Master of Science Thesis

Optimisation of direct digital mammography with the Jackknife FROC method

Tony Svahn

Supervisor: Anders Tingberg, PhD

The work has been performed at
Department of Radiation Physics
Malmö University Hospital

Medical Radiation Physics
Clinical Sciences, Lund
Lund University, 2005

PURPOSE: The aim of this study was to investigate how the diagnostic accuracy in digital mammography is affected by dose to the breast.

MATERIAL AND METHODS: Structures resembling invasive tumours and microcalcifications on x-ray images were positioned in an anthropomorphic breast phantom equivalent to a 50 mm compressed breast containing 50% glandular tissue. The average glandular dose based on phantom exposures using the automatic exposure control of the x-ray system (Siemens Mammomat Novation direct digital mammography) was 1.4 mGy, henceforth referred to as the 100% dose level. Thirty digital images were acquired of the phantom at each of three dose levels: 100%, 50% and 30% of the AEC-level. In each image there were at most 3 lesions present at random locations. Eight observers (3 radiologists and 5 non-radiologists) interpreted these 90 images on a DICOM calibrated 5 mega-pixel mammography monitor. For each image the observers were asked to identify the locations of perceived lesions and to assign an integer rating on a 4 - point confidence scale. A viewer was designed to make the evaluation easier and more accurate to perform. Observer performance was evaluated by the jackknife FROC (JAFROC) method.

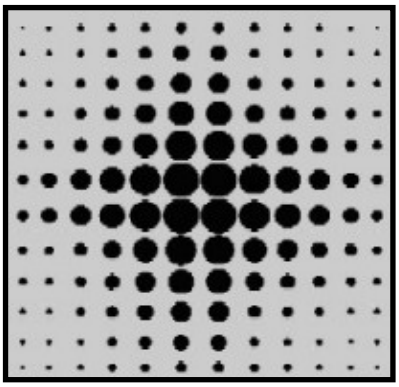
RESULTS: The mean FROC figures-of-merit (θ) for detecting the simulated lesions for the 100%, 50% and 30% dose levels, were 0.69, 0.68 and 0.63, respectively. There was no observed statistical difference in detection accuracy among observers between the 100% and the 50% dose level. The number of detected microcalcifications was almost identical (62.5%) at both dose levels, while the detected simulated invasive tumours was about 1% lower at the 50% level. At the 30% level the number of detected lesions was more than 10% lower both for the simulated invasive tumours and the microcalcifications.

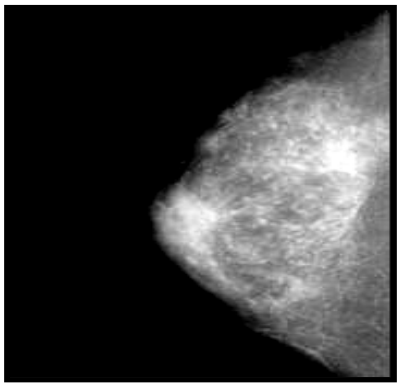
CONCLUSION: Within the precision of our measurements, it is possible to reduce the absorbed dose to the breast to half of the currently used dose level without compromising diagnostic accuracy.

Key Words: Diagnostic accuracy**,** Digital mammography, Dose level, FROC analysis method, FROC figure-of-merit.

Flyttade förändringar eftersöks

Varje år genomförs cirka 750 000 röntgenundersökningar i Sverige av bröst. Genom röntgenstrålning fås en bild fram där radiologen kan utvärdera om bröstet är friskt. Ifall detta inte är fallet ökar chanserna att framgångsrikt kunna behandla cancern om upptäckten sker i god tid. Med undersökningarna måste fördelarna vägas mot nackdelarna. Risken att någon gång i livet få strålningsinducerad cancer på grund av en mammografiundersökning är mycket liten. Den kan jämföras med en flygresa till Kanarieöarna eller liknande. Men det går inte att komma ifrån att det finns en risk och även om den är liten så är det viktigt att försöka minska den så mycket som möjligt. Studien som gjorts avsåg att ta reda på om det går att reducera den absorberade dosen till bröstet och samtidigt bibehålla en tillräckligt hög noggrannhet i bildkvalitén.





Det mänskliga ögat är komplext och fungerar närmast som en videokamera som sänder sina signaler till hjärnan som i sin tur "framkallar" och sedan tolkar den mottagna bilden. Hjärnans uppfattning om den mottagna bilden påverkas av personens psykiska och fysiska status, buller, stress mm. Ett exempel på en illusion ges i bilden till vänster där cirklarna i mitten lätt kan tolkas som de befinner sig närmre observatören än de övriga. Informationen i en bild behöver inte nödvändigtvis vara ”ärligare” bara för det enligt teorin borde vara så. Genom att studera skillnaderna mellan granskning av bilder på bröst, tagna med olika inställningar kan ett direkt mått på observatörens uppfattning fås. Den nya digitala undersökningstekniken innebär många fördelar i jämförelse med tidigare teknik. Möjligheten att få en högre bildkvalité eller minskad absorberad dos finns, men studier behöver göras som bekräftar att mer fördelaktiga inställningar kan användas kliniskt. Jag har gjort en studie med en serie bilder av ”bröst” där strukturer som liknar förändringar placerats in samt flyttats till olika positioner mellan bildtagningarna. Bilderna har varit realistiska, se bild till höger. En del av bilderna innehöll inte några förändringar medan andra innehöll flera. I slumpmässig ordning utvärderades sedan bilderna på ett kliniskt sätt av erfarna radiologer. Radiologerna sökte då i varje bild efter förändringarna och fick betygsätta dessa i skala 1-4 på hur säkra de var att de funnit en. Den samlade data med dessa betyg genomgick därefter en statistisk analys med en nyutvecklad programvara.

Resultatet av studien innebär att funna förändringar i bilder tagna med en lägre absorberad dos kan bekräfta att en minskning av den absorberade dosen borde göras och följden bli att fler strålningsinducerade bröstcancerfall i framtiden förebyggs.

Handledare: ***Anders Tingberg***
Examensarbete 20 p i Optimering av direktdigital röntgen med JAFROC analys ht 2004
Institutionen för Medicinsk Strålningsfysik
Malmö Universitets Sjukhus

Contents

1 Introduction

1.1 Optimisation in mammography

In Sweden, breast cancer accounts for almost three out of ten (29%) of all female cancers [1]. Each year, approximately 6000 women are diagnosed as having breast cancer and about 25% of these cases are expected to result in mortality [2]. Routine screening mammography has however shown to improve the survival significantly, making it possible to detect the breast cancer at an early stage [3]. The present annual number of mammographic examinations in Sweden is about 750,000 of which 80% are screening investigations and 20% are clinical investigations [4]. The female breast is particularly radiosensitive [5]. It is therefore important to evaluate the risk/benefit ratio for digital mammography if it is to be used for screening purposes. Screening mammography differs from clinical mammography quantitatively rather than qualitatively. The large number of cases handled per time unit in screening implies special demands on processing including display, reading and storage of the images [4].

The primary aim of mammography is early detection of small invasive cancers (appearing as opacities only). This is particularly difficult in breasts with a high content of glandular tissue [6]. In order to achieve this aim the best possible mammography techniques and procedures have to be used. If a cancer is suspected, the woman will be referred for further assessment that may involve core tissue biopsy that is an invasive technique and should be avoided if possible. The development of modern mammography is focused on increasing the image quality for diagnostic accuracy while minimising the absorbed dose received by the patient. With regards to patient exposure, optimisation refers to the determination of the lowest radiation dose possible that yields a sufficient level of clinical image quality. For digital mammography, the recommended maximum dose level for an examination should not exceed what was previously used for screen-film systems, and still provide a comparable image quality. Euref states an “upper acceptable level of 2.5 mGy” and an “upper achievable level of 2.0 mGy” based on standard measurements while SSI recommends 1.5 mGy, however it is up to each clinic to optimise their technique to give the best diagnostic outcome at the lowest possible radiation dose [4, 7]. It is therefore possible for a clinic to use higher values than the recommendations whenever it is motivated. In Malmö, for screen-film systems, the average glandular dose (AGD) is typically 1.2-1.4 mGy [8], while in Göteborg this is between 0.8-1.2 mGy for different types of mammography systems [9].

1.2 Complexity of receiver operating characteristics (ROC) methods

Using Receiver Operating Characteristics (ROC) methods as optimisation tools the quality of the whole imaging chain is being evaluated, including the human observer (the radiologist). The outcome of the evaluation is a measure of how good the system is for diagnostic purposes [10]. Because of the dependence of the human observer there is always a risk that the performance may be influenced by the subjectivity of the observer, for instance the experience of the radiologist, fatigue, viewing conditions etc [10]. Figure 1 shows how the difference between an experienced and an inexperienced reader might appear. The inexperienced reader usually makes decisions at more locations not knowing were to look [11]. Efforts have been made to diminish these subjectivity effects and increase the statistical power.

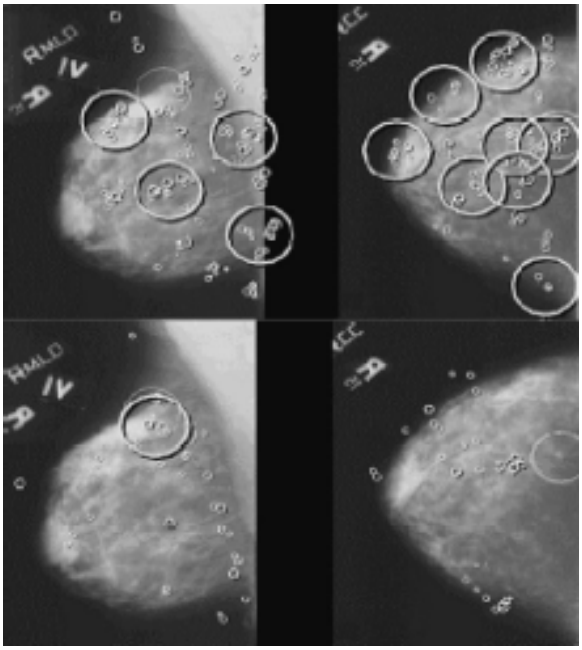


Figure 1. *A two-view mammogram soft-copy display for two readers, an inexperienced reader (upper two panels) and an expert reader (lower two panels). The small circles are individual fixations where the reader evaluated for more than 100 ms and the larger circles are clustered fixations that occurred with total dwell time of at least 1 second [11].*

In the classic ROC analysis method the lesion location is ignored [12]. The observers´ task is to decide whether an image is normal or abnormal. The lesion location is however of clinical relevance as it may guide a subsequent surgical intervention. If the observer makes a decision resulting in an abnormal for a positive image but indicates an incorrect location this event would be scored as a true positive (TP) according to the classic ROC method [12]. As shown in figure 2 (left image) the score will be as an abnormal image with one nodule present even though the reader missed the actual lesion and selected a nodule free location instead.

The classic ROC analysis method also neglects additional information whether there is more than one lesion present in the image [12]. For instance, with two lesions present and two observers evaluating, the first observer might detect only one of the lesions while the second observer detects both. Still, the scoring will be identical for both observers. The clinical consequences of missing one of the lesions could be very significant and should result in "penalizing" the reader for his judgement instead of "rewarding" him. This would in that way lead to a more true description of the image quality.

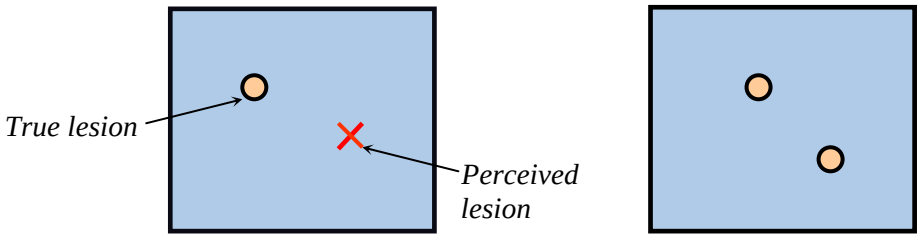


Figure 2. *The left image illustrates the incorrect location viewed by the observer believing it is a nodule. The result is scoring of a true positive as if the mark was in the correct position. The right image shows two lesions, the first observer detects one but the second observer detects them both - still the score will be the same [12].*

The neglected lesion location results in a loss of statistical power, which has to be compensated for. One way to do this is by using a large number of matched cases and readers [12], but this increases the overall cost and duration of the study.

1.3 Jackknife Analysis of FROC Data (JAFROC)

Another way to recover the loss of statistical power is by using the free-response ROC (FROC) paradigm. The observer then search each image for suspicious regions and assigns a rating to each marked region. Unlike the other methods the number of marks is completely determined by the observer. In each image there can be one, several or no lesions. To every marked lesion the observer calls a rating, depending on the confidence level of how certain he or she is in his or her decision. To increase the statistical power even more several observers may be involved and read the same image set. These multi-reader multi cases (MRMC) methods are desirable for optimal statistical efficiency and to generalise both populations of cases and readers. JAFROC is a variant of multi-reader multi cases method that deals with the problems previously mentioned in the classic ROC paradigm.

Chakraborty and Berbaum [11] have shown that JAFROC has superior statistical power in a model for simulated FROC data in comparison to earlier receiver operating characteristics methods. Figure 3 shows the result of the study, the ratio d (JAFROC-2)/d (ROC)
as a function of the incorrect localisation fraction (ILF). JAFROC-2 is the currently used and developed JAFROC method. The d-parameter is a measure of the statistical power of the analysis method. A high d-value means that it is relatively easy to detect a difference between two exposure conditions. The incorrect localisation fraction is defined as the number of times a noise site was rated higher than a signal site on an abnormal image. As the ILF increases it becomes more difficult for the observer to distinguish between signal sites and noise sites. This leads to the disadvantages of the classic ROC analysis becomes more evident. According to the study the power advantage, comparing JAFROC with ROC, is typically of a factor 2-3, but can be as large as 8 in some situations [11]. It will follow that a JAFROC study of diagnostic performance will have either greater statistical power or reduced cost due to reduced sample size compared to the classic ROC analysis.

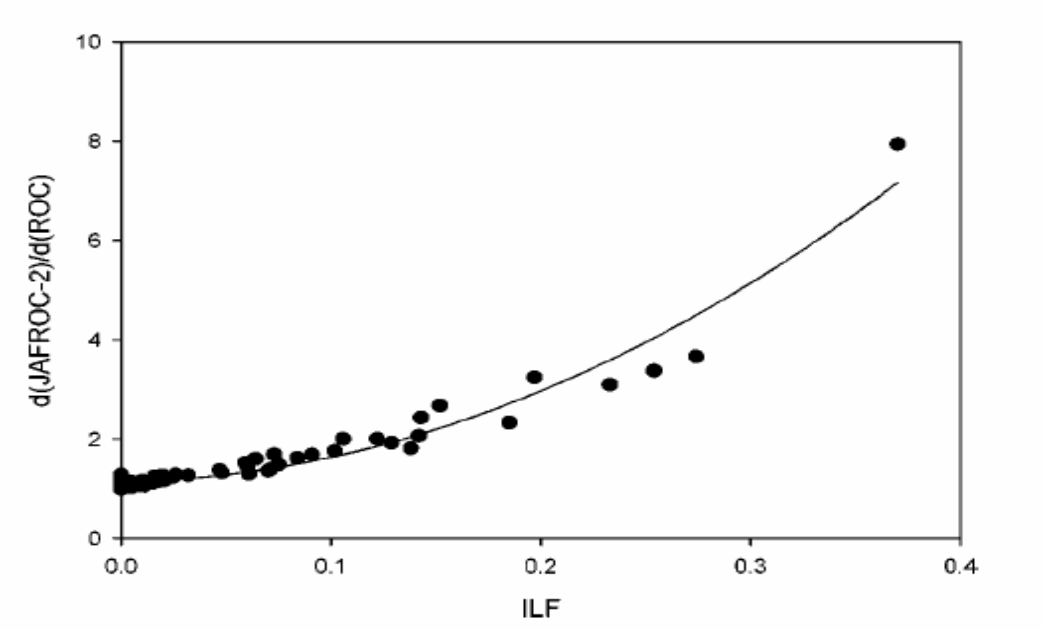


Figure 3. *A comparison between the statistical power of JAFROC analysis and classic ROC analysis. The ratio d (JAFROC-2)/d (ROC) is plotted as a function of the incorrect localisation fraction (ILF). The d-parameter is a measure of the statistical power of the analysis methods being compared [11].*

1.4 Background to the present investigation

Since 1980 mammography has continuously expanded and is in many countries one of the most common used x-ray examinations, particularly in Sweden, Finland, the Netherlands, the United Kingdom and Australia [4].

In 1982 there were only 134 mammography units in the USA increasing to about 10,000 in 1990 [4]. The increasing use of mammography and the large amount of screening examinations have inspired a lot of efforts in optimising this examination.

Since the mid 90´s there has been a great interest in digital mammography, but the shift from screen-film to digital mammography has been slowed down by technical challenges and high costs due to the demands on high image quality and low absorbed dose to the breast. Questions of image quality in radiology ultimately depend on the ability of radiologists or other trained observers to correctly decide patients' states of health and disease from clinical images [13]. Observer performance methods that employ clinical images can provide direct assessments of the image quality [13]. There is also a great potential of full-field digital mammography systems (FFDM) with significantly higher image quality or significantly lower dose than screen-film mammography (SFM), or both. Further research is therefore necessary to take advantage of this potential.

1.5 Aim

- The aim of the current study was to investigate how the diagnostic accuracy in digital mammography is affected by dose to the breast.

2 Material and methods

2.1 The anthropomorphic breast phantom – RMI-165

The breast phantom used in the measurement (shown in figure 4) was an RMI phantom - model 165, manufactured by RMI Radiation Measurement inc. [14]. The x-ray image of the phantom is similar to a mammogram of a real breast and it provides reproducible clinical type of images. The phantom consists of regions of high, medium and low proportions of glandular tissue, which all contain thin line structures of varying opacity. In the phantom there is a transverse slit (1 mm wide) where a film-substrate with structures resembling microcalcifications (μCa) and invasive tumours can be inserted. As shown in figure 4, these structures could be placed at arbitrary locations on the film-substrate.

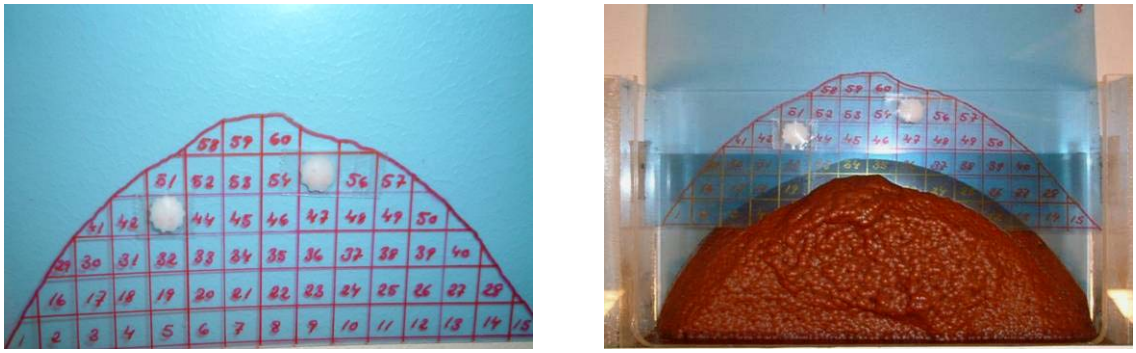


Figure 4. *The photo to the left illustrates two invasive tumours on an unexposed developed film-substrate with squares that each corresponds to determined coordinates. To the right is the film-substrate with the simulated lesions, partly inserted in the 1 mm transverse slit of the anthropomorphic breast phantom.*

2.1.1 Addition of microcalcifications

The microcalcifications were made of aluminium oxide (AlO_2), each with an approximate diameter of 200 μm and provided by the manufacturer of the phantom [14]. As shown in Figure 5, each cluster consisted of 10-12 microcalcifications distributed on an area of up to 25 mm^2, to avoid high-density spots in the images, which would be easier to detect [6].

The distances between the microcalcifications making up each cluster were varied and they were distributed randomly so that each cluster had an individual appearance on the x-ray images.

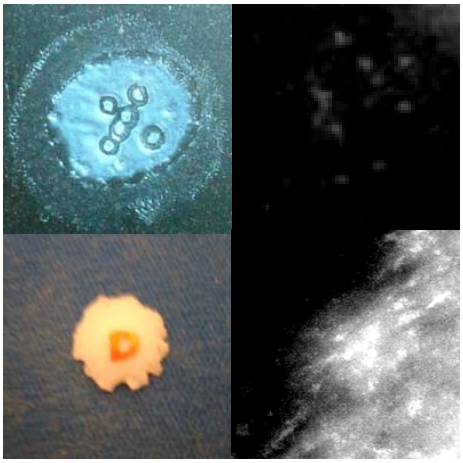


Figure 5. *The left panels illustrate photos of the simulated microcalcifications (upper image) and the simulated invasive tumour (lower image), and the corresponding x-ray images of them to the right.*

2.1.2 Addition of invasive tumours

In order to simulate the radiographic appearance of tumour masses, discs made of polytetrafluoro-ethylene (PTFE, Teflon) were placed on the film-substrate insert and were provided by the Department of Medical Physics and Biomedical Engineering at Sahlgrenska University Hospital in Gothenburg, Sweden. Eight such discs were used in the study, making up the total amount of simulated tumours per dose level, with diameters 10.5 mm – 11.5 mm. The centre thickness varied from 0.72 to 0.79 mm and decreased towards the edges where the thickness was approximately 0.12 mm. The edges of the discs were made irregular to simulate realistic looking malignant masses. The variation of the discs in diameters, centre thickness, irregularity of the edges as well as positioning all give rise to an extensive number of difficulty levels for the observer in his or her attempt to localise them in the images, shown in figure 6-7.

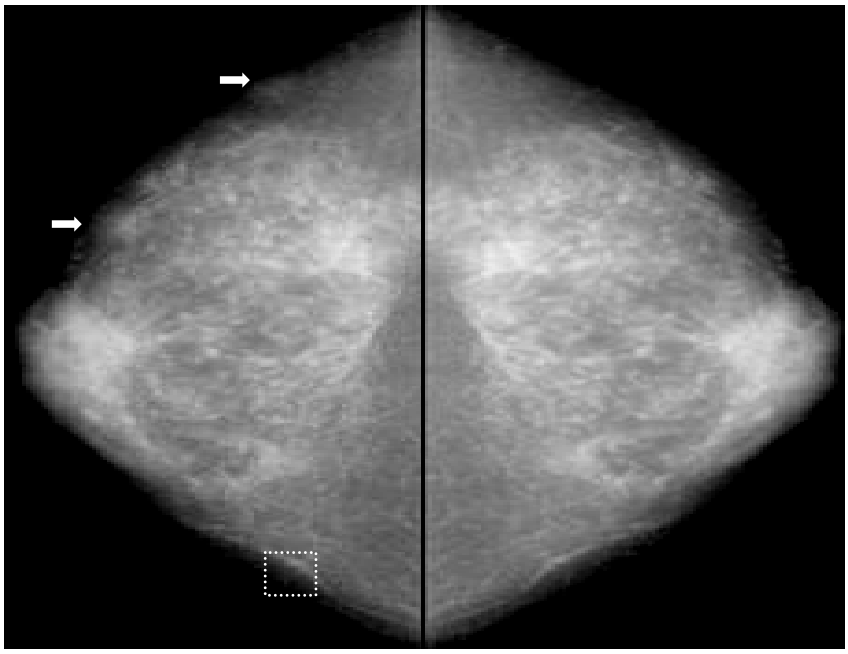


***Figure 6**. The image on the left shows a radiograph of the breast phantom with structures added to simulate pathologies. The arrows indicate added masses and the outlined region contains added microcalcifications. To the right is the corresponding image without added pathological structures.*

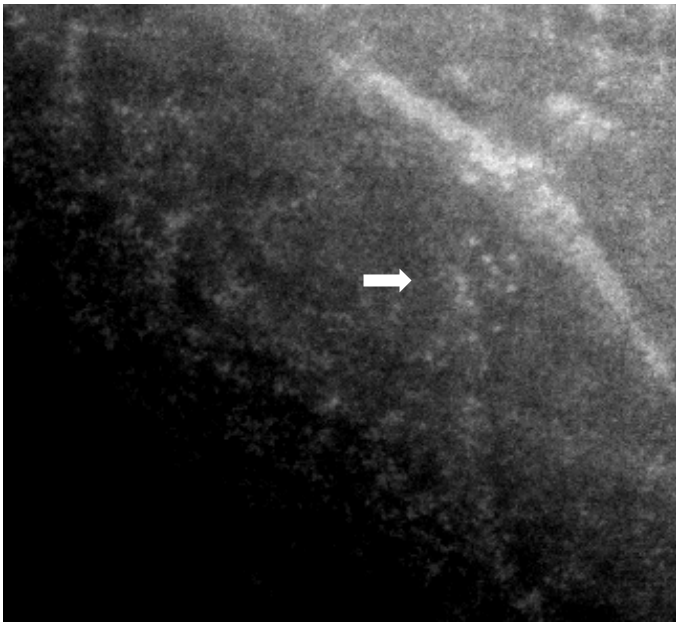


***Figure 7.** The image illustrates a close-up of one of the simulated clusters of microcalcifications in the outlined region in figure 6.*

2.2 The confidence levels

The task for the observers participating in a JAFROC study is to view images containing one, several or no signals. For each suspicious location the observer has to state the confidence of his or her decision on a scale from one to four, as shown in Table 1. For example, if the observer is totally sure that a suspicious region of an image contains a tumour he or she ranks it as a four. If the observer cannot detect any lesions in an image, this image receives a score of zero.

Table 1. *The four-level confidence scale that were used in the JAFROC study.*

Confidence level	Significance
1	Probably not a lesion
2	Possibly a lesion
3	Likely to be a lesion
4	Very likely to be a lesion

2.3 Image acquisition

In total, 90 unique images were acquired of the breast phantom for this study – thirty at each of three dose levels. The 100% dose level was determined through the automatic exposure control (AEC) of the mammography unit, and images were subsequently acquired at 50% and 30% of the AEC-determined dose level. Each of the thirty images contained zero to three lesions distributed according to Table 2. In order to minimise observer memory effects, the structures were placed in different positions in the phantom – twenty possible positions for tumours and ten for microcalcifications. The selection of positions was based on a pre-study of images acquired of the RMI-165 phantom at different dose levels with inserted structures in regions of low, medium and high proportions of glandular tissue. The positions were then distributed equally over these three regions. The 2D image coordinates of the positions were carefully recorded for each image. Similar coordinates were used both for microcalcifications and invasive tumours to make a comparison in the detection of them. The distribution and positions were used for all three dose levels so that no statistical power would be lost due to inconsistency in detection difficulty.

Table 2. *The distribution of number of lesions.*

# Images	# Masses	# Microcalcifications
16	0	0
4	3	0
3	2	0
3	0	2
3	1	1
1	2	1

The study was made with raw data x-ray images for the following reasons: the way of processing is not necessarily the same between different manufacturers of x-ray systems, processed images are typically non-linear with exposure, and the processing method is frequently changed and updated. Multi-Objective Frequency Processing (MFP) is commonly used in daily improvement of diagnostic image quality. Various spatial frequencies are enhanced using weights that are determined by each manufacture to produce images that appear to the satisfaction of radiologists [15, 16, 17]. The image signals are after processing not linear to the absorbed dose anymore, in the contrary to raw data images, as shown in figure 8. By including only raw data images the results of the observer performance should be more comparable with other radiographic systems.

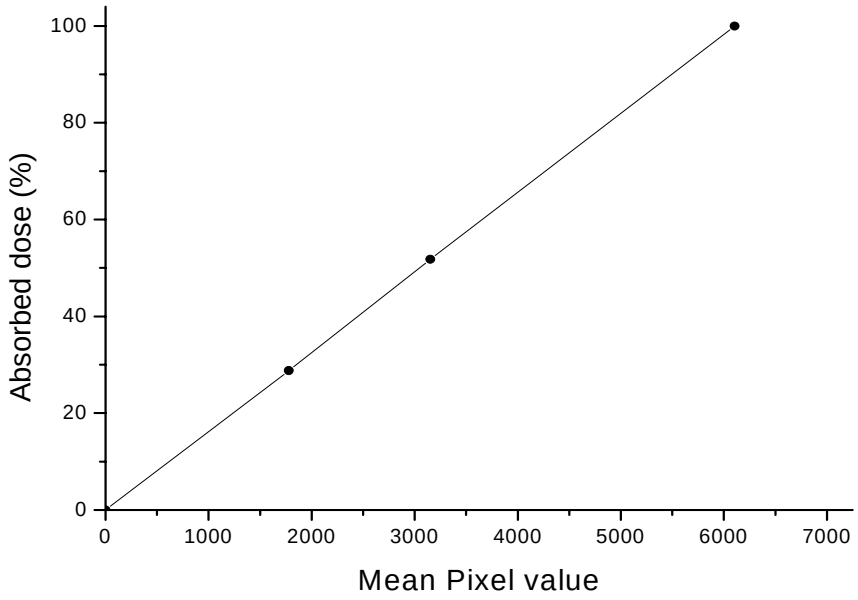


Figure 8. *The linear response between absorbed dose (%) and mean pixel value for the acquired dose levels.*

2.4 Average glandular dose (AGD)

In mammography examinations, the average glandular dose (AGD) to the breast is considered to be directly related to the radiation risk [18]. A so-called standard breast is simulated with 45 mm PMMA (polymethylmethachrylate) and defined as a 50 mm compressed breast containing about 50% glandular tissue [19]. The anthropomorphic breast phantom RMI-165 consists of material that makes it comparable to a standard breast regarding exposures [20]. The phantom used in this study is modified to contain 5 mm PMMA of a slightly different density composition than the original. This was tested using the same beam quality as in Table 3 as well as the automatic exposure control (AEC) of the mammography unit, comparing the exposure of a 45 mm PMMA phantom to the exposure of the RMI-165 breast phantom. The result was a tube loading of 114 mAs (PMMA) and 126 mAs (RMI-165) respectively, i.e. a difference of +11%. This confirms that there is a relatively small difference in exposure between these phantoms and therefore it is reasonable to use the procedures in the European protocol on dosimetry in mammography [19] for a standard breast to estimate the AGD. This protocol contains a table of conversion factors (g_{PB}) one can use to calculate the AGD from the measured entrance surface dose, as a function of half-value layer (HVL). Using the tube loading 114 mAs (and the tube output 0.0364 mGy/mAs at the entrance surface of the 45 mm PMMA phantom) as well as a measured half value layer 0.54 mm Al [18], the value of g_{PB} is 0.298, which results in an estimated AGD of 1.24 mGy for a standard breast. For W/Rh anode/filter combination, the AGD can be calculated from the determined value using the correction factor s = 1.042 [22], which accounts for the differences in beam quality. Applying this factor yields an AGD for a standard breast of 1.29 mGy. For a similar breast that might be simulated with the RMI-165 phantom used in this work, the AGD can be estimated to approximately 1.4 mGy at the tube loading 126 mAs as shown in Table 3.

Table 3. *The parameter settings of the mammography unit Mammomat Novation (Siemens).*

Parameter	Settings
Tube loadings	126, 63 and 36 mAs
Tube potential	28 kV
Anode/filter	W/Rh

2.5 Viewer for Digital Evaluation of X-ray images (ViewDEX) version 0.996

ViewDEX is an efficient software tool for conducting observer performance studies such as ROC, VGA and FROC analysis [23] in a digital environment. ViewDEX, version 0.996, was especially designed for this study by a software engineer [24] for evaluation of image quality with JAFROC analysis. As is illustrated in figure 9 the interface contains an image canvas and an answering protocol. The use of ViewDEX speeds up the process, both for the observers participating in the study and the subsequent collection and analysis of data. To score marks as either true positives (TP) or false positives (FP), a 10 mm diameter circular region was placed around the centre of each lesion – a mark placed inside or outside this circle was counted as a TP or FP, respectively. However, to avoid errors associated with random selection of a lesion the observers were instructed to place their marks in what they considered to be the centre of the lesion. These regions were implemented into the ViewDEX software tool, which then automatically calculated what markings were TPs or FPs.

The observer interpreting the images marks a suspicious location in the image and calls a confidence level grade for each suspected lesion. The marked locations were checked against the true locations by the software through pre-determined coordinates and with the scoring diameter implemented. The raw data was then automatically collected into a log file that was exported to MS Excel where it was transformed and saved into a text file suitable to be analysed by the JAFROC software [25]. In this way, the data could be analysed shortly after it was produced. The observers were allowed to zoom in and out, pan across, and alter the window/level setting of each image.

Eight observers, out of which 3 were radiologists and 5 were physicists, interpreted 90 images on a DICOM calibrated 5 MP mammography monitor. Before the observers started the image reading they were given instructions on how to use the viewer, what types of structures to search for, and that there were at most three lesions per image, but they knew nothing about the exact number, type, or positions of the structures. To avoid any learning curve effects, the observers had to evaluate 20 training images to become familiarised with the appearance of the normal structures in the mammogram of the phantom as well as the inserted structures. The values of window/level were hidden so that no information about the settings could be

found. Because raw data images were used in the study, the window/level-values were automatically adapted to the x-ray exposure. Three different window/level-settings had to be adjusted for. These settings were preset and controlled by an experienced radiologist.

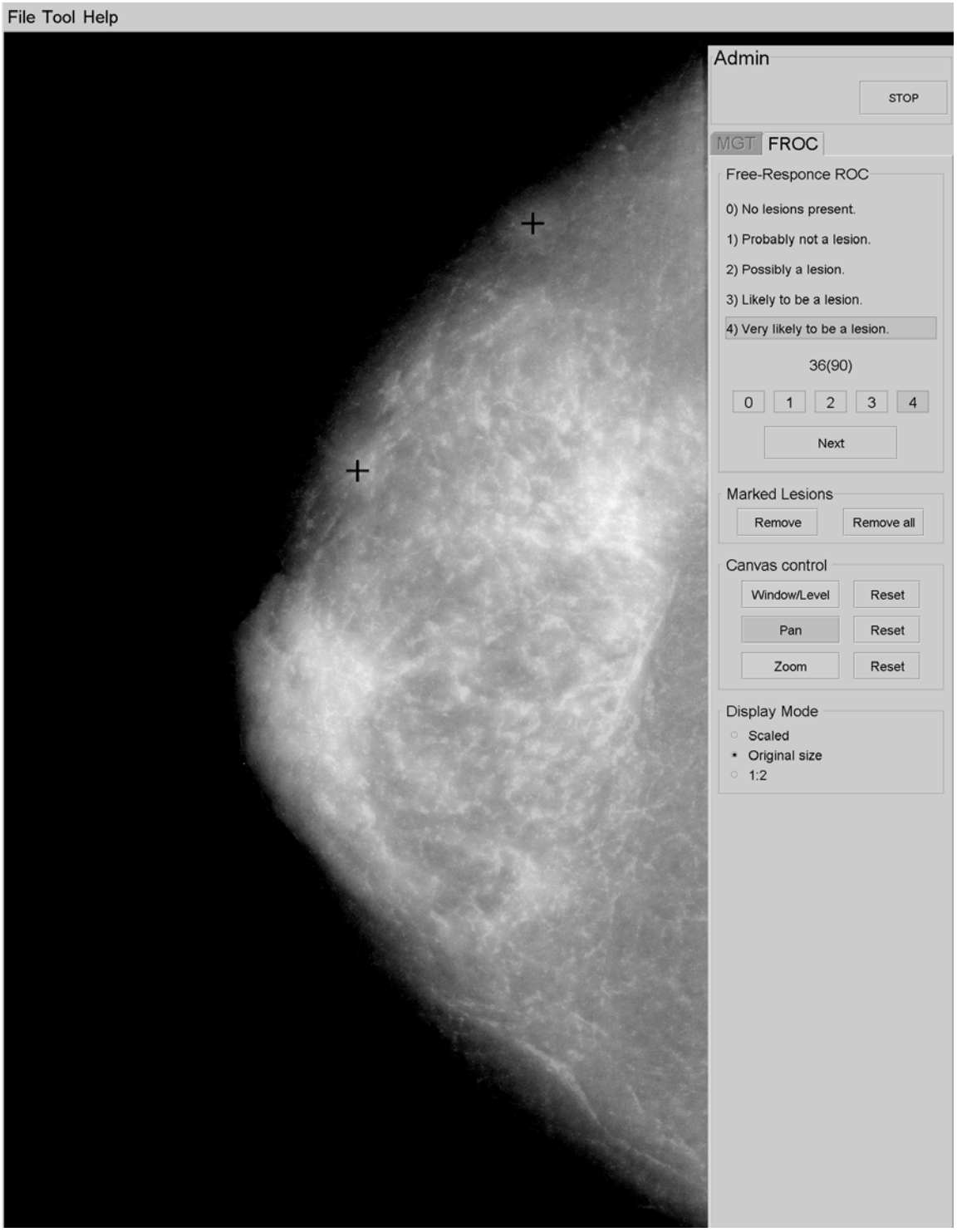


Figure 9. *The ViewDEX version 0.996 interface showing the 5 confidence levels and other available functions.*

2.6 Statistical Analysis

The FROC multi-reader multi-case (MRCM) dataset, consisting of TP and/or FP mark-ratings originating from 720 observations (3 modalities x 8 readers x 30 images), was analysed with the jackknife free-response receiver operating characteristic (JAFROC) method. There are two steps to the analysis of FROC observer data – a scoring step and a statistical analysis step. The scoring step is made in order to diminish the FROC data for a given observer and set of cases to a single value.

2.6.1 The FROC figure-of-merit (θ)

This value is the statistic FROC figure-of-merit (θ), i.e. the weighted probability that a signal rating exceeds a noise rating [11]. The FROC figure-of-merit is bounded between 0 (worst possible performance) and 1 (perfect performance). The calculation rewards the observer for making good decisions (true positives and true negatives) in the same way it penalises the observer for making bad decisions (false negatives and false positives). The 95% confidence interval (CI-95) that compares the dose levels with each other indicates no significant difference between the mean FOM's, when zero is included in this interval [25, 26].

3 Results

3.1 The observers' performance

The radiologists evaluated the whole series of 90 mammograms for 1 hour and 5 minutes on the average, while the physicists needed 1 hour 30 minutes on the average. The radiologists generally had higher values of the FROC figure-of-merit. The mean FROC figures-of-merit (θ) for detecting the simulated invasive tumours and microcalcifications for the 100%, 50% and 30% dose levels, were 0.69, 0.68 and 0.63, respectively. As shown in table 4, the JAFROC method indicated no significant difference between the 100- and the 50%- level, while it indicated a significant difference between the 100- and 30%- level, and 50- and the 30%- level.

3.1.1 Comparison between the 50% and the 100% dose level

- The detection of microcalcification was identical (62.5%) at both dose levels.
- The detection of invasive tumours was generally low: 34% for the 100% dose level and 33% for the 50% dose level.

3.1.2 Comparison between the 30% and the 100% dose level

- At the 30% dose level the detection of lesions was about 10% lower than for the 100% dose level, both for the simulated invasive tumours and the microcalcifications.

Table 4*. The 95% confidence intervals for the difference in FOM values between pairs of dose-levels for the JAFROC method. A range that includes 0 means the difference was non-significant (ns) and otherwise it was significant (sig).*

Dose levels	100%-50%	100%-30%	50%-30%
JAFROC	[-0.031, 0.054] - ns	[0.020, 0.105] - sig	[0.009, 0.094] - sig

4 Discussion

This study represents the first observer performance study in optimisation of direct digital mammography analysed by the jackknife FROC method, investigating the dependence of the diagnostic accuracy on dose to the breast. Using this method, a dose reduction of up to 50% is determined to be possible without compromising diagnostic accuracy. This would correspond to a reduction in AGD to 0.6-0.7 mGy from the presently used 1.2-1.4 mGy in Malmö.

Although the classic receiver operating characteristic (ROC) analysis is the accepted methodology for evaluation of diagnostic imaging systems, it has shortcomings in as much it is restricted to one observer report per image. The classic ROC analysis is limited to the use of one or no signal in each image and thus, in order to achieve statistical significance, a large number of images have to be used. The classic ROC analysis method has no way of dealing with tasks involving detection and localisation (D&L). A number of approaches to collect and analyse D&L data have been published each requiring different tasks of the observer, such as Localization ROC (LROC), free-response ROC (FROC), alternative free-response ROC (AFROC) etc. It has been shown that the JAFROC method has improved statistical power, which is of high relevance when performing a study such as our, decreasing both the cost and duration of the study [11, 26, 27]. Radiologists are often stressed and have lack of time; using the JAFROC method minimises this problem.

A recent study compared the performance and patient dose of full-field digital mammography units (FFDM) for clinical use [28]. Measurements of linearity and automatic exposure control stability were performed on four units. The entrance air-kerma was calculated over a sample of 800 cranio-caudal mammograms and the average glandular dose obtained, assuming two mean glandular compositions of 50% and 30%, respectively. The digital systems showed very good linearity and comparable responses. The results of the study were compared to those of other surveys, indicating that full-field digital mammography allows a significant clinical dose reduction compared to screen/film mammography (SFM).

An earlier study has compared microcalcification detectability of a full-field digital mammography to a conventional screen-film mammography by using different doses in the

digital system [29]. The investigations were performed with an FFDM (Senographe 2000 D, GEMS) and an SFM system (Senographe DMR, GEMS) and an anthropomorphic breast phantom with superimposed microcalcifications. The digital detector was exposed with standard dose of SFM and with a dose reduction of up to 75%. ROC analysis with a confidence level ranging from 1 to 5 was done with the results of the anthropomorphic phantom. The ROC analysis yielded better results for the FFDM system. The same lesion detectability in digital mammography as in the conventional method was reached at a dose reduction of about 25%, concerning spot views even at higher reduction. The same detectability as in conventional mammography was reached, however, by a dose reduction of about 50%.

In our study, a total number of 23 simulated invasive tumours and 10 simulated clusters of microcalcifications were used at each dose level, consisting of 30 digital x-ray images and consequently, several of the images contained multiple lesions. Clinically, multiple breast lesions are not that common. As mentioned earlier the observers had to evaluate example images before the actual evaluation started. This wouldn’t have happened in a clinical situation, but the study was based on learning the main structures in the repeated mammogram of the RMI-165 phantom. Otherwise the observers’ knowledge of the main structures would have increased during the evaluation and would perhaps have resulted in a higher false-positive rate in the beginning.

The simulated tumours seemed to be more difficult to detect than the microcalcifications. The simulated invasive tumours appeared with low contrast, and were reasonably easy to detect in the fatty tissue, but more difficult to detect in the glandular tissue. In some regions of the mammogram of the RMI-phantom where the proportion of glandular tissue was particularly high, the simulated invasive tumours got blurred out and thus the detection rate decreased significantly. This could be due to the shapes and sizes of the structures. The microcalcifications were more conspicuous because of their pattern and high density. However, in this study, the relation between the different levels of absorbed doses provided the information about the diagnostic accuracy and by using a large number of difficulty levels it was easier to distinguish better image quality from worse.

The results regarding the FROC Figure-of-merit, which is a measure of the observer's performance, would have been higher if detection of simulated invasive tumours wouldn't have been so low. Comparing the observers with each other, the radiologists generally evaluated more "careful" than the physicists, meaning they only marked locations when they were confident in their decisions, yielding less false positive ratings. One reason for this could be the time factor. The radiologists evaluated the whole series of 90 mammograms for 1 hour and 5 minutes on the average, while the physicists used 1 hour 30 minutes on the average. Some of the physicists did find more of the invasive tumours in the high proportioned glandular tissue than the radiologists, but they also marked out a higher amount of false positive ratings. This could be one reason, besides better skills of locating lesions, why the radiologists had higher value of the FROC figure-of-merit than the physicists.

The detector used in this study consisted of an amorphous selenium (a-Se) photoconductive material, referred to as a direct-conversion digital detector. It could, however, be interesting to make a study with a technique based on indirect conversion digital detectors, such as those using CsI(Tl) as the conversion material, or a computed radiography system such as those consisting of photo-stimulable phosphors. Also, a clinical study with many more patients may be desirable to firmly establish the preliminary conclusion of this study. However, a large-scale clinical trial was outside the scope of this work, but future work along this direction may be desirable given the public interest in mammography risk. The results indicated no significant statistical difference in detection accuracy between the standard absorbed dose (100%) and half (50%) of the standard absorbed dose. The statistical result confirms that it may be possible to reduce the absorbed dose, but there is a limit of reduction somewhere between the 50% level of the absorbed dose and the 30% level.

5 Conclusion

Based on the results from our study, it may be possible to reduce the absorbed dose to the breast to half of the currently used dose level, without compromising diagnostic accuracy.

6 Acknowledgements

First, I would like to thank my supervisor Anders Tingberg for providing me this project and guiding me through this work with helpful discussions and suggestions, for his expertise in ROC methods, and particularly for his patience. Also I would like to thank Bengt Hemdal for numerous of helpful advices and for his expertise in digital mammography. I would like to thank Anne Thilander-Klang for providing material and whose thesis was my bible, Sune Svensson for his programming skills in JAVA, Mark Ruschin for bright explanations and suggestions concerning the image processing, Peter Wallenius for increasing my computer knowledge, Ingvar Andersson for helping a lot with the evaluation and of course I would like to thank all the observers. I would also like to give special thanks to Dev Chakraborty and Hong-Jun Yoon for their consistent support in statistical matters. Finally I would like to thank all the staff at the department of Radiation Physics at Malmö University Hospital for making it a pleasant place to work at.

7 References

1. Swedish National Board of Health and Welfare (2002); Cancer incidence
2. Swedish National Board of Health and Welfare (2002); Causes of death
3. Tabar L, Yen MF, Vitak B, Chen HH, Smith R A, Duffy SW (2003) Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. The Lancet, Vol. 361 p.1405-1410
4. Hemdal B, Andersson I, Thilander-Klang A, Bengtsson G, Leitz W, Bjurstam N, Jarlman O, Mattsson S (2002) Mammography – recent technical developments and their clinical potential. SSI report 2002:8
5. ICRP 60 (1991) International Commission on Radiological Protection, Publication 60: 1990 Recommendations of the International Commission on Radiological Protection, Annals of the ICRP Vol. 21, No 1-3
6. Thilander-Klang A (1997) Diagnostic Quality and Absorbed Dose in Mammography: Influence of X-ray Spectra and Breast Anatomy Thesis
7. Van Engen R, Young KC, Bosmans H, Thijssen M (2003) Addendum on digital mammography, version 1, addendum to chapter 3 of the European Guidelines for Quality Assurance in Mammography Screening, third edition, (available at www.euref.org).
8. Hemdal B (2005). Department of Medical Radiation Physics, Lund University, Malmö University Hospital, Malmö, Sweden. Private communications
9. Thilander-Klang A (2005) Department of Medical Radiation Physics, Sahlgrenska University Hospital, Gothenburg, Sweden. Private communications
10. Tingberg A (2000) Quantifying the quality of medical x-ray images: An evaluation based on normal anatomy for lumbar spine and chest radiography Thesis
11. Chakraborty DP, Berbaum KS (2004) Jackknife Free-Response ROC Methodology Medical Physics, Vol. 31, No.8 p. 2313-2330
12. Chakraborty DP (1996) The FROC, AFROC and DROC variants of ROC analysis. Handbook of Medical Imaging. Vol. 1 p. 772-787
13. Metz CE (1996) Fundamental ROC Analysis Medical Imaging. Handbook of Medical Imaging. Vol. 1 p. 752-757
14. RMI Radiation Measurement inc (1990) Middleton, WI, USA (Caldwell and Yaffe)
15. P. Vuylsteke, E. Schoeters. Image Processing in Computed Radiography. International Symposium on Computerized Tomography for Industrial Applications and Image Processing in Radiology. Germany, 1999.
16. Stahl M, Aach T, Dippel S. Digital radiography enhancement by nonlinear multiscale processing, Med. Phys. 2000; 27, 56-65.
17. FCR (Fuji Computed Radiography). General description of image processing, Fuji Photo Film Co., Ltd. Japan, 2002.
18. Comments to the regulations and general advice of the Swedish Radiation Protection Authority on diagnostic standard doses and reference levels within x-ray diagnostics. SSI FS 2002:2
19. Zoetelief J et al. (1996) European protocol on dosimetry in mammography. Report EUR 16263 EN. Luxembourg: Office for Official Publication of the European Communities

20. Thilander-Klang A (1997) Diagnostic Quality and Absorbed Dose in Mammography: Influence of X-ray Spectra and Breast Anatomy. Thesis, Paper VI, p. 1-16

21. Hemdal B (2004). Department of Medical Radiation Physics, Lund University, Malmö University Hospital, Malmö, Sweden. Private communications

22. Dance DR, Skinner CL, Young KC, Becket JR and Kotre CJ (2000) Additional factors for the estimation of mean glandular breast dose using the UK mammography dosimetry protocol. Phys. Med. Biol. Vol. 45, p. 3225-3240

23. Börjesson et al. (2005) A tool for increased efficiency when performing observer performance studies in a digital environment, *Radiat Prot Dosimetry (in press)*

24. Svensson S (2004) Sahlgrenska University Hospital, Gothenburg, Sweden

25. The JAFROC software (version 1.05) is available for download from the web site: www.devchakraborty.com.

26. Chakraborty DP (2004) University of Pittsburgh, PA, USA. Private communications

27. Dorfman DD, Berbaum KS et al. (1992) ROC characteristic rating analysis: Generalization to the Population of Readers and Patients with the Jackknife method. Invest. Radiol. p. 723-731

28. Gennaro G, Baldelli P, Taibi A, Di Maggio C, Gambaccini M (2004) Patient dose in full-field digital mammography: an Italian survey. Eur Radiol. 14(4): 645-52

29. Obenauer S, Hermann KP, Schorn C, Fischer U, Grabbe E (2000) Full-field digital mammography: dose-dependent detectability of breast lesions and microcalcinosis. [Article published in **Germany**] Rofo. 172(12): 1052-6