Master of Science Thesis

Potential for lower doses in digital mammography

Tumor detection in hybrid images using a dose reduction simulation method

Pontus Timberg

Supervisor: Anders Tingberg, PhD

The work has been performed at the
Medical Radiation Physics
Department of Clinical Sciences, Malmö

Medical Radiation Physics
Clinical Sciences, Lund
Lund University, 2006

ABSTRACT

Pontus Timberg

Potential for lower doses in digital mammography
Tumor detection in hybrid images using a dose reduction simulation method

Purpose: To determine how image quality linked to tumor detection is affected by reducing the absorbed dose to 50% and 30% of standard dose levels, i.e. an average glandular dose of 1.26 mGy for a standard breast according to European guidelines.

Material and methods: Tumors were computer simulated and inserted into 40 out of 90 normal, unprocessed images acquired from the screening department at Malmö University Hospital using a Siemens Mammomat Novation full-field digital mammography unit. All 90 images were then dose-reduced by adding simulated quantum noise to represent images acquired at 50% and 30% of the original dose levels set by the automatic exposure control (AEC). This yielded 270 images that were subsequently processed for final display. One radiologist and 3 physicists participated in this study in which they searched for and marked the positions of the tumors and the degree of suspicion. The analysis of results were carried out using the jackknife free-response ROC (JAFROC) method together with analysis of variance (ANOVA).

Results: In the tumor detection study, the cumulative figure of merits (FOM's) calculated from JAFROC scoring were 0.61, 0.66 and 0.67 for the 100%, 50%, and 30% dose levels, respectively. ANOVA revealed an F-stat and p-value of 1.90 and 0.15, respectively, indicating no statistical difference in tumor detection between any two pairs of scores.

Conclusion: For the simulated tumors used in this experiment, there was no significant change in detection by increasing quantum noise, indicating a potential for dose reduction. Further studies could involve using different shaped masses or microcalcifications, different processing algorithms and different window/level settings.

Advisor: **Anders Tingberg**
Degree project 20 credits in medical radiation physics. Spring 2005
Department of Medical Radiation Physics, Clinical Science, Malmö
Lund University

Introduction

Screening mammography refers to the routine x-ray examination of women for presence of breast cancer in an early stage. In Sweden, women of age 40-50 years and above are called for screening examinations once every two years, and the present annual number of examinations is around 600,000 [Hemdal, 2002]. Up until relatively recently, mammography has been a predominately screen-film based imaging system. However, there has been an increasing trend in the usage of digital mammography (DM) in recent years as it has several inherent advantages over screen-film systems. These advantages include easier archiving of images (i.e. no need for film archives), increased dynamic range, increased potential for image processing, automated cancer detection techniques, and a higher detective quantum efficiency (DQE) at low and medium spatial frequencies [Muller, 1997].

One of the main limitations of screening mammography is still its relatively low sensitivity[1]. The sensitivity of screening mammography is about 70-80%, which indicates that about 20-30% of existing cancers are missed upon initial examination, most of which can be detected upon retrospective analysis of the same images [Chakraborty, 1996]. The most probable explanation for this phenomenon is that mammography is a 2D imaging modality and therefore images contain a superposition of overlying and underlying tissues that act to obscure the visibility of potential malignancies. However, there are many factors involved in the imaging chain for DM, all of which can be individually and collectively optimised to yield the best possible diagnostic image quality.

One of the most important physical factors to optimise in digital screening mammography is radiation dose. Since women, most of whom are healthy, are called in routinely for examinations they can receive a significant cumulative dose over time. The optimisation of dose is not yet established for DM, but according to European guidelines [EC, 2005] the resulting image quality should be at least as good as in screen-film mammography (SFM). One restriction of SFM is its limited dynamic range, which DM is not restricted to, i.e. the optical density (OD) or blackening of the films limits the range of absorbed doses. The average glandular dose AGD in the breast is typically around 1.0 mGy per projection [Hemdal, 2002]. Recommendations on dose vary, but for SFM the Swedish recommendation by the Swedish radiation protection authority (SSI) states that the AGD is not allowed to exceed 1.5 mGy for the appropriate OD setting used and should not exceed 1.0 mGy at OD 1.0 according to measurement set up in the European protocol [Hemdal, 2002]. The average glandular dose at the Malmö University Hospital is 1.3 mGy for a standard breast according to European guidelines, and is used as a "standard" dose level [Hemdal et. al, 2005a][Zoetelief, 1996].

Deleted: 26

[1] Sensitivity is a measure of the radiologists performance in detecting abnormalities, whereas specificity deals with the radiologists performance in generating false alarms.

The effect of further lowering the radiation dose is an increased relative fluctuation in the x-ray beam due to the statistical nature of the photon interactions, resulting in lower signal and higher quantum noise properties from the incoming detector signal. This, in turn, presents itself as quantum noise in the resulting images and serves to decrease the signal-to-noise ratio (SNR). For digital mammography, the effect of increased quantum noise on clinical image quality has not been fully evaluated. Recent studies have shown that for certain detection tasks in digital mammography, lowering the radiation dose (hence increasing noise) by amounts of up to 50% of the standard AGD has no significant effect on the assessment of the images [Hemdal et. al, 2005b][Hemdal, 2003][Gennaro, 2004][Svahn, 2004].

If the dose could be reduced by 50%, this would either spare women 50% of the absorbed dose in mammography screening over their lifetime or it would allow for more projections to be acquired (such as required in tomosynthesis) at the same AEC setting, set for the average glandular dose inherent in Malmö. In either case, there is a potential for an increased benefit-to-risk ratio.

In order to fully evaluate the effects of lowering the standard dose levels, one must have a means of evaluating the image quality resulting from images acquired at lower doses. Image quality could be measured using phantom- or clinical images, or by study imaging system properties like the DQE [EC, 2005]. Clinical image quality is measured via clinical trials such as ROC[1] [Metz, 1986], JAFROC[2] [Chakraborty, 2004a], VGA[3] [Tingberg, 2003] etc, in which observers (radiologists) study a set of images and make detection tasks, the concept being that performance for a specific task improves with higher image quality. Contrast-detail phantoms are also useful tools that give an indication of some of the physical properties of the imaging system, however they are limited in terms of clinical relevance [van Engen, 2003].

The primary limitations of ROC are that it cannot handle multiple lesions in the image or the localization of lesions [Chakraborty, 2004b]. Therefore, in order to achieve a high level statistical accuracy, one needs to include many images in a study for each test parameter (dose level, in this case) as ROC has a relatively low statistical power; a measure of the method's ability to detect true differences between modalities (dose levels in this study), without falsely declaring them to be different. To compensate for these limitations, the JAFROC method is proposed, having the ability to handle multiple lesions and takes information about the locations in the free-response ROC (FROC) paradigm with more statistical power [Chakraborty, 2004a].

[1] Reciever operating characterstics (ROC) is widely used in clinical trials but is limited in just allowing one mark per image taking no localization information into account.
[2] The jackknife free-response receiver operating characteristic (JAFROC) method is described in the theory section.
[3] Visual grading analysis (VGA) is a trial where the observer grades the visibility of defined structures in a set of images.

Recently a study has been done using JAFROC analysis, comparing unprocessed phantom images acquired at different dose levels [Svahn, 2005]. Simulated microcalcifications and malignant masses were physically inserted into an anthropomorphic breast phantom (RMI-165). The phantom was then exposed under automatic control exposure (AEC) conditions referred as the 100% dose, and at 50% and 30% of that dose, using a Siemens Mammomat Novation unit. Eight observers including 3 trained radiologists observed all the images under low ambient light conditions, with the use of an in-house graphical user interface (ViewDEX) [Börjesson, 2005], they were able to mark selections on the computer monitor indicating where they detected a lesion. Their responses were exported to analysis software (JAFROC v.1.04) [Chakraborty, 2005] where a figure of merit (FOM) was calculated for each dose level. The results indicated that there was no statistically significant difference in lesion detectability between the 100% and 50% dose levels, while the FOM was slightly yet statistically higher for the 100% dose level compared to the 30% dose level. This study represented one of the first studies in optimisation of direct digital mammography using JAFROC. Some of the limitations of this study were that only a single phantom was used, not necessarily representing the true appearance of a normal breast and that image processing was not performed on the images, which is normally done in the clinic.

In the current study, an attempt has been made to overcome these limitations and to further examine the effects of lower doses in digital mammography using the JAFROC methodology as a tool to evaluate image quality. Clinical images are used (acquired from the screening department at Malmö University Hospital, using a Siemens Mammomat Novation full-field digital mammography unit) with inserted computer simulated tumors (to yield “hybrid” images) [Ruschin, 2005]. These hybrid images are put into a dose reduction program, simulating an image taken at lower dose, using a method adapted from Båth et al. [Båth, 2005]. The noise properties in the dose reduced image are equal, i.e. yielding the same noise power spectrum as that in an image actually collected at the reduced dose level. By creating hybrid images and simulating different dose levels, it is not necessary to receive ethical approval from radiation protection committees, since no additional exposure of the patients is required. The discrete dose levels were chosen based on previous studies.

The aim of this project was to investigate tumor detection in clinical images at lower dose levels in DM (30% and 50% of the original exposure) using hybrid images.

Theory

Dose reduction

In order to study the effect of dose reduction on image quality, clinical images must exist at the dose levels one wishes to examine. Rather than acquiring these patient images directly , which would involve additional exposures, the digital mammograms used in this study were acquired from routine screening at Malmö University Hospital (always under AEC conditions) and subsequently modified to represent lower dose levels by adding simulated

quantum noise. The greater the extent of the noise added, the greater the dose reduction was compared with the original clinically acquired image.

The theory of the dose reduction simulation were developed by Båth et al. [Båth, 2005], for digital CR chest images. It is based upon the assumptions that the digital system is linear with respect to entrance dose. Under these assumptions, the noise power spectrum (NPS) of the system, which can be measured at any dose level, can be used to enhance the noise of an image acquired at any given dose level. The method is easily verified if the noise in the simulated image agrees with the noise in an original image taken at the (simulated) lower dose level. Local dose variations are taken into account, which locally increases or decreases the noise according to signal variations set by the projected object (anatomy in this case). In an inhomogeneous image, the signal to noise ratio (SNR) is higher in high dose regions and the SNR is lower for the lower dose regions.

For homogeneous images (i.e. no object present in the radiation field) with known NPS acquired at the original,'orig', and simulated dose level, 'sim', the mean pixel value, P_{mean} , of the relevant area is proportional to the dose, D, which shows a linear relationship given by

$$P_{mean,sim} = P_{mean,orig} \cdot \left(\frac{D_{sim}}{D_{orig}} \right). \quad (1)$$

Due to this relationship, it is also possible to scale the original image, Im, to the simulated level according to

$$\mathrm{Im}(x,y)_{orig,scaled} = \mathrm{Im}(x,y)_{orig} \cdot \left(\frac{D_{sim}}{D_{orig}} \right), \quad (2)$$

where x and y is the spatial coordinates. The NPS scales with the square of the dose given by

$$NPS(u,v)_{\mathrm{Im}_{orig,scaled}} = NPS(u,v)_{\mathrm{Im}_{orig}} \cdot \left(\frac{D_{sim}}{D_{orig}} \right)^2, \quad (3)$$

where u and v refers to the coordinates in the frequency domain.
If the image was actually collected at D_{sim}, then the $NPS(u,v)_{\mathrm{Im}_{noise}}$ could be added to the scaled $NPS(u,v)_{D_{orig}}$ [Båth, 2005]:

$$NPS(u,v)_{D_{sim}} = NPS(u,v)_{D_{orig}} \cdot \left(\frac{D_{sim}}{D_{orig}} \right)^2 + NPS(u,v)_{\mathrm{Im}_{noise}}. \quad (4)$$

To take the local dose variations into account over a inhomogeneous image the corrected image, corr, is given by

$$\mathrm{Im}(x,y)_{noise,corr} = \mathrm{Im}(x,y)_{noise,uncorr} \cdot \sqrt{\frac{\mathrm{Im}(x,y)_{orig,scaled}}{P_{mean,sim}}}, \qquad (5)$$

based on the assumption that the DQE is constant over the dose variations in the image [Båth, 2005]. The DQE is also kept constant for the clinical image and the flatfield image at the same dose level in the dose reduction method. The variance of every pixel value is proportional to the dose which is in turn proportional to the pixel value. According to equation (5) the variance of pixel values increases with the square of the scaling of the image. In this way the noise image is corrected for local dose variations, so it could be added to the scaled original image.

Deterministic effects like edges of the image or structural noise of the flatfield images are eliminated in the frequency domain by rotation and flipping of a central part of the noise image to create a homogeneous full scaled image. Reproducible non-uniformities such as arising from the heel effect are subtracted from the noise image by fitting a second grade polynomial to the flatfield images.

JAFROC - scoring and statistical analysis

The JAFROC data can be analysed in a scoring step and a statistical analysis step [Chakraborty, 2004b]. The scoring step reduces the JAFROC data for all readers for each modality (dose level in this study) to a single value called figure of merit (θ), FOM, given by

$$\left.\begin{aligned} \theta &= \frac{1}{N_N \cdot N_A} \sum_{i=1}^{N_N} \sum_{j=1}^{N_A} \sum_{k=1}^{n_j} W_{jk} \cdot \psi\left(X_i, Y_{jk}\right) \\ \psi(X,Y) &= \begin{cases} 1.0 \; if \; Y > X \\ 0.5 \; if \; Y = X \\ 0.0 \; if \; Y < X \end{cases} \\ \sum_{k=1}^{n_j} W_{jk} &= 1 \end{aligned}\right\}, \qquad (7)$$

where N_N is the number of normal cases, N_A is the number of abnormal cases, W_{jk} is a weighted function (clinical importance of findings, not used for this study), ψ is a function depending on X_i, the highest noise rating for normal case i and Y_{jk} is the signal rating for target k on case j [Chakraborty, 2004b]. The FOM is by equation (7) the weighted probability that a signal rating exceeds a noise rating and is bound between 0 and 1, worst and best of the observers performance, respectively.

In the statistical analysis step the cumulative observer performance scores (FOM) for each modality (dose level in this study) are compared to each other to determine if there is a statistically significant difference between any of them. This is generally known as hypothesis testing. The hypothesis test is based on a mixed model ANOVA[1] analysis of the pseudo values explicitly showing the truth dependence [Chakraborty, 2004b]. The original dose level is compared to the lower dose levels to see if they differ or could be identical without false declaration. The test statistic is modelled by two unit-normal distributions: the null hypothesis, NH, and alternative hypothesis, AH, as shown in figure 1. NH declares the modalities to be different while AH declares no difference [Chakraborty, 2004b]. There is also a critical value that rejects the NH and accepts the AH. This value is set by the probability value P provided from the F test statistic used in ANOVA i.e the probability that a modality pair is different.

Another useful statistical test is confidence interval testing. A 95% confidence interval (CI) is calculated comparing each modality pair. The difference between two FOM's is checked to see if it is inside the 95% CI, and if zero is contained in that interval then there is no significant differences between the modalities compared.

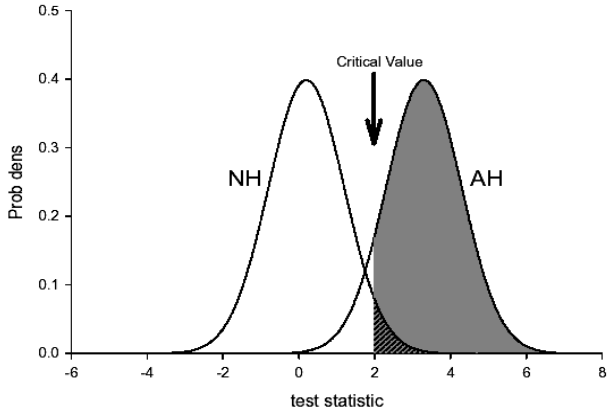

Figure 1. Null hypothesis test, showing the probability density function, where NH is null hypothesis and AH is alternative hypothesis..

Materials and methods

Image collection

It is very rare to find breast cancers in the screening program, since less than 1% of the women undergoing screening are thus diagnosed [Ruschin, 2005]. It is therefore preferable to simulate unique lesions. As the basis of this study, 90 clinical images were collected from the

[1] ANOVA tests the variances of all samples.

mammography-screening department at Malmö University Hospital. All the images were acquired with a direct-digital mammography system, Siemens Mammomat Novation. The calibration of the mammography unit was performed weekly so all 90 clinical images were collected within a one-week period to ensure consistency of the system. Women typically receive four exposures in a screening examination: R- and L-MLO (right and left breast, mediolateral oblique projection), and R- and L-CC (right and left breast, cranio–caudal projection). Only the unprocessed images were saved as they should be linear with respect to entrance dose, which is a requirement for the dose reduction and structure insertion methods.

In this study, only the R- and L-MLO projections were used, because the mammograms contains a lot of breast tissue over the mammogram area, and some radiologists are used to view R- and L-MLO only. Breast composition can be classified according to the amount of fibroglandular or dense tissue within the breast, based on the BI-RADS classification [ACR, 2003][Obenauer, 2005]. For this study, only those breast types that were categorized as fat/dense or dense by a radiologist according to their amount of glandular tissue were included in the study as these represent the most clinically challenging, and hence most relevant, cases. The x-ray tube potential ranged between 27-32 kVp, depending on the thickness and density of breast, determined by the AEC of the mammography unit. All the images were taken with the W/Rh anode filter combination. A radiologist selected the patient images collected from the screening that did not contain any visible pathologies.

Flatfield images for the dose reduction method

To calculate the noise-power spectrum (NPS) needed for the dose-reduction program, a set of flatfield images (images with only uniform objects in the beam) were collected to represent all dose levels and breast thicknesses used in this study. A homogenous polymethylmethacrylate (PMMA) (breast-equivalent density) phantom was placed on the detector covering all pixel elements with a margin of at least 1 cm outside the pixel area. The flatfield images representing 100% dose level were collected at the tube loading (in units of milliampere-seconds (mAs)) determined by the AEC setting, and the mAs settings representing the 50 % and 30 % dose level were calculated and collected in manual mode. With varying PMMA thicknesses, tube voltages and exposures, a total of 63 flatfield images were collected to represent all breast types as follows:

27 kVp, 20 to 50 mm thickness in steps of 5 mm, at 100, 50 and 30% dose,
28 kVp, 30 to 60 mm thickness in steps of 5 mm, at 100, 50 and 30% dose,
32 kVp, 40 to 70 mm thickness in steps of 5 mm, at 100, 50 and 30% dose.

Linearization of images

A prerequisite for the dose reduction method is that the images that should be dose-reduced are linear, i.e. the pixel values are linear to entrance dose (also exposure). Most digital systems are linear or could be set up to be so. The detector used in Siemens Mammomat Novation mammography unit is linear but has an offset pixel value for zero exposure to avoid

having negative pixel values, which would affect image processing. This offset value was calculated for each pair of flatfield images mentioned above. The mean pixel value for the same region of interest (ROI) in both images (each at different dose levels) was calculated and plotted as a function of mAs[1] (figure 2). The y-intercept of a linear fit made between these values represents the offset for the detector. By subtracting this offset, the system was linearized, i.e. the resulting best-fit line will have a y-intercept of zero.

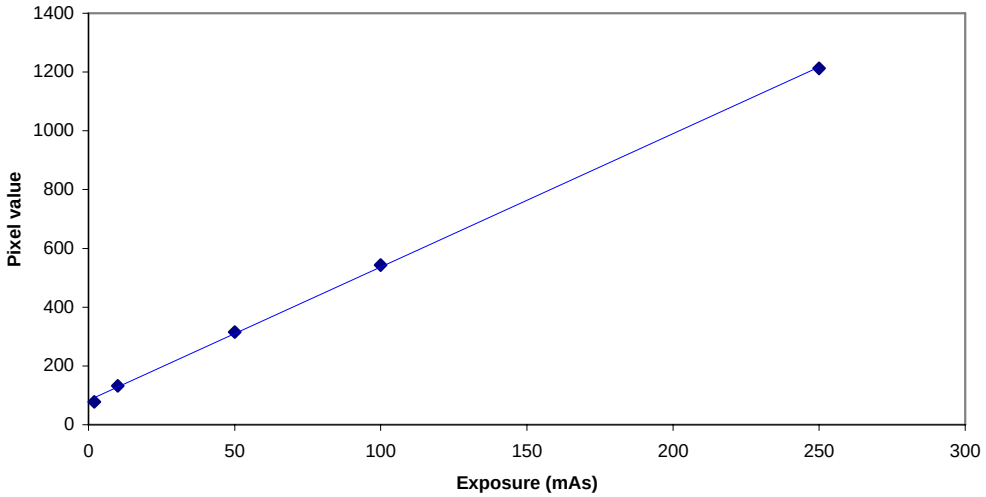

Figure 2. *Linearity test for five ROI's doing a linear fit. Notice the offset value at zero exposure.*

Insertion of structures

Computer simulated structures were inserted into the clinical unprocessed images, thus creating hybrid images (a normal image plus an added simulated lesion) [Ruschin, 2005]. In dense breasts where the tissue is dense or inhomogeneous, care should be taken when inserting lesions to make them look realistic. Lesions representing irregular and slightly spiculated malignant masses were collected from Saunders et al. [Saunders, 2004]. These structures were scaled in terms of density and size to have low image contrast and a diameter of roughly 1 cm. An ROC analysis has shown these structures to be very similar to real lesions [Saunders, 2004]. The characterization of the structures has been matched with cases from the Breast Imaging and Reporting Data System, BI-RADS [Obenauer, 2005][ACR, 2003], showing good agreements [Saunders, 2004].

Using a structure insertion computer program written in IDL [Ruschin, 2005], the researcher can control properties of the lesion like size, sharpness, margins, shape and contrast. It is also possible to insert the lesion anywhere in the linear unprocessed raw data image. The contrast

[1] The entrance dose is proportional to tube loading (in units of mAs).

of the lesion is normalized against the background (figure 3), and can be varied to represent different densities. The appearance of the lesions was simulated in log-space as x-ray transmission through an object consists of exponential attenuation. Hence the lesions are added to the x-ray images logarithmically:

$$I_{new,image}(x,y) = I_{structure,image}(x,y) \cdot e^{-c \cdot S(x,y)}, \qquad (6)$$

The used structures have been inserted with the assumption that different beam qualities do not change the appearance of the shape and contrast of the lesion. In some cases, the contrast of the structure were subsequently slightly increased to a visibility threshold set by a radiologist.

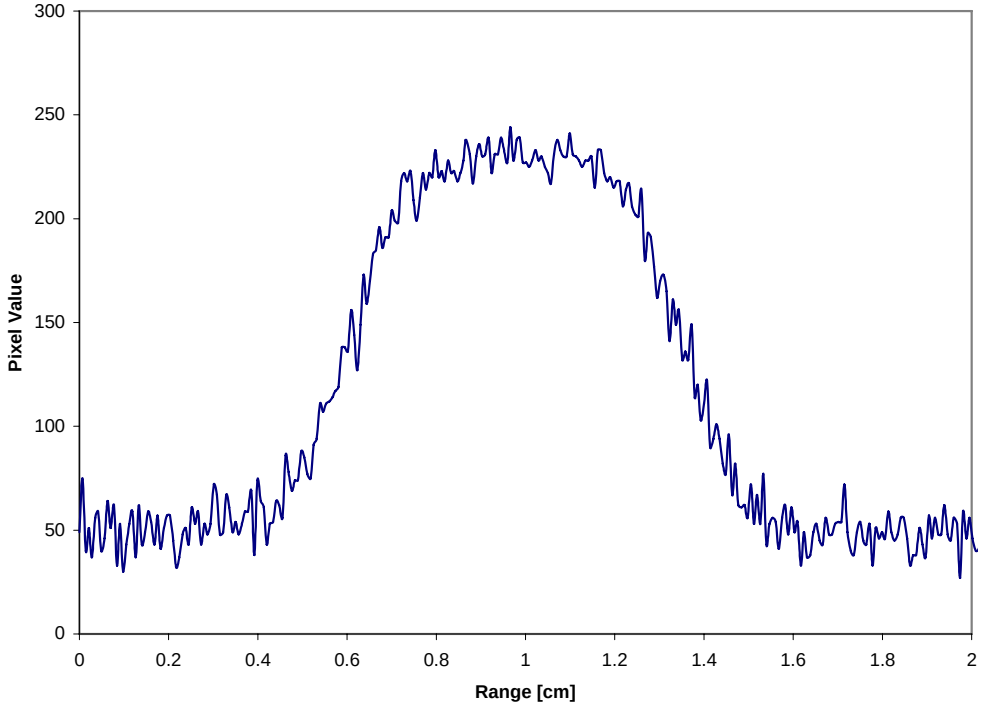

Figure 3. *Line profile of structure inserted on a flatfield image.*

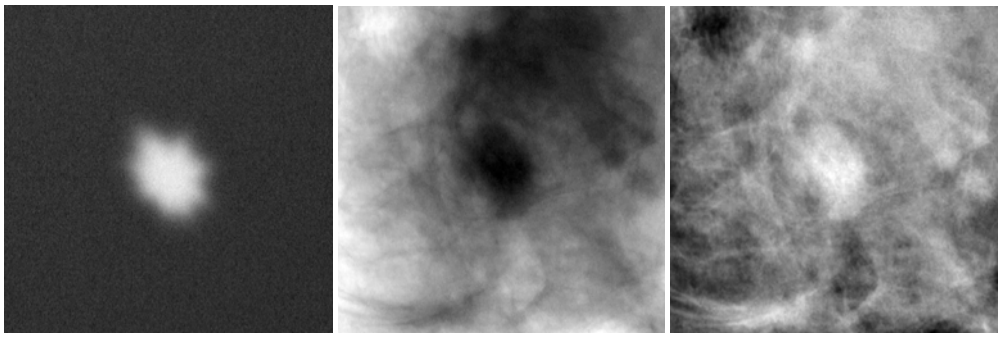

Figure 4. *Example of inserted lesion on a flatfield image (left), raw data showed in log space (middle) and processed image (right).*

The dose reduction method

To determine the $NPS(u,v)_{Im_{noise}}$ in equation (4) in practise, two flatfield images are collected at two different dose levels D_1 (close to the simulated dose level) and D_2 (close to the original dose level). The DQE is approximated to be constant for these dose levels. The $NPS(u,v)_{Im_{noise}}$ is then given by [Båth, 2005]:

$$NPS(u,v)_{\mathrm{Im}_{noise}} = NPS(u,v)_{D_1} \cdot \frac{D_{sim}}{D_1} - NPS(u,v)_{D_2} \cdot \frac{D_{sim}^2}{D_{orig} \cdot D_2}. \qquad (7)$$

To create the noise image that finally is added to the scaled original image, white noise with a normal distribution around zero is added (so the mean pixel value of the image will not change) and a filter given by the frequency components determined by the square root of the NPS is applied. The noise image is finally corrected for local dose variations according to equation (5), before it is added to the scaled original image. A normalized NPS calculated from a homogeneous area in an unprocessed phantom image is shown in figure 5. The NPS in the simulated image at 30 % dose agrees well with the image collected at 30% dose. Figure 6 visualizes the noise in processed mammograms at the simulated dose levels compared to the original exposure.

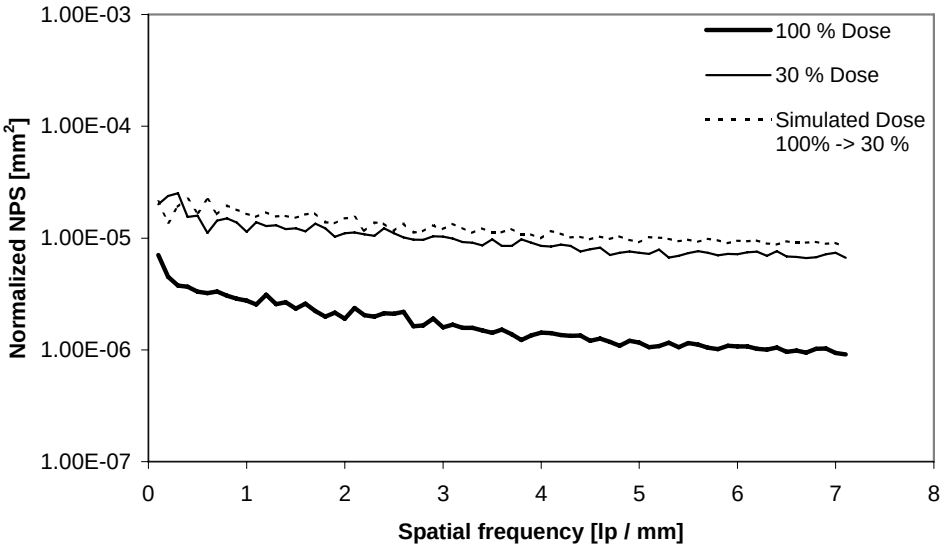

Figure 5. *Normalized NPS for a homogeneous phantom.*

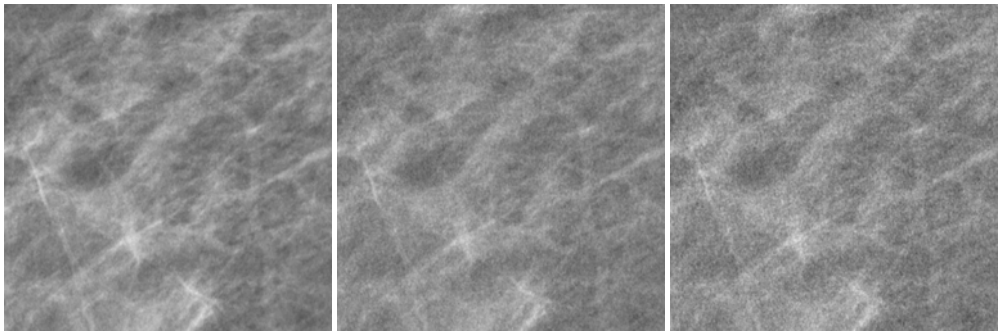

Figure 6: *Clinical exposure taken at full dose compared to the simulated images at 50 % and 30 % dose level in order left to right.*

Processing

All the physical manipulations like dose-reduction and insertion of structures are done on unprocessed images, which ideally are linear with respect to entrance dose, which is in turn linear with respect to mAs. However, since unprocessed images are not used clinically, radiologists are not accustomed to viewing them. Therefore, as the final step before viewing the images in the study, the images were processed according to Siemens image processing algorithms. The purposes of the processing are primarily:

- edge and contrast enhancement of structures,
- background suppression,
- stretching of the dynamic range to a sigmoid or log curve.

The exact processing algorithm is proprietary information held by Siemens, but operates on the principle of multi-scale frequency enhancement. The processing leads to a non-linear relationship between exposure and pixel values, and unique window-level settings are derived for each image and stored in the DICOM header for the final display on the monitor viewed by radiologists. Figure 7 illustrates the difference in appearance between an unprocessed an processed image.

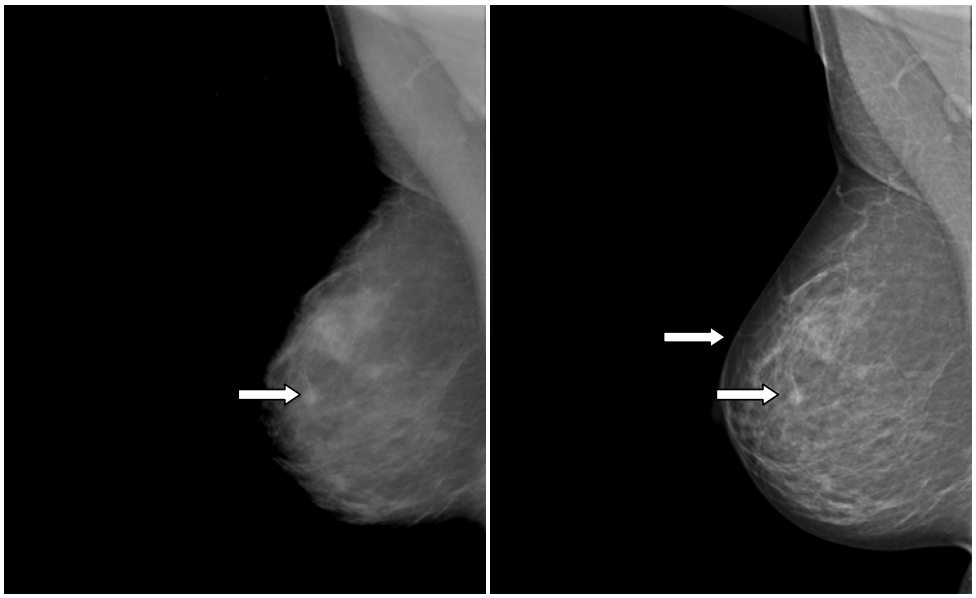

Figure 7. ***Left***: *Unprocessed image: arrow pointing at inserted structure. The skin line is not visible in this image.* ***Right****: Same image after processing: upper arrow pointing at enhanced skin line and lower at inserted structure.*

Viewing session, setup for JAFROC

Five structures collected from Saunders et al. [Saunders, 2004] were chosen randomly and inserted according to table 1 into the mammograms at the original dose level. To have a counter balance between the dose levles, 40 abnormal (table 2) and 50 normal cases were used per dose level. All the abnormal and normal cases were dose reduced to 50 % and 30 % of the original exposure, creating a total of 270 images and three dose levels.

Four observers participated in this study, one radiologist and three physicists. In order to avoid observers recalling which images they selected as abnormal or normal between the dose levels, the total viewing for each observer was split into three sessions, one week apart. In each viewing session 90 images were shown, showing 30 images per dose level.

Table 1. *The distribution of number of lesions for each dose level*

Abnormal images	Number of masses
20	1
15	2
5	3

Viewer for Digital Evaluation of X-ray images, ViewDEX

In order to give the observers a user friendly basis for the FROC study, a digital viewer, ViewDEX, was used (figure 8), developed by Börjesson et al. [Börjesson, 2005]. Tools like zoom, pan and window/level were provided to facilitate the observers´ evaluation of the images. ViewDEX also handles multi-reader multi cases, MRMC, methods (as intended for JAFROC) that increases the statistical efficiency and generalizes differences between observers such as radiologists and physicists. The mammography images were scaled to fit the entire display of the monitor. The monitor used was a 5 MP, 8 bit CRT monitor calibrated according to the DICOM standard and EUREF [van Engen, 2003]. The window/level setting for each image was set using parameters found in the DICOM headers. The observers marked the lesions (0–3 per image), and gave each mark a rating according to a confidence level, described in table 2.

Table 2. *The four-level confidence scale used in the FROC 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

The centre coordinates of the inserted masses were stored in a so called property file. Any mark made by an observer within a fixed radius of these coordinates resulted in a true positive rating. The structures used in this study had a mean diameter of 1.1 cm so the radius of the ROI was set to 0.7 cm to represent a clinically relevant distance. A mark placed within an ROI lead to a true positive, TP, whereas a false positive, FP, indicated marks placed outside this region. The observers were given a training session at each viewing occasion, to get used to view hybrid mammograms and using ViewDEX. Information about the number of TP / FP and their confidence values were assembled in a log file, that was used as the input file for the JAFROC software [Chakraborty, 2005].

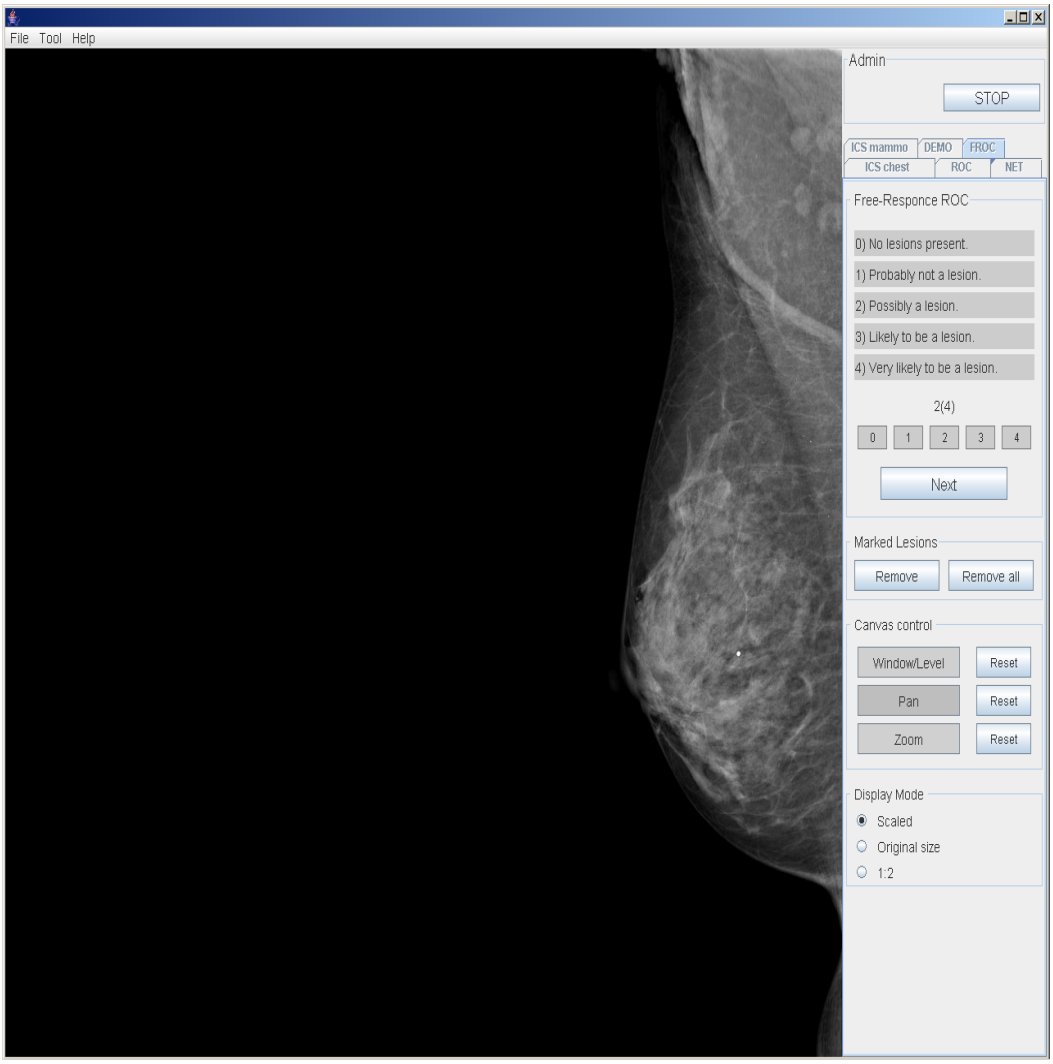

Figure 8. The ViewDEX interface [Börjesson, 2005].

Results

The figure-of-merit (FOM) resulting from the JAFROC scoring step is shown in table 3 for each reader and dose level. The results of JAFROC statistical analysis when comparing all three dose levels gives a p-value of 0.15 and a F-stat of 1.90. This result indicates that there is no significance at the 5% level.

Table 3: *Figure of merit for all readers at each dose level. Last row gives the mean FOM for all readers.*

Reader	100 % dose level	50 % dose level	30 % dose level
Radiologist	0.71	0.73	0.72
Physicist 1	0.62	0.72	0.68
Physicist 2	0.43	0.49	0.54
Physicist 3	0.70	0.72	0.73
All readers	**0.61**	**0.66**	**0.67**

The difference of each pair of FOM's is calculated, and the 95% CI limits are calculated around that difference (table 4). The 95% CI indicates no significance in the difference between mean FOM's, when zero is included in this interval:

Table 4: *Differences in mean FOM's and corresponding 95% CI's.*

	FOM(100% dose) - FOM(50% dose)	FOM(100% dose) - FOM(30% dose)	FOM(50% DOSE) - FOM(30% DOSE)
FOM difference	-0.048	-0.054	-0.006
95% CI	[-0.109, 0.012]	[-0.116, 0.006]	[-0.067, 0.054]

Discussion

The results from the JAFROC analysis indicate no statistically significant differences between the dose levels studied. As shown in table 4, the mean FOM for all readers indicates that there is a slight and statistically insignificant increase in lesion detection at the 30% dose level compared to the 100% dose level. This could be caused by the increased noise in normal images obscuring anatomical regions that may otherwise be considered suspicious. This would lead to a decrease in the number of false-positives in the normal images at the lower dose images, and hence an increase in figure-of-merit scores. Once more observers have participated, this can be tested by analysing the FP rate in the normal images at each dose level. Three more radiologists are expected to participate in the FROC study to increase statistical accuracy.

Even though there is no statistically significant difference found between the FOM's for the different dose levels, this does not automatically mean that the currently used dose levels should be reduced without further thought. More studies need to be done to confirm this finding. One such study would be to collect a large amount of rawdata images from the screening department, especially cases that include actual cancer findings. These images could then be dosreduced and be used in a study in detectability of the cancer findings. High workflow and the limitations in data storage via the PACS system makes this task difficult. In Malmö, exposures are taken at four orientations (R-MLO, L-MLO, R-CC and L-CC), resulting in four processed images and four rawdata files. All the files are in DICOM format, and the processed images are sent and archived in the PACS. For future studies it would be helpful to be able to process the raw data locally at any time. Currently, the unprocessed hybrid images had to be processed by Siemens, which created delays when sending images back and forth for evaluation of the method.

There is an increasing amount of publications indicating that system noise (including quantum noise) is of lesser importance compared to the anatomical background in detection tasks [Bochud, 1999]. However, detection tasks can still be affected by quantum noise depending on the properties of the lesions such as shape, size, localization and density. Aside from masses, microcalcifications are also highly indicative of malignant cancer and hence clinically relevant structures. Since microcalcifications are very small (typically less than 0.5 mm in diameter), they can be more likely to disappear in the presence of a large amount of system noise at low dose levels. Characteristics of tumors should also be evaluated at different dose levels, to see if the observers can make a diagnosis according to their finding.

The normalized NPS in figure 5 shows that the simulated 30% dose level and the true 30% dose level, differs not in shape of spectra but in amplitude. The simulated 30% dose level is actually representing a somewhat lower dose level. This is probably caused by the setup and method used in this study i.e the system was not perfectly linearized, or that the flatfield images were not perfectly representing all breast types. The manual system settings were also user limited when choosing mAs, so the closest setting was used to come as close to the 30% and 50% dose level.

Even with rather high contrast (20 – 30% of the background), edges of the inserted structures were blurred out. In a real dense breast structures appeared quite well even at low contrast, due to homogeneity in background. The anatomical background sometimes could be deceiving when a low contrast mass is inserted even if it was normalized against the background. Breast nodules are often mismatched against the inserted lesions under these conditions.

In clinical situations, radiologists never view just one projection of a breast, but rather multiple projections simultaneously and even supplementary material such as patient history, and ultrasound and MR images if they exist. To consider a mass and interpret its localization

and to overcome overlying structures two projections of the breast are needed. In this study, only one image at a time was showed to the observer. This is not a perfect simulation of the clinical situation. However, in image quality trials the authenticity must sometimes be compromised. To avoid having a bias from the somewhat artificial viewing conditions a pre-trial was conducted so that the observers learned how the trial was going to be conducted as well as instructing them about the characteristics of the mass.

Image processing, like non-linear contrast amplification, results in non-linear systems, which are not easy to physically interpret. The processed images are good in enhancing structures and contrast, but it could be misleading in terms of the true information it presents. When the structures were inserted in the original unprocessed images they structures were sometimes easier to find than in the processed images. This could be that the background details are blurred out, and that the structures had relatively sharper margins at the moment of insertion, which became toned down when the images were processed. When zooming out or screwing ones eyes, it was easier to see the density region corresponding to the inserted masses. The zooming out effect could be of a higher intensity per pixel area. For future studies, conspicuity[1] should be taken into account when inserting the structures.

Another possibility for future optimisation studies could involve searching for optimal window/level settings. With present processing techniques, and the great amount of intensity levels provided in a 12-bit mammogram (4096 intensity levels), there is a whole range of possibilities to optimise image quality. The human eye can only detect about 150 grey levels [Hemdal, 2002], so a 8 bit monitor containing 256 intensity levels should be sufficient, although there is information in the image containing 4096 intensity levels. Several presets of window level settings could also be set up for different detection tasks in the screening. It is therefore of importance to optimise digital mammograms to get correct interpretation of the image for correct diagnostic outcomes, and at the same time reduce the dose for the patients. ICRP [ICRP, 1996], and the European Medical Exposure Directive [European Union, 1997] gives recommendations for this task.

Before any dose reduction could be implemented in routine screening, ethical approval must be obtained for acquiring additional exposures of patients with and without confirmed malignant disease for use in a clinical trial comparing detection of pathologies between full dose and lower dose images. It should be obvious that this study on its own does not confirm a dose reduction to a certain level, but it indicates a possibility to do so.

Conclusions

There were no significant difference in tumor detection between the 100, 50 and 30% dose levels in this study, indicating a potential for dose-reduction. Further studies could involve

[1] Conspicuity is the contrast devided by the complexity of the background.

different shaped masses, microcalcifications, different processing algorithms, and window/level settings.

Acknowledgements

Thank you gentlemen: Mark Ruschin, Magnus Båth, Bengt Hemdal, Ingvar Andersson, Anders Tingberg, Daniel Fischer, Thomas Mertelmeier, Dev P. Chakraborty, Tony Svahn, Sune Svensson, Sören Mattsson, Ûnal Öhren,Bob Smulders and all staff at the Department of Medical Radiation Physics in Malmö University Hospital and at the Department of Medical Physics in Sahlgrenska University Hospital.

References

American College of Radiology (ACR). (2003) Illustrated breast imaging reporting and data system (BI-RADS), 4th edn. ACR, Reston, VA.

Bochud FO, Valley JF, Verdun FR, Hessler C and Schnyder P. (1999) Estimation of the noisy component of anatomical backgrounds. Med. Phys. 26 (7), 1365-70.

Båth M, Håkansson M, Tingberg A, and Månsson LG. (2005) Method of simulating dose reduction for digital radiographic systems. Radiat. Prot. Dosim. 114(1-3), 253-259.

Börjesson S, Håkansson M, Båth M, Kheddache S, Svensson S, Tingberg A, Grahn A, Ruschin M, Hemdal B, Mattsson S and Månsson LG. (2005) A software tool for increased efficiency in observer performance studies in digital radiology. Radiat. Prot. Dosim. 114(1-3), 45-52.

Chakraborty DP. (2005) JAFROC software, http://jafroc.radiology.pitt.edu.

Chakraborty DP and Berbaum KS. (2004a) Jackknife Free-Response ROC Methodology; Medical Imaging. SPIE Vol. 5372.

Chakraborty DP and Berbaum KS. (2004b) Methodologists for Observers Studies Involving Detection and Localization: Modelling, Analysis and validation. Med. Phys. 31, 2313-2330.

Chakraborty DP. (1996) Observers Performance Methodology in Medical Imaging. Notes for SPIE-96 Tutorial Course, 1-16.

EC (European Comission). (2005) European Protocol for the quality control of the physical and technical aspects of mammography screening. Draft Edition, Brussels, Belgium.

EC (European Comission). (1996) European Guidelines on quality criteria for diagnostic radiographic images. EUR. 16260, Brussels, Belgium.

van Engen R, Young K, Bosmans H and Thijssen M. (2003) Addendum on digital mammography to chapter 3 of the European Guidelines for quality assurance in mammography screening, version 1.0, EUREF.

European Union. (1997) Council Directive 97/43/Euratom on Health Protection of Individuals against the Dangers of Ionising Radiation in Relation to Medical Exposure. Official Journal of the European Communities No. L180/22-27.9.07.97.

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

Hemdal B. private communications (2005a).

Hemdal B, Herrnsdorf L, Andersson I, Bengtsson G, Heddson B and Olsson M. (2005b) Average glandular dose in routine mammography screening using a Sectra MicroDose Mammography unit. Radiat. Prot. Dosim. 114(1-3), 436-43.

Hemdal B, Bay TH, Bengtson G, Gangeskar L, Martinsen, AC, Pedersen K, Thielander-Klang A and Mattson S. (2003) Comparasion of screen-film, image plate and direct mammography with CD phantoms in Digital Mammography. The 6th international workshop on digital mammography, IWDM 2002, Bremen, Germany. Edited by Peitgen, H.-O. Springer verlag Berlin, Heidelberg, New York, 105-107.

Hemdal B, Andersson I, Thilander-Klang A, Bengtsson G, Leitz W, Bjurstam N, Jarlman O and Mattsson S. (2002) Mammography – recent technical developments and their clinical potential. SSI report 2002:08.

ICRP Publication 73 (1996) Radiological protection and safety in medicine. Ann. ICRP 26.

Metz CE. (1986) ROC methodology in radiologic imaging. Invest. Radiol., 21, 720-733.

Muller S. (1997) Full-field digital mammography designed as a complete system. Eur. J. Radiol. 31, 25-34.

Obenauer S, Hermann KP and Grabbe E. (2005) Applications and litterature review of the BI-RADS classification. Eur. Radiol. 15, 1027-1036.

Ruschin M, Tingberg A, Båth M, Grahn A, Håkansson M, Hemdal B and Andersson I. (2005) Using Simple Mathematical Functions to Simulate Pathological Structures – Input for Digital Mammography Clinical Trial. Radiat. Prot. Dosim. Vol. 114, No. 1-3, 424-431.

Saunders RS and Samei E. (2004) Characterization of Breast Masses for Simulation Purposes Proc. SPIE 5372, 242-250.

Svahn T, Ruschin M, Hemdal B, Chakraborty D, Andersson I, and Tingberg A. (2005) Dose optimisation of direct digital mammography using the JAFROC and standard ROC method. Manuscript in preparation.

Tingberg, A. (2000) Quantifying the quality of medical x-ray images. An evaluation based on normal anatomy for lumbar spine and chest radiography. Thesis, Lund University, Malmö, Sweden.

Zoetelief J, Fitzgerald M, Leitz W and Säbel M. (1996) European protocol on dosimetry in mammography. EUR. 16263 (Luxemburg: EC).