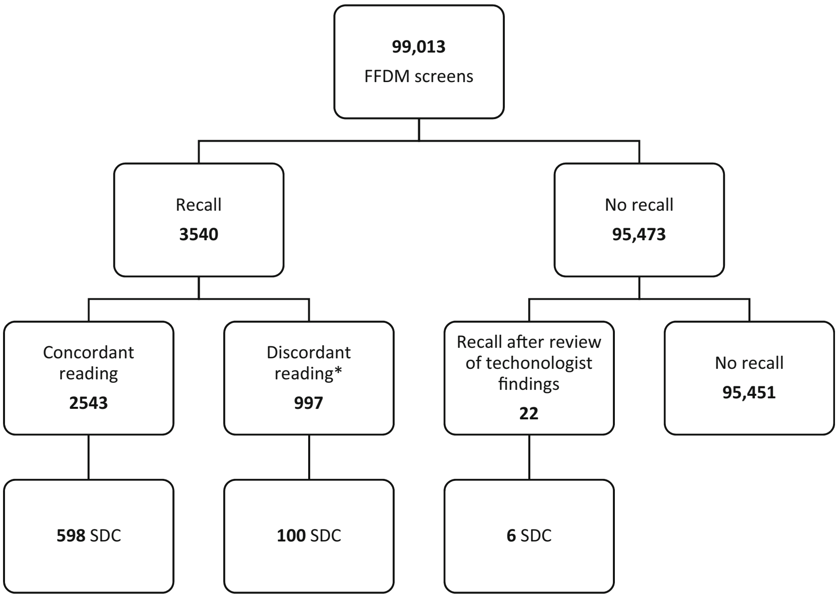

Fig. 1 Concordant versus discordant recalls at radiologist blinded double reading: screening outcome at 2-year follow-up. FFDM = full-field digital mammography; SDC = screen-detected cancer. *At radiologist blinded double reading, a woman was recalled if the mammogram was considered positive by both radiologists or, in case of a discordant reading, if at least one radiologist considered recall necessary. Thus, all discordant readings were recalled, without consensus reading or arbitration being performed

Discussion

To our knowledge, this is the first study comparing screening mammography findings and a wide range of tumour characteristics between concordant and discordant recalls at radiologist blinded double reading in digital mammography screening. We found that a discordant reading between radiologists was present in more than a quarter of all recalls and the majority of these discordant recalls comprised BI-RADS 0 mammographic abnormalities. Screen-detected cancers tended to be more favourable following discordant recalls, expressed in a more favourable tumour grading and a trend to smaller tumours, less often associated with nodal involvement.

The percentage of discordant recalls was lower than the one found in a previous study by Klompenhouwer et al (28% vs. 57%, respectively) [4], set between 2009 and 2011 in the same screening region and population and also using FFDM (with the transition from SFM just completed). This difference might, at least partly, be explained by the slightly higher recall rate (3.6% vs. 3.3%) in recent years. Also, screening radiologists have since then gained an additional 2 years of experience reading digital mammograms, are being continuously

Table 1 Concordant versus discordant recalls at radiologist blinded double reading: mammographic characteristics and screening outcome

	Concordant recall		Discordant recall		*p* value
	n=2,543		*n=997*		
Screening round, no. (%)					0.143
Initial	602	(23.7)	213	(21.4)	
Subsequent	1941	(76.3)	784	(78.6)	
Mammographic abnormality*, no. (%)					< 0.001
Mass	1556	(61.2)	548	(55.0)	
Microcalcifications	389	(15.3)	115	(11.5)	
Mass with microcalcifications	96	(3.8)	24	(2.4)	
Asymmetry	336	(13.2)	217	(21.8)	
Architectural distortion	166	(6.5)	93	(9.3)	
Recall BI-RADS, no. (%)					< 0.001
BI-RADS 0	1431	(56.3)	755	(75.7)	
BI-RADS 4	940	(37.0)	237	(23.8)	
BI-RADS 5	172	(6.8)	5	(0.5)	
Diagnosis, no. (%)					< 0.001
True positive	598	(23.5)	100	(10.0)	
False positive	1945	(76.5)	897	(90.0)	

*Dominant mammographic abnormality in case of multiple recalled lesions

Table 2 Concordant versus discordant recalls at radiologist blinded double reading: tumour characteristics of screen-detected cancers

	Concordant recall		Discordant recall		*p* value
	n = 598		*n = 100*		
Type of cancer, no. (%)					0.094
DCIS*	104	(17.4)	25	(25.0)	
Invasive	494	(82.6)	75	(75.0)	
DCIS grade, no. (%)					0.800
Low	17	(16.3)	5	(20.0)	
Intermediate	38	(36.5)	10	(40.0)	
High	49	(47.1)	10	(40.0)	
Histology of invasive cancers, no. (%)					0.320
Ductal	389	(78.7)	60	(80.0)	
Lobular	54	(10.9)	4	(5.3)	
Mixed ductal/lobular	14	(2.8)	5	(6.7)	
Other	33	(6.7)	6	(8.0)	
Unknown	4	(0.8)	0	(0.0)	
Tumour size of invasive cancers, no. (%)					0.086
T1 (≤20 mm)	391	(79.1)	66	(88.0)	
T2+ (>20 mm)	103	(20.9)	9	(12.0)	
Lymph node status of invasive cancers, no. (%)					0.117
N+	116	(23.5)	12	(16.0)	
N-	371	(75.1)	60	(80.0)	
Nx	7	(1.4)	3	(4.0)	
Nottingham grade, no. (%)					0.022
I	197	(39.9)	41	(54.7)	
II	230	(46.6)	22	(29.3)	
III	60	(12.1)	12	(16.0)	
Unknown	7	(1.4)	0	(0.0)	
Oestrogen receptor status, no. (%)					0.785
Positive	448	(90.7)	68	(90.7)	
Negative	43	(8.7)	7	(9.3)	
Unknown	3	(0.6)	0	(0.0)	
Progesterone receptor status, no. (%)					0.297
Positive	359	(72.7)	49	(65.3)	
Negative	132	(26.7)	26	(34.7)	
Unknown	3	(0.6)	0	(0.0)	
Her2/Neu receptor status, no. (%)					0.263
Positive	42	(8.5)	10	(13.3)	
Negative	446	(90.3)	65	(86.7)	
Unknown	6	(1.2)	0	(0.0)	
Triple negative tumour, no. (%)					0.553
Yes	28	(5.7)	3	(4.0)	
No	466	(94.3)	72	(96.0)	

*DCIS = Ductal carcinoma in situ

trained and receive regular feedback on their performance. It is likely that this has improved interobserver agreement and thus lowered the number of discordant recalls. The PPV of discordant recalls was significantly lower compared with concordant recalls. The effect of third reader arbitration of discordant double readings on screening outcome has been documented previously. For example, Ciatto et al [10] studied the effect of arbitration of discordant double readings and concluded that it reduces recall rates with a limited reduction in cancer detection rate. A study by Klompenhouwer et al [11] showed that arbitration of all discordant readings by a third reader improves the recall rate and increases the PPV of recall, but unfortunately also decreases programme sensitivity. Only arbitration of discordant BI-RADS 0 readings appeared to be a better strategy [12].

The proportions of DCIS and invasive cancers were comparable for screen-detected cancers detected by either concordant or discordant readings. In particular, the percentage of low-grade DCIS, a potential candidate for over-diagnosis [13, 14], did not differ significantly between the two groups. We found that in both groups most screen-detected cancers were small (T1a-c) invasive cancers and overall there was a high percentage of oestrogen and progesterone receptor positive tumours. The proportion of triple-negative tumours was very low and comparable in both groups. Recent studies suggest that these small invasive cancers with favourable biological behaviour (mainly hormone receptor-positive tumours) in many cases do not progress to larger, more aggressive cancers during the lifetime of the patient and therefore also represent over-diagnosis [15, 16]. Two of the major harms of over-diagnosis are the over-treatment that results and the anxiety and fear that a cancer diagnosis engenders [16]. It has also been suggested that, compared with other women, breast cancer patients have an increased risk to die from diseases of pulmonary circulation, various external causes and several heart diseases, which is attributed to both breast cancer treatment and breast cancer itself [17]. Tumour characteristics were comparable for both groups, except for differences in tumour grade; discordant invasive cancers were mostly well differentiated grade I tumours whereas concordant cancers were mostly grade II tumours. However, 45.3% of invasive cancers detected after a discordant recall were grade II or III cancers and 80% of DCIS were intermediate and high grade. Therefore, with concordant and discordant recalls, screen-detected cancers appear to represent a mix of over-diagnosis and early detection.

In our study, the majority of recalls were classified as BI-RADS 0 and the proportion of BI-RADS 0 recalls was also significantly higher in the discordant than the concordant reading group. In the Dutch screening setting, BI-RADS 0 represents a mammographic abnormality requiring further work-up. It is, however, generally considered to be a lesion with a relatively low malignancy risk of approximately 7% [12] and the PPV of 5.9% of a BI-RADS 0 recall found in our study is in line with this report. A possible explanation for the high proportion of BI-RADS 0 recalls in the discordant reading group could be that two readers are more likely to agree on a more obvious (BI-RADS 4 or 5) abnormality [18]. This hypothesis is supported by the fact that the percentage of 'subtle' abnormalities such as asymmetries and, to a lesser degree, architectural distortions was indeed higher in the discordant reading group. Conversely, the proportion of suspicious masses and microcalcifications was higher in the concordant reading group. Various studies have also shown that there is a substantial inter- and intra-observer variability in using the BI-RADS lexicon among radiologists [19, 20]. As the second reader is not informed about the first reader's opinion at blinded double reading, this inter-observer variability is likely to have more of an impact on screening outcome than at non-blinded double reading.

The differences in type of mammographic characteristics provide possible strategies for improvement. In training new and experienced screening radiologists, additional emphasis could be placed on the 'subtle' mammographic abnormalities such as asymmetries and architectural distortions. A recent study by Houssami et al showed that a single reading of 3D mammography (digital breast tomosynthesis) detected more breast cancers and had a lower FPR compared with the current practice of double-reading 2D mammography alone [21]. Another recent study by Dibble et al showed that digital breast tomosynthesis decreases inter-observer variability and increases reader confidence in the detection of architectural distortion [22]. Previously, Durand et al showed that use of digital breast tomosynthesis is associated with a lower recall rate of screening mammography, most often for asymmetries [23]. Digital breast tomosynthesis might be implemented in the Dutch nationwide breast cancer screening programme in the future. Preliminary results of artificial intelligence and deep learning for the assessment of mammograms are also promising [24]. In the future, these computer algorithms may aid the screening radiologist in determining which women should be recalled at screening mammography.

The strength of our study lies in the large study population with virtually complete follow-up and the fact that we provide information on a wide range of tumour characteristics. However, our study also has several limitations. First, the recall rate of the Dutch nationwide breast cancer screening programme is among the lowest worldwide. Data from our study might therefore not necessarily be transposable to other screening programmes. Second, although we did not perform a cost-effectiveness analysis, since double reading of screening mammograms is already standard in The Netherlands,blinded double reading is not likely to increase screening costs apart from the higher costs associated with a higher recall rate. Double reading, either non-blinded (independent) or blinded, might however not be deemed cost effective in other screening programmes.

In conclusion, discordant recalls proved to be malignant in 10% of women and these cancers show a more favourable tumour grade than cancers diagnosed after a concordant recall. Differences in mammographic characteristics, mainly the higher proportion of asymmetries and architectural distortions in the discordant reading group, provide a possible target for improving breast cancer screening programmes, for example, by additional training of screening radiologists and also in the implementation of digital breast tomosynthesis.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Lucien E.M. Duijm.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval Institutional Review Board approval was not required.

We have previously reported our screening studies, derived from our screening database, to the Medical Ethics Review Committee of the Catharina Hospital Eindhoven, The Netherlands (www.catharina-ziekenhuis.nl/metc/). This committee reviewed our studies and replied that according to the Medical Research Involving Human Subject Acts (WMO; www.ccmo.nl) our kind of study does not need the approval of the committee. We have contacted CCMO in the past to obtain an official letter from this institute, stating that our descriptive screening studies do not warrant ethical approval for the trial.

Methodology

- prospective
- diagnostic or prognostic
- performed at one institution

References

1. Duijm LEM, Groenewoud JH, Fracheboud J, van Ineveld BM, Roumen RMH, de Koning HJ (2008) Introduction of additional double reading of mammograms by radiographers: Effects on a biennial screening programme outcome. Eur J Cancer 44(9): 1223–1228
2. Caumo F, Brunelli S, Tosi E et al (2011) On the role of arbitration of discordant double readings of screening mammography: experience from two Italian programmes. Radiol Med 116(1):84–91
3. Azavedo E, Zackrisson S, Mejàre I, Heibert Arnlind M (2012) Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. BMC Med Imaging 12(1):22
4. Klompenhouwer EG, Voogd AC, Den Heeten GJ et al (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
5. Posso MC, Puig T, Quintana MJ, Solá-Roca J, Bonfill X (2016) Double versus single reading of mammograms in a breast cancer screening programme: a cost-consequence analysis. Eur Radiol 26(9):3262–3271
6. Duijm LEM, Louwman MWJ, 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
7. European Commission (2013) European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition. Office for Official Publications of the European Communities. 138 p.
8. Gur D, Sumkin JH, Hardesty LA et al (2004) Recall and detection rates in screening mammography: a review of clinical experience - implications for practice guidelines. Cancer 100(8):1590–1594
9. Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 19(5):403–410
10. Ciatto S, Ambrogetti D, Risso G et al (2005) The role of arbitration of discordant reports at double reading of screening mammograms. J Med Screen 12(3):125–127
11. Klompenhouwer EG, Voogd AC, den Heeten GJ et al (2015) Discrepant screening mammography assessments at blinded and non-blinded double reading: impact of arbitration by a third reader on screening outcome. Eur Radiol 25(10):2821–2829
12. Klompenhouwer EG, Weber RJP, Voogd AC et al (2015) Arbitration of discrepant BI-RADS 0 recalls by a third reader at screening mammography lowers recall rate but not the cancer detection rate and sensitivity at blinded and non-blinded double reading. Breast 24(5):601–607
13. Bluekens AMJ, Holland R, Karssemeijer N, Broeders MJM, den Heeten GJ (2012) Comparison of Digital Screening Mammography and Screen-Film Mammography in the Early Detection of Clinically Relevant Cancers: A Multicenter Study. Radiology 265(3):707–714
14. van Luijt PA, Heijnsdijk EAM, Fracheboud J et al (2016) The distribution of ductal carcinoma in situ (DCIS) grade in 4232 women and its impact on overdiagnosis in breast cancer screening. Breast Cancer Res 18(1):47
15. Welch HG, Prorok PC, O'Malley AJ, Kramer BS (2016) Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. N Engl J Med 375(15):1438–1447
16. Lannin DR, Wang S (2017) Are small breast cancers good because they are small or small because they are good? N Engl J Med 376(23):2286–2291
17. Riihimäki M, Thomsen H, Brandt A, Sundquist J, Hemminki K (2012) Death causes in breast cancer patients. Ann Oncol 23(3): 604–610
18. Lee AY, Wisner DJ, Aminololama-Shakeri S J et al (2017) Inter-reader variability in the use of BI-RADS descriptors for suspicious findings on diagnostic mammography: a multi-institution study of 10 academic radiologists. Acad Radiol 24(1):60–66
19. Ciatto S, Houssami N, Apruzzese A et al (2006) Reader variability in reporting breast imaging according to BI-RADS® assessment categories (the Florence experience). Breast 15(1):44–51
20. Redondo A, Comas M, Macià F et al (2012) Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. Br J Radiol 85: 1465–1470
21. Houssami N, Bernardi D, Pellegrini M et al (2017) Breast cancer detection using single-reading of breast tomosynthesis (3D-mammography) compared to double-reading of 2D-mammography: Evidence from a population-based trial. Cancer Epidemiol 47:94–99
22. Dibble EH, Lourenco AP, Baird GL, Ward RC, Maynard AS, Mainiero MB (2018) Comparison of digital mammography and digital breast tomosynthesis in the detection of architectural distortion. Eur Radiol 28(1):3–10

23. Durand MA, Haas BM, Yao X et al (2015) Early clinical experience with digital breast tomosynthesis for screening mammography. Radiology 274(1):85–92

24. Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A et al (2017) Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 35:303–312