

New approaches to improve the evaluation of mammograms

Citation for published version (APA):

van den Biggelaar, F. J. H. M. (2009). *New approaches to improve the evaluation of mammograms*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20091106fb>

Document status and date:

Published: 01/01/2009

DOI:

[10.26481/dis.20091106fb](https://doi.org/10.26481/dis.20091106fb)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

**New approaches to improve
the evaluation of mammograms**

© F.J.H.M. van den Biggelaar, Maastricht 2009

ISBN 978-90-5278-863-0

Cover design: Linda de Beukelaer, inspired by 'The Eyes'
by Louise Bourgeois

Printed by: Datawyse / Universitaire Pers Maastricht

Lay-out: Ine Kengen / Frank van den Biggelaar



**New approaches to improve
the evaluation of mammograms**

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus
Prof. mr. G.P.M.F. Mols
volgens het besluit van het College van Decanen
in het openbaar te verdedigen
op vrijdag 6 november 2009 om 10.00 uur

door

Franciscus Johannes Henricus Maria van den Biggelaar
Geboren op 8 december 1980 te Biest-Houtakker

Promotor

Prof. dr. J.M.A. van Engelshoven

Copromotor

Dr. K. Flobbe

Beoordelingscommissie

Prof. dr. V.C.G. Tjan-Heijnen (voorzitter)

Prof. dr. R.G.H. Beets-Tan

Dr. L.E.M. Duijm

Prof. dr. R. de Wit

Voor mijn moeder

Contents

Chapter 1	
Introduction.....	9
Chapter 2	
Performance of technologists in mammogram interpretation: a systematic review.....	15
Chapter 3	
Diagnostic performance of breast technologists in reading mammograms in a clinical patient population.....	27
Chapter 4	
Costs and effects of using specialised breast technologists in pre-reading mammograms in a clinical patient population.....	43
Chapter 5	
Pre-reading mammograms by specialised breast technologists: legal implications for technologist and radiologist in the Netherlands.....	57
Chapter 6	
Computer-aided detection in full-field digital mammography in a clinical population: performance of radiologist and technologists.....	67
Chapter 7	
Strategies for digital mammography interpretation in a clinical patient population.....	83
Chapter 8	
General discussion and conclusions.....	101
Chapter 9	
Summary.....	113
Samenvatting.....	117
Dankwoord.....	123
Curriculum Vitae.....	125
List of publications.....	127

1

Introduction

Breast cancer in the Netherlands

Breast cancer is the most common malignancy among women. With an incidence of more than 12,000 cases in the Netherlands in 2006, breast cancer includes 31% of all female cancers¹. Currently, the cumulative lifetime risk of breast cancer among Dutch women is approximately 13%². Over the past decades, incidence of breast malignancies has increased. An enlarged exposure to known risk factors like younger age at menarche, older age at birth of first child, shorter lactation and increased alcohol consumption, have probably contributed substantially to the growth³. In addition, improved detection methods are likely to have raised the incidence of breast cancer. Particularly the implementation of the Dutch Nationwide Breast Cancer Screening Programme showed a clear increase in the incidence rates⁴.

The mortality rate of breast cancer has decreased over the years. In 2006, 30 women per 100,000 died of breast cancer in the Netherlands, compared to 38 per 100,000 in 1990 (age-adjusted using European Standardised Rate, ESR)⁵. Two factors are mainly responsible for this decline. First, due to early diagnosis, the detected malignancies are often smaller and low-grade tumours. Second, the treatment of breast cancer has improved substantially over the last decades³.

The improving survival rates combined with the increasing incidence rates have resulted in an increased prevalence of breast cancer cases. In the Netherlands, an enlargement in breast cancer prevalence is expected from 119,000 patients in 2005 to 194,000 in 2015. Many of these patients will need health care services, which will lead to a serious burden in health care³.

Diagnostic versus screening mammography

For early detection and diagnosis of breast cancer, mammography is the basic imaging modality. In a clinical situation, mammograms can be used for diagnostic as well as for screening reasons. Diagnostic mammography is performed in women with clinical signs or symptoms suggestive for breast cancer, like a palpable breast lump or nipple abnormalities, women referred by the Dutch Nationwide Breast Cancer Screening Programme and in women with a personal history of breast cancer⁶.

Screening mammography is performed in asymptomatic women to detect early, clinically unsuspected breast cancer. Indications for screening mammography in a hospital radiology department could be asymptomatic reasons like family history of breast cancer, a genetic predisposition or fear of breast cancer.

In addition, a large part of the screening mammograms is obtained by the Dutch Nationwide Breast Cancer Screening Programme in mobile mammography units outside a clinical centre.

Breast imaging in a clinical patient population

According to national guidelines for breast cancer diagnosis and treatment, routine breast imaging in women referred to a radiology department of a hospital consists of clinical breast examination and mammography. Ultrasonography is recommended as an adjunct to mammography in patients referred for a palpable breast mass, patients referred with an abnormal screening mammogram and in patients with an abnormal diagnostic mammogram that requires further diagnostic imaging⁷.

All patients receive a standard two-view unilateral or bilateral mammogram. The radiologist on duty evaluates the examinations and decides whether additional work-up is required. All findings of the breast imaging examinations are recorded in a report which should include a BI-RADS Assessment Category⁶. This score indicates the opinion of the radiologist regarding the level of suspicion of a breast abnormality. The BI-RADS scoring system consists of six classifications: 1 = negative examination; 2 = benign finding; 3 = probably benign finding; 4 = suspicious abnormality; 5 = highly suggestive of malignancy; 6 = proven malignancy. Furthermore, BI-RADS 0 is assigned when additional imaging evaluation or comparison to prior mammograms is required to make a final assessment.

New approaches to improve the evaluation of mammograms

In the setting of an increasing demand for breast imaging due to the rising breast cancer incidence rates and an increasing shortage of well-experienced radiologists, it is getting harder to assure the high efficacy and quality of breast imaging evaluation in daily clinical practice. In order to deal with this issue, new approaches in the evaluation of mammograms in a clinical patient population could be considered.

Deployment of breast technologists

The quality of breast imaging evaluation in daily clinical practice may be improved by the deployment of breast technologists trained in mammogram reading.

The involvement of technologists could be effectuated through pre-reading or double reading procedures. Double reading mammograms with a radiologist and a technologist will most probably increase the sensitivity in the detection of malignancies in daily clinical practice. On the other hand, double reading has the risk of overdiagnosing patients, whereby creating large numbers of false-positive results and unnecessary invasive biopsy procedures. Furthermore, as in daily clinical practice mammogram reading is performed by a single radiologist, double reading would not reduce the workload of radiologists and diagnostic costs.

Therefore, it might be argued to employ technologists as pre-readers. Studies have shown that about 70% of the patients receiving a mammogram in clinical

practice, are discharged without further evaluation⁸⁻¹¹. When technologists are able to select the cases that require further evaluation and can exclude all other patients from further assessment by a radiologist without jeopardising the cancer detection rate, the workload of the radiologist could be reduced significantly. Furthermore, costs could be lowered by reducing the time required by radiologists to review large numbers of negative mammograms.

Computer-aided detection (CAD) software

Another innovation that has the potential to improve the quality of the evaluation of mammograms, is the application of computer-aided (CAD) software. This technique have been developed to mark suspicious areas that otherwise may be missed by the mammogram reader, which could potentially result in the detection of more malignancies.

Several studies have focused on the application of CAD in screening mammography. A systematic literature review shows that CAD is controversial on improving cancer detection rates but does increase recall rates in screening programmes for breast cancer¹². However, little is known about the use of CAD software in a clinical patient population.

PERSPECT study

In order to study the diagnostic performance of breast technologists in reading mammograms in a clinical patient population and in order to evaluate different reading strategies, the PERSPECT study (PERformance of SPECialised breast Technologists in breast imaging) was performed at Maastricht University Medical Center (MUMC). The study was financially supported by the Netherlands Organisation for Health Research and Development (ZonMw) as part of the subprogram "Early Evaluation of Medical Innovation" of the "Health Care Efficiency Research Program".

During nine months, two technologists with special interest in mammography were trained in mammogram reading under the supervision of specialised breast radiologists. Subsequently, a prospective study was performed in which the performance of the technologists in evaluating mammograms was compared to the performance in daily clinical practice with mammogram reading by the radiologist on duty. All consecutive women referred for mammography between January and August 2007, were included in the study. The reference standard for the presence or absence of breast cancer was determined by pathology results during a follow-up of twelve months.

In addition, the mammograms obtained in the PERSPECT study, were used in order to evaluate the impact of CAD software on the detection of breast malignancies. In this study, a third technologist was involved who is qualified as a physician assistant in breast imaging in our radiology department.

Objectives of the thesis

This thesis addresses the feasibility of the deployment of specialised breast technologists in evaluating mammograms in a clinical patient population. Reader performance, different reading strategies, costs and legal implications were studied. Furthermore, the impact of CAD software in digital mammography on the performance of mammogram readers was evaluated.

Outline of the thesis

Chapter 2 describes a systematic literature review of the performance of technologists in mammogram interpretation and the effects of training on the performance.

In **chapter 3**, results of the PERSPECT study are reported. Two breast technologists trained in mammogram reading, prospectively evaluated 1994 mammograms. The performance of the technologists in pre-reading mammograms is studied in terms of sensitivity and specificity, by selecting either normal mammograms and mammograms with clearly benign findings.

Chapter 4 describes the costs and effects of the deployment of breast technologists in pre-reading mammograms in a clinical patient population. Four different reading strategies were analysed by decision analytic modelling and compared to the conventional strategy of standard mammogram evaluation in daily clinical practice.

Chapter 5 emphasises the legal implications of pre-reading by technologists. The legal preconditions that are required, and the risk of being sued for medical malpractice are discussed.

A study evaluating the application of CAD software in a consecutive patient population was subject of **chapter 6**. Mammograms of 1048 patients were evaluated with and without application of the CAD system by four observers, consisting of a radiologist and three technologists (two trained technologists and one physician assistant). The impact of CAD software on the detection of malignancies was studied.

Chapter 7 discusses different strategies for digital mammogram reading in a clinical patient population. The effect of using a CAD system was evaluated as well as the deployment of technologists as pre-readers and double readers. In order to study the effects of these different reading strategies, ROC analysis was used.

In **chapter 8**, the results of the studies presented in this thesis are reviewed and discussed.

References

1. <http://www.ikcnet.nl>. 20th of June 2009
2. Kiemeney LA, Lemmers FA, Verhoeven RH, et al. [The risk of cancer in the Netherlands]. *Ned Tijdschr Geneesk* 2008; 152:2233-2241
3. Louwman WJ, Voogd AC, van Dijck JA, et al. On the rising trends of incidence and prognosis for breast cancer patients diagnosed 1975-2004: a long-term population-based study in southeastern Netherlands. *Cancer Causes Control* 2008; 19:97-106
4. Fracheboud J, Otto SJ, van Dijck JA, Broeders MJ, Verbeek AL, de Koning HJ. Decreased rates of advanced breast cancer due to mammography screening in The Netherlands. *Br J Cancer* 2004; 91:861-867
5. www.ikcnet.nl/cijfers. 20th of June 2009
6. D'Orsi CJ, Bassett LW, Berg WA, et al. *Breast Imaging Reporting and Data System: ACR BI-RADS-Mammography*. Reston (VA): American College of Radiology (ACR), 2003
7. Vercauteren LD, Kessels AG, van der Weijden T, et al. Clinical impact of the use of additional ultrasonography in diagnostic breast imaging. *Eur Radiol* 2008; 18:2076-2084
8. Flobbe K, Bosch AM, Kessels AG, et al. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. *Arch Intern Med* 2003; 163:1194-1199
9. Flobbe K, van der Linden ES, Kessels AG, van Engelshoven JM. Diagnostic value of radiological breast imaging in a non-screening population. *Int J Cancer* 2001; 92:616-618
10. Zonderland HM, Coerkamp EG, Hermans J, van de Vijver, MJ, van Voorthuisen AE. Diagnosis of breast cancer: contribution of US as an adjunct to mammography. *Radiology* 1999; 213:413-422
11. Zonderland HM, Pope, TL, Nieborg, AJ. The positive predictive value of the breast imaging reporting and data system (BI-RADS) as a method of quality assessment in breast imaging in a hospital population. *Eur Radiol* 2004; 14:1743-1750
12. Taylor P, Potts HW. 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

2

Performance of technologists in mammogram interpretation: a systematic review

F.J.H.M. van den Biggelaar, P.J. Nelemans, K. Flobbe

Published as: "Performance of radiographers in mammogram interpretation: a systematic review"

The Breast 2008; 17(1): 85-90

Abstract

Radiologists may be relieved from work that could be performed by technologists. This systematic literature review focuses on the performance of technologists (also referring to radiographers and physician assistants) compared with radiologists in the interpretation of mammograms; the effect of training; and the question whether there are any studies evaluating the effects of involving technologists in the interpretation of diagnostic mammograms in daily clinical practice on the sensitivity and specificity of cancer detection in breast imaging. Six studies met the inclusion criteria (primary aim of the study has to be the evaluation of the performance of technologists, sensitivity and specificity have to be reported or calculable and there has to be a sufficient gold standard). The results showed that, in a screening setting, technologists scored higher false-positive rates with a similar sensitivity in the detection of malignancies, compared with radiologists. Furthermore, results suggested that training could improve their performance. No studies were reported assessing the performance of technologists interpreting diagnostic mammograms in a consecutive patient population in a daily clinical setting. This indicates a need for a well-designed diagnostic study using an adequate gold standard, in order to evaluate the feasibility of deploying technologists in the interpretation of diagnostic mammograms in a clinical setting.

Introduction

In the setting of an increasing workload for radiologists, the possibility of trained radiologic technicians taking over duties previously reserved for radiologists, is being explored. This concept is known as "skill mix", and implies the utilisation of expertise from individuals in related fields to complement or increase the expertise available to patients while providing financial savings to the economy¹. It has been shown that technologists can accurately report plain radiographs in clinical practice and that training improves their results². In the United Kingdom, radiologic technologists have been utilised to identify fractures in the emergency department, to perform ultrasonography, barium enema examinations and upper gastrointestinal examinations, and to double read screening mammograms¹.

Many countries have introduced a breast cancer-screening programme for symptom-free women above 50 years of age, in order to reduce the morbidity and mortality associated with breast cancer by means of early detection. Double reading of screening mammograms has shown to increase the cancer detection, which in the long term would result in a reduction of mortality from breast cancer³. As double reading by radiologists is becoming difficult to maintain because of increasing workloads, the possibility of technologists as second readers has been explored⁴⁻⁶.

This development of deploying technologists in reading screening mammograms raises the question whether they could be employed in a diagnostic setting as well by reading diagnostic mammograms.

The aim of this study was to explore to what extent the literature provides evidence on the performance of non-radiologists in reading mammograms in clinical practice. A systematic literature review was performed in order to study three research questions.

First, what is the performance of technologists (also referring to technologists and physician assistants) in the interpretation of mammograms compared with radiologists? Second, what is the effect of training on the performance of technologists in mammogram interpretation? Third, are there any studies evaluating the effects of involving technologists in the interpretation of diagnostic mammograms in daily clinical practice on the sensitivity and specificity of cancer detection in breast imaging?

Methods

Study selection

PUBMED and EMBASE databases were searched with the following terms: technologist(s), technician(s), radiographer(s), mammogram(s), mammographic and mammography. Furthermore, smaller databases such as Scirus, PsychINFO and Oxford Journals were searched with the same terms. The search was conducted until December 2006 and no publication date limit was used. A first screening of titles and abstracts resulted in the exclusion of papers that

were clearly not relevant and in the selection of eligible articles. Three reviewers (FB, PN, KF) independently read the selected articles and independently assessed whether the studies met the following inclusion criteria: the primary aim of the study was to evaluate the performance of a technologist; sensitivity and specificity had to be reported or had to be calculable; and the gold standard should be a follow-up of at least one year and/or a pathology report or the use of a validated test set. Only English, German, French or Dutch papers were considered eligible for review.

Data extraction

From the papers which met the inclusion criteria, data were extracted with respect to study characteristics, details of mammogram evaluation and reader characteristics. Study characteristics included: year of publication; country where the study was performed; setting (screening versus diagnostic); gold standard used; the total number of mammograms that were evaluated; and the prevalence of malignant abnormalities. Details of mammogram evaluation included: double- or pre-reading; blinded or non-blinded reading; use of a test set versus consecutive series of patient selection; and pre- and post-training measurement. Furthermore, reader characteristics included: profession; number of professionals; years of experience; and specific training in mammogram reading. Finally, sensitivity and specificity were recorded.

Data analysis

For each study, the diagnostic performance was expressed in terms of sensitivity, specificity, positive predictive value, negative predictive value and diagnostic odds ratio (DOR). The DOR is a measure for the diagnostic performance of a test, which combines sensitivity and specificity in one measure. A DOR of 1 implies that the test has no discriminatory power at all; the larger the DOR, the better the test discriminates between patients with and without the disease of interest⁷.

Results

The literature search in PUBMED provided 141 publications. After a crude screening of all titles and abstracts, twelve publications were considered to be eligible for inclusion^{3,5-6,8-16}. Reference screening and citation tracking resulted in another seven studies^{4,17-22}. A search through EMBASE and smaller databases did not increase the number of relevant studies. After extensive reading of the papers, five did meet the inclusion criteria^{3-5,8-9}. Ten papers were excluded because of another primary aim of study, two studies did not have an appropriate gold standard and two publications did not provide valid statistics. One paper reported data on two different populations, so the results were included as two separate studies⁸. Therefore, the final data set consisted of six studies.

The study characteristics, the details of mammogram evaluation and the reader characteristics are shown in Table 1, 2 and 3, respectively.

Table 1: Study characteristics

	Year	Country	Setting	Gold standard	Number of mammograms	Prevalence of malignancies (%)
Bassett ^{8-1a}	1995	USA	Screening	Biopsy + 1 year follow-up	627	7.7
Bassett ^{8-2b}	1995	USA	Screening	Biopsy + 1 year follow-up	611	6.9
Haiart ⁹	1991	UK	Screening	Pathology + 1 year follow-up	3362 / 6080 ^c	0.45 / 0.38
Hillman ⁴	1987	USA	Screening	Pathology + 3 years follow-up	727	25
Pauli ⁵	1996	UK	Screening	Pathology + 1.5 years follow-up	17202	0.8
Pauli ³	1996	UK	Screening	Validated test set	79	20.3

^a institution 1

^b institution 2

^c different N for technologist and radiologist

Table 2: Details of mammogram evaluation

	Double / pre-reading	Blinded	Patient selection	Pre- and post-training
Bassett ^{8-1+2a} pre-test	Double	Yes	Test set	Yes
Bassett ^{8-1+2b} post-test	Double	Yes	Test set	Yes
Haiart ⁹	Pre	Yes	Consecutive series of patients	No
Hillman ⁴	Double	Yes	Test set	No
Pauli ⁵	Double	Yes	Consecutive series of patients	No
Pauli ³	Single	N.M. ^b	Test set	Yes

^a institution 1 and institution 2

^b N.M. = Not Mentioned

The publications originated from the UK (n=3) and the USA (n=3) and the year of publication ranged from 1987⁴ to 1996^{3,5}. The number of mammograms studied ranged from 79³ to 17,202⁵. The prevalence of breast cancer varied between 0.38%⁹ and 45.5%⁴, with the higher prevalences originating from selected test sets of mammograms.

In four studies, technologists were employed in double reading of mammograms^{4-5,8}, in one study they were pre-reading the images⁹ and in one study

they were the single observer³. The different non-radiologist observers in the studies were named radiographers^{3,5,9}, technologists⁸, and physician assistants⁴.

Table 3: Reader characteristics

	Radiologist		Non-radiologist		Training
	no. ^a	experience	no.	experience	
Bassett ⁸⁻¹ ^b	4	Experienced in interpreting mammograms	4	>1 year in performing mammography	2x 0.5 days + 6 weeks
Bassett ⁸⁻² ^c	4	Experienced in interpreting mammograms	3	>1 year in performing mammography	2x 0.5 days+ 10 weeks
Haiart ⁹	1	10 years interpreting screening and symptomatic mammograms	1	N.M. ^d	6 months
Hillman ⁴	6	1-20 years, no specialists in mammography	4	N.M.	5 weeks
Pauli ⁵	N.M.	N.M.	7	6-121 months working in screening mammography	2 weeks
Pauli ³	--	--	7	6-121 months working in screening mammography	2 weeks

^a number of professionals

^b institution 1

^c institution 2

^d N.M. = Not Mentioned

Mammogram interpretation performance

Five publications provided data on the comparability of sensitivity and specificity between radiologists and technologists, whereas these data were missing for the study of Pauli et al.³ (Table 4). When data were reported for a pre- and a post-training evaluation by technologists, the numbers of the post-training evaluation are used.

Four studies^{5,8-9} demonstrated that the performance of technologists in the interpretation of mammograms was poorer than for radiologists, because of higher numbers of false-positive results. Specificity ranged from 81% to 95% for radiologists and from 64% to 91% for technologists. The rates of sensitivity were comparable, ranging from 73% to 86% for radiologists and 73% to 90% for technologists.

Consequently, the DORs were lower for technologists compared with radiologists. One study reported a higher sensitivity for technologists compared with radiologists⁴.

Table 4: Mammogram interpretation performance

	Radiologist					Technologist				
	Se ^a (%)	Sp ^b (%)	DOR ^c	PPV ^d (%)	NPV ^e (%)	Se (%)	Sp (%)	DOR	PPV (%)	NPV (%)
Bassett ^{8-1f}	86 (80-91) ^g	83 (81-85)	30	34 (30-39)	98 (98-99)	90 (85-94)	75 (73-77)	27	27 (24-31)	99 (98-99)
Bassett ^{8-2h}	85 (78-90)	81 (79-83)	24	31 (27-36)	98 (97-99)	84 (78-89)	64 (62-66)	9	19 (18-22)	97 (96-98)
Haiart ⁹	83 (61-95)	86 (85-87)	29	2 (1-3)	100 (99-100)	80 (52-96)	78 (76-79)	14	2 (1-3)	100 (100-100)
Hillman ⁴	61 (54-68)	92 (89-94)	18	72 (64-79)	88 (85-90)	78 (71-84)	91 (88-93)	36	74 (68-80)	93 (90-95)
Pauli ⁵	73 (65-80)	95 (95-95)	51	11 (9-13)	100 (100-100)	73 (65-80)	86 (85-87)	16	4 (3-5)	100 (100-100)

^a Se = sensitivity; ^b Sp = specificity; ^c DOR = diagnostic odds ratio; ^d PPV = positive predictive value;

^e NPV = negative predictive value; ^f institution 1; ^g numbers in brackets are 95% confidence intervals;

^h institution 2

Effects of training

In five studies, a specific training programme was offered to the mammogram readers, which varied between twice a half-day and a five-weeks training period (Table 3). The training consisted of didactic lectures, self-study of teaching files and daily reading of mammograms. In three studies, a performance measurement was done before and after the training period. In Table 5, the mammogram interpretation performance of technologists was reported before and after the training programmes. Two studies reported an increased DOR mainly because of an increased specificity^{3,8} and, despite a decreased specificity, Bassett et al. reported at institution 1 an increased DOR, because of an increased sensitivity⁸.

Table 5: Effects of training on mammogram interpretation performance of technologists

	Pre-training measurement					Post-training measurement				
	Se ^a (%)	Sp ^b (%)	DOR ^c	PPV ^d (%)	NPV ^e (%)	Se (%)	Sp (%)	DOR	PPV (%)	NPV (%)
Bassett ^{8-1f}	78 (72-84) ^g	87 (85-89)	24	38 (33-43)	97 (97-98)	90 (85-94)	75 (73-77)	27	27 (24-31)	99 (98-99)
Bassett ^{8-2h}	89 (84-93)	44 (42-46)	6	14 (12-16)	98 (96-98)	84 (78-89)	64 (62-66)	9	19 (18-22)	97 (96-98)
Pauli ³	77 (68-84)	68 (64-72)	7	38 (32-45)	92 (89-95)	83 (75-89)	80 (76-84)	20	51 (44-59)	95 (92-97)

^a Se = sensitivity; ^b Sp = specificity; ^c DOR = diagnostic odds ratio; ^d PPV = positive predictive value;

^e NPV = negative predictive value; ^f institution 1; ^g numbers in brackets are 95% confidence intervals;

^h institution 2

Effects in daily clinical practice

All studies were performed in a screening setting; no studies were found that were performed in a diagnostic patient population. Furthermore, there were two publications that studied consecutive series of patients from breast cancer screening programmes^{5,9}. No publications were found that studied the effects on sensitivity and specificity of employing technologists in the reading of diagnostic mammograms in a consecutive patient population in clinical practice.

Discussion

This literature review shows that there are only a few well-designed studies that evaluate the effects of involving technologists in mammogram interpretation on the sensitivity and specificity of breast cancer diagnosis. The results showed that technologists scored higher false-positive rates with a similar sensitivity in the detection of malignancies, compared with radiologists. Furthermore, it was indicated that training programmes could improve the performance by reducing the number of false-positive results and increasing the specificity. No publications were found studying the effects on sensitivity and specificity of employing technologists in the reading of diagnostic mammograms in a clinical setting.

In four studies, technologists were employed in double reading of mammograms, whereas only one study considered pre-reading by technologists. Double reading of mammograms has been shown to increase the number of cancers detected in population-based breast cancer screening programmes^{5,23-30}. Double interpretation has been performed in one of two ways: mammograms are interpreted by each observer independently without discussion of the finding; or mammograms are interpreted in consensus, in which recall occurs only with agreement of the observers involved³⁰. Moreover, the study of Duijm et al.³¹ demonstrated that another 11% of the cases with findings that remained undecided after the consensus meeting of the two observers, were breast cancer positive.

Conventionally, double reading is performed by two radiologists, but the additional costs, the shortage of radiologists and the large volume of mammograms has resulted in the exploration of the feasibility to employ technologists as double readers. Several studies showed that, in a screening setting, double reading by technologists improved the cancer detection rate^{4-6,16}. Furthermore, the study of Hillman et al.⁴ reported that the interpretations by physician assistants were even more sensitive and as specific as those of radiologists. They suggest that motivation and specific training are likely explanations for this result. The study of Hillman⁴ also demonstrated that the involvement of technologists as second readers leads to cost reduction. However, only direct costs were included. Costs of training, overhead and logistics were excluded.

Pre-reading is an alternative way of involving non-radiologists in the interpretation of mammograms. When pre-reading mammograms, technologists group mammograms into two basic categories: cases that require further evaluation

and cases that can be discharged, because they have either negative or clearly benign findings¹³. Mucci et al.²⁰ studied the role of technologists in identifying those patients with an abnormal mammogram which, in their opinion, needed further assessment, by placing a "red dot" on the patients form indicating the site of the lesion. However, their sensitivity was low compared with that of the radiologists. Furthermore, the study of Haiart showed that pre-reading cannot be justified in a screening setting, neither in terms of performance, nor on economic grounds.

Several studies have shown the ability of technologists to identify abnormalities on films in a screening setting. The feasibility of the involvement of technologists in reading mammograms in a diagnostic setting has not yet been explored. One of the questions is whether, in such a setting, technologists can best be employed as double readers or as pre-readers of mammograms. Double reading of screening mammograms has been proven superior to increase sensitivity in the detection of malignancies, when compared with single reading. The deployment of technologists is expected to increase sensitivity. Although the sensitivity of readers of mammograms in a diagnostic setting is higher than in a screening setting (about 90%³²) and the use of additional diagnostic examinations, like ultrasonography and needle biopsy, increases the sensitivity of the detection of malignancies³³, the risk of missing malignant lesions is still existing. Therefore, double reading with a technologist will most probably increase the sensitivity of the detection of malignancies in a diagnostic setting.

On the other hand, it can be argued that in clinical practice, in a diagnostic patient population, technologists might be employable as pre-readers. In the Netherlands, about 70% of the patients receiving a mammogram in the hospital, are discharged without further evaluation³⁴. When technologists are able to select the cases that require further evaluation and can discharge the patients with either negative or clearly benign findings, the workload of the radiologist could be reduced significantly. In addition, this review shows that technologists could be as sensitive as radiologists in the detection of malignancies. Therefore, in a diagnostic setting, pre-reading might be more cost-effective than double reading.

In this review, studies that were performed outside a screening setting in consecutive series of patients are found to be lacking. Most studies evaluated double reading of a selected set of mammograms. As a selected test set does not represent daily clinical practice and often has an overrepresentation of abnormal cases, it would be important to document studies in a more real life setting.

The effects of employing technologists in the interpretation of diagnostic mammograms is studied most accurately with reference to daily practice, by determining the effects on the number of cancers detected and the number of mammograms interpreted as abnormal. Prior to the study, the technologists have to

receive a sufficient training, not only to learn difficult cancer patterns in order to increase sensitivity, but also to identify benign lesions as benign. This would imply more extended training programmes in future studies.

This review illustrates the need for a well-designed diagnostic and cost-effectiveness study to demonstrate the feasibility of deploying technologists in the interpretation of diagnostic mammograms, using an appropriate gold standard consisting of known pathology and follow-up of at least one year. Such a study could also serve to evaluate which strategy (double reading or pre-reading) will be most sufficient in daily clinical practice.

References

1. FriedenberG RM. The role of the supertechnologist. *Radiology* 2000; 215:630-633
2. Brealey S, Scally A, Hahn S, Thomas N, Godfrey C, Coomarasamy A. Accuracy of radiographer plain radiograph reporting in clinical practice: a meta-analysis. *Clin Radiol* 2005; 60:232-241
3. Pauli R, Hammond S, Cooke J, Ansell J. Radiographers as film readers in screening mammography: an assessment of competence under test and screening conditions. *The British Journal of Radiology* 1996; 69:10-14
4. Hillman BJ, Fajardo LL, Hunter TB, et al. Mammogram interpretation by physician assistants. *AJR Am J Roentgenol* 1987; 149:907-912
5. Pauli R, Hammond S, Cooke J, Ansell J. Comparison of radiographer/radiologist double film reading with single reading in breast cancer screening. *Journal of Medical Screening* 1996; 3:18-22
6. Wivell G, Denton ER, Eve CB, Inglis JC, Harvey I. Can radiographers read screening mammograms? *Clin Radiol* 2003; 58:63-67
7. Glas AS, Lijmer JG, Prins MH, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; 56:1129-1135
8. Bassett LW, Hollatz-Brown AJ, Bastani R, Pearce JG, Hirji K, Chen L. Effects of a program to train radiologic technologists to identify abnormalities on mammograms. *Radiology* 1995; 194:189-192
9. Haiart DC, Henderson J. A comparison of interpretation of screening mammograms by a radiographer, a doctor and a radiologist: results and implications. *Br J Clin Pract* 1991; 45:43-45
10. Luo P, Qian W, Romilly P. CAD-aided mammogram training. *Acad Radiol* 2005; 12:1039-1048
11. Nodine CF, Kundel HL, Lauver SC, Toto LC. Nature of expertise in searching mammograms for breast masses. *Acad Radiol* 1996; 3:1000-1006
12. Nodine CF, Kundel HL, Mello-Thoms C, et al. How experience and training influence mammography expertise. *Acad Radiol* 1999; 6:575-585
13. Sumkin JH, Klamon HM, Graham M, et al. Prescreening mammography by technologists: a preliminary assessment. *AJR Am J Roentgenol* 2003; 180:253-256
14. Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H. Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography. *Health Technol Assess* 2005; 9:1-58
15. Taylor PM, Champness J, Given-Wilson RM, Potts HW, Johnston K. An evaluation of the impact of computer-based prompts on screen readers' interpretation of mammograms. *Br J Radiol* 2004; 77:21-27
16. Tonita JM, Hillis JP, Lim CH. Medical radiologic technologist review: effects on a population-based breast cancer screening program. *Radiology* 1999; 211:529-533
17. Balleyguier C, Kinkel K, Fermanian J, et al. Computer-aided detection (CAD) in mammography: does it help the junior or the senior radiologist? *Eur J Radiol* 2005; 54:90-96
18. Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K. Improving breast cancer diagnosis with computer-aided diagnosis. *Acad Radiol* 1999; 6:22-33
19. Mello-Thoms C. Perception of breast cancer: eye-position analysis of mammogram interpretation. *Acad Radiol* 2003; 10:4-12
20. Mucci B, Lawson S, Athey G, Scarisbrick G. Radiographers as readers in breast screening: experience with a 'red dot' method. *The Breast* 1997; 6:183-185

21. Nodine CF, Mello-Thoms C, Kundel HL, Weinstein SP. Time course of perception and decision making during mammographic interpretation. *AJR Am J Roentgenol* 2002; 179:917-923
22. Rao MC, Griffith CDM, Griffiths AB. Can breast surgeons read mammograms of symptomatic patients in the one-stop breast clinic? *Ann R Coll Surg Engl* 2001; 83:108-109
23. Anderson ED, Muir BB, Walsh JS, Kirkpatrick AE. The efficacy of double reading mammograms in breast screening. *Clin Radiol* 1994; 49:248-251
24. Thurfjell EL, Lernevall KA, Taube AA. Benefit of independent double reading in a population-based mammography screening program. *Radiology* 1994; 191:241-244
25. Laming D, Warren R. Improving the detection of cancer in the screening of mammograms. *J Med Screen* 2000; 7:24-30
26. 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
27. 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; 6:152-158
28. Mucci B, Athey G, Scarisbrick G. Double reading of screening mammograms: the use of a third reader to arbitrate on disagreements. *The Breast* 1999; 8:63-65
29. Anttinen I, Pamilo M, Soiva M, Roiha M. Double reading of mammography screening films--one radiologist or two? *Clin Radiol* 1993; 48:414-421
30. Harvey SC, Geller B, Oppenheimer RG, Pinet M, Riddell L, Garra B. Increase in cancer detection and recall rates with independent double interpretation of screening mammography. *AJR Am J Roentgenol* 2003; 180:1461-1467
31. 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; 231:564-570
32. Flobbe K, van der Linden ES, Kessels AG, van Engelshoven JM. Diagnostic value of radiological breast imaging in a non-screening population. *Int J Cancer* 2001; 92:616-618
33. Flobbe K, Nelemans PJ, Kessels AG, Beets GL, von Meyenfeldt MF, van Engelshoven JM. The role of ultrasonography as an adjunct to mammography in the detection of breast cancer. a systematic review. *Eur J Cancer* 2002; 38:1044-1050
34. Flobbe K, Bosch AM, Kessels AG, et al. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. *Arch Intern Med* 2003; 163:1194-1199

3

Diagnostic performance of breast technologists in reading mammograms in a clinical patient population

F.J.H.M. van den Biggelaar, A.G.H. Kessels,
J.M.A. van Engelshoven, K. Flobbe

*Accepted for publication in
International Journal of Clinical Practice*

Abstract

In the setting of an increasing workload for radiologists, this study focuses on the feasibility of skill mixing in breast imaging in a hospital radiology department.

Two radiologic technologists were trained in pre-reading mammograms in order to select the cases that require further evaluation by a radiologist. Mammograms of 1994 consecutive patients were evaluated by the technologists, next to the standard clinical interpretation by the radiologist on duty. Mammographic findings were recorded and a BI-RADS classification was assigned for each breast.

In total, 93 breast cancers were found in 91 patients, leading to a prevalence of 4.6% (91/1994). The sensitivity and specificity in selecting mammographic findings (cut-off point between BI-RADS 1 and BI-RADS 0,2-5 and the radiologist's diagnosis as reference standard) was 98% and 74% for technologist 1 and 98% and 78% for technologist 2. In distinguishing normal and benign mammograms from those with abnormalities that are probably benign, suspicious or highly suggestive for malignancy (cut-off point BI-RADS 1-2 and BI-RADS 0,3-5 and pathology results as reference standard), the sensitivity decreased to 89% and 91%, respectively. The specificity, however, increased to 82% for both technologists.

In a subgroup of 1389 patients with no immediate indication for additional imaging with the involvement of a radiologist, technologists obtained a mean sensitivity and specificity of 98% and 77% in detecting mammographic findings, and a mean sensitivity and specificity of 78% and 88% in detecting suspicious abnormalities.

Concluding, the employment of technologists in pre-reading mammograms seems to be an effective working strategy in daily clinical practice. However, its position in clinical practice remains indistinct as a continuous availability of radiologists still needs to be guaranteed. Nevertheless, as a substantial proportion of mammograms could be evaluated without the attention of a radiologist, the employment of technologists in pre-reading mammograms seems a promising new working strategy.

Introduction

An increasing demand for diagnostic imaging tests and an increasing shortage of radiologists have led to studies evaluating the possibility of radiologic technologists taking over duties previously reserved for radiologists. This so-called "skill mixing" could free the radiologist from uncomplicated routine work, so that they are able to devote more time to specialised and complex examinations¹.

Routine breast imaging in a clinical population consists of clinical breast examination and mammography, which can be extended with an ultrasound examination when indicated. Skill mixing could be applied in this diagnostic work-up through the means of pre-reading of mammographic images. In the pre-reading method, a technologist groups mammograms into two basic categories: mammograms that require further evaluation by a radiologist and mammograms that have either negative or clearly benign findings which would not need further attention of a radiologist².

Although studies have shown that technologists could be as sensitive as radiologists in detecting malignancies in screening mammograms, evaluations regarding the deployment of technologists in reading mammograms in a clinical setting are found to be lacking³. In the Netherlands, hospital radiology departments perform both diagnostic and screening mammograms. Diagnostic examinations are carried out in women with clinical signs and symptoms suggestive for breast cancer, women referred by the Dutch Nationwide Breast Cancer Screening Programme and in women with a personal history of breast cancer. Furthermore, mammograms with a screening nature are performed in asymptomatic women referred for a family history of breast cancer, a genetic predisposition or for reassurance. These screening examinations are not yet part of the Dutch Nationwide Breast Cancer Screening Programme.

As an absolute majority of patients is discharged with a mammogram that is negative for breast cancer, it may be assumed that the workload of a breast radiologist in a clinical setting could be reduced substantially when technologists select the cases that primarily require the attention of a radiologist based on mammographic findings. Moreover, technologists could be potentially helpful in distinguishing mammographic findings that appear clearly benign from those that are suspicious for malignancy. In this strategy, even more patients could be excluded from further assessment by a radiologist, namely those with normal mammograms and those with clearly benign mammographic findings.

In a clinical patient population, mammography is often completed with ultrasonography when indicated. According to evidence-based guidelines in clinical practice⁴, a 30% of women referred for mammography will require additional ultrasonography or fine needle aspiration cytology examination based on their reason for referral, such as a palpable breast mass⁵⁻⁶. As these patients need to be seen by a radiologist anyway, it could be argued to limit the role of breast technologists in pre-reading primarily to patients with no immediate indication for further testing based on their reason for referral.

The main objective of this study was to assess the feasibility of training and employing radiologic technologists in pre-reading mammograms in a clinical patient population. Furthermore, the diagnostic reading performance of two technologists was studied in selecting patients with no mammographic findings as well as patients with clearly benign mammographic findings. Finally, their reading performance was studied in a subgroup of patients with no immediate indication for further work-up.

Methods

Study design

Between January and August 2007, all consecutive women referred for mammography to the radiology department of Maastricht University Medical Center were included in the study. Patients were informed by written brochures and approved the use of their mammogram images for the purpose of this study. In addition, approval from the institutional ethics committee was obtained.

The performance and evaluation of mammograms took place according to routine daily practice. All mammograms were performed on a full-field digital system (Giotto Image FFDM, IMS, Bologna, Italy) and were read on a digital work station (Raffaello Review Workstation, IMS, Bologna, Italy).

The radiologist on duty evaluated the mammogram of each consecutive patient, initiated additional imaging work-up when necessary and recorded the findings in a breast imaging report. An imaging conclusion was expressed as a BI-RADS score which is an internationally accepted grading scale for mammography with an increasing degree of suspicion for malignancy: BI-RADS 0 = need additional imaging evaluation; BI-RADS 1 = negative examination; BI-RADS 2 = benign finding; BI-RADS 3 = probably benign finding; BI-RADS 4 = suspicious abnormality; BI-RADS 5 = abnormality is highly suggestive of malignancy⁷.

In addition, two trained technologists who had full information on patient characteristics and reason for referral, evaluated all mammograms. They had no knowledge of the results of additional breast imaging and were blinded for the evaluations of the other observers. Each technologist registered the mammographic findings on a standardised case report form (CRF). Abnormalities were marked on a schematic representation of the breast in craniocaudal and medio-lateral oblique views. For each breast, an imaging conclusion was classified as a BI-RADS score. Furthermore, it was indicated on the CRF whether the technologist advised any additional work-up.

Readers and training

Two radiologic technologists with more than ten years of experience in performing mammograms and assessing image quality in our radiology department did participate in the study. In a training programme of nine months, they received about 500 hours of specific training in mammogram interpretation under the supervision of breast radiologists. Training consisted of a theoretical part,

including the evaluation of up-to-date literature, instruction in breast anatomy and mammographic features of benign and malignant breast disease, study of pathology and radiology manuals like the BI-RADS Atlas of The American College of Radiology⁷, and attendance of breast cancer conferences. Furthermore, practical training was provided by daily reading of diagnostic mammograms, evaluation of difficult clinical cases with breast radiologists on duty and participation in the regular clinical oncology and pathology meetings. Additionally, the technologists received three days of extensive training at the National Expert and Training Centre for Breast Cancer Screening (NETCB) in Nijmegen, which was based on the national training of screening radiologists. This training consisted of teaching files of mammograms, attendance of assessment clinics, and study of literature. Finally, a one-week practical training was received in the mammography and breast unit of the VieCuri Medical Centre Venlo, a top clinical teaching hospital.

In this study, the mammograms were interpreted by six different radiologists on duty, depending on their working schedule. They were all radiologists of our department and skilled in breast imaging, having a variable 1-25 years of experience in reading mammograms.

Analysis

Based on the actual data from the breast imaging reports of the radiologists on duty and the imaging conclusions given by the technologists, the different pre-reading scenarios were analysed using clinical decision rules. Two different cut-off points of BI-RADS classifications were applied on the data.

First, a cut-off point between BI-RADS 1 (considered negative for analysis) and BI-RADS 0,2-5 (considered positive for analysis) was used to study the performance in distinguishing mammograms with mammographic findings from those without mammographic findings. Mammograms that are defined as positive for analysis are referred to the radiologist for further evaluation. A reference standard for the presence or absence of mammographic findings was used which was determined by the radiologist's evaluation of the mammogram in the breast imaging report. Both benign findings and those suspicious for malignancy were regarded as positive here. Any findings detected by additional breast imaging, but not on the two-view mammogram, were excluded from analysis.

Second, the performance of the technologists in discriminating mammograms with negative and benign findings from those that are probably benign, suspicious or highly suggestive for malignancy, was evaluated by using a cut-off point between BI-RADS 1-2 (considered negative for analysis) and BI-RADS 0,3-5 (considered positive for analysis). Mammograms with BI-RADS classifications 0,3-5 were assumed to be positive for analysis, as further action by imaging, tissue analysis or follow-up, and thus interference by a radiologist, would be recommended⁴. A reference standard for the presence or absence of a malignancy was used which was determined by the pathologic results from core needle biopsies and surgical

excisions during a follow-up of 12 months. Breast cancer status was considered negative when no pathologic condition was reported in the PALGA system. Lobular carcinomas in situ were excluded as malignancies. Multiple foci of cancer in one breast were counted as one malignancy. Pathology data were retrieved from PALGA, a nation wide network and registry of histopathology and cytopathology in the Netherlands, to which all Dutch hospital pathology departments are linked.

Although mammography is the basic imaging modality to evaluate patients with breast abnormalities, some malignant abnormalities may not be visible mammographically. Therefore, in order to exclude the possibility of falsely classifying an imaging examination as false-negative, the mammographic visibility of all pathology-proven malignancies was assessed by an independent radiologist.

The primary outcome measures were sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). As the statistical measures were calculated on patient level, a single BI-RADS category was determined for each woman, based on the most malignant score of the two breasts. All the performance characteristics with 95% confidence intervals (CIs) were calculated for each technologist. Furthermore, a diagnostic odds ratio (DOR) with 95% CIs was calculated. The DOR is a measure for the diagnostic performance of a test, which combines sensitivity and specificity in one measure. A DOR of 1 implies that the test has no discriminatory power at all; the larger the DOR, the better the test discriminates between patients with and without the disease of interest⁸.

Statistical analyses were performed using SPSS 16.0 for Windows (SPSS Inc., Chicago, USA) and Stata 9.0 statistical software package (StataCorp LP, Texas, USA). McNemar's test was used to test differences between sensitivity and specificity of the technologists. Furthermore, in order to test differences between the DORs of the technologists, the bootstrap technique was used⁹. P value less than 0.05 was considered to be statistically significant.

Analysis was performed for the overall clinical patient population as well as for the subgroup of patients with no immediate indication for further work-up.

Results

General

Of 2034 mammograms assessed for eligibility, 35 (2%) were excluded because data were not complete. Another five patients with a proven breast malignancy at the time of examination were excluded as well. Consequently, 1994 mammograms (98%) were included. The mean age of the study population was 51 years (range 20–92).

Diagnostic mammography was performed in 1572 patients (79%), whereas 422 women (21%) were referred for screening mammography. Indications for referral for diagnostic breast imaging were: follow-up of prior breast malignancy (n=533, 27%), including 325 examinations after lumpectomy and 208 after mastectomy, the occurrence of a palpable breast lump (n=524, 26%), other

symptomatic complaints like pain or nipple abnormalities (n=342, 17%), follow-up of a prior benign abnormality (n=92, 5%) and referral by the Dutch Nationwide Breast Cancer Screening Programme (n=81, 4%). Indications for screening mammography were: family history of breast cancer, including BRCA gene mutation (n=319, 16%), and other asymptomatic reasons for referral (n=103, 5%).

The radiologist on duty reported mammographic findings in 717 patients (36%), of which 564 were classified as benign (BI-RADS 2), 75 as probably benign (BI-RADS 3) and 64 as suspicious or highly suggestive for malignancy (BI-RADS 4-5). In 14 patients, a BI-RADS 0 was scored.

In the study population, 93 breast cancers were found in 91 patients, leading to a prevalence of 4.6% (91/1994). Two patients had bilateral breast cancer. Eighty-four of these patients with a malignancy were referred for diagnostic mammography; seven were referred for screening mammography. The histopathological classification of breast cancer included 14 ductal carcinomas in situ, 56 invasive ductal malignancies, 13 invasive lobular malignancies and 8 other invasive malignancies.

Based on their reason for referral, all patients with a palpable breast lump or an abnormal screening mammogram had an immediate indication for ultrasonography (n=605, 30%). Consequently, 1389 patients (70%) were included for subgroup analysis, being the group with the largest potential to be evaluated by pre-reading by technologists in clinical practice. Among this subgroup, mammographic findings were reported by the radiologist on duty in 485 patients (35%) of which 443 were classified as benign (BI-RADS 2), 23 as probably benign (BI-RADS 3) and 12 as suspicious or highly suggestive for malignancy (BI-RADS 4-5). In 7 patients, a BI-RADS 0 was scored. Furthermore, a total of 20 pathology-proven malignancies was found, leading to a prevalence of 1.4% (20/1389) in this subgroup.

Performance in an overall clinical patient population

Table 1 shows the performance of the technologists in reading mammograms in the overall study population (n=1994). The sensitivity in detecting mammographic findings (cut-off point between BI-RADS 1 and BI-RADS 0,2-5 and the radiologist's diagnosis as reference standard) was 98% for both technologists. Technologist 1 scored a higher number of false-positive results than technologist 2 (n=339 and n=285, respectively) at a lower number of true-negative results (n=938 and n=992, respectively). This was illustrated by a significantly lower specificity of technologist 1 (74%) compared to technologist 2 (78%), McNemar P=0.001. Furthermore, technologist 1 scored a lower DOR score (107) than technologist 2 (188).

Table 1: Performance of technologists in reading mammograms in the overall clinical population (n=1994)

	Cut-off point BI-RADS 1-2 ¹		Cut-off point BI-RADS 2-3 ²	
	Technologist 1	Technologist 2	Technologist 1	Technologist 2
TP (n)	699	704	81	83
FP (n)	339	285	335	347
TN (n)	938	992	1568	1556
FN (n)	18	13	10 ⁴	8 ⁵
Total	1994	1994	1994	1994
Se (%)	98 (96-98)	98 (97-99)	89 (81-94)	91 (84-96)
Sp (%)	74 (73-74) ³	78 (77-78) ³	82 (82-83)	82 (81-82)
DOR	107 (67-174)	188 (108-329)	38 (20-73)	47 (23-96)
PPV (%)	67 (66-68)	71 (70-72)	20 (18-21)	19 (18-20)
NPV (%)	98 (97-99)	99 (98-99)	99 (99-100)	99 (99-100)

TP = true-positives; FP = false-positives; TN = true-negatives; FN = false-negatives; Se = sensitivity; Sp = specificity; DOR = diagnostic odds ratio; PPV = positive predictive value; NPV = negative predictive value; numbers in brackets are 95% confidence intervals

¹ radiologist as reference standard

² pathology as reference standard

³ significant difference between technologist 1 and technologist 2 (McNemar, P=0.001)

⁴ overlooked/misinterpreted (n=6); not visible on mammogram (n=4)

⁵ overlooked/misinterpreted (n=5); not visible on mammogram (n=3)

Of the 91 patients with a pathology-proven breast malignancy, technologist 1 detected 81 cases and technologist 2 detected 83 malignancies. Consequently, their sensitivities in detecting malignancies (cut-off point between BI-RADS 1-2 and BI-RADS 0,3-5 and pathology as reference standard) were 89% and 91% respectively. Furthermore, technologist 1 scored 335 and technologist 2 scored 347 false-positive findings, resulting in a specificity of 82% for both technologists. The DOR score was 38 and 47 for technologist 1 and technologist 2, respectively.

Table 2: Cases considered positive for analysis for technologist 1 and 2

	Cases (n)	Malignancies (n)	PPV ¹
Technologist 1			
BI-RADS 0	3	0	0%
BI-RADS 3	326	12	4%
BI-RADS 4-5	87	69	79%
Total	416	81	19%
Technologist 2			
BI-RADS 0	3	0	0%
BI-RADS 3	331	11	3%
BI-RADS 4-5	96	72	75%
Total	430	83	19%

¹ PPV = positive predictive value

Table 3: Disagreement between radiologists and technologists on the presence of mammographic findings (cut-off point between BI-RADS 1 and BI-RADS 0,2-5 and the radiologist’s diagnosis as reference standard, n=1994)

BI-RADS score radiologist and type of finding	Only undetected by technologist 1	Only undetected by technologist 2	Undetected by both technologists	Total
0			-	
High breast density	-	1	1	2
2				
Cyst	3	2	-	5
Fibroadenoma	1	2	-	3
Surgery area	2	2	1	5
Compression artefact	-	-	1	1
Calcifications	4	-	-	4
Fibroglandular density	2	1	1	4
3				
Fibrosis after porth-a-cath	1	-	-	1
Increase of fibroglandular density	1	1	-	2
Total	14	9	4	27

In Table 2, details of all cases considered as positive for analysis are given for both technologists. It is shown that most mammograms were classified as BI-RADS 3, whereas most malignancies were found in BI-RADS category 4-5. Technologist 1 detected 69 cancers in 87 BI-RADS 4-5 cases, leading to a PPV of 79%, whereas technologist 2 detected 72 malignancies in 96 cases, resulting in a PPV of 75%. Due to lower PPVs in BI-RADS 0 and BI-RADS 3 cases, the overall PPV is 19% for both technologists.

In Table 3, the disagreement between the radiologists and technologists on the presence of mammographic findings was analysed. It is shown that four findings were missed by both technologists. Another three were missed by a single technologist and were scored as BI-RADS 3 by the radiologist.

Table 4 shows the characteristics of the false-negative imaging results in detecting malignancies. The ten false-negative imaging results of technologist 1 and the eight false-negative imaging results of technologist 2 (Table 1) presented in a total of twelve patients. In seven patients, the reason for referral was the occurrence of a palpable breast lump. It is shown that six malignancies were missed by both technologists, and another six malignancies were missed by either one of the technologists. In two and four false-negative results, the mammogram was evaluated as negative (BI-RADS 1) by technologist 1 and 2, respectively. Furthermore, the radiologist on duty failed to detect eleven of these twelve cases; the technologists missed the remaining cancer mammographically, although they did advise further imaging through ultrasound examination. In eight cases, the mammographic signs of malignancy were overlooked

or misinterpreted, whereas in the other four patients, the malignancy was mammographically not visible.

Table 4: Characteristics of false-negative results in detection of malignancies (cut-off point between BI-RADS 1-2 and BI-RADS 0,3-5 and pathology as reference standard, n=1994)

Reason for referral	Type of the malignancy ¹	Age	Missed by techn. 1	Missed by techn. 2	Missed by the radiologist	Ultrasound advised	Classification ³
Palpable lump	IDC	42	Yes	Yes	Yes	Yes	A
Palpable lump	IDC	45	Yes	Yes	No	Yes	A
Palpable lump	ILC	56	Yes	No	Yes	Yes	A
Palpable lump	IDC	46	Yes	No	Yes	Yes	B
Palpable lump	IDC	45	Yes ²	No	Yes	Yes	A
Palpable lump	IDC	35	No	Yes ²	Yes	Yes	A
Palpable lump	DCIS	50	No	Yes ²	Yes	Yes	A
Family history	Medullary carcinoma	37	Yes ²	No	Yes	Yes	A
Nipple retraction	DCIS	69	Yes	Yes ²	Yes	Yes	B
Nipple discharge	DCIS	66	Yes	Yes ²	Yes	Yes	B
Nipple discharge	DCIS	76	Yes	Yes	Yes	Yes	B
Follow-up after malignancy	IDC	52	Yes	Yes	Yes	No	A

¹ IDC = invasive ductal carcinoma; ILC = invasive lobular carcinoma; DCIS = ductal carcinoma in situ

² mammogram evaluated as negative (BI-RADS 1)

³ classification: A = mammographic signs of malignancy were overlooked or misinterpreted; B = malignancy was mammographically not visible

Performance in a subgroup of patients with no immediate indication for further work-up

Table 5 shows the performance of the technologists in pre-reading mammograms in patients with no indication for further work-up based on their reason for referral. The sensitivity in detecting mammographic findings (cut-off point between BI-RADS 1 and BI-RADS 0,2-5 and the radiologist's diagnosis as reference standard) was 97% for technologist 1 and 98% for technologist 2. Furthermore, technologist 1 had a significantly lower specificity of 75% compared to 79% for technologist 2, resulting in a lower DOR score (107) compared to technologist 2 (196).

The sensitivity in detecting malignancies (cut-off point between BI-RADS 1-2 and BI-RADS 0,3-5 and pathology as reference standard) was 75% and 80% for technologist 1 and 2, respectively, whereas a specificity of 88% and 87% was found for technologist 1 and 2, respectively.

Table 5: Performance of technologists in reading mammograms in a subgroup of patients with no immediate indication for further work-up (n=1389)

	Cut-off point BI-RADS 1-2 ¹		Cut-off point BI-RADS 2-3 ²	
	Technologist 1	Technologist 2	Technologist 1	Technologist 2
TP (n)	472	476	15	16
FP (n)	229	192	161	175
TN (n)	675	712	1208	1194
FN (n)	13	9	5 ⁴	4 ⁵
Total	1389	1389	1389	1389
Se (%)	97 (96-98)	98 (97-99)	75 (54-89)	80 (59-92)
Sp (%)	75 (74-75) ³	79 (78-79) ³	88 (88-88)	87 (87-87)
DOR	107 (61-188)	196 (101-382)	23 (8-60)	27 (9-79)
PPV (%)	67 (66-68)	71 (70-72)	9 (6-10)	8 (6-10)
NPV (%)	98 (97-99)	99 (98-99)	99 (99-100)	99 (99-100)

TP = true-positives; FP = false-positives; TN= true-negatives; FN = false-negatives; Se = sensitivity; Sp = specificity; DOR = diagnostic odds ratio; PPV = positive predictive value; NPV = negative predictive value; numbers in brackets are 95% confidence intervals

¹ radiologist as reference standard

² pathology as reference standard

³ significant difference between technologist 1 and technologist 2 (McNemar, P=0.003)

⁴ overlooked/misinterpreted (n=2); not visible on mammogram (n=3)

⁵ overlooked/misinterpreted (n=1); not visible on mammogram (n=3)

Discussion

The two radiologic technologists in this study were successful in pre-reading mammograms in a clinical patient population by separating mammograms without mammographic findings from those with mammographic findings. With a sensitivity of 98%, the two technologists detected 699 and 704 of the 717 mammographic findings (BI-RADS 0,2-5) that were scored by the radiologist. Moreover, the technologists were able to identify 89% and 91% of all breast cancers that were found in this study population.

When excluding the proportion of patients that would need immediate attention by a radiologist, based on an indication for further imaging based on their reason for referral, the sensitivity of detecting mammographic findings remained a comparable 97% and 98%, respectively. However, the detection rate of malignancies in this subgroup decreased to 75% and 78%, respectively. It should be noted that the breast cancer prevalence in the subgroup analysis (1.4%) was much lower than for the overall patient population (4.6%), which indicates that the group of patients referred directly to the radiologist for additional imaging is indeed the group with a higher risk of breast cancer.

These results suggest that the employment of technologists in pre-reading of mammograms can be an effective innovation that seems to be feasible in clinical practice.

Strength of this study is the prospective design with a standardised data collection. Study monitoring by the principle investigator did motivate the observers to complete the case record forms which resulted in very few missing data. Furthermore, as a digital mammography system was used with a workstation dedicated to this study only, the technologists could evaluate the images at any time without being impeded by logistic problems.

Furthermore, both technologists performed very well in this study, which can be expounded further. First, the two technologists both have more than ten years of experience in performing mammograms in clinical practice, in which they review the images for technical quality and attend the radiologist when evaluating the mammograms. It may be assumed that these working procedures have increased their knowledge and their ability to recognise mammographic abnormalities.

Secondly, it may be concluded that the nine-month training programme that was offered, has been an effective way of increasing their knowledge and skills. In the literature, several studies evaluated the effect of training on the performance of technologists in the interpretation of screening mammograms¹⁰⁻¹⁴. They showed an increased sensitivity and specificity of the mammogram readers after a training programme varying between two weeks^{11,14} and six months¹². In the present study, the skills and performance of the technologists in evaluating mammograms were measured before and after the training period using a set of 108 digital mammograms including 27 benign abnormalities and 27 malignancies. The overall performance of the technologists, expressed as a DOR score, increased significantly from 3 and 7 before the training programme to 12 and 20 after training for technologist 1 and 2, respectively, by using a cut-off point between BI-RADS 1 and BI-RADS 0,2-5. In the detection of malignancies (cut-off point between BI-RADS 1-2 and BI-RADS 0,3-5), the DOR increased significantly from 7 to ∞ for technologist 1 and from 16 to 22 for technologist 2.

These findings do suggest that this new working procedure will function best when the technologists are well-experienced in their current work and much attention is given to the design of an effective training programme.

The overall prevalence of breast cancer in this population was 4.6% using a follow-up of 12 months, and is somewhat lower compared to prevalences reported in other studies in consecutive clinical populations. However, differences in patient inclusion criteria and length of follow-up may explain the differences in prevalence¹⁵⁻¹⁸.

Some limitations of this study need to be addressed as well. First, despite the effectiveness of the pre-reading procedures as suggested, in clinical practice a continuous availability of the radiologist on duty has to be guaranteed as the

proportion of mammograms that will need to be re-evaluated after a first interpretation by the technologist is unpredictable. This implies that radiologists need to stay around and should be involved in other activities that may be interrupted easily. Possibly, scheduling patients' appointments according to their reason for referral may facilitate the planning of additional work-up such as ultrasound examination that needs the involvement of the radiologist. Also, by expanding the activities of the technologist with the independent performance of ultrasonography could potentially improve efficiency of the involvement of technologists in breast imaging even more. In the USA, specialised technologists are certified to perform breast ultrasonography and in our medical centre, a physician assistant has been trained to perform breast ultrasound examinations. However, in the Netherlands, structural application of ultrasonography by radiologic technologists still falls outside the legal scope of practice. Therefore, in order to make this innovation feasible in clinical practice, adaptation of legislation could be considered.

Another limitation of the study was the choice of the reference standard for the cut-off point between BI-RADS 1 and BI-RADS 0,2-5, which was generated from the mammogram diagnosis of the radiologist. Although this standard is subjective and a pathological confirmation of these benign diagnoses would be preferable, histological confirmation of benign findings is no common clinical practice and would be unethical to execute for the purpose of this study only. Therefore, it was decided that the radiologist's diagnosis is an acceptable alternative.

Furthermore, the technologists were regularly present at the performance of the mammographic images as part of their regular duties in the department. In these cases, their prior knowledge of the patient and mammogram characteristics may have biased the evaluation of the images. However, work schedules were registered in the study and further analysis showed that there were no differences in performance between cases where the technologist had been present and had not been present at the time of construction (results are not shown).

The subgroup analysis of patients without an immediate indication for further testing showed that the technologists were well capable to select the patients with mammograms classified as BI-RADS 1, which did lead to high sensitivity levels. However, in this scenario the number of examinations that were classified as positive, was high, 707 for technologist 1 and 675 for technologist 2, and will obviously include a large number of mammograms with benign findings, which do not need additional imaging follow-up requiring a radiologist.

Therefore, it could be argued to deploy technologists in the selection of patients with mammograms classified as BI-RADS 1 and 2. Based on these data, however, the technologists would then miss a substantial number of five and four pathology-proven malignancies, leading to a sensitivity of 75% and 80%, respectively, which is lower compared to the mean sensitivity of 90% in the overall study population. However, excluding four false-negative results which

are mammographically not visible, leads to a mean sensitivity of 94% in the overall study population compared to 93% in the subgroup. Furthermore, it needs to be mentioned that in 11 of 12 patients with false-negative imaging results, an additional ultrasound examination was recommended by the technologists.

Concluding, the results of this study suggest that pre-reading of mammograms by technologists in a clinical setting can be an effective working strategy in daily practice. A nine-month training programme has shown to be effective in acquiring mammography interpretation skills. Nevertheless, despite the high reading performance results, the position of pre-reading in clinical practice remains indistinct as a continuous availability of radiologists still needs to be guaranteed. However, as a substantial proportion of mammograms could be evaluated without the attention of a radiologist, the employment of technologists in pre-reading mammograms seems a promising new working strategy.

References

1. Friedenberg RM. The role of the supertechnologist. *Radiology* 2000; 215:630-633
2. Sumkin JH, Klamon HM, Graham M, et al. Prescreening mammography by technologists: a preliminary assessment. *AJR Am J Roentgenol* 2003; 180:253-256
3. van den Biggelaar FJ, Nelemans PJ, Flobbe K. Performance of radiographers in mammogram interpretation: a systematic review. *Breast* 2008; 17:85-90
4. Zonderland HM, Van Veghel T, Van Asperen CJ, et al. Richtlijn Screening en diagnostiek van het mamma carcinoom. Utrecht: CBO, 2007
5. Flobbe K, Bosch AM, Kessels AG, et al. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. *Arch Intern Med* 2003; 163:1194-1199
6. Vercauteren LD, Kessels AG, van der Weijden T, et al. Clinical impact of the use of additional ultrasonography in diagnostic breast imaging. *Eur Radiol* 2008; 18:2076-2084
7. D'Orsi CJ, Bassett LW, Berg WA, et al. Breast Imaging Reporting and Data System: ACR BI-RADS-Mammography. Reston (VA): American College of Radiology (ACR), 2003
8. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; 56:1129-1135
9. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall, 1993
10. Bassett LW, Hollatz-Brown AJ, Bastani R, Pearce JG, Hirji K, Chen L. Effects of a program to train radiologic technologists to identify abnormalities on mammograms. *Radiology* 1995; 194:189-192
11. Pauli R, Hammond S, Cooke J, Ansell J. Radiographers as film readers in screening mammography: an assessment of competence under test and screening conditions. *The British Journal of Radiology* 1996; 69:10-14
12. Haiart DC, Henderson J. A comparison of interpretation of screening mammograms by a radiographer, a doctor and a radiologist: results and implications. *Br J Clin Pract* 1991; 45:43-45
13. Hillman BJ, Fajardo LL, Hunter TB, et al. Mammogram interpretation by physician assistants. *AJR Am J Roentgenol* 1987; 149:907-912
14. Pauli R, Hammond S, Cooke J, Ansell J. Comparison of radiographer/radiologist double film reading with single reading in breast cancer screening. *Journal of Medical Screening* 1996; 3:18-22
15. Zonderland HM, Coerkamp EG, Hermans J, van de Vijver MJ, van Voorthuisen AE. Diagnosis of breast cancer: contribution of US as an adjunct to mammography. *Radiology* 1999; 213:413-422
16. Flobbe K, van der Linden ES, Kessels AG, van Engelshoven JM. Diagnostic value of radiological breast imaging in a non-screening population. *Int J Cancer* 2001; 92:616-618
17. Flobbe K, Nelemans PJ, Kessels AG, Beets GL, von Meyenfeldt MF, van Engelshoven JM. The role of ultrasonography as an adjunct to mammography in the detection of breast cancer. a systematic review. *Eur J Cancer* 2002; 38:1044-1050
18. Zonderland HM, Pope TL, Jr., Nieborg AJ. The positive predictive value of the breast imaging reporting and data system (BI-RADS) as a method of quality assessment in breast imaging in a hospital population. *Eur Radiol* 2004; 14:1743-1750

4

Costs and effects of using specialised breast technologists in pre-reading mammograms in a clinical patient population

F.J.H.M. van den Biggelaar, A.G.H. Kessels,
J.M.A. van Engelshoven, K. Flobbe

*Accepted for publication in
International Journal of Technology Assessment in Health Care*

Abstract

The objective of this study was to assess the costs and effects of using specialised breast technologists in pre-reading mammograms in order to reduce the increasing workload of radiologists in daily clinical practice.

A total of 1389 mammograms of consecutive patients were evaluated by two technologists trained in mammogram interpretation. The costs and effects of four different experimental strategies of pre-reading mammograms by technologists were analysed by decision analytic modelling and compared to the conventional strategy of standard evaluation by the radiologist on duty.

Overall, the employment of technologists in this patient population resulted in a potential time saving up to 73% (1019/1389) for the radiologist. No additional false-negative imaging results were found as compared to the conventional strategy.

The total diagnostic costs in the conventional strategy were determined at €150,602. The experimental strategies resulted in cost savings up to 17.2% (range €122,494–€139,781).

It can be concluded that the employment of technologists in pre-reading mammograms in a clinical patient population could be effective to reduce the workload of radiologists without jeopardising the detection of malignancies. Furthermore, diagnostic costs can be reduced considerably.

Introduction

In order to reduce the increasing workload of radiologists, the employment of radiologic technologists taking over duties previously reserved for radiologists is an upcoming solution.

In breast imaging, mammography is the most widely used modality for early detection and diagnosis of breast malignancies. Deployment of specialised breast technologists in the interpretation of mammograms provides the radiologist the opportunity to devote more time to specialised and complex breast examinations. This is expected to provide cost savings to the health care system. Technologists could be involved in two different ways, through double reading and pre-reading procedures.

In the Dutch Nationwide Breast Cancer Screening Programme, mammograms are routinely evaluated by two radiologists. Studies have shown that double reading of screening mammograms by both a radiologist and a technologist could also increase the number of cancers detected¹⁻⁴. With the method of pre-reading mammograms, a technologist selects those examinations that require further evaluation by a radiologist and discharges patients with either negative or clearly benign findings from further evaluation. Haiart et al.⁵ showed that pre-reading could not be justified in screening mammograms, neither in terms of performance, nor on economic grounds. However, all studies involving technologists in interpreting mammograms have been performed in a screening setting. Evaluations regarding the deployment of technologists in reading mammograms in a clinical patient population with both diagnostic and screening examinations are found to be lacking⁶.

In daily clinical practice, there has been a considerable increase in demand for radiologic services, which has not been met by a commensurate increase in radiologist staffing. Furthermore, it has been demonstrated that, in the Netherlands, more than half the patients who undergo mammography in a clinical setting, are discharged without further imaging work-up⁷⁻⁸. These two facts provide a strong basis for the deployment of technologists in pre-reading mammograms.

According to evidence-based guidelines in clinical practice⁹, about 30% of women referred for mammography will require additional ultrasonography, fine needle aspiration cytology examination or core needle biopsy based on their reason for referral, such as a palpable breast mass⁷⁻⁸. It can be argued to exclude these patients from a pre-selection interpretation by a technologist, as they need to be seen by a radiologist anyway. In all other patients referred for mammography to a clinical radiology department, technologists are expected to be useful in selecting the cases that require further evaluation. This method could reduce the workload of the breast radiologist substantially and could provide cost savings by reducing the time they require to review large numbers of negative mammograms.

The purpose of this study was to assess the costs and effects of deploying specialised breast technologists in pre-reading mammograms in a clinical

population without an immediate indication for further testing at referral. Four different experimental pre-reading strategies were analysed in terms of costs and effects by decision analytic modelling and were compared to the conventional strategy of standard mammogram evaluation by a radiologist.

Methods

Study design

A decision analytic model was designed to compare a conventional strategy of mammogram evaluation by a radiologist with four different experimental strategies of pre-reading mammographic images by breast technologists. The model was based on a large prospective study in which all consecutive mammograms of patients referred to the radiology department of Maastricht University Medical Center between January and August 2007 were included. Patients were informed on the study by written brochures and approved the use of their mammogram images for the purpose of this study. The institutional ethics committee approved the study.

For the purpose of this cost analysis, all patients referred for a palpable breast mass and patients referred with an abnormal mammogram from the Dutch Nationwide Breast Cancer Screening Programme, were excluded, as, according to evidence-based guidelines, immediate additional ultrasound examination is recommended, which requires the input of a radiologist⁸.

All mammograms were performed on a full-field digital system (Giotto Image FFDM, IMS, Bologna, Italy) and read on a digital workstation (Raffaello Review Workstation, IMS, Bologna, Italy), conforming daily clinical practice. Of each consecutive patient, the radiologist on duty evaluated the examination and recorded the findings in a breast imaging report. An imaging conclusion was given as a BI-RADS score which is based on a grading reporting scale for mammography with an increasing degree of suspicion for malignancy: 0 = need additional imaging evaluation; 1 = negative examination; 2 = benign finding; 3 = probably benign finding; 4 = suspicious abnormality; 5 = highly suggestive of malignancy¹⁰. Furthermore, additional imaging work-up was initiated when necessary.

In addition, all mammograms were independently evaluated by two technologists, trained in mammogram interpretation during a 500-hour training programme under the supervision of breast radiologists. They had full information on the patient characteristics and reason for referral, but had no information on the results of additional imaging tests and were blinded for the evaluations of the other observers. Each technologist registered the mammographic findings on a standardised case report form. Abnormalities were marked on a schematic representation of the breast in craniocaudal and mediolateral oblique views. Furthermore, for each breast a BI-RADS score was assigned. Finally, it was indicated on the case report form whether the technologist advised additional work-up.

The reference standard for the presence or absence of breast cancer was determined by the pathologic results from core needle biopsies and surgical excisions during a follow-up of 12 months. Pathology data were retrieved from PALGA, a nation wide network and registry of histopathology and cytopathology in the Netherlands, to which all Dutch hospital pathology departments are linked. Breast cancer status was considered negative when no pathologic condition was reported in the PALGA system within one year. Lobular carcinomas in situ were excluded as malignancies.

Strategies and data sources

Four different experimental strategies of pre-reading mammograms by breast technologists were analysed and compared to a conventional strategy of mammogram evaluation by the radiologist on duty ($n = 6$ well-experienced radiologists). The design of the decision model is shown in Figure 1, in which the different strategies are outlined and each possible clinical pathway is indicated with a branch number (1-22).

In the conventional strategy, the numbers of patients in each pathway (1-4) were retrieved from actual data from daily clinical practice. According to clinical guidelines¹⁰, patients were either discharged (BI-RADS 1-2) or referred for additional work-up (BI-RADS 0,3-5).

In the experimental strategies, the distribution of patients among the different pathways (5-22) was determined by applying clinical decision rules based on the BI-RADS classification. Strategy 1 represents mammogram evaluation by one technologist, who discharged patients with no mammographic findings (BI-RADS 1), and referred all patients with any mammographic findings (BI-RADS 2-5) or requiring additional imaging (BI-RADS 0) to the radiologist. The strategy is subdivided into 1a and 1b, representing the results of the two separate technologists participating in the study.

In experimental strategy 2, two technologists read all mammograms independently. When mammographic findings (BI-RADS 2-5) were reported by at least one technologist or at least one technologist recommended additional imaging evaluation (BI-RADS 0), the mammogram was referred to the radiologist for re-evaluation. Patients without mammographic findings (BI-RADS 1) were discharged.

Summarising, in strategy 1 and 2 a cut-off point between BI-RADS 1 and 2 was used for the decision whether to refer patients to the radiologist. Additionally, in strategy 3 and 4, a cut-off point between BI-RADS 2 (benign finding) and 3 (probably benign finding) was used for the decision whether to refer. Furthermore, the assumptions regarding the different readers were similar in strategy 3 and 4 to those in strategy 1 and 2 respectively, resulting in strategy 3a and 3b reporting the results of the two separate technologists, and strategy 4 reporting the results of both technologists.

In all strategies, it was assumed that the classification of a technologist would never overrule the classification of the radiologist.

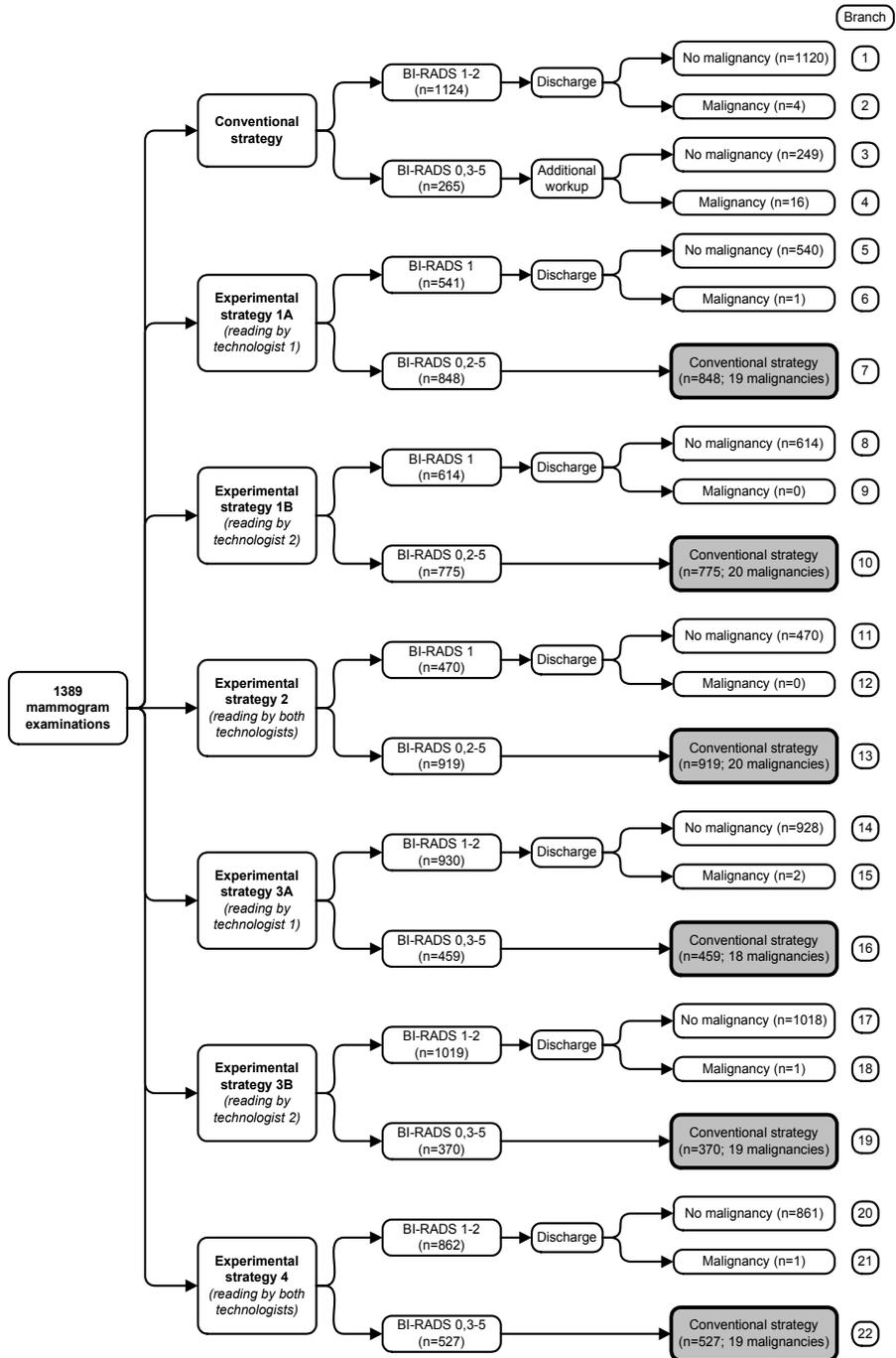


Figure 1: Pre-reading strategies by technologists

Cost data

The costs of all radiological breast procedures were acquired according to 2008 national reimbursement rates which contain hospital facility charges and a fee for the physician(s) involved¹¹.

In the conventional strategy, the costs of a mammogram were equal to the reimbursement rate for mammography (€86.80), consisting of €59.20 for hospital facility charges and €27.60 for radiologists' fee. To assess the costs of the evaluation of a mammogram by a technologist, equal hospital facility charges were assumed and physician's fees were replaced by a proportional fee for technologists. According to the Dutch guidelines for economic evaluations¹², an hourly rate of €160 for a physician and €33 for a technologist could be counted. Assuming an equal time unit needed for the evaluation of one mammogram, the costs for a technologist evaluating a mammogram were determined at €5.70 ($(€33 / €160) * €27.60$). Consequently, in strategy 1 and 3 the total costs of a mammogram evaluated by one technologist followed by discharge were established at €64.90 (€59.20 + €5.70). When the technologist recommends a re-evaluation by the radiologist, the costs of a mammogram increase with the fee for the radiologist, leading to €92.50 (€59.20 + €5.70 + €27.60).

In strategy 2 and 4, two technologists are involved, resulting in costs of a mammogram of €70.60 (€59.20 + €5.70 + €5.70) in case of discharge of the patient, and €98.20 (€59.20 + €5.70 + €5.70 + €27.60) when referring the patient to a radiologist. Costs of additional mammograms in patients with recommendation for short-term follow-up of a probably benign lesion were charged at the reimbursement rate for mammography (€86.80). In Table 1, an overview is given of the different cost items.

Table 1: Costs of diagnostic procedures (in euros)

Diagnostic procedure	Costs (€)
Mammography	
By radiologist (conventional strategy and short-term follow-up)	86.80
By one technologist (strategy 1 and 3, discharging patient)	64.90
By two technologists (strategy 2 and 4, discharging patient)	70.60
By one technologist and a radiologist (strategy 1 and 3, re-evaluation by radiologist)	92.50
By two technologists and a radiologist (strategy 2 and 4, re-evaluation by radiologist)	98.20
Ultrasound examination	84.00
Fine needle aspiration cytology	114.40
Core needle biopsy	164.20

National reimbursement rates from ctg.bit-ic.nl/Nzatarieven/top.do

Furthermore, treatment costs for breast cancer were assessed. Probabilities and direct medical costs of surgery, adjuvant therapy, local recurrence, palliative care and follow-up care in disease-free patients were adapted from Flobbe et al.¹³. In Figure 2, probabilities and costs are shown for all branches in a decision

tree for surgery of lesions. Surgery of a lesion could reveal a true-positive (probability of 0.983) or false-positive imaging result (probability of 0.017). Further distinction was made between cases with a bad prognosis leading directly to palliative care, surgery with and without adjuvant chemotherapy and hormonal therapy, and occurrence of a local or systemic recurrence. All costs in the decision tree are presented as year-2003 euros, whereas the mean costs of a malignancy were discounted at an annual rate of 4% and presented in year-2008 euros as well. Only direct medical costs were calculated; direct non-medical costs and indirect costs were excluded from analyses. The mean costs of a malignancy resulted in €17,565 (year-2008 euros).

For the purpose of this study, it was assumed that the costs of malignancy resulting from the discharge groups (pathways 6, 9, 12, 15, 18 and 21) are comparable to the costs of malignancies resulting from the patients re-evaluated by the radiologist (pathways 7, 10, 13, 16, 19 and 22).

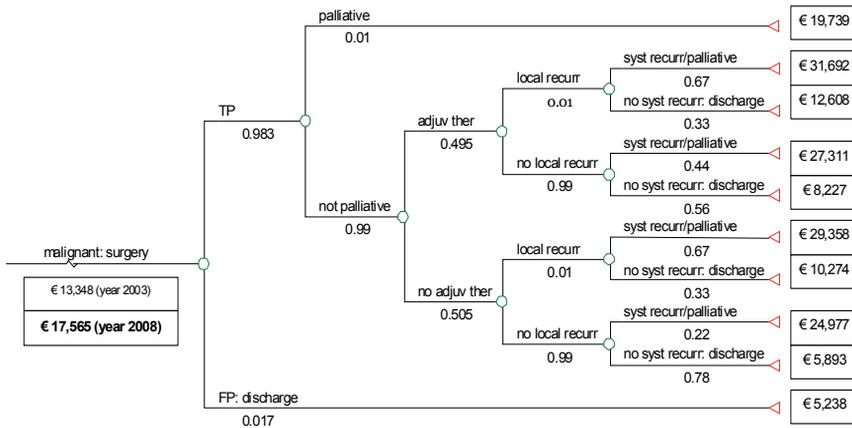


Figure 2: Decision tree for surgery of malignancies

TP=true-positive; FP=false-positive; palliative=palliative care; adjuv ther=adjuvant chemotherapy; recurr=recurrence; syst recurr=systemic recurrence
 adapted from Flobbe, et al., 2004

Outcome measures

The main outcome measure was the number of malignancies failed to be detected by the technologists in each experimental strategy, as compared to the number of malignancies failed to be detected in the conventional strategy (false-negative imaging results).

Furthermore, for the cost analysis the main outcome measures included the total costs of each experimental strategy and the proportion of costs in relation to the conventional strategy.

Threshold analysis

In order to evaluate the influence of the personnel costs of the technologist on the cost-effectiveness of the different experimental strategies, a threshold analysis was performed. In each experimental strategy, the hourly rate for the radiologist was assumed to be constant (€160), whereas the rate for the technologist was raised to identify the values at which total costs of the experimental strategy was equal to the total costs of the conventional strategy.

Results

General

In the prospective clinical study that formed the basis of this decision analysis, 2034 consecutive mammograms were assessed for eligibility. Thirty-five exams (2%) were excluded because data were not complete. Another five patients with a proven breast malignancy at the time of performing the study mammogram were excluded as well. Furthermore, patients were excluded from the present cost-effectiveness study because they were referred for a palpable breast mass (n=524) or an abnormal screening mammogram (n=81). Consequently, 1389 patients were analysed in this study.

The mean age of the study population was 53 years (range 21–90). Diagnostic mammography was performed in 967 patients (70%), whereas 422 women (30%) were referred for screening mammography. Indications for referral for diagnostic breast imaging were: follow-up of prior breast malignancy (n=533, 38%), including 325 examinations after lumpectomy and 208 after mastectomy, symptomatic complaints like pain or nipple abnormalities (n=342, 25%) and follow-up of a prior benign abnormality (n=92, 7%). Indications for screening mammography were: family history of breast cancer, including BRCA gene mutation (n=319, 23%), and other asymptomatic reasons for referral (n=103, 7%).

After a follow-up of 12 months during which pathology results from core needle biopsies and surgical excisions were retrieved, a total of 20 malignancies have been detected, leading to a breast cancer prevalence of 1.4% (20/1389) in this population.

Effects

In the decision model represented in Figure 1 (page 46), the distribution of patients over the different pathways in the decision model is shown. In the conventional strategy of mammogram evaluation by the radiologist on duty (branch 1-4), additional diagnostic work-up was performed in 265 patients (19%), whereas 1124 patients (81%) were discharged. Four malignancies were detected in the group of patients who were discharged initially (branch 2), and returned for further imaging or surgical procedure in a later stage of the follow-up (after 2, 4, 7 and 12 months, respectively). Furthermore, 16 malignancies were detected in the group referred for additional work-up (branch 4).

In experimental strategy 1 (branch 5-10), 541 (39%) and 614 (44%) patients were discharged by the technologists in strategy 1a and 1b, respectively. Another 848 (61%) and 775 (56%) patients were referred to the radiologist. Using the BI-RADS classifications of both technologists in experimental strategy 2 (branch 11-13) resulted in a discharge of 470 patients (34%) and re-evaluation by the radiologist in 919 patients (66%). Using a cut-off point between BI-RADS 2 and 3, the patients referred to the radiologist, decreased to 459 (33%) in strategy 3a (branch 14-16), 370 (27%) in strategy 3b (branch 17-19) and 527 (38%) in strategy 4 (branch 20-22).

In the experimental strategies 1a, 3a, 3b and 4, a few patients were discharged who were proven to have breast cancer later (branch, 6, 15, 18 and 21). However, the malignancies that were not detected in the experimental strategies were also not identified in the conventional strategy. Therefore, no additional false-negative results were reported in the experimental strategies compared to the conventional strategy.

Costs

Table 2 shows the diagnostic procedures and costs of each strategy. For each experimental strategy, the total number of patients referred for further evaluation by the radiologist and the total number of patients discharged are shown. Furthermore, the total number of each diagnostic procedure performed is given per strategy.

In the conventional strategy, a total of 1399 mammograms (1389 initial and 10 short-term follow-up mammograms), 268 ultrasonography examinations, 18 fine needle aspiration cytology and 28 core needle biopsies were performed. In the experimental strategies, the number of diagnostic procedures was lower, due to the discharge of many patients after mammography. In particular, the number of ultrasound examinations decreased in the experimental strategies (range 184–236) as compared to the conventional strategy (268 examinations). The total diagnostic costs in the conventional strategy were determined at €150,602. Strategies 1a, 1b, 3a, 3b and 4 were cheaper than the conventional strategy (range €122,494–€139,781), because fewer mammogram re-evaluations and fewer ultrasound examinations by a radiologist were needed. Strategy 2 was slightly more expensive (€150,612), which was caused by a relatively high number of patients referred to the radiologist for further evaluation as well as the high costs for personnel, as two technologists were involved in the pre-selection procedures.

The total number of malignancies was equal in each strategy (n=20). As no differences were assumed between the breast cancers in the different strategies in terms of staging, prognosis and surgical interventions, the total costs of breast cancer treatment resulted in €351,300 for each strategy. As these costs were no determinants of cost-effectiveness, they were not included in Table 2.

Table 2: Diagnostic procedures and costs

	Convent.	Exp. 1a	Exp. 1b	Exp. 2	Exp. 3a	Exp. 3b	Exp. 4
		<i>Mean</i>			<i>Mean</i>		
Diagnostic procedures(n)							
Mammogram by radiologist	1389	-	-	-	-	-	-
Mammogram by technologist(s)	-	541	614	470	930	1019	862
Mammogram by radiologist and technologist(s)	-	848	775	919	459	370	527
Mammogram for short-term follow-up	10	10	10	10	10	8	10
Ultrasound examination	268	226	214	236	203	184	216
Fine needle aspiration cytology	18	17	17	18	16	15	17
Core needle biopsy	28	27	27	27	26	26	27
Costs (€)							
Total costs per strategy	150,602	139,781	136,758	150,612	126,834	122,494	137,999
		<i>138,270</i>			<i>124,664</i>		
Difference compared with conventional strategy	-	10,821	13,844	-10	23,768	28,108	12,603
		<i>12,332</i>			<i>25,938</i>		
% of conventional strategy	-	7.2	9.2	0	15.8	18.7	8.4
		<i>8.2</i>			<i>17.2</i>		

Threshold analysis

Table 3 shows the results of the threshold analysis in which the hourly rates for technologists were varied, given a constant hourly rate for the radiologist of €160. The total costs in experimental strategy 2 were equal to the total costs in the conventional strategy, resulting in an hourly rate of the technologist of €33 as threshold value, which was 21% of the hourly rate of the radiologist. The costs of experimental strategy 4 were equal to the costs of the conventional strategy at an hourly rate of €59 (37% of the rate of the radiologist), whereas the costs of experimental strategy 1 were equal at a rate of €84 (53%). In experimental strategy 3, the threshold value was determined at €141 (88% of the hourly rate of the radiologist).

Table 3: Results of threshold analysis

Experimental strategy	Hourly rate technologist (€)	% of hourly rate radiologist (€160)
2	33	21%
4	59	37%
1	84	53%
3	141	88%

Discussion

This study shows that the employment of specialised breast technologists in the pre-selection of mammograms in a clinical patient population can be an effective tool, which reduces diagnostic costs. Although the number of diagnostic procedures was decreased, no additional false-negative results were recorded in the experimental strategies compared to the conventional strategy. Savings are largest in the strategy in which one technologist separates the normal and benign mammograms (BI-RADS 1-2) from those with abnormalities that are probably benign, suspicious or highly suggestive for malignancy (BI-RADS 3-5) or examinations requiring additional imaging (BI-RADS 0), leading to a cost reduction of 17.2% compared to the conventional strategy. Using two technologists in pre-reading mammograms was only cost-effective in strategy 4 (cut-off point between BI-RADS 2 and 3), resulting in cost savings of 8.4%.

Strength of this cost-analysis is that it was based on empirical data collected in a prospective, clinical study. Radiologists on duty evaluated all 1389 mammograms according to daily clinical practice. Additionally, the two technologists in this study read the images in another room on a separate workstation, but under similar viewing conditions as the radiologists.

The results showed that in 470 to 1019 patients (in strategy 2 and 3b respectively), the technologists decided that re-evaluation of the mammogram by a radiologist would not be necessary, resulting in a time saving up to 73% (1019/1389) for the radiologist in this patient population. Furthermore, some of the additional imaging tests that are performed regularly in the conventional strategy were not recommended in the experimental strategies. The number of advised ultrasound examinations was 268 in the conventional strategy compared to 184-236 in the experimental strategies. The number of additional mammograms, fine needle aspiration cytology and biopsies would be reduced by 1-3 in different strategies. Although it should be noted that the quality of care in these patients could be affected seriously as potential pathology may be missed, no breast cancers were found in these specific group of mammograms and the effect on the quality of care seems to be minimal here.

However, in experimental strategies 1a, 3 and 4, the technologists would have discharged patients that were diagnosed with breast cancer at a later stage. As these malignancies were also not diagnosed by mammography in the conventional strategy, these procedures would not lead to further delay in diagnosis in daily clinical practice. Although there is concern that delay by providers in the diagnosis of breast cancer would result in a significant progress of the malignancy, studies also show that this delay has no significant effect on stage, treatment or survival¹⁴⁻¹⁶. Nevertheless, delay in diagnosis does distress patient and clinician, and must be avoided as much as possible. Our data suggest that technologists are well able to assist radiologists in avoiding oversights whereby decreasing the miss rate in detecting malignancies in daily clinical practice.

In order to evaluate the cost savings of pre-reading mammograms by technologists, only integral diagnostic costs were used, including costs for personnel, material, capacity and departmental overhead. As it was assumed that non-medical costs were comparable for all strategies, these were excluded. Furthermore, the number of malignancies was equal in all strategies and it was assumed that the costs of breast cancer treatment were comparable among the strategies. Therefore, these costs were excluded in the calculations of cost savings.

Before the start of the prospective study, the technologists received a 500 hours training programme in mammogram reading during nine months under the supervision of well-experienced breast radiologists. Training consisted of evaluation of up-to-date literature on mammography and anatomy and pathology of the breast, daily reading of diagnostic mammograms and evaluation of difficult clinical cases with a specialised breast radiologist. Furthermore, the technologists participated in pathology and oncology meetings, attended at mammography symposia and received practical and theoretical training in other medical centres. The total costs of the training programme were estimated to be €47,500, including personnel costs of technologists and radiologists, costs for material and costs for symposia and training elsewhere. This must be taken into consideration in the decision to employ technologists in reading mammograms. Since in strategy 3 and 4, the technologists need to discriminate the certainly benign (BI-RADS 2) from the probably benign (BI-RADS 3) lesions, it could be argued that these are also the strategies in which the required performance and confidence of the technologists is highest. Furthermore, it illustrates the need for a good and proper training programme in order to reach this level of expertise and knowledge.

Also, these findings suggest the feasibility of incorporating the performance of breast ultrasonography into the task responsibilities of breast technologists. Doing so, even more patients could be discharged based on a negative mammogram and negative ultrasound examination, which would reduce the workload of the breast radiologist even further. Finally, the mammograms of the groups of patients that were excluded from the current analysis because of their direct indication for breast ultrasonography, could then be seen by technologists as well. However, as in the Netherlands, independent performance of ultrasonography by technologists will fall outside their legal scope of practice, adaptation of legislation should be considered.

Concluding, the results of this study indicate that the employment of breast technologists in pre-reading mammograms in a clinical patient population could be an effective tool to reduce the workload of radiologists without jeopardising the detection of breast malignancies. Beside its effect on the clinical pathways of the patients referred for mammography, diagnostic costs can be reduced considerably.

References

1. Hillman BJ, Fajardo LL, Hunter TB, et al. Mammogram interpretation by physician assistants. *AJR Am J Roentgenol* 1987; 149:907-912
2. Pauli R, Hammond S, Cooke J, Ansell J. Comparison of radiographer/radiologist double film reading with single reading in breast cancer screening. *Journal of Medical Screening* 1996; 3:18-22
3. Tonita JM, Hillis JP, Lim CH. Medical radiologic technologist review: effects on a population-based breast cancer screening program. *Radiology* 1999; 211:529-533
4. Wivell G, Denton ER, Eve CB, Inglis JC, Harvey I. Can radiographers read screening mammograms? *Clin Radiol* 2003; 58:63-67
5. Haiart DC, Henderson J. A comparison of interpretation of screening mammograms by a radiographer, a doctor and a radiologist: results and implications. *Br J Clin Pract* 1991; 45:43-45
6. Van den Biggelaar FJ, Nelemans PJ, Flobbe K. Performance of radiographers in mammogram interpretation: a systematic review. *Breast* 2008; 17:85-90
7. Flobbe K, Bosch AM, Kessels AG, et al. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. *Arch Intern Med* 2003; 163:1194-1199
8. Vercauteren LD, Kessels AG, van der Weijden T, et al. Clinical impact of the use of additional ultrasonography in diagnostic breast imaging. *Eur Radiol* 2008; 18:2076-2084
9. Zonderland HM, Van Veghel T, Van Asperen CJ, et al. Richtlijn Screening en diagnostiek van het mamma carcinoom. Utrecht: CBO, 2007
10. D'Orsi CJ, Bassett LW, Berg WA, et al. Breast Imaging Reporting and Data System: ACR BI-RADS-Mammography. 4th ed. Reston (VA): American College of Radiology (ACR), 2003
11. Nederlandse Zorgautoriteit: National reimbursement rates. <http://ctg.bit-ic.nl/Nzatarieven/top.do>. Date accessed: 19-08-2008
12. Oostenbrink J, Bouwmans C, Koopmanschap M, Rutten F. Handleiding voor kostenonderzoek: methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg. Diemen: College voor zorgverzekeringen, 2004
13. Flobbe K, Kessels AG, Severens JL, et al. Costs and effects of ultrasonography in the evaluation of palpable breast masses. *Int J Technol Assess Health Care* 2004; 20:440-448
14. Hardin C, Pommier S, Pommier RF. The relationships among clinician delay of diagnosis of breast cancer and tumor size, nodal status, and stage. *Am J Surg* 2006; 192:506-508
15. Sainsbury R, Johnston C, Haward B. Effect on survival of delays in referral of patients with breast-cancer symptoms: a retrospective analysis. *Lancet* 1999; 353:1132-1135
16. Tartter PI, Pace D, Frost M, Bernstein JL. Delay in diagnosis of breast cancer. *Ann Surg* 1999; 229:91-96

5

Pre-reading mammograms by specialised breast technologists: legal implications for technologist and radiologist in the Netherlands

F.J.H.M. van den Biggelaar, K. Flobbe,
J.M.A. van Engelshoven, N.P.Y.M. de Bijl

European Journal of Health Law 2009; 16(3): 271-279(9)

Abstract

This paper focuses on the legal implications in terms of duties and responsibilities for radiologists and technologists of independent pre-reading of mammograms by technologists, so patients could be discharged without being seen by a radiologist. Although pre-reading is not a reserved procedure, it could be considered as a potentially risky procedure. Therefore, the criteria for performing reserved procedures by non-authorised professionals as stated in the Individual Health Care Professions Act should be applied.

Furthermore, compliance with a protocol or code of conduct in combination with adequate training and supervision should be sufficient to disprove potential claims.

It is likely that pre-reading mammograms by technologists will expand from a local initiative. For a wide implementation, pre-reading should be well-embedded in legal rules and should answer the professional standard of care. In addition, in order to protect patients, the legislator should consider to formulate general conditions that skill mixing initiatives of both potentially risky procedures as well as reserved procedures should meet.

Introduction

In several countries, staff shortages, rising health care costs and efforts to improve quality of care have led to the utilisation of expertise from individuals in related fields in medicine to complement or increase the expertise available to patients. This so-called "skill mixing" have been evaluated thoroughly in nursing studies¹, highlighting the scope for effective deployment of nurses in advanced roles. In the United Kingdom, the scope of practice for new professional roles like nurse specialists, practice nurses and physician assistants, is well-defined and unambiguous. However, many skill-mix initiatives are less elucidated.

In radiology practice, skill mixing could be effectuated through the deployment of radiologic technologists. In the setting of an increasing demand for diagnostic imaging tests and an increasing shortage of radiologists, trained technologists taking over duties could free the radiologist from uncomplicated routine work, so they are able to devote more time to specialised and complex examinations. This is thought to provide cost savings for managed health care and provide time savings for the radiologists. In the United Kingdom, technologists have been utilised to identify fractures in the emergency department, to perform and interpret ultrasonography, barium enema examinations and upper gastrointestinal examinations, and to double read screening mammograms². In the Dutch Nationwide Breast Cancer Screening Programme, it was shown that mammogram reading by a technologist in addition to standard double reading by two radiologists increased the detection rate of breast malignancies³.

In this paper, attention will be focused on the deployment of specialised breast technologists as pre-readers of mammograms in a clinical patient population. Generally in patients referred to the radiology department of a Dutch hospital, a breast technologist performs a mammogram and checks the technical quality of the image. Following, a radiologist evaluates the images and decides whether additional imaging is indicated. In the concept of pre-reading, a technologist evaluates all mammograms and groups them into two basic categories: cases that require further evaluation and cases that can be discharged, because they have either negative or clearly benign findings. Consequently, patients without a reason for further evaluation by a radiologist, could be discharged by a technologist without being seen by a radiologist, which will considerably reduce the workload of the radiologist. Van den Biggelaar et al.⁴ showed that trained technologists are well capable in pre-reading mammograms and that this strategy seems to be a feasible intervention in daily clinical practice.

However, the practice of independent pre-reading by a technologist raises the question what are the legal implications in terms of duties and responsibilities for both radiologists and technologists.

Therefore, this paper focuses on the issue whether employment of technologists in pre-reading mammograms fits in the current Dutch legal framework. It was studied which legal preconditions are required. Furthermore, the risk of being sued for medical malpractice was discussed for the technologist, radiologist and hospital.

Legal framework

When pre-reading mammograms by technologists is considered, pertinent legislation must be taken into account. In this paragraph, the Individual Health Care Professions Act, the Medical Contract Act and relevant case law are discussed which are most relevant for the transfer of tasks between technologist and radiologist in the Netherlands. Legislation such as the Health Care Institution Quality Act and legislation regarding health finances are discounted.

Individual Health Care Professions Act

In 1997, the Individual Health Care Professions Act (IHCP Act, in Dutch: Wet op de Beroepen in de Individuele Gezondheidszorg) was introduced in the Netherlands. The Act provides conditions necessary for fostering and monitoring the quality of professional practice in the individual health care. In section 3 of the IHCP Act, a system is described to protect the titles of a limited number of professional groups (e.g. physicians, dentists and nurses). These professionals are registered in the IHCP register and could be subject to disciplinary procedures and Codes of Conduct. For registration, they must meet a number of statutory requirements (e.g. finishing training that is required).

Furthermore, in section 34 of the IHCP Act professional title protection of specified other professionals (e.g. radiologic technologists, optometrists and dieticians) is regulated. For these professionals, only requirements for education and the area of professional competence are legislated by the government. They are not listed in a register and no disciplinary procedures can be started against them. However, the professional bodies may institute registration and disciplinary procedures themselves.

The basic principle behind the IHCP Act is to open up the practice of medicine instead of restricting it. Nevertheless, certain procedures (so-called reserved procedures) that could cause serious health risks to patients when performed by individuals lacking suitable professional competences, are restricted to authorised professionals specified in the IHCP Act (e.g. physicians). The reserved procedures consist of fourteen categories, including surgical procedures, punctures and injections, cardioversion and procedures involving the use of radioactive substances and ionising radiation⁵.

According to the Act, it is possible that other professionals than the ones specified in the IHCP Act, perform reserved procedures, using a construction formerly known as the "extended arm" construction. Extended performance is allowed when four conditions, stated in section 38 of the Act, are met. First, the order has to be assigned by a qualified physician. Second, if necessary, an authorised person has to give instructions. Furthermore, the possibility of intervention and supervision by a physician has to be arranged. Finally, both the qualified professional who gives the order, and the unauthorised professional may reasonably assume that the unauthorised person has the proficiency required for performing the procedure properly.

Furthermore, section 39 of the IHCP Act stated that, for some reserved procedures, nurses have a functionally independent status which do not require the condition of possible intervention and supervision by a physician.

Medical Contract Act

The Medical Contract Act (in Dutch: Wet op de Geneeskundige Behandelings-overeenkomst) is part of the Dutch Civil Code (section 7:446 – 7:468) and regulates the rights and obligations of both health care providers and patients in order to strengthen the position of patients. According to this act, the relationship between the doctor and the patient is treated as a civil law relationship. Usually, patients have a medical treatment contract with the health care professional who or institution that should provide medical services directly to the patient or to a specific third party (section 7:446 Civil Code). The Act requires that a professional must provide care of a conscientious health care provider and must act in accordance with the responsibility following from the professional standard (section 7:453 Civil Code).

For a patient, it could be complicated to determine whom they have a contract with and who should be held liable in the event of any deficiency for any treatment carried out in the institution. Therefore, the Medical Contract Act stated that the institution will be jointly liable as if it were itself a party to the contract (section 7:462 Civil Code). Furthermore, health care provider's liability may not be limited or excluded (section 7:463 Civil Code).

Case law

From Dutch case law, it appears that judges take into account both the legal rules and the professional standard of care when they pass their judgement on skill mixing. Judges also consider education, professionalism and individual skills and experience of the professional. In general, certain conditions have to be met when taking over duties. Mentoring, supervision, orders and instructions, clear agreements, comparing information and checking the competence and experience of the unauthorised person are considered highly relevant by judges⁶.

Both the person delegating a task and the person accepting the task are responsible for good care. For instance, judges considered that a general practitioner could not be held responsible for a doctor's assistant who ignored the protocol⁷⁻⁸. On the other hand, a general practitioner is responsible for a well-described and clear protocol, and the general practitioner have to be sure that the instructions are clear to the doctor's assistant⁹⁻¹⁰.

Another judge decided that nurses are allowed to perform triage at an emergency department, although this is highly medical in nature. However, they must be sufficiently qualified and authorized, and possibly be subject to an after-check by a medical practitioner¹¹.

Preconditions for pre-reading mammograms

The evaluation of a mammogram could be regarded as an assessment of the patient's state of health. Although, according to the IHCP Act, this is not a reserved procedure, it could be considered as part of "individual health care procedures" (section 1), which practice is restricted to the area of expertise of a doctor (section 19). Therefore, a technologist who is involved in pre-reading mammograms acts outside the area of expertise of a technologist. In addition, as a malignancy that failed to be detected on the mammogram could cause serious health problems to patients, evaluating mammograms can be regarded as a potentially risky procedure. Therefore, the professional standard of care in this situation requires the delivery of care of high quality. Consequently, when transferring the evaluation of mammograms from radiologists to technologists, the criteria for performing reserved procedures by non-authorized professionals should be applied.

The first criterion is that the order has to be assigned by a qualified radiologist. This radiologist should be well-experienced in evaluating mammograms. The order must be explicit and definite to avoid misunderstandings. Therefore, although not required, it is recommended to lay down the order on paper¹².

The second criterion is to give instructions when necessary. A radiologist has to instruct the technologists in pre-reading mammograms.

Third, the possibility of intervention and supervision by a physician has to be arranged. Pre-reading mammograms by technologists implicates that all patients with abnormal examinations will be evaluated by a radiologist. In addition, to guarantee quality over time and to validate the legal responsibilities of the radiologists, a randomly selected part of the images which are interpreted as normal by the technologists should be re-evaluated by a radiologist every month.

The fourth criterion is that the technologists have to be proficient to evaluate mammograms. Therefore, the technologists should be well-experienced in performing mammograms. In addition, it is clearly important that the technologists should receive a training programme to learn to evaluate mammograms. Attending refresher courses regularly is needed to preserve the interpretation skills.

In conclusion, according to the IHCP Act, pre-reading mammograms by technologists is allowed when the preconditions for delegating reserved procedures are met.

Liability in case of malpractice

In case of malpractice of the technologist in evaluating mammograms, a liability claim could be filed in a disciplinary, civil and/or criminal court. Both technologist and radiologist as well as the hospital could be subject of the claim.

In this paragraph, disciplinary and civil liability will be discussed. Criminal charges can only arise when a health care professional is guilty of a gross lack of competence or indifference to a patient's safety. As this occurs very rarely, the criminal liability of technologist, radiologist and hospital will not expound here.

Disciplinary liability

As technologists are professionals as mentioned in section 34 of the IHCP Act, no disciplinary procedures could be started against them. Radiologists, however, could be subject to medical disciplinary law based on section 3 of the IHCP Act. Disciplinary law aims to secure the quality of care and to define standards for good professional practice. In article 47 of the IHCP Act, two disciplinary norms are described. The first includes any act or omission in conflict with the care that the professional should provide to the patient. The second norm covers all other activities which discord with the interests of good practice of individual health care.

Disciplinary court will take into account the rules and norms for good professional practice, including law, guidelines and protocols of the professional societies, and reports of scientific advisory body like The Health Council of the Netherlands.

Civil liability

The ground of civil liability could be default of the contract for medical services or wrongful act.

Generally in Dutch hospitals, radiologists practice as independent professionals. Consequently, the patient has a contract with the hospital for nursing and care and a contract with the radiologist for medical services. In case of malpractice, both the hospital as well as the radiologist could be held liable on the ground of default of the contract (section 6:74 Civil Code in conjunction with article 7:453 Civil Code).

In Dutch University Medical Centres, radiologists are employed on the basis of an employment agreement. The hospital is the contractual partner of the patient, not only regarding nursing and care but also regarding medical treatment. Consequently, in case of malpractice, only the hospital could be held liable on the ground of default of the contract for medical services (section 6:74 Civil Code in conjunction with article 7:453 Civil Code).

In addition, a wrongful act (section 6:162 Civil Code) could be a ground for a hospital liability claim. Furthermore, both radiologist and technologist could be held liable based on a wrongful act.

In case of a civil liability claim, the judge will test whether the professional or institution have met the standard of care that a reasonably prudent professional should provide to a patient.

Professional standard of care

A health care professional or institution should provide "responsible care" (according to the IHCP Act) and "care of a conscientious health care provider" (according to the Medical Contract Act). Therefore, both civil and disciplinary court will test the claim against the professional standard of care. In order to define this concept more clearly, codes, quality assurance policies, quality

programmes and procedures are determined by the medical profession, often based on evidence-based clinical guidelines. These guidelines do not have direct legal status, but have gained indirect legal significance as acting outside guidelines may expose physicians to liability. Furthermore, the Dutch Supreme Court stated that protocols are a part of the medical professional standard¹³.

Future prospects

It has been assumed that the redefinition of professional roles in health care could improve quality standards and could result in more efficient care. Pre-reading mammograms by technologists has that potential as well.

At this moment, pre-reading is not common practice yet. It is likely that the concept will expand from a local initiative, in which pre-reading could be safely regulated when following the preconditions for transferring reserved procedures. A local protocol could be formulated, including clear definitions of responsibilities and liabilities. From case law, it appeared that such a protocol needs to be followed¹⁴. However, the scope of such a protocol is limited to the local professionals. For a wider implementation, the concept should be well-embedded in legal rules and should answer the professional standard of care. Therefore, several issues need to be addressed.

First, technologists should receive sufficient training in pre-reading mammograms. This could be arranged as an extra training programme for technologists who are interested in the new task, or it could be incorporated in the regular education programme for technologists.

Second, the procedures regarding pre-reading mammograms by technologists should be concretised. Professional associations should be stimulated to formulate and implement guidelines, protocols and professional codes with regard to pre-reading mammograms. It may be feasible to deploy technologists as physician assistants, who are well-organised. The Netherlands Association of Physician Assistants (NAPA) formulated a Professional Profile which reflects the professional duties performed in medical practice. Furthermore, a Master Training has been developed and physician assistants have to be registered in a certification register¹⁵.

Third, in addition to self-regulation through professional associations, the legislator should consider whether the procedure of pre-reading mammograms by technologists could be subject to legal regulations. Namely, in order to protect patients, the legislator should consider to formulate general conditions that skill mixing initiatives of both potentially risky procedures as well as reserved procedures should meet. Although this does not fit in the system of the current IHCP Act, it could elucidate the conditions for skill-mixing in a more general sense. Now that the concept of skill-mixing is increasingly widespread, more attention should be given to the protection of patients. In this view, it may be argued that all professionals involved in skill mixing initiatives of potentially risky and reserved procedures should be subjected to disciplinary law. Although

this does not fit in current law either and it may not be reasonable to change the system to that extent, it could serve the purpose to protect the patient.

Conclusions

Skill mixing with new professionals like nurse practitioners and physician assistants becomes more and more standard practice in health care. In order to reduce the workload of radiologists and to improve the quality of patient care, pre-reading mammograms by technologists could be considered. This paper discussed the legal implications of pre-reading. It was shown that pre-reading mammograms by technologists could be effectuated when preconditions to perform reserved procedures by unauthorised professionals are met. All responsibilities and liabilities should be laid down in an unambiguous protocol.

Technologists as well as radiologists and the hospital could be subject to a liability claim, which will be tested against the professional standard of care. Compliance with the conditions mentioned in a protocol or code of conduct regarding pre-reading mammograms by technologists, in combination with adequate training and supervision should be sufficient to disprove potential claims.

For a wide implementation of the concept of pre-reading, adaptation of legislation is needed.

References

1. Buchan, J, Dal Poz, MR. Skill mix in the health care workforce: reviewing the evidence. *Bulletin of the World Health Organization* 2002; 80:575-580
2. Friedenberg, RM. The role of the supertechnologist. *Radiology* 2000; 215:630-633
3. Duijm, LE, Groenewoud, JH, Fracheboud, J, Ineveld, BM, Roumen, RM, De Koning, HJ. Introduction of additional double reading of mammograms by radiographers: effects on a biennial screening programme outcome. *Eur J Cancer* 2008; 44:1223-1228
4. Van den Biggelaar, FJHM, Kessels, AG, Van Engelshoven, JMA, Flobbe, K. Diagnostic performance of breast technologists in pre-reading diagnostic mammograms (chapter 3 of this thesis) and Van den Biggelaar, FJHM, Kessels, AG, Van Engelshoven, JMA, Flobbe, K. Costs and effects of using specialised breast technologists in pre-reading mammograms in a clinical patient population (chapter 4 of this thesis)
5. In addition to the reserved procedures, section 40 provides conditions to regulate the quality of professional practice. This section defines that professionals described in section 3 and section 34 who practise their profession other than in an institution as meant in the Health Care Institution Quality Act, should deliver responsible care, which is of a high level, effective, suitable and orientated on the patient, and meets patients' realistic needs. Comparable conditions for health care institutions are addressed in the Health Care Institution Quality Act.
6. Raad voor de Volksgezondheid en Zorg (Council for Public Health and Health Care). *Juridische aspecten van taakherschikking in de gezondheidszorg (Legal aspects of shifting tasks in health care)*. Zoetermeer: Raad voor de Volksgezondheid en Zorg, 2002
7. Centraal Tuchtcollege voor de Gezondheidszorg (Central Medical Disciplinary Board), 18 December 2001. *Medisch Contact* 2002; 57(14)
8. Centraal Tuchtcollege voor de Gezondheidszorg (Central Medical Disciplinary Board), 1 February 2005. *Medisch Contact* 2005; 60:583-585
9. Centraal Tuchtcollege voor de Gezondheidszorg (Central Medical Disciplinary Board), 25 May 2000. *Medisch Contact* 2001; 56(1)
10. Centraal Tuchtcollege voor de Gezondheidszorg (Central Medical Disciplinary Board), 4 March 2008. *Medisch Contact* 2008; 63:866-868
11. Tuchtcollege voor de Gezondheidszorg Amsterdam (Medical Disciplinary Board Amsterdam), 16 February 1999. *Tijdschrift voor Gezondheidsrecht (Journal of Health Law)* 1999; 23:259-262
12. Biesart, CIH, Kalkman-Bogerd, LE, Sluijters, B. *T&C Gezondheidsrecht*. Deventer: Kluwer, 2008
13. Hoge Raad (Dutch Supreme Court), 2 March 2001. *Nederlandse Jurisprudentie* 2001; 649
14. Kastelein, WR, Sluijters, B, Giesen, I, Van Dijk, Chr, Kottenhagen, RJP, Gevers, JKM. *Medische aansprakelijkheid*. Den Haag: Sdu Uitgevers BV, 2003
15. www.napa.nl. Date accessed: 20th of June 2009

6

Computer-aided detection in full-field digital mammography in a clinical population: performance of radiologist and technologists

F.J.H.M. van den Biggelaar, A.G.H. Kessels,
J.M.A. van Engelshoven, C. Boetes, K. Flobbe

*Accepted for publication in
Breast Cancer Research and Treatment*

Abstract

The purpose of the study was to evaluate the impact of a computer-aided detection (CAD) system on the performance of mammogram readers in interpreting digital mammograms in a clinical population. Furthermore, the ability of a CAD system to detect breast cancer in digital mammography was studied in comparison to the performance of radiologists and technologists as mammogram readers.

Digital mammograms of 1048 consecutive patients were evaluated by a radiologist and three technologists. Abnormalities were recorded and an imaging conclusion was given as a BI-RADS score before and after CAD analysis. Pathology results during 12 months follow-up were used as a reference standard for breast cancer. Fifty-one malignancies were found in fifty patients. Sensitivity and specificity were computed before and after CAD analysis and provided with 95% CIs. In order to assess the detection rate of malignancies by CAD and the observers, the pathological locations of the 51 breast cancers were matched with the locations of the CAD marks and the mammographic locations that were considered to be suspicious by the observers.

For all observers, the sensitivity rates did not change after application of CAD. A mean sensitivity of 92% was found for all technologists and 84% for the radiologist. For two technologists, the specificity decreased (from 84% to 83% and from 77% to 75%). For the radiologist and one technologist, the application of CAD did not have any impact on the specificity rates (95% and 83%, respectively). CAD detected 78% of all malignancies. Five malignancies were indicated by CAD without being noticed as suspicious by the observers.

In conclusion, the results show that systematic application of CAD in a clinical patient population failed to improve the overall sensitivity of mammogram interpretation by the readers and was associated with an increase in false-positive results. However, CAD marked five malignancies that were missed by the different readers.

Introduction

Mammography is the most widely used method for early detection and diagnosis of breast cancer and has shown to result in a significant reduction in mortality. However, despite its effectiveness, a number of mammographically detectable breast cancers may still be missed¹.

Therefore, interest is growing in the use of computer-aided detection (CAD) systems which mark suspicious regions that may otherwise be overlooked by the mammogram reader. Many studies have validated CAD techniques in screening populations and found a sensitivity ranging from 73%² to 96%³. Moreover, the impact of a CAD system on the performance of mammogram observers was evaluated in several studies^{2,4-14}, showing that there is insufficient evidence to claim that CAD improves cancer detection rates but that it does increase recall rates in screening programs for breast cancer¹⁵.

Little is known on the feasibility of CAD software in mammography in a clinical population. In the Netherlands, hospital radiology departments perform both diagnostic and screening mammography. Diagnostic examinations are carried out in women with clinical signs and symptoms suggestive for breast cancer, women referred by the Dutch Nationwide Breast Cancer Screening Programme and in women with a personal history of breast cancer. Furthermore, screening examinations are performed in asymptomatic women referred for a family history of breast cancer, a genetic predisposition or for reassurance. The application of CAD in such a clinical population is as a screening tool, indicating areas that need special interest in screening mammograms and indicating additional abnormalities in the ipsilateral and the contralateral breast in diagnostic mammograms.

In addition to the use of CAD systems to increase the cancer detection rate, another performance improving development in diagnostic breast imaging includes the deployment of radiographic technologists in double reading or pre-reading mammograms. In the setting of an increased workload for radiologists, this alternative working practice, known as skill mixing, has been explored in many studies¹⁶⁻²¹. A systematic review showed that technologists could be as sensitive as radiologists in detecting breast malignancies, but with higher false-positive rates²². This raises the question whether the application of CAD software could also enhance the reading performance of technologists.

The aim of the current study was twofold. First, the impact of a CAD system on the performance of mammogram readers in interpreting digital mammograms in a clinical population was evaluated. Second, the ability of a CAD system to detect breast cancer in digital mammography was studied in comparison to the performance of radiologists and technologists as mammogram readers.

Methods

Patient selection

Digital mammograms of 1050 consecutive women referred to the radiology department of Maastricht University Medical Centre between January 2007 and July 2007, were selected for this study. Two patients with a proven breast malignancy at the time of examination were excluded. Consequently, mammograms of 1048 patients with a mean age of 51 years (range 20-90) were included in the study. In 829 patients (79%), diagnostic mammography was performed, whereas 219 women (21%) were referred for screening mammography. Indications for referral for diagnostic breast imaging were: follow-up of prior breast malignancy (n=285, 27%), including 164 examinations after lumpectomy and 121 after mastectomy, the occurrence of a palpable breast lump (n=255, 24%), other symptomatic complaints like pain or nipple abnormalities (n=189, 18%), follow-up of a prior benign abnormality (n=62, 6%) and referral by the Dutch Nationwide Breast Cancer Screening Programme (n=38, 4%). Indications for screening mammography were: family history of breast cancer, including BRCA gene mutation (n=174, 17%), and other asymptomatic reasons for referral (n=45, 4%).

In all patients, a standard two-view unilateral or bilateral mammogram was performed, using a full-field digital mammography system (Giotto Image FFDM, IMS, Bologna, Italy).

The institutional medical ethics committee approved the study.

Reference standard

The reference standard for the presence or absence of breast cancer was determined by the pathologic results from core needle biopsies and surgical excisions within a follow-up of 12 months. Breast cancer status was presumed to be negative when no pathologic condition was reported after one year. Lobular carcinomas in situ were excluded as malignancies.

Pathology data were retrieved from PALGA, a nationwide network and registry of histopathology and cytopathology in the Netherlands, to which all Dutch hospital pathology departments are linked.

In the study population, 51 breast cancers were found in 50 patients, leading to a prevalence of 4.8% (50/1048). One patient had bilateral breast cancer. Forty-six patients with a malignancy (92%) were referred for diagnostic mammography; four were referred for screening mammography.

The histopathological classification of breast cancer included 6 ductal carcinomas in situ (DCIS), 35 invasive ductal malignancies, 8 invasive lobular malignancies and 2 other invasive malignancies.

Breast density categories were scored for all patients with breast cancer by a well-experienced radiologist (CB). Twenty-one mammograms were classified as <25% dense, 17 examinations as 25-50%, 12 as 50-75% and 1 as >75%.

CAD system

All digital images were analysed using the Second Look Digital CAD system (iCAD Inc.; Beavercreek, OH, USA; version 7.2-H; date of installation: Dec 4th 2006). The system uses artificial intelligence and pattern recognition technology to highlight suspicious regions on a mammogram with size-varying marks¹⁰. Possible masses on the mammogram are marked by ellipses; possible micro-calcifications are indicated by squares. The software uses binary thresholds in order to decide whether a mark is displayed.

Instruction of the observers in the use of the Second Look CAD system was provided by the manufacturer by onsite training after installation. Furthermore, nine months of clinical experience with the software was realised before the onset of this study. During this period, all readers used the software according to their own judgment in each specific patient.

Impact of CAD on reader performance

In order to evaluate the impact of the CAD system on the performance of the observers, the set of 1048 consecutive mammograms was evaluated before and after CAD analysis. Each examination was interpreted by four observers, consisting of the radiologist on duty and three technologists.

Two radiologists were involved in the study and according to their work schedule in daily practice, they both evaluated half of the set of mammograms. They have, respectively, 5 and 20 years of experience in reading over 1000 mammograms a year in the department. The technologists involved had one year experience in mammogram interpretation, as part of a project on radiology skill mixing in which they were trained as mammogram readers²³. The observers had full information on patients' age and reason for referral, and had access to prior screening and diagnostic mammographic images. All were blinded to the evaluations of the other observers.

All mammographic findings were registered on case record forms. Abnormalities were sketched in a representation of each breast in craniocaudal and medio-lateral oblique views. A BI-RADS score was given, which is based on a grading reporting scale for mammography with an increasing degree of suspicion for malignancy²⁴. Furthermore, it was indicated whether the observer advised additional diagnostic work-up.

The impact of CAD on the performance of the observers was evaluated in special reading sessions, in which the mammograms were read again with the help of the CAD software. All observers indicated whether the CAD marks added valuable diagnostic information to their own original evaluation, sketched in the relevant CAD marks on the representation of the breasts and assigned a BI-RADS score for that abnormality. Furthermore, it was indicated whether additional diagnostic work-up was advised by the readers, based on the new findings of the CAD system.

To evaluate the diagnostic performance of the observers, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were com-

puted before and after CAD analysis and provided with 95% CIs. Furthermore, a diagnostic odds ratio (DOR) with 95% CIs was calculated before and after application of the CAD system. The DOR is a measure for the diagnostic performance of a test as it combines sensitivity and specificity in one measure. The DOR can be derived by dividing the odds of a positive test result among diseased persons by the odds of a positive test result among non-diseased persons $[(\text{sensitivity}/(1-\text{sensitivity})) * (\text{specificity}/(1-\text{specificity}))]$. A DOR of 1 implies that the test has no discriminatory power at all; the larger the DOR, the better the test discriminates between patients with and without the disease of interest²⁵. The statistical measures were calculated on patient level with a single overall BI-RADS assessment per patient that was based on the most suspicious score of the two breasts. BI-RADS 1 (normal) and 2 (benign finding) assessments were interpreted as negative for analysis. In addition, mammograms given a BI-RADS 3 (probably benign finding) assessment with a recommendation of short-term or normal follow-up were recoded as BI-RADS 3- and were considered as negative for analysis.

Mammograms with a BI-RADS 3 assessment with a recommendation of immediate work-up were recoded as BI-RADS 3+ and were considered as positive for analysis. This sub-classification of BI-RADS 3 assessments was adopted from the Breast Cancer Surveillance Consortium²⁶, and has been used in several mammography studies^{5,26-28}. Furthermore, BI-RADS assessments 0 (need additional imaging evaluation), 4 (suspicious abnormality) and 5 (highly suggestive of malignancy) were considered as positive for analysis.

Statistical analyses were performed using SPSS 16.0 for Windows (SPSS Inc., Chicago, USA) and Stata 9.0 statistical software package (StataCorp LP, Texas, USA). McNemar's test was used to test differences between sensitivity and specificity before and after CAD analysis. Furthermore, in order to test differences between the DORs before and after CAD application, the bootstrap technique was used, which is a resampling method with replacement from the original sample²⁹. P value less than 0.05 was considered to be statistically significant.

Validity of CAD software

In order to determine the ability of the CAD system to detect breast cancer in mammograms, all images with proven malignancies (n=51) were evaluated by a well-experienced breast radiologist (CB), who was not involved in the primary data collection of this study. First, the breast quadrant with a proven malignancy was obtained from the pathology reports from core needle biopsies and surgical excisions. Then, the pathological location of the breast cancers were matched visually with the corresponding location on the mammogram images. Consequently, it was defined whether the CAD system had correctly indicated these concerning locations. Similarly, the detection rates of the observers in the study were defined and were compared to the detection rate of the CAD system.

All cases with locations of CAD marks corresponding with the centre of the malignancy were considered to be true-positive. Furthermore, when the location of the tumor was confirmed by CAD in at least one mammographic view of the examination, the result was regarded as true-positive.

Consequently, the sensitivity of the CAD software was computed as the number of malignancies correctly marked by the system divided by the total number of malignancies.

Furthermore, in order to determine the rate of false-positive results, CAD was applied to mammograms of 100 consecutive patients with normal mammograms (BI-RADS 1, n=64) or mammograms with clearly benign findings (BI-RADS 2, n=36). All CAD marks in these images were counted and considered to be false-positive. Subsequently, the average number of false-positive marks of the CAD system per patient was calculated with 95% CIs.

Results

Impact of CAD on reader performance

In Table 1, the performance of mammogram interpretation by the observers is presented before and after CAD analysis, thus with and without knowledge of the findings of the CAD software. For all observers, the sensitivity rates did not change after application of the CAD software. A mean sensitivity of 92% was found for all technologists, as compared to the sensitivity of 84% for the radiologist.

For technologist 1 and 3, the specificity decreased from 84% to 83% and from 77% to 75%, respectively. For the radiologist and technologist 2, the application of CAD did not have any impact on the specificity rates.

Furthermore, the DOR rates decreased after the use of CAD software. For the radiologist, the DOR decreased from 96 to 90, whereas the mean DOR for the three technologists decreased from 52 to 49.

Table 2 shows the effect of CAD in all cases that were found to be positive for analysis, with a BI-RADS score of 0, 3+, 4 and 5 given by each observer. Before CAD analysis, the radiologist classified 49 cases as BI-RADS 3+ (5% of all mammograms), whereas the technologists scored a mean number of 165 patients as BI-RADS 3+ (16% of all mammograms). Furthermore, it was shown that application of CAD did lead to an increase in the number of positive cases, but mainly in the BI-RADS 3+ category (n=30). The highest increase was found for technologist 3, from 178 to 200 BI-RADS 3+ cases. However, no additional malignancies were detected after CAD analysis, so these cases can be regarded as false-positive results.

Further analysis of these 30 cases showed that before CAD, these mammograms had a BI-RADS classification of either benign (n=6) or normal (n=24).

The two other changes in BI-RADS category after the application of CAD included a shift of BI-RADS 2 into 0 for technologist 1 and BI-RADS 2 into 4 for technologist 3.

Table 1: Reader performance (50 patients with breast cancer in 1048 patients)

		Before CAD analysis	After CAD analysis
Radiologist	Sens (%)	84 (71–93)	84 (71–93)
	Spec (%)	95 (93–96)	95 (93–96)
	PPV (%)	45 (34–55)	44 (34–54)
	NPV (%)	99 (98–100)	99 (98–100)
	DOR	96 (43–210)	90 (42–193)
Technologist 1	Sens (%)	92 (81–98)	92 (81–98)
	Spec (%)	84 (82–86)	83 (81–86)*
	PPV (%)	23 (17–29)	22 (16–28)
	NPV (%)	99 (99–100)	99 (99–100)
	DOR	62 (23–167)	57 (21–155)*
Technologist 2	Sens (%)	92 (81–98)	92 (81–98)
	Spec (%)	83 (81–85)	83 (81–85)
	PPV (%)	22 (16–28)	21 (16–27)
	NPV (%)	99 (99–100)	99 (99–100)
	DOR	57 (21–154)	56 (21–152)
Technologist 3	Sens (%)	92 (81–98)	92 (81–98)
	Spec (%)	77 (74–79)	75 (72–77)**
	PPV (%)	17 (12–22)	15 (11–20)
	NPV (%)	99 (99–100)	99 (99–100)
	DOR	38 (14–103)	34 (13–91)**

Sens = sensitivity; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value; DOR = diagnostic odds ratio; numbers in brackets are 95% confidence intervals

*difference between before and after data (P=0.002)

**difference between before and after data (P<0.001)

Validity of CAD software

Analysing all CAD marks on the 51 mammograms with proven breast cancer showed that the CAD system correctly marked 40 malignancies. Consequently, the sensitivity of the detection of malignancies by the CAD software in this specific population was 78%. In 11 of the 51 mammograms (22%), however, CAD analysis missed the malignancy (false-negative result). In three mammograms, no CAD marks were given at all, whereas marks were placed outside the location of the malignancy in eight mammograms.

Furthermore, in the 100 consecutive patients with normal (n=64) and benign findings (n=36), CAD software indicated 225 marks in 79 patients, leading to an average number of 2.9 false-positive marks per patient (95% CIs 2.4–3.3). Of these 79 patients, 60% were scored as BI-RADS 1 and 40% as BI-RADS 2. In all 100 patients, the CAD system displayed an average number of 2.3 false-positive marks per patient (95% CIs 1.8–2.7) with 0.7 calcification marks (95% CIs 0.4–1) and 1.6 mass marks (95% CIs 1.3–1.8).

Table 2: Effect of CAD in all cases regarded to be positive for analysis

	Radiologist		Technologist 1		Technologist 2		Technologist 3	
	Before	After	Before	After	Before	After	Before	After
BI-RADS 0								
Total (n)	7	7	2	3	3	3	41	41
Malignancies (n)	1	1	0	0	0	0	0	0
BI-RADS 3+								
Total (n)	49	51	155	160	161	162	178	200
Malignancies (n)	9	9	6	6	4	4	7	7
BI-RADS 4-5								
Total (n)	38	38	46	46	50	50	57	58
Malignancies (n)	32	32	40	40	42	42	39	39
Total positive cases								
Total (n)	94	96	203	209	214	215	276	299
Malignancies (n)	42	42	46	46	46	46	46	46

Breast cancer detection by CAD versus observers

Table 3 (page 76) presents detailed information of the performance of the observers and the CAD system in detecting the 51 malignancies in this study. The mean detection rate of the observers in diagnostic mammograms was 87% (range 79%-89%).

In the screening mammograms, the three technologists detected all four malignancies (100%), whereas the radiologist detected three. Distinguishing the performance by breast density levels, the mean detection rate of the observers was 94% in non-dense breasts (<50% density) compared with 67% in dense breasts (>50% density). The CAD system detected 84% of the malignancies in the non-dense breasts, compared with 62% in dense breasts.

When specifying the histopathological type of breast cancer, it is shown that all patients with DCIS were correctly identified by all observers and the CAD system. Furthermore, the observers and CAD detected about 85% of all patients with invasive ductal carcinoma. Women with invasive lobular carcinomas were found by the radiologist and technologists in six to eight out of twelve cases, whereas CAD identified only four.

Analysing the CAD marks given by the system in the 51 mammograms with a histology-proven breast cancer, showed that 5 malignancies were indicated by a CAD mark without being noticed as suspicious by the observers. The radiologist failed to identify four of these malignancies, one technologist failed in two cases and two technologists failed to detect three. Four of these malignancies were found in diagnostic mammograms; one malignancy was detected in screening mammograms. Three patients had non-dense breasts (<50% density) and two patients had dense breasts (>50% density). All malignancies were invasive ductal carcinomas.

Table 3: Detection of malignancies (n=51)

	n malignancies	n detected by					
		Radiologist	Technologist 1	Technologist 2	Technologist 3	CAD	CAD alone ¹
Total	51 (100%)	42 (82%)	45 (88%)	46 (90%)	45 (88%)	40 (78%)	5 (10%)
Setting							
Screening	4 (8%)	3 (75%)	4 (100%)	4 (100%)	4 (100%)	3 (75%)	1
Diagnostic	47 (92%)	39 (83%)	41 (87%)	42 (89%)	41 (87%)	37 (79%)	4
Density							
<25%	21 (41%)	19 (90%)	20 (95%)	20 (95%)	20 (95%)	19 (90%)	2
25-50%	17 (33%)	16 (94%)	16 (94%)	16 (94%)	16 (94%)	13 (76%)	1
51-75%	12 (24%)	7 (58%)	9 (75%)	10 (83%)	9 (75%)	8 (67%)	2
>75%	1 (2%)	0	0	0	0	0	-
Type							
DCIS	6 (12%)	6 (100%)	6 (100%)	6 (100%)	6 (100%)	6 (100%)	-
Invas duct	35 (69%)	29 (83%)	30 (86%)	30 (86%)	29 (83%)	30 (86%)	5
Invas lob	8 (16%)	6 (75%)	7 (88%)	8 (100%)	8 (100%)	4 (50%)	-
Other invas	2 (4%)	1 (50%)	2 (100%)	2 (100%)	2 (100%)	0	-
CAD analysis							
Detected	40 (78%)	36 (90%)	38 (95%)	37 (93%)	37 (93%)	-	-
Missed	11 (12%)	6 (55%)	7 (64%)	9 (82%)	8 (73%)	-	-

DCIS = ductal carcinoma in situ; Invas duct = invasive ductal carcinoma; Invas lob = invasive lobular carcinoma; Other invas = other invasive carcinomas

¹ CAD alone: malignancies detected by CAD, but missed by at least one of the observers

Discussion

This study demonstrated that the application of CAD was associated with an increased number of false-positive cases in all observers and failed to improve the sensitivity of mammogram interpretation. Furthermore, the CAD system correctly indicated 78% of the 51 cancers which is consistent with results in other studies with consecutive patient series in screening mammography, reporting sensitivities ranging from 73% to 96%^{2-3,13,30}.

The use of a consecutive patient series is a strength of this study, as it represents the patient population referred for diagnostic and screening mammography in daily clinical practice. However, the use of special reading sessions in order to evaluate the mammograms with and without the application of CAD software can be regarded as a potential limitation, as this does not represent daily practice. The lack of direct clinical consequences of the BI-RADS assessments given, combined with the large volume of mammograms in these reading sessions, may possibly have led to a loss of focus and a potential decrease in

performance. However, in order to provide equal evaluation circumstances for all observers and to monitor the data collection closely, this study design was thought to be the most appropriate.

Another limitation of this study is the lack of information on a potential learning curve in using the CAD software. Although the software was available to all observers for nine months before the onset of the study, detailed information about the actual experience of the observers with the CAD system was missing.

In this study, the sensitivity of mammogram evaluation without including information on additional imaging tests was 84% for the radiologist and 92% for the three technologists (McNemar, P value=0.125). However, the higher sensitivity of the technologists comes at the cost of a lower specificity. The technologists' mean specificity was significantly lower than for the radiologist (81% and 95% respectively, McNemar: P value<0.001). Apparently, technologists were more cautious and uncertain in their evaluation which did lead to a higher rate of false-positive results. These findings are in accordance with the results of a systematic literature review indicating that technologists could be as sensitive as radiologists in detecting malignancies on mammograms, but with lower specificity²².

It needs to be mentioned that the sensitivity of the radiologist for mammogram evaluation in our study is in concordance with the performance found in other studies on similar patient populations^{31,32}. In addition, the DOR score which combines sensitivity and specificity into one measure, was higher for the radiologist (96) as compared to the technologists (range 38-62). This means that the overall performance in mammogram evaluation was better for the radiologist as compared to the technologists.

In most published studies, CAD was applied to conventional mammographic images that were digitised. A few studies evaluated CAD in digital mammograms. In a study of Kim et al.¹², initial and follow-up mammograms of 93 patients with a malignancy were read by a CAD system. An overall sensitivity of 90% was found, whereas the reproducibility of true-positive CAD marks was high. Yang et al.³ reported a CAD sensitivity of 96% in 103 screening breast cancers. In a study of Wei et al.³³, CAD was evaluated in full-field digital mammograms and digitised screen-film mammograms. In a set with 131 masses, no significant difference was found between the performance of CAD in the digital images and the screen-film images.

It can be assumed that the availability of digital mammography, like in the present study, will facilitate the applicability of CAD software in daily clinical practice. Evaluating mammograms with CAD software in digital images avoids loss of image quality and reproducibility of the CAD system because of a mechanical scanning digitisation process. Furthermore, in digital mammography CAD software can be applied immediately after image acquisition, whereas in

conventional analogue mammography, the digitisation of mammograms will provide a time delay.

Most studies report the application of CAD software in screening mammograms. Only a few studies evaluated the feasibility of CAD software in a population including both symptomatic patients as well as women referred for screening mammography. Brancato et al.³⁴ included 3425 women attending a self-referral breast centre for digital mammography. Of all women, 25% presented with symptoms and 75% presented for screening. In this population, 102 malignancies (88 in symptomatic women and 14 in women having screening) were detected by the radiologist without the use of CAD. After application of the CAD software, three additional malignancies were found in patients with symptoms and two in women presenting for screening.

Furthermore, in a prospective study, Dean and Ilvento³⁰ evaluated the impact of CAD on cancer detection in 9520 film-screen mammograms. Of all mammograms, 60% were screening studies and 40% were diagnostic. 104 malignancies were detected and, after CAD application, a 10.8% increase in cancer detection was found.

In the present study, only 21% of all women attended for screening mammography. The other 79% were referred for diagnostic mammography. In this population, no additional malignancies were found by the observers after CAD analysis. However, a close analysis of the CAD marks that were given in mammograms with histology-proven malignancies showed that CAD indicated five additional malignancies that were missed by one or more observers. This indicates that the readers in this study did not always respond to a correct mark of the CAD system. The large number of false-positive marks resulting in a low specificity of the CAD system, may be an explanation. On average, over two false-positive marks were displayed per patient, so more than 99% of the marks should be ignored to find one true-positive finding that was missed initially by the observer. The interpretation of a CAD mark is based on mammographic features of the finding and the experience of the observer. Furthermore, patient characteristics, reason for referral and clinical findings influence the observers' decision to act on CAD marks. Therefore, it can be argued that it is difficult to improve sensitivity while maintaining the high specificity of the observer.

The application of CAD decreased the performance of two technologists in reading mammograms in a clinical population because of higher false-positive rates. It can be assumed that the technologists were indecisive of the CAD marks because of the large number of marks. Furthermore, it is likely that the degree of experience will play a role. Studies evaluating the association between radiologists' experience and accuracy in reading screening mammograms, demonstrate a higher specificity in radiologists who routinely evaluate large numbers of mammograms^{35,36}. Therefore, it may be assumed that the use of CAD software will influence inexperienced readers, like the technologists in this study, more than well-experienced readers, like the radiologist. The results of

this study support this assumption, as two technologists produce much more additional false-positive results after CAD analysis, compared to the radiologist.

Concluding, this study has disclosed valuable information about the application of CAD software in digital mammography in a daily clinical population. Although the results show that systematic application of CAD in a clinical patient population failed to improve the overall sensitivity of mammogram interpretation by the readers and was associated with an increase of false-positive results, CAD marked five malignancies that were missed by the different readers. However, the usefulness of CAD software in a clinical population remains limited.

References

1. Collins MJ, Hoffmeister J, Worrell SW. Computer-aided detection and diagnosis of breast cancer. *Semin Ultrasound CT MR* 2006; 27:351-355
2. Taylor PM, Champness J, Given-Wilson RM et al. An evaluation of the impact of computer-based prompts on screen readers' interpretation of mammograms. *Br J Radiol* 2004; 77:21-27
3. Yang SK, Moon WK, Cho N et al. Screening mammography-detected cancers: sensitivity of a computer-aided detection system applied to full-field digital mammograms. *Radiology* 2007; 244:104-111
4. Gilbert FJ, Astley SM, Gillan MG et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med* 2008; 359:1675-1684
5. Fenton JJ, Taplin SH, Carney PA et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 2007; 356:1399-1409
6. Freer TW, Ulisse MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 2001; 220:781-786
7. 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; 189:1135-1141
8. Gromet M. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *AJR Am J Roentgenol* 2008; 190:854-859
9. Hadjiiski L, Sahiner B, Helvie MA et al. Breast masses: computer-aided diagnosis with serial mammograms. *Radiology* 2006; 240:343-356
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. Khoo LA, Taylor P, Given-Wilson RM. Computer-aided detection in the United Kingdom National Breast Screening Programme: prospective study. *Radiology* 2005; 237:444-449
12. Kim SJ, Moon WK, Cho N et al. Computer-aided detection in full-field digital mammography: sensitivity and reproducibility in serial examinations. *Radiology* 2008; 246:71-80
13. Ko JM, Nicholas MJ, Mendel JB et al. Prospective assessment of computer-aided detection in interpretation of screening mammography. *AJR Am J Roentgenol* 2006; 187:1483-1491
14. Malich A, Schmidt S, Fischer DR et al. The performance of computer-aided detection when analyzing prior mammograms of newly detected breast cancers with special focus on the time interval from initial imaging to detection. *Eur J Radiol* 2009; 69:574-578
15. Taylor P, Potts HW. 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
16. Haiart DC, Henderson J. A comparison of interpretation of screening mammograms by a radiographer, a doctor and a radiologist: results and implications. *Br J Clin Pract* 1991; 45:43-45
17. Hillman BJ, Fajardo LL, Hunter TB et al. Mammogram interpretation by physician assistants. *AJR Am J Roentgenol* 1987; 149:907-912

18. Pauli R, Hammond S, Cooke J et al. Radiographers as film readers in screening mammography: an assessment of competence under test and screening conditions. *The British Journal of Radiology* 1996; 69:10-14
19. Pauli R, Hammond S, Cooke J et al. Comparison of radiographer/radiologist double film reading with single reading in breast cancer screening. *Journal of Medical Screening* 1996; 3:18-22
20. Sumkin JH, Klamann HM, Graham M et al. Prescreening mammography by technologists: a preliminary assessment. *AJR Am J Roentgenol* 2003; 180:253-256
21. Wivell G, Denton ER, Eve CB et al. Can radiographers read screening mammograms? *Clin Radiol* 2003; 58:63-67
22. van den Biggelaar FJ, Nelemans PJ, Flobbe K. Performance of radiographers in mammogram interpretation: a systematic review. *Breast* 2008; 17:85-90.
23. van den Biggelaar FJHM, Kessels AG, van Engelshoven JM et al. Diagnostic performance of breast technologists in reading mammograms in a clinical patient population. Submitted for publication (chapter 3 of this thesis)
24. D'Orsi CJ, Bassett LW, Berg WA et al. *Breast Imaging Reporting and Data System: ACR BI-RADS-Mammography*. Reston (VA): American College of Radiology (ACR), 2003
25. Glas AS, Lijmer JG, Prins MH et al. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; 56:1129-1135
26. Barlow WE, Lehman CD, Zheng Y et al. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. *J Natl Cancer Inst* 2002; 94:1151-1159
27. Poplack SP, Tosteson AN, Grove MR et al. Mammography in 53,803 women from the New Hampshire mammography network. *Radiology* 2000; 217:832-840
28. Smith-Bindman R, Chu PW, Miglioretti DL et al. Comparison of screening mammography in the United States and the United Kingdom. *Jama* 2003; 290:2129-2137
29. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. New York: Chapman & Hall, 1993
30. 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; 187:20-28
31. Flobbe K, Bosch AM, Kessels AG et al. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. *Arch Intern Med* 2003; 163:1194-1199
32. Zonderland HM, Coerkamp EG, Hermans J et al. Diagnosis of breast cancer: contribution of US as an adjunct to mammography. *Radiology* 1999; 213:413-422
33. Wei J, Hadjiiski LM, Sahiner B et al. Computer-aided detection systems for breast masses: comparison of performances on full-field digital mammograms and digitized screen-film mammograms. *Acad Radiol* 2007; 14:659-669
34. Brancato B, Houssami N, Francesca D et al. Does computer-aided detection (CAD) contribute to the performance of digital mammography in a self-referred population? *Breast Cancer Res Treat* 2008; 111:373-376
35. Barlow WE, Chi C, Carney PA et al. Accuracy of screening mammography interpretation by characteristics of radiologists. *J Natl Cancer Inst* 2004; 96:1840-1850
36. Moss SM, Blanks RG, Bennett RL. Is radiologists' volume of mammography reading related to accuracy? A critical review of the literature. *Clin Radiol* 2005; 60:623-626

7

Strategies for digital mammography interpretation in a clinical patient population

F.J.H.M. van den Biggelaar, A.G.H. Kessels,
J.M.A. van Engelshoven, K. Flobbe

*Accepted for publication in
International Journal of Cancer*

Abstract

Mammography is the basic imaging modality for early detection of breast cancer. The aim of this prospective study was to evaluate the impact of different mammogram reading strategies on the diagnostic yield in a consecutive patient population referred for digital mammography to a hospital. First, the effect of using computer-aided detection (CAD) software on the performance of mammogram readers was studied. Furthermore, the impact of employing technologists as either pre-readers or double readers was assessed, as compared to the conventional strategy of single reading by a radiologist.

Digital mammograms of 1048 consecutive patients were evaluated by a radiologist and three technologists with and without the use of CAD software. ROC analysis was used to study the effects of the different strategies.

In the conventional strategy, an overall area under the curve (AUC) of 0.92 was found, corresponding to a sensitivity of 84% and a specificity of 94%. When applying CAD software, the AUCs were similar before and after CAD for all readers (mean of 0.95). Employing technologists in pre-reading and double reading of mammograms resulted in a mean AUC of 0.91 and 0.96, respectively. In the pre-reading strategy, the corresponding sensitivity and specificity were 81% and 96%; in the double reading strategy they were 96% and 79%, respectively.

Concluding, in this clinical population systematic application of CAD software by either radiologist or technologists failed to improve the diagnostic yield. Furthermore, employing technologists as double readers of mammograms was the most effective strategy to improve breast cancer detection in daily clinical practice.

Introduction

Mammography is the basic breast imaging modality for early detection and diagnosis of breast cancer. It has been demonstrated that breast cancer screening programs with mammography can reduce mortality by as much as 30%¹⁻⁴. However, despite its effectiveness, a number of mammographically detectable breast malignancies may be missed⁵. In order to increase the cancer detection rate, independent double reading by two radiologists has been recommended⁶⁻⁸. However, as double reading is expensive and there is an increasing shortage of radiologists, studies have explored the feasibility to deploy radiologic technologists as double readers.

In screening mammography, studies have demonstrated that double reading by technologists in addition to radiologists may increase the detection rate of breast malignancies⁹⁻¹¹. Furthermore, it has been shown that technologists could be as sensitive as radiologists in detecting breast malignancies, but with higher false-positive rates¹². In addition to double reading, the employment of technologists in pre-reading screening mammograms has been evaluated. The pre-reading method includes a technologist grouping mammograms into two basic categories: mammograms that require further evaluation by a radiologist and mammograms that have either negative or clearly benign findings which would not need further attention of a radiologist¹³. However, Haiart showed that pre-reading cannot be justified in a screening setting, neither in terms of performance, nor on economic grounds¹⁴. To date, no information is available on the effectiveness of technologists in pre-reading diagnostic mammograms.

In the last decade, interest is growing in the application of computer-aided detection (CAD) software¹⁵⁻²². A CAD system marks suspicious regions that may otherwise be overlooked by the radiologist, which could potentially result in the detection of more malignancies. A systematic literature review showed insufficient evidence to claim that CAD improves cancer detection rates but concludes that it does increase recall rates in screening programs for breast cancer²³.

In the Netherlands, hospital radiology departments perform mammograms in a clinical population referred for breast imaging, with a diagnostic and screening nature. Diagnostic examinations are performed in women when a problem-solving indication was present, like clinical signs and symptoms suggestive for breast cancer, referral by the Dutch Nationwide Breast Cancer Screening Programme and a personal history of breast cancer. Furthermore, screening examinations are performed in asymptomatic women referred for a family history of breast cancer, a genetic predisposition or for reassurance. These screening examinations are not yet part of the Dutch Nationwide Breast Cancer Screening Programme.

Although several studies have focused on the diagnostic value of the use of breast technologists and the application of CAD software in breast cancer screening programs, the diagnostic value of these interventions in a clinical patient population is not well established. Therefore, the purpose of this study was to evaluate the impact of CAD software on the performance of mammogram readers in a clinical patient population. Furthermore, the impact on the

diagnostic yield of employing technologists in either pre-reading or double reading of mammograms in a clinical patient population was evaluated and compared to the conventional strategy of standard mammogram evaluation by a radiologist.

Methods

Patient selection

Digital mammograms of 1050 consecutive women referred to the radiology department of Maastricht University Medical Center between January 2007 and July 2007, were eligible for this study. Two patients were excluded as their mammograms were performed of a proven malignancy for monitoring of responses to neoadjuvant chemotherapy prior to surgical excision.

Consequently, mammograms of 1048 women with a mean age of 51 years (median=50, range=20-90) were included in the study. All patients underwent a standard two-view unilateral or bilateral mammogram, using a full-field digital mammography system (Giotto Image FFDM, IMS, Bologna, Italy). Based on the reason for referral, in 829 patients (79%) the nature of the examination was considered "diagnostic", whereas in 219 women (21%) its nature was "screening". Indications for referral for diagnostic breast imaging were: follow-up of prior breast malignancy (n=285, 27%), including 164 lumpectomies and 121 mastectomies, the occurrence of a palpable breast lump (n=255, 24%), other symptomatic complaints like pain or nipple abnormalities (n=189, 18%), follow-up of a prior benign abnormality (n=62, 6%) and referral by the Dutch Nationwide Breast Cancer Screening Programme (n=38, 4%). Indications for mammography with a screening nature were: family history of breast cancer, including BRCA gene mutation (n=174, 17%), and other asymptomatic reasons for referral (n=45, 4%).

The institutional medical ethics committee approved the study.

Reference standard

The reference standard for the presence or absence of breast cancer was determined by the pathologic results from core needle biopsies and surgical excisions within a follow-up of 12 months. Pathology data were retrieved from PALGA, a nation wide network and registry of histopathology and cytopathology in the Netherlands, to which all Dutch hospital pathology departments are linked. Breast cancer status was presumed to be negative when no pathologic condition was reported in the PALGA system within 12 months. Lobular carcinomas in situ were excluded as malignancies.

In the study population, 51 breast cancers were found in 50 patients, leading to a prevalence of 4.8% (50/1048). One patient had bilateral breast cancer. In 46 patients, the reason for referral was of a diagnostic nature, whereas in 4 patients the reason for referral had a screening nature.

The histopathological classification of breast cancer included 6 ductal carcinomas in situ, 35 invasive ductal malignancies, 8 invasive lobular malignancies and 2 other invasive malignancies.

Study design

In Figure 1, a flow chart of the study design is shown. Each mammogram was interpreted by four observers, consisting of one radiologist on duty and three technologists. Two radiologists were involved in the study and each evaluated about half of the study cases, according to their work schedule in daily practice. They have, respectively, 5 and 20 years of experience in reading over 1000 mammograms per year in the department.

All three technologists had one year experience in mammogram interpretation, as part of a project on radiology skill mixing in which they were trained as mammogram readers. The observers had full information on patients' age and reason for referral, and had access to prior mammograms. All were blinded for the evaluations of the other observers. The technologists evaluated the mammograms in special mammogram reading sessions.

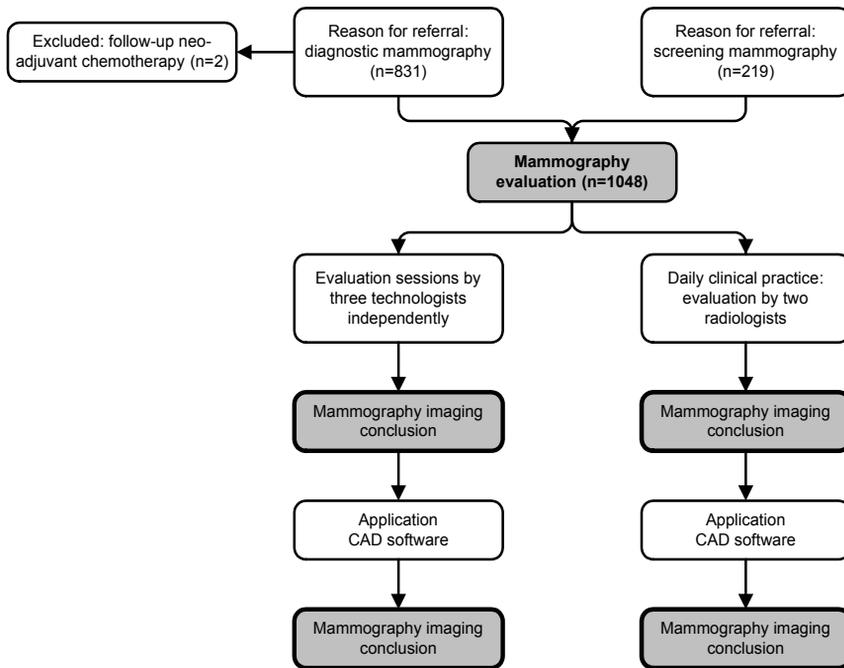


Figure 1: Flow chart study design

All mammographic findings were registered on case record forms. Abnormalities were sketched in a representation of each breast in craniocaudal and medio-lateral oblique views. For each breast, a BI-RADS (Breast Imaging Reporting And Data System) score was given, which is based on a grading reporting scale for mammography with an increasing degree of suspicion for malignancy: 0 = need additional imaging evaluation; 1 = negative examination; 2 = benign finding; 3 = probably benign finding; 4 = suspicious abnormality; 5 = highly suggestive of malignancy²⁴. Furthermore, it was recorded whether the observer advised additional diagnostic work-up.

All mammograms were evaluated and scored by the observers before and after analysis with the Second Look Digital CAD system (iCAD, Inc., Beavercreek, OH, USA).

Analysis

To evaluate the performance of the mammogram readers, receiver operating characteristics (ROC) curves were created. A ROC curve is a plot of the true-positive rate (or sensitivity) versus the false-positive rate (or 1 – specificity) of a test at different cut-off levels²⁵. For the purpose of this study, the BI-RADS classifications of the observers were used to define positive and negative test results for breast malignancy at different cut-off points. As analysis is performed on patient level, the BI-RADS score per patient was determined on the most suspicious BI-RADS evaluation of the separate breasts.

The area under the ROC curve (AUC) represents a measure for the accuracy of the observer, ranging from 0 to 1. A higher AUC indicates a better overall performance of the observer in the detection of malignancies on the mammogram²⁶.

In order to compare the AUC in the different strategies, Stata 9.0 statistical software package (StataCorp LP, Texas, USA) was used. P value less than 0.05 was considered to be statistically significant.

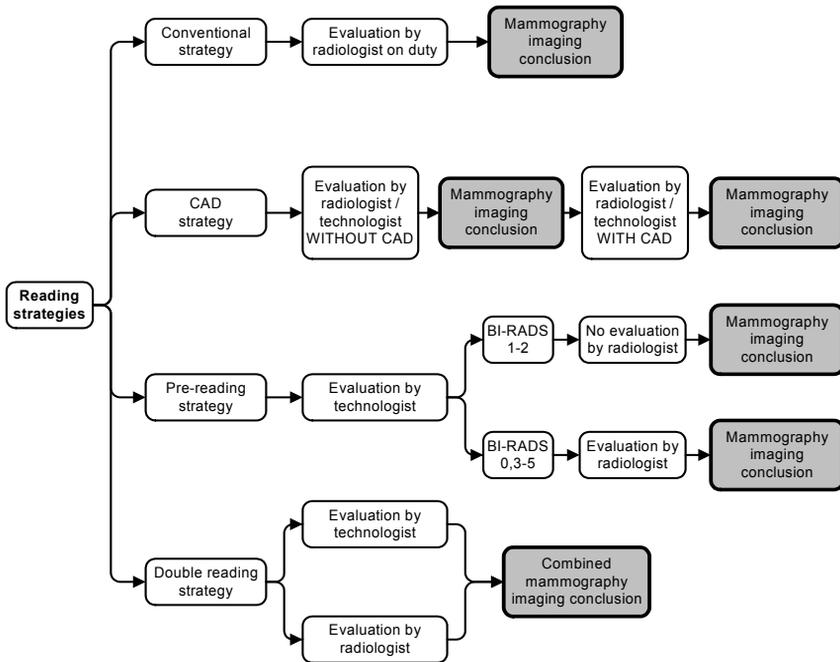
Furthermore, sensitivity and specificity were calculated using a cut-off point between BI-RADS 1-2 (considered negative for analysis) and BI-RADS 0,3-5 (considered positive for analysis).

The 95% confidence intervals (CIs) were calculated using exact binomial confidence intervals (<http://statpages.org/ctab2x2.html>). Mammograms with BI-RADS classifications 0,3-5 are assumed to be positive for analysis, as further action by imaging, tissue analysis or follow-up is recommended.

McNemar's test was used to test differences between sensitivity and specificity in the different strategies. Statistical significance was set at $P < 0.05$. Statistical analyses were performed using SPSS (version 16.0 for Windows).

Reading strategies

In Figure 2, four different strategies of reading mammograms in a clinical patient population are shown.



Figur 2 Reading strategies

First, the **conventional strategy** represents mammogram interpretation by the radiologist on duty, according to daily clinical practice. Actual data are used from the clinical imaging report and the BI-RADS classification of the radiologist on duty was used to obtain a ROC curve.

Second, in the **CAD strategy** the impact of CAD software on the performance of mammogram readers is evaluated. All digital mammograms were evaluated and scored by all four observers with and without help of the CAD software. AUCs were calculated for the observers before and after CAD analysis.

Third, a **pre-reading strategy** using a technologist is analysed and compared to the conventional strategy. In this strategy, decision rules were applied on the actual data: only mammograms with BI-RADS classifications 0,3-5 as diagnosed by the technologist were referred for further evaluation by a radiologist. Mammograms with negative or clearly benign findings (BI-RADS 1 and 2) were not re-read by a radiologist. Consequently, a ROC curve was constructed for each technologist-radiologist combination involved, based on BI-RADS classifications 1 and 2 of the technologists and the concerning BI-RADS classification of the radiologist in the remaining cases.

Finally, a **double reading strategy** by both radiologist and technologist is evaluated and compared to the conventional strategy. In this strategy, decision rules were applied on the actual data: it was assumed that all mammograms

are evaluated by one technologist and a radiologist, resulting in an overall conclusion, consisting of the highest BI-RADS score of either the radiologist or the technologist. A ROC curve is obtained for each technologist involved.

Results

Conventional strategy

The AUC of the radiologists before CAD analysis (0.92) represents the conventional strategy, which corresponds with a mammography sensitivity of 84% and a specificity of 94%, using a cut-off point between BI-RADS 1-2 (assumed to be negative) and BI-RADS 0,3-5 (assumed to be positive).

CAD strategy

Table 1 shows the AUCs of the four observers before and after application of the CAD software. Before CAD analysis, an AUC of 0.92 was found for the radiologist, whereas the AUC was 0.96 and 0.94 for technologists 1 and 2, and technologist 3, respectively. After application of the CAD software, the AUCs were constant for the radiologist and technologists 1 and 2, and slightly decreased for technologist 3.

Table 1: AUCs in CAD strategy

Observer	Before CAD analysis		After CAD analysis	
	AUC	95% CIs	AUC	95% CIs
Radiologist	0.92	0.87-0.98	0.92	0.87-0.98
Technologist 1	0.96	0.94-0.99	0.96	0.93-0.99
Technologist 2	0.96	0.92-0.99	0.96	0.92-0.99
Technologist 3	0.94	0.88-0.98	0.93	0.89-0.97

Pre-reading strategy

Figure 3 displays the ROC curves in the pre-reading strategy, based on BI-RADS scores 1 and 2 of the technologists and BI-RADS scores of the radiologist in the remaining cases. Furthermore, the curve of the conventional strategy is shown. The AUCs were 0.91, 0.92 and 0.91 when pre-reading with technologists 1, 2 and 3, respectively, which was comparable to the AUC of 0.92 in the conventional strategy.

Using a cut-off point between BI-RADS 1-2 and 0,3-5, the number of false-negative results in the pre-reading strategy was higher compared to the conventional strategy, resulting in a lower sensitivity of 80% using technologists 1 and 3, and 82% using technologist 2, compared to 84% in the conventional strategy (Table 2). The mean specificity was 96%, compared to 94% in the conventional strategy.

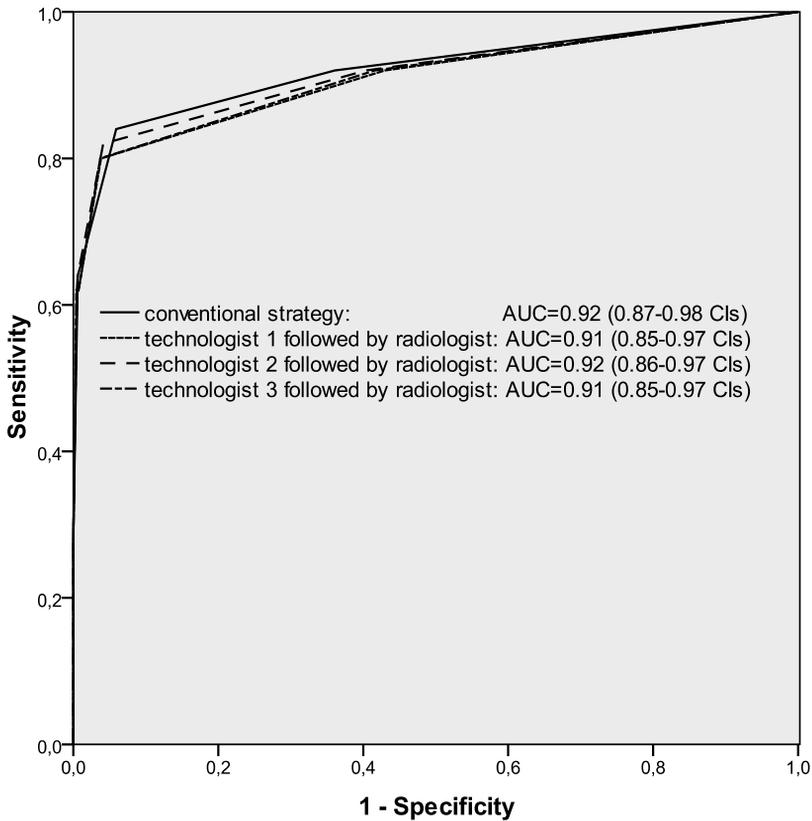


Figure 3: ROC curves in pre-reading strategy

Double reading strategy

In Figure 4 (page 92), the ROC curves of the conventional strategy and the double reading strategy are presented. Double reading with technologist 1 shows an AUC of 0.97, compared to an AUC of 0.92 in the conventional strategy (P value=0.05), whereas double reading with technologists 2 and 3 results in an AUC of 0.96 (P value=0.09).

Table 2 demonstrates that the number of true-positive results was higher in the double reading strategy compared to the conventional strategy, leading to a significantly higher sensitivity of 96% in the double reading strategy with technologist 1 and 3, compared to a sensitivity of 84% in the conventional strategy (P value=0.03).

On the other hand, specificity in the double reading strategy was significantly lower (mean specificity of 79%), compared to the specificity of 94% in the conventional strategy (P value<0.001).

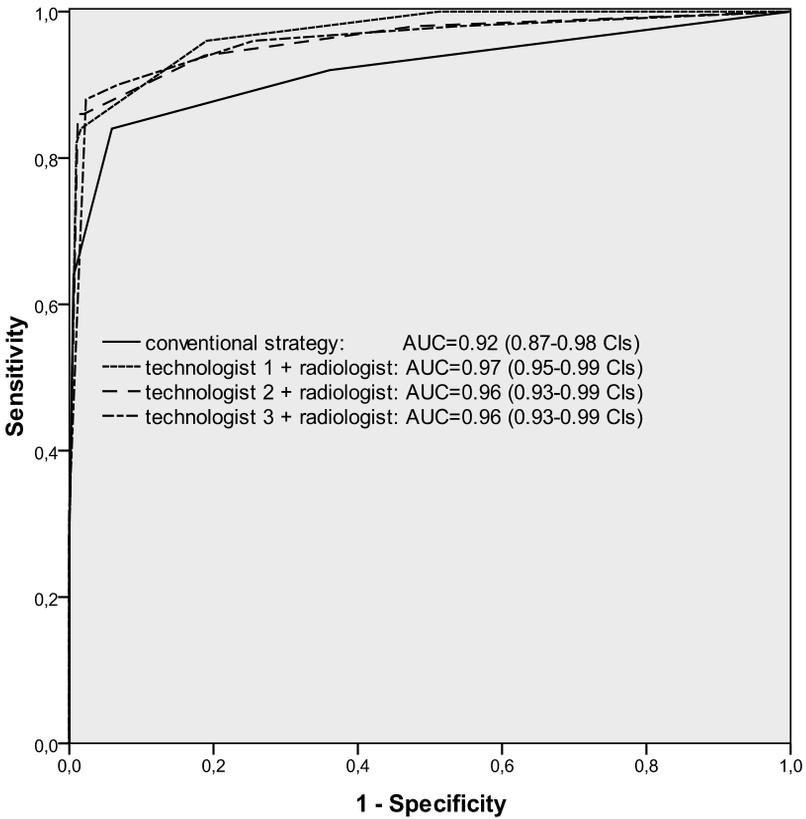


Figure 4: ROC curves in double reading strategy

Discussion

In this study, ROC curve analysis demonstrated that the strategy of double reading mammograms by a radiologist and a technologist obtained the highest diagnostic yield in this patient population, as compared to the strategy of pre-reading by technologists or the conventional strategy of mammogram reading by a single radiologist. Combined mammogram evaluation of radiologist and technologist resulted in an average AUC of 0.96 as compared to 0.92 for the conventional strategy.

Comparing the findings in the different reading strategies showed that double reading resulted in a higher sensitivity at the cost of a lower specificity, whereas pre-reading resulted in a higher specificity at the cost of a lower sensitivity.

It could be argued to prefer a high sensitivity above a high specificity in mammography performed in a clinical patient population. In screening mammography, a subtle balance between referral rate and cancer detection rate is needed. On the one side, a high referral rate may result in more cancers detected.

Table 2: Performance with cut-off point between BI-RADS 1-2 and BI-RADS 0,3-5

	TP (n)	FP (n)	TN (n)	FN (n)	Sens (%)	Spec (%)
<i>Conventional strategy</i>						
Radiologist	42	59	939	8	84 (71-93)	94 (92-95)
<i>Pre-reading strategy</i>						
Technologist 1 followed by radiologist	40	38	960	10	80 (66-90)	96 (95-97)
Technologist 2 followed by radiologist	41	41	957	9	82 (69-91)	96 (94-97)
Technologist 3 followed by radiologist	40	37	961	10	80 (66-90)	96 (95-97)
<i>Double reading strategy</i>						
Technologist 1 + radiologist	48	190	808	2	96 (86-100) ¹	81 (78-83) ²
Technologist 2 + radiologist	47	187	811	3	94 (83-99)	81 (79-84) ²
Technologist 3 + radiologist	48	255	743	2	96 (86-100) ¹	74 (72-77) ²

TP = true-positives; FP = false-positives; TN= true-negatives; FN = false-negatives; Sens = sensitivity; Spec = specificity; numbers in brackets are 95% confidence intervals

¹ difference compared to conventional strategy (P=0.03)

² difference compared to conventional strategy (P<0.001)

However, on the other side this may result in an increase in the rate of false-positive referrals, which will lead to more healthy women undergoing unnecessary additional work-up. In a clinical population, however, most women have breast problems, and sensitivity should be as high as possible. In addition, the prevalence of breast cancer in a clinical population (4.9% in this study) is approximately 10-fold higher compared to a screening population⁹ and the stage of disease is most likely more advanced. Furthermore, lower mammogram specificity could be justified in this clinical setting as additional work-up, like ultrasonography and biopsy, is easily available which could limit the discomfort for patients with a false-positive imaging result.

The use of a consecutive patient series is a strength of this study, as it represents the patient population referred for diagnostic and screening mammography in daily clinical practice. However, the use of special reading sessions in order to evaluate the mammograms can be regarded as a potential limitation, as this does not represent daily practice. The lack of direct clinical consequences of the BI-RADS assessments given, combined with the large volume of mammograms in these reading sessions, may possibly have led to a loss of focus for the observers and a potential decrease in performance. However, in order to provide equal evaluation circumstances for all observers and to monitor the data collection closely, this study design was thought to be the most appropriate.

Although the overall prevalence of breast cancer in this population is consistent with results of other studies in consecutive patients in a clinical setting for mammography (4.1-7.1%)²⁷⁻³¹, the number of 51 cancers is relatively small to

know whether they represent all breast cancers detected in the radiology department in general. The performance of similar studies in different settings may provide more information about the generalisability.

Furthermore, the inclusion of screening examinations in this study may be confusing. It should be noted that in order to study the effects of these evaluation strategies in common daily practice in our radiology departments, both mammograms of diagnostic and screening nature should be included. An examination was considered diagnostic when a problem-solving indication was present, such as symptoms suggestive for breast cancer, whereas examinations in asymptomatic women were considered as screening examinations. However, both exams are technically performed equally and the screening examinations are not performed as part of the Dutch Nationwide Breast Cancer Screening Programme. As only four breast cancer cases were found in the screening group, inclusion of these cases is not expected to bring bias into the results of this study.

The results of the present study demonstrated that systematic application of CAD software in a clinical population failed to improve the performance of both radiologist and technologist readers.

In screening mammography, several studies evaluated the accuracy of single reading with CAD compared to double reading^{18,20,32-34}. A systematic review of Bennett et al.³⁵ showed that the majority of the eight studies included were performed using a selected test set of mammograms. Furthermore, sufficient training in the use of CAD was found to be lacking in several studies and the evaluation of double reading was simulated in all but two studies. As a result, Bennett et al. stated that evidence is limited for either single reading with CAD and double reading screening mammograms. In the study of Gromet¹⁸, a set of over 230,000 screening examinations was evaluated, showing an increase in sensitivity for both double reading and single reading with CAD.

In the present study in a consecutive clinical patient population, it is demonstrated that double reading could increase the cancer detection rate. However, the use of CAD software failed to improve the performance of the observers.

ROC curves can be used to assess the diagnostic accuracy of an observer, independent of the prevalence of the disease. ROC curves of different observers can be visualised in one plot. Furthermore, the area under the curve (AUC) could be evaluated, resulting in a measure of the overall performance of the observer to distinguish between patients with a malignancy and those without a malignancy across the full range of cut-off points. However, the AUC provides no information on the sensitivity and specificity on a single cut-off point. Therefore, the sensitivity and specificity in the different strategies were evaluated using a cut-off point between BI-RADS 1-2 and BI-RADS 0,3-5 which is commonly used in studies evaluating mammography performance^{28,36-37}.

Table 2 shows that the sensitivity and specificity in the conventional strategy in this study were 84% and 94%, respectively. This performance of the radiologist

is comparable to studies of Flobbe et al.²⁸ and Zonderland et al.³⁰ who reported a mammography sensitivity of 83% and a specificity of 92% and 97%, respectively. It needs to be mentioned that the diagnostic value of additional work-up, like ultrasonography and fine needle aspiration cytology, was excluded in the current study. Therefore, the performance of the whole process of breast imaging in daily clinical practice would be higher than presented in this paper.

In a pre-reading strategy, the technologist selects mammograms with negative and clearly benign findings that do not need further work-up, whereas all other patients would require further attention of a radiologist. Table 2 displays the performance in the pre-reading strategy with a cut-off point between BI-RADS 1-2 and BI-RADS 0,3-5, showing a number of ten false-negative results when pre-reading with technologists 1 and 3, and nine false-negative results when pre-reading with technologist 2, compared to 8 false-negative results in the conventional strategy. The number of false-positive results was 38, 41 and 37 using technologist 1, 2 and 3, respectively, compared to 59 in the conventional strategy. Consequently, the mean specificity of 96% in the pre-reading strategy was higher, compared to the conventional strategy (94%). The mean sensitivity of 81% was lower as compared to the conventional strategy (84%), although this was not statistically significant.

Nevertheless, this lower sensitivity could partly be explained by the design of the pre-reading strategy in this study, as the final classification in patients that need additional evaluation by a radiologist, was based on the BI-RADS score of the radiologist. Therefore, all malignancies that failed to be diagnosed in the conventional strategy, are not expected to be observed in the pre-reading strategy. In addition, the technologists would miss malignancies in the patient group that was scored as BI-RADS 1 and 2, leading to a lower sensitivity and lower AUCs in the pre-reading strategy, compared to the conventional strategy. Pre-reading mammograms by technologists is no standard care in daily clinical practice and currently falls outside the legal scope of practice of technologists. In current law in the Netherlands, the supervising radiologists have a final responsibility for the actions of the technologists and the consequences of their actions.

In order to increase the cancer detection rate, it can be argued to use a double reading strategy. This strategy obtained the highest AUC (mean of 0.96) as compared to the AUC of 0.92 in the conventional strategy. Furthermore, Table 2 shows only two false-negative results using technologists 1 and 3, and three false-negative results using technologist 2, resulting in sensitivities of 96% and 94%, respectively. Compared to the conventional strategy, six additional malignancies could be detected, which would increase the cancer detection rate by 14%. On the other hand, the number of false-positive results in the double reading strategy (mean of 211) is much higher compared to 59 in the conventional strategy. As all positive mammography results would be followed by additional

work-up, the application of a double reading strategy leads to an increasing workload for the radiologic staff as well as increased health care costs. It needs to be mentioned, however, that a significant number of the patients with a false-positive result would receive additional imaging anyway, as according to evidence-based guidelines, all patients referred for a palpable breast mass and patients referred with an abnormal screening mammogram from the Dutch Nationwide Breast Cancer Screening Programme, have an indication for an additional ultrasound examination following mammography. However, more research needs to be done in order to study the costs and effects associated with these different reading strategies in a clinical population.

It can be concluded that in a clinical population systematic application of CAD software in the evaluation of mammograms does not improve the diagnostic yield as compared to standard daily practice. Furthermore, the employment of technologists in a double reading strategy in addition to a radiologist for the evaluation of mammograms in a clinical patient population may be an effective approach to improve interpretive performance.

References

1. Feig SA, D'Orsi CJ, Hendrick RE, Jackson VP, Kopans DB, Monsees B, Sickles EA, Stelling CB, Zinnering M, Wilcox-Buchalla P. American College of Radiology guidelines for breast cancer screening. *AJR Am J Roentgenol* 1998; 171:29-33
2. Nystrom L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Ryden S, Andersson I, Bjurstam N, Fagerberg G, Frisell J, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993; 341:973-978
3. Shapiro S. Screening: assessment of current studies. *Cancer* 1994; 74:231-238
4. Tabar L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A, Smith RA. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. *Cancer* 1995; 75:2507-2517
5. Collins MJ, Hoffmeister J, Worrell SW. Computer-aided detection and diagnosis of breast cancer. *Semin Ultrasound CT MR* 2006; 27:351-355
6. Anttinen I, Pamilo M, Soiva M, Roiha M. Double reading of mammography screening films--one radiologist or two? *Clin Radiol* 1993; 48:414-421
7. Brown J, Bryan S, Warren R. Mammography screening: an incremental cost effectiveness analysis of double versus single reading of mammograms. *Bmj* 1996; 312:809-812
8. Thurfjell EL, Lernevall KA, Taube AA. Benefit of independent double reading in a population-based mammography screening program. *Radiology* 1994; 191:241-244
9. Duijm LE, Groenewoud JH, Fracheboud J, de Koning HJ. Additional double reading of screening mammograms by radiologic technologists: impact on screening performance parameters. *J Natl Cancer Inst* 2007; 99:1162-1170
10. Pauli R, Hammond S, Cooke J, Ansell J. Comparison of radiographer/radiologist double film reading with single reading in breast cancer screening. *Journal of Medical Screening* 1996; 3:18-22
11. Tonita JM, Hillis JP, Lim CH. Medical radiologic technologist review: effects on a population-based breast cancer screening program. *Radiology* 1999; 211:529-533
12. van den Biggelaar FJ, Nelemans PJ, Flobbe K. Performance of radiographers in mammogram interpretation: a systematic review. *Breast* 2008; 17:85-90
13. Sumkin JH, Klamon HM, Graham M, Ruskauff T, Gennari RC, King JL, Klym AH, Ganott MA, Gur D. Prescreening mammography by technologists: a preliminary assessment. *AJR Am J Roentgenol* 2003; 180:253-256
14. Haiart DC, Henderson J. A comparison of interpretation of screening mammograms by a radiographer, a doctor and a radiologist: results and implications. *Br J Clin Pract* 1991; 45:43-45
15. Birdwell RL, Bhandokar P, Ikeda DM. Computer-aided detection with screening mammography in a university hospital setting. *Radiology* 2005; 236:451-457
16. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, Berns EA, Cutter G, Hendrick RE, Barlow WE, Elmore JG. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 2007; 356:1399-1409
17. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 2001; 220:781-786
18. Gromet M. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *AJR Am J Roentgenol* 2008; 190:854-859

19. Gur D, Sumkin JH, Rockette HE, Ganott M, Hakim C, Hardesty L, Poller WR, Shah R, Wallace L. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *J Natl Cancer Inst* 2004; 96:185-190
20. Karssemeijer N, Otten JD, Verbeek AL, Groenewoud JH, de Koning HJ, Hendriks JH, Holland R. Computer-aided detection versus independent double reading of masses on mammograms. *Radiology* 2003; 227:192-200
21. Kim SJ, Moon WK, Cho N, Cha JH, Kim SM, Im JG. Computer-aided detection in full-field digital mammography: sensitivity and reproducibility in serial examinations. *Radiology* 2008; 246:71-80
22. Marx C, Malich A, Facius M, Grebenstein U, Sauner D, Pfeleiderer SO, Kaiser WA. Are unnecessary follow-up procedures induced by computer-aided diagnosis (CAD) in mammography? Comparison of mammographic diagnosis with and without use of CAD. *Eur J Radiol* 2004; 51:66-72
23. Taylor P, Potts HW. 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
24. D'Orsi CJ, Bassett LW, Berg WA, Feig SA, Jackson VP, Kopans DB, Linver MN, Mendelson EB, Moss LJ, Sickles EA. *Breast Imaging Reporting and Data System: ACR BI-RADS-Mammography*, 4th ed. Reston (VA): American College of Radiology (ACR), 2003
25. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr* 2007; 96:644-647
26. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol* 2004; 5:11-18
27. Duijm LE, Guit GL, Zaat JO, Koomen AR, Willebrand D. Sensitivity, specificity and predictive values of breast imaging in the detection of cancer. *Br J Cancer* 1997; 76:377-381
28. Flobbe K, Bosch AM, Kessels AG, Beets GL, Nelemans PJ, von Meyenfeldt MF, van Engelshoven JM. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. *Arch Intern Med* 2003; 163:1194-1199
29. Flobbe K, van der Linden ES, Kessels AG, van Engelshoven JM. Diagnostic value of radiological breast imaging in a non-screening population. *Int J Cancer* 2001; 92:616-618
30. Zonderland HM, Coerkamp EG, Hermans J, van de Vijver MJ, van Voorthuisen AE. Diagnosis of breast cancer: contribution of US as an adjunct to mammography. *Radiology* 1999; 213:413-422
31. Zonderland HM, Pope TL, Jr., Nieborg AJ. The positive predictive value of the breast imaging reporting and data system (BI-RADS) as a method of quality assessment in breast imaging in a hospital population. *Eur Radiol* 2004; 14:1743-1750
32. Georgian-Smith D, Moore RH, Halpern E, Yeh ED, Rafferty EA, D'Alessandro HA, Staffa M, Hall DA, McCarthy KA, Kopans DB. Blinded comparison of computer-aided detection with human second reading in screening mammography. *AJR Am J Roentgenol* 2007; 189:1135-1141
33. Khoo LA, Taylor P, Given-Wilson RM. Computer-aided detection in the United Kingdom National Breast Screening Programme: prospective study. *Radiology* 2005; 237:444-449
34. Gilbert FJ, Astley SM, McGee MA, Gillan MG, Boggis CR, Griffiths PM, Duffy SW. 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

35. 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; 61:1023-1028
36. Molins E, Macia F, Ferrer F, Maristany MT, Castells X. Association between radiologists' experience and accuracy in interpreting screening mammograms. *BMC Health Serv Res* 2008; 8:91
37. Balleyguier C, Kinkel K, Fermanian J, Malan S, Djen G, Taourel P, Helenon O. Computer-aided detection (CAD) in mammography: does it help the junior or the senior radiologist? *Eur J Radiol* 2005; 54:90-96

8

General discussion and conclusions

Worldwide, mammography is the primary modality for early detection and diagnosis of breast malignancies. However, in the setting of an increasing demand for breast imaging and an increasing shortage of well-experienced radiologists, it is hard to assure the high efficacy and quality of breast imaging in daily clinical practice. In order to deal with this issue, new approaches in the evaluation of mammograms in a clinical patient population could be considered. In this thesis, the deployment of specialised breast technologists in reading mammograms and the application of computer-aided detection (CAD) software in digital mammography are discussed.

In a systematic review of the literature (chapter 2), it was shown that all studies included were performed in screening mammograms. No studies were found evaluating the performance of technologists in reading mammograms in a clinical patient population. In response to this lack of evidence in the literature, the PERSPECT study was performed in which two specialised breast technologists were trained in mammogram reading. Subsequently, their performance was evaluated in a prospective effectiveness study. The results of this study were presented in chapter 3. The two technologists evaluated 1994 digital mammograms. A sensitivity rate of 90% in the detection of breast cancer in the overall study population was found. In addition, chapter 4 demonstrated that pre-reading mammograms by technologists in order to select suspicious cases could provide costs savings up to 17.2% without jeopardising the cancer detection rate, as compared to regular mammogram reading in daily clinical practice. Furthermore, legal implications of using technologists as independent pre-readers were discussed, showing that pre-reading could be effectuated when the criteria are met for performing reserved procedures by non-authorised professionals as stated in the Individual Health Care Professions Act (chapter 5). Another intervention to improve the evaluation of mammograms could be the application of computer-aided detection (CAD) software which has the potential to increase the detection of malignancies. In chapter 6, the impact of a CAD system on the performance of mammogram readers in interpreting digital mammograms in a clinical patient population was evaluated. It was shown that application of CAD in 1048 consecutive patients failed to improve the sensitivity in detecting breast malignancies and was associated with an increase in false-positive results. In chapter 7, different strategies of mammogram reading with technologists and CAD software were compared, using ROC analysis. It was shown that the employment of a technologist as double reader in addition to a radiologist obtained the highest overall performance (AUC=0.96), compared to standard single reading by a radiologist (AUC=0.92) and pre-reading by a technologist (AUC=0.91). The use of CAD did not increase the diagnostic yield. Based on the findings presented in this thesis, it was concluded that breast technologists could be employed as double or pre-readers of mammograms in a clinical patient population without jeopardising the detection of malignancies. Furthermore, it was shown that the application of CAD software did not improve the detection of breast malignancies in daily clinical practice.

Performance in mammogram reading

In the studies discussed in this thesis, the performance of the technologists is compared to standard mammogram reading by a single radiologist. In the overall clinical patient population (n=1994) as described in chapter 3, the sensitivity in the detection of malignancies was 81% for the radiologist and 90% for the technologists (McNemar, P value=0.02). However, the higher sensitivity comes at the cost of a lower specificity of the technologists which was significantly different as compared to the specificity of the radiologist (82% and 96%, respectively; McNemar, P value<0.001). In addition, the Diagnostic Odds Ratio (DOR) which combines sensitivity and specificity into one measure was higher for the radiologist (DOR score of 102) as compared to the technologists (DOR score of 41). This means that the overall performance in mammogram evaluation was better for the radiologist as compared to the technologists.

These findings are in accordance with results found in the literature. The performance of the radiologists is comparable to the performance found in other studies on similar patient populations¹⁻². Furthermore, the systematic literature review presented in chapter 2, showed that technologists could be as sensitive as radiologists in detecting malignancies, but with more false-positive results. Apparently, the technologists involved in our studies were more cautious and uncertain in their evaluation which did lead to a lower specificity. Furthermore, as the technologists evaluated the mammograms under experimental conditions, a false-positive result would not lead to unnecessary additional work-up and emotional burden for the patient. This may have resulted in more positive mammogram evaluations for the technologists as compared to the radiologists.

It should be noted that, in this thesis, the performance of the technologists and radiologists is based on mammogram interpretation alone, without including the results of additional diagnostic testing. In daily clinical practice, radiologists will also use results of additional imaging like ultrasonography and fine needle aspiration cytology, in their evaluation of the patient, whereas during this study the technologists were only trained in mammogram reading. Therefore, more subtle findings of malignancy on the mammogram might be mentioned by the technologists as compared to the radiologist which could result in a higher detection rate for the technologists. Furthermore, it need to be mentioned that the radiologists participated in these studies are well-experienced but not necessarily the most or least expert observers. The performance of the radiologist can be assumed to be an average performance resulting from clinical practice in general.

Use of the BI-RADS

In this thesis, the Breast Imaging Reporting And Database System (BI-RADS) was used as a reporting system for breast imaging³. This system is designed by the American College of Radiology in order to standardise breast imaging reporting and to facilitate outcome monitoring. The BI-RADS lexicon includes seven standardised assessment categories. In our studies, the following six

categories were used: 0 (need additional imaging evaluation), 1 (negative examination), 2 (benign finding), 3 (probably benign finding), 4 (suspicious abnormality), 5 (highly suggestive of malignancy). According to the BI-RADS lexicon, patients with BI-RADS category 6 in whom mammography was performed of a proven malignancy for monitoring of responses to neoadjuvant chemotherapy prior to surgical excision, should be excluded from measurement of outcome parameters.

Although the BI-RADS guidelines aim to standardise the terminology in mammography reporting, literature shows that variability in the assessment of findings and in management recommendations is still existing. Particularly, the use of BI-RADS category 3 has been the subject of debate. According to the BI-RADS lexicon, BI-RADS 3 should only be given after complete diagnostic imaging evaluation in probably benign findings with a risk of malignancy less than 2%. Furthermore, in these patients, a short-term follow-up (six months) is recommended for confirmation. However, studies have shown that this recommendation is not always in concordance with the work-up advised in daily clinical practice. In a study of Geller et al.⁴ in 51,673 diagnostic mammograms, category 3 cases had the most variability in work-up recommendations. In only 40% of these patients, a short term follow-up was advised. Taplin et al.⁵ found similar results in 292,795 screening mammograms.

Also in our study populations with clinical mammograms made for both screening and diagnostic purposes, the use of BI-RADS 3 was found to be inconsistent with the BI-RADS lexicon. In more than 15% of the patients, the technologists scored a BI-RADS 3, compared to 4% of the radiologist. The prevalence of breast cancer in these patients was 4% for the technologists and 21% for the radiologist. Furthermore, short-term follow-up was recommended by all observers in less than 10% of the BI-RADS 3 patients, whereas ultrasound work-up was advised in more than 90%.

The incompatibility with the BI-RADS lexicon may be explained by the design of the studies. According to the BI-RADS lexicon, evaluation of further diagnostic work-up, like additional mammographic images or ultrasonography, should be taken into consideration for a complete assessment. However, the observers in our studies were asked to assign a BI-RADS assessment category based on the findings on the two-view mammogram only. In order to obtain "final" assessments as much as possible, observers were stimulated to use BI-RADS 1-5, combined with a recommendation for additional work-up. Therefore, it could be assumed that a significant number of BI-RADS 3 should actually be a BI-RADS 0 (need additional imaging evaluation and/or prior mammograms for comparison).

Furthermore, it can be assumed that the technologists were indecisive on many cases. Probably, this resulted in a high number of BI-RADS 3 evaluations for the technologists.

Pre-reading versus double reading of mammograms

The employment of breast technologists in evaluating mammograms in a clinical patient population could be effectuated through pre-reading or double reading procedures.

In a double reading strategy, two observers evaluate the mammograms independently. Discrepant readings could be managed in different ways: a patient may require for additional work-up when only one observer considers the mammogram positive, without discussion of findings; a consensus between the two observers may be reached about the decision whether additional work-up is required; or a third observer may decide⁶.

In chapter 7 of this thesis, a double reading strategy with a radiologist and a technologist is described. The highest BI-RADS assessment of either the radiologist or the technologist was used for analysis and discrepant readings were not discussed. The mean sensitivity in detecting malignancies in this strategy was 95%. Furthermore, it was shown that the overall performance as represented in an AUC in this strategy is 0.96, compared to an AUC of 0.91 in a pre-reading strategy by technologists or an AUC of 0.92 in a strategy of standard mammogram reading by one radiologist. However, double reading in daily clinical practice would raise the number of unnecessary additional work-up and is no solution for the increasing workload of the radiologists.

Therefore, it could be argued to employ technologists as pre-readers of mammograms. Based on the findings described in this thesis, it can be concluded that technologists were successful in selecting mammograms with negative or clearly benign findings (BI-RADS 1 and 2), which could relieve the radiologist from evaluating these mammograms. Furthermore, chapter 4 showed that pre-reading by technologists can reduce diagnostic costs considerably without jeopardising the detection rate of malignancies.

In conclusion, this thesis presented that double reading of mammograms with a radiologist and a technologist can increase the detection of malignancies (at the price of high false-positive rates), but does not decrease the workload of radiologists. The method of pre-reading mammograms by a technologist can reduce the workload of radiologists and diagnostic costs.

Moreover, our studies show that the sensitivity rates of the technologists were higher compared to sensitivity in daily clinical practice. Therefore, it could be reasoned to use technologists in assisting radiologists in order to avoid oversights.

In order to take the best of the two reading strategies, it could be argued to implement a technologist pre-reading strategy with double reading of the positive evaluations of the technologists instead of classifying these positive mammograms by a single radiologist only, as evaluated in this thesis. The decision whether to perform additional work-up should then be based on the highest BI-RADS assessment or should be made in a consensus meeting with radiologist and technologist. Such a strategy has the potential to increase the

detection rate of malignancies in daily clinical practice, to reduce the workload of the radiologists and to reduce diagnostic costs.

Implementation of pre-reading in daily clinical practice

Based on the results of this thesis, it could be argued to deploy technologists in pre-reading mammograms in daily clinical practice. For an effective implementation, a few issues need to be addressed.

First, technologists who participate as pre-readers, should be well trained in the evaluation of images by supervising radiologists who are well-experienced in breast imaging. The two technologists involved in the PERSPECT study, received about 500 hours of specific training in mammogram interpretation during nine months. Training consisted of evaluation of up-to-date literature, daily reading of mammograms and evaluation of difficult cases with a specialised breast radiologist. Furthermore, the technologists participated in pathology and oncology meetings, attended at mammography symposia and received practical and theoretical training in external medical centres. Evaluation of the training programme showed that the technologists had substantially increased their knowledge and performance in mammogram interpretation. Therefore, it could be stated that an intensive training programme is a precondition for the employment of technologists in evaluating mammograms. Additionally, attending refresher courses are needed for regular preservation of interpretation skills. Furthermore, the involvement of technologists that are well-experienced in performing mammograms in clinical practice, like in our studies, may be most feasible.

Second, in order to reduce the workload of the radiologists, patients with negative or clearly benign findings should be discharged autonomously by the technologists. Therefore, in addition to the competence of evaluating mammograms, technologists have to be skilled in recording the findings of the mammogram in a breast imaging report. Furthermore, in order to improve efficacy of practice even more, it could be reasoned to expand the activities of the technologist with the independent performance of ultrasonography. For an independent deployment of technologists, legal preconditions must be met. It is recommended to lay down all responsibilities and liabilities in an unambiguous protocol. For a wide implementation of the concept of pre-reading, adaptation of legislation is needed.

Third, the deployment of technologists as mammogram readers must fit in the organisational structures. The technologists must be scheduled for pre-reading which would reduce their participation in regular technologist's duties. In order to achieve a continuous availability, a minimum number of two technologists should be trained as pre-readers. Furthermore, a radiologist should be available for consultation and to perform additional work-up at any time. This implies that the concerning radiologists should be involved in other activities that may be interrupted easily. In daily clinical practice, this continuous availability of a radiologist could be a serious limitation for an effective implementation of the

deployment of technologists as mammogram readers, as this may not result in a decreased workload of radiologists and a reduction in costs.

Finally, dedication and support of motivated radiologists and technologists are essential for this innovation to be successful.

Quality monitoring and assurance

In order to consolidate the level of knowledge and skills of the technologists and radiologists, their performance should be monitored thoroughly. A possible design of monitoring in daily clinical practice is through a continuous computer registration of procedures and outcomes on a local, regional or national level. The findings in the breast imaging report should be linked to the pathology results of the patient in order to determine the performance of the individual mammogram reader and the inter-reader variability. When a false-negative result is found by the registration system, the observer involved should be alerted by the system. In addition, true-positive, false-positive and false-negative results should be discussed on a regular basis with professionals involved in breast imaging.

Future prospects

Although our studies showed promising results for the deployment of technologists in reading mammograms, this is based on data obtained in a single hospital location with a relatively small number of professionals and patients involved. In order to further evaluate the performance of technologists in daily clinical practice, multi-centre trials are needed. Research should be focused on the implementation of technologist reading strategies.

In addition, more research needs to be done on legal consequences of evaluating mammograms by technologists. Professional associations of radiologists and technologists should stimulate an effective deployment of technologists in breast imaging by formulating protocols and by developing training programmes.

Conclusions of this thesis

The conclusions of the studies discussed in this thesis, are:

- Based on the literature, only a few well-designed studies have been conducted evaluating the performance of technologists in interpreting screening mammograms. So far, no studies were reported assessing the performance of technologists in reading mammograms in a clinical patient population (chapter 2).
- In a clinical study population of 1994 consecutive patients, specialised breast technologists are well capable of evaluating mammograms (chapter 3).
- Pre-reading mammograms by technologists in a clinical patient population is an effective approach in order to reduce the workload of radiologists without jeopardising the detection of breast malignancies (chapter 3 + chapter 4).

- Pre-reading mammograms by technologists will reduce diagnostic costs (chapter 4).
- Pre-reading mammograms by technologists could be effectuated when the criteria are met for performing reserved procedures by non-authorised professionals as stated in the Individual Health Care Professions Act (chapter 5).
- The application of CAD software in mammography in a clinical patient population failed to improve the cancer detection rate (chapter 6).
- In order to increase the detection of breast cancer in a clinical population, the employment of a technologist as double reader in addition to a radiologist is an effective strategy (chapter 7).
- A technologist pre-reading strategy with double reading of suspicious mammograms would be most effective in daily clinical practice (chapter 8).

References

1. Flobbe K, Bosch AM, Kessels AG, Beets GL, Nelemans PJ, von Meyenfeldt MF, et al. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. *Arch Intern Med* 2003; 163:1194-1199
2. Zonderland HM, Coerkamp EG, Hermans J, van de Vijver MJ, van Voorthuisen AE. Diagnosis of breast cancer: contribution of US as an adjunct to mammography. *Radiology* 1999; 213:413-22
3. D'Orsi CJ, Bassett LW, Berg WA, Feig SA, Jackson VP, Kopans DB, et al. Breast Imaging Reporting and Data System: ACR BI-RADS-Mammography. 4th ed. Reston (VA): American College of Radiology (ACR), 2003
4. Geller BM, Barlow WE, Ballard-Barbash R, Ernster VL, Yankaskas BC, Sickles EA, et al. Use of the American College of Radiology BI-RADS to report on the mammographic evaluation of women with signs and symptoms of breast disease. *Radiology* 2002; 222:536-542
5. Taplin SH, Ichikawa LE, Kerlikowske K, Ernster VL, Rosenberg RD, Yankaskas BC, et al. Concordance of breast imaging reporting and data system assessments and management recommendations in screening mammography. *Radiology* 2002; 222:529-535
6. Duijm LE, Louwman MW, Groenewoud JH, van de Poll-Franse LV, Fracheboud J, Coebergh JW. Inter-observer variability in mammography screening and effect of type and number of readers on screening outcome. *Br J Cancer* 2009; 100:901-917

9

Summary

Samenvatting

Dankwoord

Curriculum Vitae

List of publications

Summary

With an incidence of more than 12,000 cases in the Netherlands in 2006, breast cancer is the most common malignancy among women. In the setting of an increasing demand for breast imaging due to rising breast cancer incidence and prevalence rates combined with an increasing shortage of well-experienced radiologists, it is getting harder to assure the high efficacy and quality of breast imaging evaluation in daily clinical practice. In order to deal with this issue, new approaches in the evaluation of mammograms in a clinical patient population could be considered.

Therefore, this thesis addresses the feasibility of the deployment of specialised breast technologists in evaluating mammograms in a clinical patient population. Reader performance, different reading strategies, costs and legal implications were studied. Furthermore, the impact of CAD software in digital mammography on the performance of mammogram readers was evaluated.

In chapter 2, a systematic literature review was performed, focusing on the performance of technologists compared with radiologists in the interpretation of mammograms; the effect of training; and the question whether there are any studies evaluating the effects of involving technologists in the interpretation of diagnostic mammograms in daily clinical practice on the sensitivity and specificity of cancer detection in breast imaging. Six studies met the inclusion criteria (primary aim of the study has to be the evaluation of the performance of technologists, sensitivity and specificity have to be reported or calculable and there has to be a sufficient gold standard). The results showed that, in a screening setting, technologists scored higher false-positive rates with a similar sensitivity in the detection of malignancies, compared with radiologists. Furthermore, results suggested that training could improve their performance. No studies were reported assessing the performance of technologists interpreting diagnostic mammograms in a consecutive patient population in a daily clinical setting. This indicates a need for a well-designed diagnostic study using an adequate gold standard.

Chapter 3 presents the results of such a study, the PERSPECT study (PERformance of SPECialised breast Technologists in breast imaging), performed at Maastricht University Medical Center (MUMC). The study focused on the feasibility of skill mixing in breast imaging in a hospital radiology department. During nine months, two technologists with special interest in mammography were trained in mammogram reading under the supervision of specialised breast radiologists. Subsequently, a prospective study was performed in which 1994 consecutive women referred for mammography between January and August 2007, were evaluated by the technologists, next to the standard clinical interpretation by the radiologist on duty. Mammographic findings were recorded and a BI-RADS classification was assigned for each breast. A BI-RADS Assessment Category indicates the opinion of the observer regarding the level of suspicion of a breast

abnormality. The scoring system consists of six classifications: 1 = negative examination; 2 = benign finding; 3 = probably benign finding; 4 = suspicious abnormality; 5 = highly suggestive of malignancy; 6 = proven malignancy. Furthermore, BI-RADS 0 is assigned when additional imaging evaluation or comparison to prior mammograms is required to make a final assessment.

In the study population of 1994 mammograms, 93 breast cancers were found in 91 patients, leading to a prevalence of 4.6% (91/1994). The sensitivity and specificity in selecting mammographic findings (cut-off point between BI-RADS 1 and BI-RADS 0,2-5 and the radiologist's diagnosis as reference standard) was 98% and 74% for technologist 1 and 98% and 78% for technologist 2. In distinguishing normal and benign mammograms from those with abnormalities that are probably benign, suspicious or highly suggestive for malignancy (cut-off point BI-RADS 1-2 and BI-RADS 0,3-5 and pathology results as reference standard), the sensitivity decreased to 89% and 91%, respectively. The specificity, however, increased to 82% for both technologists.

In a subgroup of 1389 patients with no immediate indication for additional imaging with the involvement of a radiologist, technologists obtained a mean sensitivity and specificity of 98% and 77% in detecting mammographic findings, and a mean sensitivity and specificity of 78% and 88% in detecting suspicious abnormalities.

Concluding, the employment of technologists in pre-reading mammograms seems to be an effective working strategy in daily clinical practice. However, its position in clinical practice remains indistinct as a continuous availability of radiologists still needs to be guaranteed. Nevertheless, as a substantial proportion of mammograms could be evaluated without the attention of a radiologist, the employment of technologists in pre-reading mammograms seems a promising new working strategy.

In order to study the costs and effects of using specialised breast technologists in pre-reading mammograms, a decision analytic model which was based on the mammogram evaluations obtained in the PERSPECT study, is presented in chapter 4. Mammograms of 1389 patients without immediate indication for additional imaging were included in the model. The costs and effects of four different experimental strategies of pre-reading mammograms by technologists were analysed and compared to the conventional strategy of standard evaluation by the radiologist on duty.

Overall, the employment of technologists in this patient population resulted in a potential time saving up to 73% (1019/1389) for the radiologist. No additional false-negative imaging results were found as compared to the conventional strategy. The total diagnostic costs in the conventional strategy were determined at €150,602. The experimental strategies resulted in cost savings up to 17.2% (range €122,494–€139,781).

It can be concluded that the employment of technologists in pre-reading mammograms in a clinical patient population could be effective to reduce the work-

load of radiologists without jeopardising the detection of malignancies. Furthermore, diagnostic costs can be reduced considerably.

Chapter 5 focuses on the legal implications in terms of duties and responsibilities for radiologists and technologists of independent pre-reading of mammograms by technologists. Although pre-reading is not a reserved procedure, it could be considered as a potentially risky procedure. Therefore, the criteria for performing reserved procedures by non-authorized professionals as stated in the Individual Health Care Professions Act (in Dutch: Wet op de Beroepen in de Individuele Gezondheidszorg) should be applied. Furthermore, compliance with a protocol or code of conduct in combination with adequate training and supervision should be sufficient to disprove potential claims.

It is likely that pre-reading mammograms by technologists will expand from a local initiative. For a wide implementation, pre-reading should be well-embedded in legal rules and should answer the professional standard of care. In addition, in order to protect patients, the legislator should consider to formulate general conditions that skill mixing initiatives of both potentially risky procedures as well as reserved procedures should meet.

Beside the employment of technologists in reading mammograms, another innovation that has the potential to improve the quality of the evaluation of mammograms, is the application of computer-aided detection (CAD) software. In chapter 6, the impact of a CAD system on the performance of mammogram readers in interpreting digital mammograms in a clinical population was evaluated. Furthermore, the ability of a CAD system to detect breast cancer in digital mammography was studied in comparison to the performance of radiologists and technologists as mammogram readers.

Digital mammograms of 1048 consecutive patients were evaluated by a radiologist and three technologists. Abnormalities were recorded and an imaging conclusion was given as a BI-RADS score before and after CAD analysis. Pathology results during 12 months follow-up were used as a reference standard for breast cancer. Fifty-one malignancies were found in fifty patients. Sensitivity and specificity were computed before and after CAD analysis and provided with 95% CIs. In order to assess the detection rate of malignancies by CAD and the observers, the pathological locations of the 51 breast cancers were matched with the locations of the CAD marks and the mammographic locations that were considered to be suspicious by the observers.

For all observers, the sensitivity rates did not change after application of CAD. A mean sensitivity of 92% was found for all technologists and 84% for the radiologist. For two technologists, the specificity decreased (from 84% to 83% and from 77% to 75%). For the radiologist and one technologist, the application of CAD did not have any impact on the specificity rates (95% and 83%, respectively). CAD detected 78% of all malignancies. Five malignancies were indicated by CAD without being noticed as suspicious by the observers.

In conclusion, the results show that systematic application of CAD in a clinical patient population failed to improve the overall sensitivity of mammogram interpretation by the readers and was associated with an increase in false-positive results. However, CAD marked five malignancies that were missed by the different readers.

In chapter 7, the employment of technologists as either pre-readers or double readers of mammograms and the impact of CAD software on the performance of mammogram readers were evaluated and compared to the conventional strategy of single reading by a radiologist. In order to study the effects of the different strategies, ROC analysis was used.

Digital mammograms of 1048 consecutive patients were evaluated by a radiologist and three technologists with and without the use of CAD software. In the conventional strategy, an overall area under the ROC curve (AUC) of 0.92 was found, corresponding to a sensitivity of 84% and a specificity of 94%. When applying CAD software, the AUCs were similar before and after CAD for all readers (mean of 0.95). Employing technologists in pre-reading and double reading of mammograms resulted in a mean AUC of 0.91 and 0.96, respectively. In the pre-reading strategy, the corresponding sensitivity and specificity were 81% and 96%; in the double reading strategy they were 96% and 79%, respectively.

Concluding, in this clinical population systematic application of CAD software by either radiologist or technologists failed to improve the diagnostic yield. Furthermore, employing technologists as double readers of mammograms was the most effective strategy to improve breast cancer detection in daily clinical practice.

Finally, in chapter 8 the studies presented in this thesis are reviewed and discussed. It can be concluded that the application of CAD software failed to improve the cancer detection rate in digital mammograms in a clinical patient population. Furthermore, it was shown that the employment of technologists can be effectuated through pre-reading and double reading procedures. Pre-reading mammograms can be effective to reduce the workload of radiologists and diagnostic costs without jeopardising the detection of breast malignancies in daily clinical practice whereas double reading can be effective to increase the cancer detection in a clinical patient population.

Samenvatting

Borstkanker is de meest voorkomende kankersoort bij vrouwen in Nederland. In 2006 zijn meer dan 12.000 nieuwe gevallen ontdekt. Doordat mammacarcinomen steeds vroeger ontdekt worden en er steeds meer vrouwen de ziekte overleven door verbeterde behandelingen, neemt het aantal vrouwen dat ooit borstkanker heeft gehad toe. In combinatie met een groeiend tekort aan ervaren radiologen wordt het steeds moeilijker om een hoge kwaliteit van beeldvormend onderzoek van de borsten te behouden.

Mammografie (onderzoek van de borsten door middel van röntgenstraling) is op dit moment het belangrijkste hulpmiddel om borstkanker in een vroeg stadium te ontdekken. Naast mammografie die aangeboden wordt in het kader van het bevolkingsonderzoek naar borstkanker (screeningsmammografie), wordt mammografie ook in het ziekenhuis verricht (voornamelijk zogenaamde diagnostische mammografie bij vrouwen met klachten).

In dit proefschrift worden nieuwe benaderingen in de beoordeling van mammogrammen in een ziekenhuissetting beschreven. Er wordt ingegaan op het inzetten van radiodiagnostisch laboranten in het beoordelen van mammogrammen en de toepassing van software als hulpmiddel in het detecteren van maligniteiten (kwaadaardige afwijkingen).

Hoofdstuk 2 beschrijft een systematisch literatuuronderzoek dat ingaat op de prestaties van radiodiagnostisch laboranten in het beoordelen van mammogrammen vergeleken met de prestaties van radiologen. Ook het effect van training van laboranten is geanalyseerd, evenals de vraag of er studies zijn die de effecten meten van het inzetten van laboranten in de dagelijkse klinische praktijk in een ziekenhuissetting. Zes studies zijn meegenomen in de review. De resultaten laten zien dat laboranten in de screeningsmammografie een gelijke sensitiviteit hebben in het detecteren van maligniteiten als radiologen, maar met een lagere specificiteit. De sensitiviteit geeft aan welk percentage van de bewezen maligniteiten wordt gevonden door de beoordelaar van het mammogram. De specificiteit betreft het percentage van alle patiënten zonder een bewezen maligniteit dat ook als zodanig is gekwalificeerd door de beoordelaar van het mammogram.

Uit het literatuuronderzoek blijkt verder dat het trainen van laboranten hun prestaties kan verbeteren. Er is geen enkele studie gevonden die de prestaties van laboranten in het beoordelen van mammogrammen in de dagelijkse klinische praktijk heeft bestudeerd. Om de mogelijkheden te onderzoeken om laboranten in te zetten in het beoordelen van mammogrammen in een ziekenhuissetting dient een goed opgezette diagnostische studie uitgevoerd te worden.

Een dergelijke studie, de PERSPECT study (PERformance of SPECialised breast Technologists in breast imaging), is uitgevoerd in het academisch ziekenhuis Maastricht (azM). In deze studie zijn twee radiologisch laboranten met interesse

voor mammografie getraind in het beoordelen van mammogrammen en is prospectief gekeken hoe zij presteren in een klinische setting.

De resultaten van de studie worden beschreven in hoofdstuk 3. Mammogrammen van 1994 vrouwen die tussen januari en augustus 2007 zijn doorverwezen voor mammografie naar de afdeling radiologie, zijn beoordeeld door de dienstdoende radioloog en de twee laboranten. Afwijkingen op het mammogram zijn geregistreerd en aan iedere borst is een BI-RADS score toegekend. Een BI-RADS score geeft de mening aan van de beoordelaar over de mate van kwaadaardigheid van een borstafwijking: 1 = normaal mammogram / geen afwijkingen; 2 = zeker benigne (goedaardige) afwijking; 3 = waarschijnlijk benigne afwijking; 4 = verdacht maligne (kwaadaardige) afwijking; 5 = zeker maligne afwijking. Daarnaast kan een BI-RADS 0 worden gegeven wanneer extra onderzoek of een vergelijking met eerder gemaakte mammogrammen nodig is om tot een uiteindelijke beslissing te komen.

In totaal zijn 93 maligniteiten gevonden bij 91 patiënten (prevalentie van 4.6%). In het selecteren van mammografische bevindingen (afkappunt tussen BI-RADS 1 en BI-RADS 0,2-5 en de beoordelingen van de radioloog als gouden standaard [=referentie standaard]) hadden de laboranten een sensitiviteit van 98%. De specificiteit was 74% voor laborant 1 en 78% voor laborant 2. Bij het onderscheiden van zeker goedaardige van meer kwaadaardige afwijkingen (afkappunt tussen BI-RADS 1-2 en BI-RADS 0,3-5 en pathologie uitslagen gedurende een jaar follow-up als gouden standaard) daalde de sensitiviteit naar 89% bij laborant 1 en 91% bij laborant 2. De specificiteit steeg echter tot 82% bij beide laboranten.

In een subpopulatie van 1389 patiënten die volgens richtlijnen geen indicatie hadden voor echografisch vervolgonderzoek bij binnenkomst op de afdeling, werd een gemiddelde sensitiviteit van 98% en een gemiddelde specificiteit van 77% gevonden in het selecteren van mammografische bevindingen en een gemiddelde sensitiviteit van 78% en een gemiddelde specificiteit van 88% in het onderscheiden van zeker benigne van meer kwaadaardige afwijkingen.

Geconcludeerd kan worden dat het inzetten van laboranten in het voorselecteren van mammogrammen een effectieve strategie kan zijn in de dagelijkse klinische praktijk. Precieze toepasbaarheid blijft echter onduidelijk omdat een continue beschikbaarheid van de radioloog gegarandeerd moet zijn. Desondanks zou een aanzienlijk deel van de mammogrammen beoordeeld kunnen worden zonder tussenkomst van een radioloog, waardoor het inzetten van laboranten in het voorselecteren van mammogrammen een veelbelovende nieuwe werkwijze kan worden.

Om de kosten en effecten van het inzetten van laboranten in het voorselecteren van mammogrammen te bestuderen is in hoofdstuk 4 een besliskundig model uiteengezet dat is gebaseerd op de mammogram beoordelingen die in de PERSPECT studie verkregen zijn.

Mammogrammen van 1389 patiënten zonder een indicatie voor echografisch vervolgonderzoek bij binnenkomst op de afdeling zijn meegenomen. De kosten en effecten van vier verschillende experimentele strategieën van het voorselecteren van mammogrammen door laboranten zijn geanalyseerd en vergeleken met de gebruikelijke strategie van standaard beoordeling door de dienstdoende radioloog.

Het inzetten van laboranten in deze patiëntenpopulatie kan leiden tot een potentiële tijdsbesparing tot 73% voor de radioloog (1019 van de 1389 mammogrammen hoeven niet door de radioloog beoordeeld te worden). Er zijn in vergelijking met de gebruikelijke strategie geen extra fout-negatieve mammogram beoordelingen (gemiste maligniteiten) gevonden. De totale diagnostische kosten in de gebruikelijke strategie kwamen neer op €150.602. Het toepassen van een experimentele strategie kan resulteren in kostenbesparingen tot 17% (range in totale kosten €122.494-€139.781).

Concluderend kan gesteld worden dat het inzetten van laboranten in het voorselecteren van mammogrammen in een klinische patiëntenpopulatie effectief kan zijn om de werklust van de radiologen te verminderen zonder het ontdekken van maligniteiten in gevaar te brengen. Daarnaast kunnen diagnostische kosten aanzienlijk gereduceerd worden.

Hoofdstuk 5 gaat in op de juridische voorwaarden in termen van aansprakelijkheid en verantwoordelijkheden om het zelfstandig voorselecteren van mammogrammen door laboranten mogelijk te maken, zodat patiënten kunnen worden ontslagen zonder gezien te zijn door een radioloog. Hoewel voorselectie van mammogrammen geen voorbehouden handeling is, kan het worden aangemerkt als een potentieel risicovolle handeling. Daarom dienen de criteria voor het overdragen van voorbehouden handelingen aan niet zelfstandig bevoegden zoals beschreven in de Wet Beroepen in de Individuele Gezondheidszorg (Wet BIG), te worden toegepast.

Daarnaast zou het naleven van een protocol of gedragscode in combinatie met een deugdelijke training en toezicht voldoende moeten zijn om potentiële claims te kunnen weerleggen.

Het is waarschijnlijk dat het concept van het voorselecteren van mammogrammen door laboranten zich uitbreid vanuit een lokaal initiatief. Voor een verdere effectuering dient het concept te worden vastgelegd in wetgeving en moet het passen binnen de professionele standaard. Ter bescherming van patiënten kan de wetgever daarnaast overwegen om algemene voorwaarden te formuleren voor het overdragen van potentieel risicovolle en voorbehouden handelingen.

Naast het inzetten van laboranten in het beoordelen van mammogrammen is het toepassen van computersoftware (computer-aided detection system, CAD systeem) een innovatie die mogelijk de kwaliteit in het beoordelen van mammogrammen verbetert.

Hoofdstuk 6 beschrijft de invloed van een dergelijk CAD systeem om verdachte afwijkingen op digitale mammogrammen aan te duiden op de prestaties van beoordelaars in een klinische populatie. Daarnaast wordt de bekwaamheid van CAD in het aantonen van maligniteiten getest en afgezet tegen de prestaties van radiologen en laboranten.

Digitale mammogrammen van 1048 opeenvolgende patiënten zijn beoordeeld door een radioloog en drie laboranten. Afwijkingen op het mammogram zijn geregistreerd en aan iedere borst is een BI-RADS score toegekend vóór en ná het toepassen van het CAD systeem. Als gouden standaard voor de aanwezigheid van borstkanker zijn alle pathologie resultaten gedurende één jaar gebruikt. Er werden 51 maligniteiten gevonden in vijftig patiënten. Sensitiviteit en specificiteit werden berekend vóór en ná het toepassen van CAD. Om de prestaties in het ontdekken van maligniteiten van de beoordelaars en het CAD systeem te kunnen bepalen zijn de locaties van de 51 kankers zoals die vermeld staan in het pathologieverslag, vergeleken met de markers die zijn afgegeven door het CAD systeem en de locaties op het mammogram die als verdacht zijn aangewezen door de beoordelaars.

Bij alle beoordelaars was de sensitiviteit in het ontdekken van maligniteiten gelijk vóór en ná het toepassen van CAD. Een gemiddelde sensitiviteit van 92% werd gevonden voor de laboranten en een sensitiviteit van 84% voor de radioloog. De specificiteit werd lager bij twee laboranten na toepassing van CAD (van 84% naar 83% en van 77% naar 75%). Bij de radioloog en één laborant had CAD geen invloed op de specificiteit (respectievelijk 95% en 83%). Het CAD systeem detecteerde 78% van alle maligniteiten. Vijf maligniteiten werden gevonden door CAD zonder te worden beoordeeld als verdacht door de beoordelaars.

Er kan geconcludeerd worden dat systematische toepassing van CAD in een klinische patiëntenpopulatie de sensitiviteit in het detecteren van maligniteiten niet verbeterd en dat het leidt tot meer fout-positieve beoordelingen (onterecht als maligniteit aangemerkt). Niettemin werden vijf maligniteiten door het CAD systeem gemarkeerd die gemist werden door de verschillende beoordelaars.

In hoofdstuk 7 worden het inzetten van laboranten in het voorselecteren danwel dubbel beoordelen van mammogrammen in aanvulling op de beoordeling van de radioloog, en de invloed van CAD software bestudeerd en vergeleken met de gebruikelijke strategie van beoordeling door één radioloog. Om de effecten van de verschillende strategieën visueel zichtbaar te maken werden ROC curven gemaakt.

Digitale mammogrammen van 1048 patiënten werden beoordeeld vóór en ná het toepassen van CAD door een radioloog en drie laboranten. In de gebruikelijke strategie werd een oppervlakte onder de curve (area under the curve, AUC) gevonden van 0.92 hetgeen overeen komt met een sensitiviteit van 84% en een specificiteit van 94%. De AUC vóór en ná toepassing van CAD was hetzelfde bij alle beoordelaars (gemiddelde van 0.95). Het inzetten van laboranten in het voorselecteren van mammogrammen resulteerde in een AUC van

0.91, een sensitiviteit van 81% en een specificiteit van 96%. Een dubbele beoordeling van de mammogrammen door radioloog en laborant gaf een AUC van 0.96, een sensitiviteit van 96% en een specificiteit van 79%.

Geconcludeerd kan worden dat in een klinische populatie een systematische toepassing van CAD software de prestaties van mammogram beoordeling niet verbeterd. Daarnaast is gebleken dat het inzetten van laboranten als tweede beoordelaar naast een radioloog de meest effectieve strategie is om het ontdekken van maligniteiten in de dagelijkse klinische praktijk te verbeteren.

Tenslotte worden in hoofdstuk 8 de verschillende studies uit dit proefschrift besproken. Er kan geconcludeerd worden dat toepassing van CAD software niet bruikbaar is om meer maligniteiten te ontdekken in de dagelijkse klinische praktijk. Het inzetten van laboranten in het voorselecteren van mammogrammen in de dagelijkse klinische praktijk kan effectief zijn om de werklast van de radiologen alsmede de diagnostische kosten te verminderen zonder daarmee het ontdekken van maligniteiten in gevaar te brengen. Om de kankerdetectie te verhogen kunnen laboranten als tweede beoordelaar worden ingezet.

Dankwoord

Een proefschrift schrijven is een hele klus; het dankwoord wellicht een nog veel grotere. Hoewel aanlokkelijk, ga ik me er niet vanaf maken met: bedankt allemaal!

Veel mensen zijn belangrijk geweest in de totstandkoming van dit proefschrift. Professor Van Engelshoven, mijn promotor, u heeft de PERSPECT studie geïnitieerd en heeft altijd meegedacht over elk onderdeel van de studies en het proefschrift. Bedankt voor uw goede begeleiding.

Beste Karin, geen twijfel over mogelijk dat het zonder jouw doorzettings- en oplossend vermogen nooit tot zo'n mooi proefschrift was gekomen. Jij zag altijd weer het positieve ook als ik er zelf even wat minder in geloofde. Dank voor je enthousiasme en gezelligheid.

Fons, er waren in mijn onderzoek de nodige hobbels te nemen (en niet alleen statistisch gezien). Jij hebt met verve bijgedragen aan oplossingen. Ik denk met veel plezier terug aan de dinsdagochtendbesprekingen met Karin en jou, die eigenlijk maar voor hooguit de helft van de tijd over mijn onderzoek gingen en voor de andere helft opgingen aan allerlei bijzaken en geroddel, maar daardoor wel erg 'leerzaam' waren. Ik hoop dat ik je nog vaak kan lastig vallen met statistische of andere moeilijkheden.

Patty, dank voor je hulp bij mijn review en voor het introduceren van het woord 'jeopardise' dat nu veelvuldig in mijn proefschrift terugkomt.

Jenny en Marion, alle studies zouden nooit zo goed zijn gegaan zonder jullie. Altijd enthousiast, altijd meedenkend en vooral altijd gezellig! Bedankt daarvoor.

Jolanda, in de CAD studie pas echt betrokken geraakt, maar daarin nog veel werk verzet, zodat ik een derde observer kon toevoegen in mijn papers. Bedankt daarvoor.

Ine, dank voor de opmaak van mijn proefschrift. Je hebt me veel tijd bespaard! Carla, helaas raakte jij pas na de dataverzameling betrokken bij mijn onderzoek. Toch heb je met zinnig commentaar een nuttige bijdrage kunnen leveren aan dit proefschrift. Fijn dat je altijd alle tijd voor me had.

Verder wil ik alle radiologen bedanken die de moeite hebben genomen de 'blauwe formulieren' in te vullen. Vaak compleet en zonder klagen. Ed en Elvira, bedankt voor de medewerking in de CAD-studie!

Daarnaast dank aan alle andere mensen van de afdeling die in meer of mindere mate mijn onderzoek mogelijk hebben gemaakt. Ik denk daarbij aan de mensen van het archief, staffbureau, technische dienst, ICT en secretariaat. En natuurlijk al mijn naaste collega's daar beneden in de gang. Dankzij jullie heb ik een heel leuke tijd gehad. Marlies en Robbert-Jan (of Patrick of hoe je dan ook mag heten), leuk dat we het afgelopen jaar min of meer dezelfde koers volgden, waardoor we elkaar her en der konden helpen (en de eerlijkheid gebiedt me te zeggen dat ik meer aan jullie heb gehad dan andersom volgens mij...). En Marlies, bedankt voor de TIFF-tips!

Met mijn paranimfen Jaap en Joost ben ik nu alweer een aantal jaren herenigd. Eigenlijk min of meer toevallig zijn we elkaar in Maastricht weer tegengekomen, maar blijken we het erg goed met elkaar te kunnen vinden. Jaap, bedankt dat je de moeite neemt om helemaal vanuit New York over te komen om letterlijk en figuurlijk achter me te staan. Joost, ik geniet altijd van jouw levendige verhalen als je weer eens wat hebt meegemaakt (en dat gebeurt nogal eens). Leuk dat je nog in Maastricht bent (voorlopig tenminste), zodat we elkaar nog regelmatig zien.

Bedankt ook alle vrienden bij de harmonie, voetbalmaatjes en vrienden in Brabant die al die tijd voor de broodnodige ontspanning hebben gezorgd. Job, bedankt dat je nog naar mijn voorkant hebt gekeken.

Broers van me, Mark en Hans, fijn dat we zo goed met elkaar kunnen opschieten. Het is altijd erg gezellig als we met de hele familie (inclusief aanhang!) bij elkaar zijn, ook al komt dat veel te weinig voor.

Pap en mam, het is altijd vanzelfsprekend geweest dat jullie onvoorwaardelijk achter me staan en ik probeer me daar altijd bewust van te zijn. Niets is jullie teveel geweest om ons gelukkig te maken en dat is in mijn geval heel erg goed gelukt. Mam, ik ben erg blij dat je dit nog kunt meemaken en ik hoop dat je nog een hele tijd bij ons kunt blijven.

Lieve Lin, waar zou ik zijn zonder jou? Na alle drukke weken, avonden en weekenden aan dat proefschrift komen er nu eindelijk weer tijden die ik volledig met jou kan besteden. Ik hoop samen (en met de mops!) nog lang te genieten van het goeie leven!

Curriculum Vitae

Frank van den Biggelaar werd geboren op 8 december 1980 in het Brabantse dorp Biest-Houtakker, gemeente Hilvarenbeek. Hij doorliep het gymnasium aan het St. Odulphuslyceum te Tilburg (1993-1999), waarna hij Gezondheidswetenschappen in Maastricht ging studeren. Daar studeerde hij in 2005 af in de richtingen Beleid en Beheer, en Zorgwetenschappen. Tevens heeft hij in 2004-2005 de opleiding Facility Management aan de Hogeschool Zuyd in Heerlen voltooid.

Na zijn afstuderen is hij op de afdeling Radiologie van het MUMC als junior onderzoeker begonnen binnen de PERSPECT studie wat is uitgemond in een promotieonderzoek. Momenteel is hij als HTA onderzoeker werkzaam op de afdeling Oogheelkunde van het MUMC.

Daarnaast is hij sinds 2006 penningmeester en momenteel interim-voorzitter van de landelijke Alopecia Areata Patiëntenvereniging (AAPV) en zit hij in de Raad van Toezicht van het Facilitair Bureau Patiëntenorganisaties Nederland (FBPN).

List of publications

Van den Biggelaar, FJHM, Nelemans, PJ, Flobbe, K. Performance of radiographers in mammogram interpretation: a systematic review. *Breast*. 2008;17(1):85-90.

Van den Biggelaar, FJHM, Kessels, AGH, Van Engelshoven, JMA, Flobbe, K. Diagnostic performance of breast technologists in reading mammograms in a clinical patient population. *International Journal of Clinical Practice*, accepted for publication.

Van den Biggelaar, FJHM, Kessels, AGH, Van Engelshoven, JMA, Flobbe, K. Costs and effects of using specialised breast technologists in pre-reading mammograms in a clinical patient population. *International Journal of Technology Assessment in Health Care*, accepted for publication.

Van den Biggelaar, FJHM, Flobbe, K, Van Engelshoven, JMA, De Bijl, NPYM. Pre-reading mammograms by specialised breast technologists: legal implications for technologist and radiologist in the Netherlands. *European Journal of Health Law*. 2009;16(3):271-279(9).

Van den Biggelaar, FJHM, Kessels, AGH, Van Engelshoven, JMA, Boetes, C, Flobbe, K. Computer-aided detection in full-field digital mammography in a clinical population: performance of radiologist and technologists. *Breast Cancer Research and Treatment*, accepted for publication.

Van den Biggelaar, FJHM, Kessels, AGH, Van Engelshoven, JMA, Flobbe, K. Strategies for digital mammography interpretation in a clinical patient population. *International Journal of Cancer*, accepted for publication.

Van den Biggelaar, FJHM, Smolders, JJFM, Jansen, JFA. Complementary and alternative medicine in Alopecia Areata. *American Journal of Clinical Dermatology*, accepted for publication.

