THE USE OF A FIGURE-OF-MERIT (FOM) FOR OPTIMIZATION IN DIGITAL MAMMOGRAPHY:

An Exploratory Study in Malta.

Mark Borg

PhD Thesis

30th September, 2015

Department of Medical Physics and Biomedical Engineering,

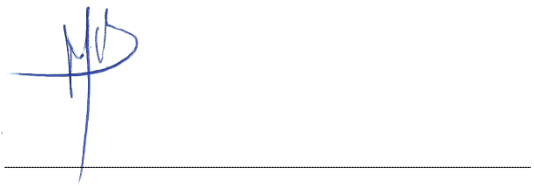
Malet Place Engineering Building,

University College London

DECLARATION

I, Mark Borg confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed,



Mark Borg

30th September, 2015

ABSTRACT

This PhD thesis comprises an exploratory study in digital mammography physics that portrays two essential components. The first component (1) presents the first national survey of the technical performance of mammography equipment in Malta using the *European Protocol* [1-3]. This demonstrated considerable differences in the technical performance of the mammography units across the country with a wide range in performance, patient dose and image quality. A common problem was that many clinics had implemented computed radiography (CR) systems to replace existing film-screen (FS) systems without due consideration to optimization. All direct digital (DR) mammography units met current international technical performance standards and the effectiveness of DR mammography in reducing patient dose and maintaining high image quality compared to CR has been confirmed. The second component (2) was to explore the use of a figure-of-merit (FOM) for optimization and characterisation in digital mammography. The use of image quality parameters in digital mammography such as contrast-to-noise ratio (CNR) or signal-difference-to-noise ratio (SDNR), signal-to-noise ratio (SNR) and detective quantum efficiency (DQE) have been traditionally used for the quantitative evaluation of the system performance against international standards or guidelines. The use of FOMs is relatively new and may be considered as a new quality assurance tool in digital mammography permitting the quantitative and simultaneous assessment of image quality and patient dose. The main objective in having a FOM is to have a numerical expression representing the efficiency and efficacy of a given system gauging how good or poor a system is performing. This may be useful in optimization and in predicting a predetermined or expected image quality with a given amount of radiation dose for a given system. The most interesting aspect of the FOMs in this work will be to investigate and explore the possibility for inter-system comparison.

DEDICATION AND ACKNOWLEDGEMENTS

I dedicate this work to my family, especially…

..to my lovely wife Joanne for her love, patience and understanding. I am truly thankful for having her in my life;

..to my two special sons, Benjamin and Gabriel who have grown into two lovely boys in spite of their father spending so much time away from them working on this thesis;

..to my late Mom for her love and who inspired me for this work and research;

..to my Dad for instilling the importance of hard work and higher education;

..to my Mom and Dad-in-law for their constant encouragement and motivation.

I am grateful to Professor Gary Royle and Dr. Ishmail Badr for their constant guidance who have generously given their time and expertise to better my work. I thank them for their contribution and their good-natured support.

I am also grateful for the support and advice from Dr. Anastasios Konstantinidis (Tasos) who offered unflagging support and wise advice.

TABLE OF CONTENTS

LIST OF EQUATIONS

$$HVL = a\,(tube\ voltage)^2 + b\,(tube\ voltage) + c$$ 19

$$MGD = DgN * K$$ 20

$$Dance_{predicted} = \frac{(Displayed + 0.564)}{1.737}$$ 21

$$Wu_{predicted} = 0.9623 Displayed - 0.0323$$ 22

$$Boone_{predicted} = 0.8788 Displayed + 0.0192$$ 23

$$Dance_{predicted} = \frac{(Displayed + 0.312)}{1.3}$$ 24

$$Boone_{predicted} = 1.1099 Displayed - 0.0367$$ 25

$$Threshold\quad Contrast_{measured} * CNR_{measured} = Threshold\quad Contrast_{limiting} * CNR_{limiting}$$ 26

$$Relative\ CNR = \frac{CNR_{measured}}{CNR_{limiting\ value}}$$ 27

$$NPS(u_n, v_k) = \frac{\Delta x \cdot \Delta y}{M \cdot 256 \cdot 256} \sum_{m=1}^{M} \left| \sum_{i=1}^{256} \sum_{j=1}^{256} (I(x_i, y_j) - - S(x_i, y_j)) exp - 2(\pi i(u_n x_i) + (v_k y_j)) \right|^2$$ 28

$$DQE(u) = \frac{MTF^2(u)}{q.E.NNPS(u)}$$ 29

$$NEQ(u) = DQE(u) * q$$ 30

$$MPV(Raw) = 16765.897 e^{-0.00121 * MPV(Processed)}$$ 31

$$SNR(Raw) = 16.165 + (6.349 * SNR(\text{Processed}))$$ 32

$$SNR\ (Raw) = -\ 47.453 + (13.343 * SNR\ (Processed\ ))$$ 33

$$H_T = \frac{1}{Contrast * \sqrt{Area(\det ail)}}$$ 34

$$Contrast(\%) = -1.78x^2 + 15.13x - 0.21$$ 35a

$$Contrast\ (\%) = 0.78x^2 + 13.05x + 0.01$$ 35b

$$Tc = a + b.x^{-1} + c.x^{-2} + d.x^{-3}$$ 36

$$Q(Qfactor1) = \frac{1}{n} . \sum_{i=1}^{n} \frac{H_T(Ai)}{Href(Ai)}$$ 37

$$Q_{norm}1 = \frac{Qfactor_1^4}{MGD}$$ 38

$$NEQ(u') = \frac{MTF(u')^2}{NNPS(u')}$$ 39

$NEQ_{MGD}(u') = \frac{MTF(u')^2}{NNPS(u')} . \frac{1}{MGD}$	40
$NEQ_{DC}(u') = NEQ_{MGD}(u').C\%$	41
$eNEQ(u') = \frac{MTF(u')^2(1-SF)^2}{NNPS(u')}$	42
$eNEQ_{MGD}(u') = \frac{MTF(u')^2(1-SF)^2}{NNPS(u')} . \frac{1}{MGD}$	43
$eNEQ_{DC}(u') = eNEQ_{MGD}(u').C\%$	44
$C\% = \frac{PV_{background} - PV_{Alumium}}{PV_{background}}$	45
$eDQE(u') = \frac{MTF(u')^2(1-SF)^2}{NNPS(u').TF.E.q}$	46
$FOM\ X = \frac{Qfactor1.CNR.DQE}{MGD}$	47
$FOM\ X1 = \frac{Qfactor2.IQF_{INV}.CNR\ (using\ CDMAM)}{MGD}$	48
$SDNR = 2C_{RAD}SNR$	49
$SNR_{IN} = \sqrt{n_0}$	50
$DQE = \frac{SNR_{OUT}^2}{SNR_{IN}^2}$	51
$NEQ(f) = \frac{SPS(f)}{NPS(f)}$	52
$f_N = \frac{1}{2p}$	53
$d' = \frac{speration}{spread}$	54
$d' = z(H) - z(F)$	55

Note: The equations in the text are not in numerical order. Therefore reference to the equation list should be made from the equation number quoted throughout the document. Each equation within the document is hyperlinked to the equation list.

ABBREVIATIONS

ACR - American College of Radiology

ACRIN - American College of Radiology Imaging Network

AEC - Automatic Exposure Control

Ag – Silver

Al - Aluminium

BI-RADS - Breast Imaging-Reporting and Data System

CC - Cranio Caudal

CD - Contrast Detail

CDMAM – Contrast detail mammography phantom

CIRS - Computerized Imaging Reference Systems

CNR - Contrast-to-noise ratio

CR - Computed Radiography

Cs(I) – Caesium iodide

d' - d prime

DMIST - Digital Mammographic Imaging Screening Trial

DQE - Detective quantum efficiency

DR - Digital Radiography

DR – Direct radiography

DRL - Diagnostic Reference Level

DRL_{IQ} - Diagnostic Reference Level (Image Quality)

E - Exposure (mGy)

eDQE - Effective Detective quantum efficiency

EDR - Exposure Data Recognizer

eNEQ - Effective Noise equivalent quanta

$eNEQ_{DC}$ - Effective Noise equivalent quanta (normalized to MGD multiplied by object contrast)

$eNEQ_{MGD}$ - Effective Noise equivalent quanta (normalized to MGD)

ESAK or K - Entrance Surface Air Kerma (mGy)

FFDM - Full Field Digital Mammography

FOM – Figure-of- Merit

FS - Film Screen

FSM - Film Screen Mammography

HVL - Half Value Layer

IAEA - International Atomic Energy Agency

ICRP - International Commission on Radiological Protection

IEC - International Electro technical Commission

IPEM - Institute of Physics and Engineering in Medicine

IQF_{INV} – Image Quality Figure Inverse

kV - Kilovoltage

MGD – Mean Glandular Dose (mGy)

MLO - Medio-lateral Oblique

Mo – Molybdenum

MPV - Mean Pixel Value

MTF – Modulation Transfer Function

NA - Not available

NDRL - National Diagnostic Reference Level

NEG_{DC}- Noise equivalent quanta (normalized to MGD multiplied by object contrast)

NEQ - Noise Equivalent Quanta

NEQ_{MGD} - Noise equivalent quanta (normalized to MGD)

NHSBSP - NHS Breast Screening Programme

NNPS – Normalised Noise Power Spectrum

NPS – Noise Power Spectrum

NS - Non-significant

OD – Optical Density

PMMA – Polymethyl Methacrylate

PPV - Positive Predictive Value

Q factor – Quality Factor

Q norm – Normalised Quality Factor

Rh – Rhodium

RMS - Root Mean Square

ROI - Region of Interest

SD - Standard Deviation

SDNR – Signal-difference-to Noise Ratio

Se - Selenium

SF - Scatter Fraction

SNR – Signal-to-noise Ratio

TCDD - Threshold Contrast Detail Detectability
TF - Transmission Factor
TFT – Thin Film Transistor
W - Tungsten

LIST OF JOURNAL PUBLICATIONS RELATED TO THIS THESIS

1. Borg, M., I. Badr, and G.J. Royle, *Mammography equipment performance, image quality and mean glandular dose in Malta.* Radiat Prot Dosimetry, 2013 [4].

2. Borg, M., I. Badr, and G. Royle, *Should Processed or Raw Image Data Be Used in Mammographic Image Quality Analyses? A Comparative Study of Three Full-Field Digital Mammography Systems.* Radiat Prot Dosimetry, 2014 [5].

3. Borg, M., I. Badr, and G.J. Royle, *A Study to Determine the Differences between the Displayed Dose Values for Two Full-Field Digital Mammography Units and Values Calculated Using a Range of Monte Carlo-Based Techniques: A Phantom Study.* Radiat Prot Dosimetry, 2012 [6].

4. Borg, M., I. Badr, and G.J. Royle, *The Use of a Figure-of-Merit (FOM) for Optimization in Digital Mammography: A Literature Review.* Radiat Prot Dosimetry, 2012 [7].

CHAPTER 1

INTRODUCTION AND BACKGROUND

Introduction

The Malta National Breast Screening Programme (MNBSP) is a service dedicated to breast screening and separate from the symptomatic services within the main general hospital, Mater Dei Hospital, Malta. It is sited at the National Screening Centre at Lascaris Wharf, Valletta, Malta. The MNBSP provides free breast screening every three years for all women residing in the Maltese Islands aged 50 to 60 years. The MNBSP is an effective part of the Maltese Government's effort to reduce the death toll from breast cancer. Research has shown that screening programmes in other countries have lowered mortality rates from breast cancer [8].

The programme was set up by the Health Care Services Division within the Ministry for Social Policy in late 2007 with the first direct digital (DR) mammography unit being operational in 2009, in response to the recommendations of a specific working group. The Government appointed a group of experts who have advised on setting up the programme according to European quality standards [1, 2]. Further still, the MNBSP is nationally coordinated and sets national standards which are monitored through European quality assurance networks.

The demand for mammography screening in Malta is on the rise mainly due to media awareness. Consequently, women of all age groups attend either the MNBSP or private clinics for their routine mammogram. Although, this is something good, the author of this work as a medical physicist felt that mammography practice in Malta should be explored and audited so that the existing digital mammography units could be evaluated, characterized, optimized and compared in a simple and effective way.

This exploratory work commenced with the (1) first Maltese national mammography survey confirming the effectiveness of DR mammography in breast cancer screening. Patient data was

made available from three clinics out of the participating nine. A dose survey of mean glandular dose (MGD) for 759 patients examined in the state-owned mammography facilities was performed. A MGD national diagnostic reference level (DRL) was set at 1.87 mGy for patients with breast compression thicknesses (BCT) between 5.0 cm and 7.0 cm. This wide range in BCT was chosen since patient data was retrieved from three clinics only and the results showed that other international BCT reference levels may be unsuitable for the Maltese population. The survey results have shown that the technical standard of mammographic equipment in the MNBSP is at par with other countries, including its Western European counterparts. However, the survey also showed that there is still need for optimization, especially the in units installed within the private sector.

The (2) second step of this exploration was done to evaluate the displayed MGD vis-à-vis the calculated values and whether the displayed values can be confidently used for DRL and optimization purposes. Modern DR units display the MGD and the entrance surface or incident air kerma (K or ESAK) to the breast following each exposure. Information on how these values are calculated is limited or absent and knowing how displayed MGD values compare and correlate to conventional Monte Carlo-based methods may be useful. From measurements done on polymethyl methacrylate (PMMA) phantoms, it has been shown that displayed and calculated MGD values are similar for thin to medium thicknesses and appear to differ with larger PMMA thicknesses. Consequently, a multiple linear regression analysis on the data was performed to generate models by which displayed MGD values on two DR units within MNBSP included in the study may be converted to the Monte Carlo values calculated by conventional methods. These models should be a useful tool for medical physicists requiring MGD data from similar DR units and should reduce survey time spent on dose calculations.

The (3) next step was to compare a number of measured image quality parameters using processed and unprocessed or raw images in the same DR units and one Computed Radiography (CR) mammography system. This part of the study was essential as it showed that difference between raw and processed image data are system-specific. The results have shown that there are no significant differences between raw and processed data in the mean threshold-contrast values using the CDMAM phantom in all the systems investigated, however these results cannot

be generalized to all available systems. Differences were noted in contrast-to-noise ratio (CNR) and in other tests including: response function, Modulation Transfer Function (MTF), Noise Equivalent Quanta (NEQ), Normalised Noise Power Spectra (NNPS) and Detective Quantum Efficiency (DQE) as specified in IEC 62220-1-2 [9]. Consequently, it was concluded and strongly recommended to use raw data for all image quality analyses in digital mammography. The (4) final step is addressed in the next section.

THE FIGURE-OF-MERIT – NOVELTY IN CHARACTERISATION AND OPTIMIZATION IN DIGITAL MAMMOGRAPHY

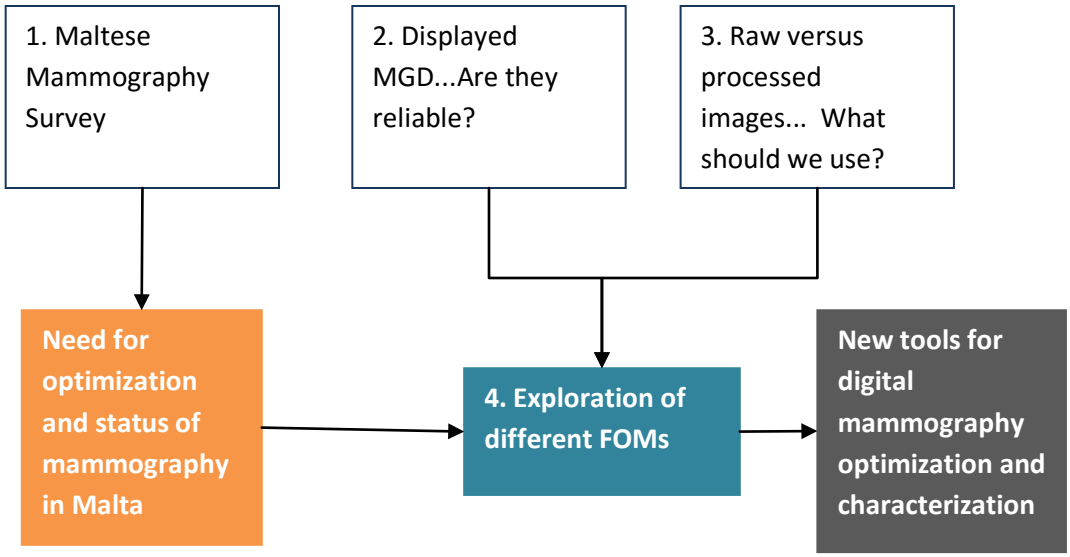
The use of image quality parameters in digital mammography such as CNR, signal-to-noise ratio (SNR) and DQE have been widely used and employed with the intention of detector evaluation and/or quantitative evaluation of the system performance. These parameters are useful in ensuring adequate system performance when tests are done against international standards or guidelines. CNR and SNR are relative quantities that lie within a range which is manufacturer and system-dependent. *The use of a figure-of-merit (FOM) is a relatively new concept as a tool in digital mammography permitting the quantitative assessment in terms of image quality and patient dose*.

Since radiographic contrast is less important in digital mammography due to the separation of image acquisition and display, the optimization processes between film-screen (FS) and digital mammography differ too. The contrast measure of choice in digital mammography should be the CNR [10]. The *FOM is seen as a complimentary parameter and as an objective means of optimization in digital mammography*.

Early breast cancer detection is very desirable however this is heavily influenced by many factors that require optimization and evaluation in mammography. These factors mainly include the beam spectra, MGD, detector performance and detector characteristics, image processing, image display and the radiographic technique. The objective of optimization is to establish standardized imaging protocols by determining the optimal trade-off between image quality and dose, which is especially important for screening mammography given the lifetime risk to women who undergo annual mammography examinations [11]. *The use of a FOM presents a very attractive way in performing such an optimization and evaluation process in any available digital mammography*

unit. The general concept behind the FOM is that a number is calculated for any particular system, the magnitude of which reflects its performance. The larger the FOM for any particular system the better is its performance. The FOM should be independent of the dose level employed, given that the system used is quantum limited [12].

Flowchart 1.1. Design of this exploration work.



EXPERIMENTAL ERRORS

Experimental error is the difference between a measurement and the true value or between two measured values. Experimental error, itself, is measured by its accuracy and precision. Errors are normally classified in three categories: systematic errors, random errors, and blunders [13]. The knowledge of the magnitude of a measurement's uncertainty can be used to gauge the confidence that should be placed in the measurement, especially when it is compared to another measurement produced using different test equipment or protocols [14].

Systematic Errors

Systematic errors are due to identified causes and can, in principle, be eliminated. Errors of this type result in measured values that are consistently too high or consistently too low. Systematic errors may be of four kinds:

1. Instrumental: for example, a poorly calibrated instrument such as a dosimeter or ion chamber.
2. Observational: for example, parallax in reading a meter scale.
3. Environmental: for example, the pressure, temperature and humidity that might affect the dose reading in ion chambers.
4. Theoretical: due to simplification of the model system or approximations in the equations describing it. For example, if your theory says that the temperature of the surrounding will not affect the readings taken when it actually does, then this factor will introduce a source of error.

Random Errors

Random errors are positive and negative fluctuations that cause about one-half of the measurements to be too high and one-half to be too low. Sources of random errors cannot always be identified. Possible sources of random errors could be:

1. Observational: for example, errors in judgment of an observer when reading the scale of a measuring device to the smallest division.
2. Environmental: for example, unpredictable fluctuations in line voltage, temperature or mechanical vibrations of equipment.

Random errors, unlike systematic errors, can often be quantified by statistical analysis, therefore, the effects of random errors on the quantity or physical law under investigation can often be determined.

Repeated measurements produce a series of data that are all slightly different. They vary in random about an average value.

Blunders

A final source of error, called a blunder, is an outright mistake. A person may record a wrong value, misread a scale, forget a digit when reading a scale or recording a measurement, or make a similar blunder. Multiple measurements or if one person checks the work of another blunders should be avoided.

Errors in this work

In this work it is envisaged that only random errors were the source of uncertainty in the data collected. The electronic measuring devices used e.g. ion chambers, dosimeters and kV meters, were fully calibrated as recommended by the respective manufacturer. Further still, data collected in this thesis was retrieved from mammography rooms which had controlled ambient conditions. Repeated readings for each specified condition was not always feasible, especially in chapters 4, 5 and 6. It must be emphasised that each mammography unit included in this study is used clinically and the time allocated for this study was very limited. Nevertheless, any measurements taken were done using the strict instructions laid down by relevant published protocols [1, 2, 15, 16].

CHAPTER 2

LITERATURE REVIEW

1.0 Mammography practice in Malta: evaluation of breast cancer screening and clinical mammography.

INTRODUCTION AND BACKGROUND

Malta has recently launched its first national breast screening program, the MNBSP, in October, 2009 in Lascaris, Valletta. A second mammography unit serving the program was introduced in the neighbouring island, Gozo, in September 2010. Both systems employed consist of DR systems. There are also other various private institutions around the islands that are actively involved in breast cancer screening. Clinical mammography is mainly performed in Mater Dei Hospital, Malta and also in the mentioned private institutions. Although, the majority of these systems are digital (CR and DR), FS systems are still available and operational. To-date, no national technical evaluation has been done to benchmark the level of image quality and radiation dose for mammography practice in Malta. There is neither the presence of a national protocol for the evaluation of the performance of mammography systems. Consequently, there has been no formal optimization process and evaluation of the technical performance of the available mammographic systems in breast cancer screening and clinical mammography.

The establishment of reference levels is very important for the intra-comparison of mammographic systems and allows direct comparison with the specific recommendations made by the European protocol [2]. Future quality assurance and/or medical physics-related tests on mammographic systems serving the breast screening program or clinical mammography could be compared to the mentioned established benchmark values. This will ensure the maintenance of image quality and radiation doses to patients undergoing mammography examinations. The establishment of reference levels is also very important as it provides women of breast screening

age with very important information on which mammography unit performs best, permitting an informed choice in their selection if desired.

Following technical evaluation and optimization of the mammography systems, breast radiation doses amongst Maltese women of breast screening age can be measured, categorized and audited in both breast cancer screening and clinical mammography centres. Hence, breast doses in Malta were compared to the doses in other European countries. In conjunction to this, the influence of patient characteristics, e.g. type, breast density, and implants on radiation dose and image quality can be evaluated amongst the Maltese women of breast screening age. In addition, a risk-benefit analysis may be conducted in the future which will determine the efficacy of breast cancer detection in the Maltese population and determine the clinical efficiency of the MNBSP. The technical evaluation of the used mammographic systems was also useful to evaluate the systems employed i.e. whether the move from FS to digital is justified and whether CR should be phased out and replaced by DR technology.

EQUIPMENT PERFORMANCE

Film-screen mammography performance testing and quality assurance.

To-date mammography is the most efficient and available modality for the detection of breast cancer both in the clinical and the screening environments. In countries categorized as *first level health care countries* [17], several QC protocols have been made [1, 2, 16, 18-21]. There seems to be a general consensus amongst European mammographic institutions, whether involved in breast cancer screening or clinical mammography to adhere to the technical recommendations made by the *European guidelines for quality assurance in breast cancer screening and diagnosis* [1, 2]. However, other protocols are available, valid and used in various countries [22-24].

A national pilot project on the implementation of the European protocol for the QC of the technical aspects of mammography screening as well as the European protocol for dosimetry in mammography was conducted in Bulgaria [20]. The survey demonstrated considerable differences in the technical condition of the mammography units that resulted in varying image quality. The

measured values of K showed significant variations. Doses for 45% of the units were found to be below the European reference levels [1, 2]. The values for MGD ranged from 0.35 to 3.47 mGy. In this study, the main problems found were film processing, optical density (OD) control settings and AEC adjustment. The results showed the importance of film OD measurements and dose assessment. It was concluded that X-ray mammography in the Bulgaria needs optimization. Comprehensive quality assurance mammography programs should be adopted in all departments covering permanent QC of the equipment, image quality and breast dose [20].

Serbia and Croatia are relatively small European countries with a majority of FS mammography units serving clinical and screening services. A national audit of mammography equipment performance was carried out in Croatia in 2008 with the purpose of introducing quality control based on the European Commission guidelines [25]. In addition, patient doses and image quality were audited to evaluate how the European recommendations and principles of optimization could be implemented in Croatia. The MGD to the standard breast (45 mm of PMMA) using clinical exposure factors, ranged from 0.43 to 2.4 mGy. Two of the 121 units investigated had a MGD above the European reference levels [1, 2]. No image quality problems were reported in this audit apart from unsatisfying viewing conditions including ambient light and viewing boxes in the majority of the systems analyzed. The survey showed that the main problem in Croatia is the lack of written QA/QC procedures within imaging departments across the country. Another problem is that there is a lack of trained breast radiologists and medical physicists in the country. Consequently, equipment performance, image quality and MGD are unstable and activities to aid optimization are not evidence-based.

Similar problems were reported in a more recent study in Croatia and Serbia [26]. The purpose of this study was to investigate the need for and the possible achievements of a comprehensive QA program and to look at effects of simple corrective actions based on the European guidelines [1, 2]. In Croatia although QA is mandatory there is still no detailed framework with a lack of national guidelines on QC. This study consisted of a two-phase project, the initial stage being the technical evaluation of the mammography systems included in breast cancer screening in Serbia and Croatia. The second stage consisted of the evaluation of the image quality of images form a radiological perspective using the European guidelines [1, 2] after the corrective actions taken in

the first stage. The images were rated as A (full acceptability), B (acceptable with some remarks) or C (rejected). Images rated B or C, were further ranked according to the remarks made by the reporting radiologists. A reject analysis was further performed and monitored over a two-week period to estimate the rejection-rate in the mammography systems assessed. In general, the performance of the systems was found to be satisfactory and conformant with the European guidelines. Major issues were again found to be present in OD and viewing conditions. After the implementation of any possible corrective action, images rated B and C were reduced and further still the image rejection rate was significantly decreased. In this study no reference is made to optimization in digital mammography systems however, it has been demonstrated how simple quantitative assessments of clinical image quality can be used for optimization purposes.

Another study, conducted exclusively in Serbia was intended to determine the appropriateness of mammography practice for both diagnostic and potential screening services [27]. The authors claim that this is the first assessment of mammography practice in Serbia. This study has introduced the concept of a *simple three-level quantitative assessment model* for optimization purposes. Again in this study the European guidelines [1, 2] have been adopted for the assessment of patient doses, image quality and other technical aspects of the mammography imaging chain. The objective was to identify any weak points and suggest methods of improvement. Similar to the previous study [26], this work consisted of a two-phase project. The first phases included the assessment of the technical performance of the mammography units and to identify areas needing optimization and corrective action. The second phase included a simple reassessment of the mammography units tested in the first phase. The same A, B and C method for image quality assessment was used as described before [26]. Therefore, the first phase was the determination of the actual situation, while the second phase was the evaluation of the effect of corrective actions. As with other studies conducted in Serbia and Croatia [25, 26], no reference is made to digital mammography systems despite their availability. The same problems that were identified in previous studies were noted. Further still, application of the European guidelines [1, 2] was once again proven useful for the optimization of mammography practice. Doses to the breast in the participating three major hospitals in the first phase were 2.84, 1.77 and 2.20 mGy. The corresponding values in the second phase were 2.55, 1.40 and 1.08 mGy respectively. Image quality was also significantly improved in both craniocaudal (CC) and mediolateral oblique (MLO)

views. This was evident in both the radiological image quality assessment (i.e. the reduction of images rated 'C') and also in terms of basic image quality parameters like spatial and contrast resolution using a test phantom like the TOR MAS (Leeds Test Object, UK).

Digital Mammography

Different technologies

Digital mammography is expected to replace current conventional FS systems in screening programs in the near future. An important advantage in digital mammography is that it separates image acquisition and display, allowing the degree of contrast in the image to be adapted to the local density of the breast [28]. Computer-aided diagnosis (CAD) is an exceptionally useful tool and add-on to digital mammography, helping radiologists in their diagnoses. The major disadvantage of digital mammography, especially direct digital is its cost, reaching up to four times the cost of a FS system [29]. Several different types of detectors are employed in digital mammography. These are briefly described in the following sections based on the literature by Yaffe in the book by Bick and Diekmann [30].

Phosphor-Flat Panel

Phosphor flat panel detector systems (figure 2.1) are based on a large-area glass plate. Using solid-state manufacturing techniques, a rectangular array of light sensitive photodiodes is deposited onto the plate. These are interconnected with an array of control and data lines as well as a thin film transistor (TFT) switch adjacent to each photodiode. These electronic components are fabricated using amorphous silicon technology. X-rays are absorbed by a layer of thallium-activated caesium iodide phosphor CsI(Tl) deposited onto the photodiodes. The photodiodes serve as the dels of the detector, detect the light emitted by the phosphor and create an electrical charge signal that is stored on each del. Since CsI can be manufactured to have a needle like or columnar crystal structure, it can provide a better compromise between quantum efficiency and spatial resolution than is possible with the granular phosphors used in FS imaging (figure 2.1b). In a conventional phosphor, the light quanta produced on X-ray absorption readily moves laterally, leading to increased width of the line spread function. The CsI crystals act as fibre optics or *light pipes* to reduce lateral spread. This allows the detector to be made thicker without as much resolution loss as would occur in conventional phosphors.

The arrangement of the individual dels with a photodiode and TFT switch is shown in figure 2.1a. Control lines for each row of the array are energized one at a time and activate all the switches in that row. A readout line for each column transfers the signal from the del at the activated row to an amplifier and digitizer. When a given row is activated, the signals from all of the dels on that row are collected along the readout line for all columns simultaneously. In the system of this type, produced by General Electric Medical Systems (Milwaukee WI) (figure 2.2), the del pitch is 100 µm, the field size is 24 × 30 cm and the digitization is carried out at 14 bits [31]. A comparison between the performance of the original detector system and one with an improved scintillating phosphor and reduced noise characteristics [32] was published by Ghetti et al [33]. For flat-fielding correction, an offset value and gain is measured for each del in the detector. Therefore, the number of such constants is equal to twice the number of dels in the detector, about 7.2 million values. It is typical to re-measure offset values between images; however, the gain matrix generally need only be measured occasionally.

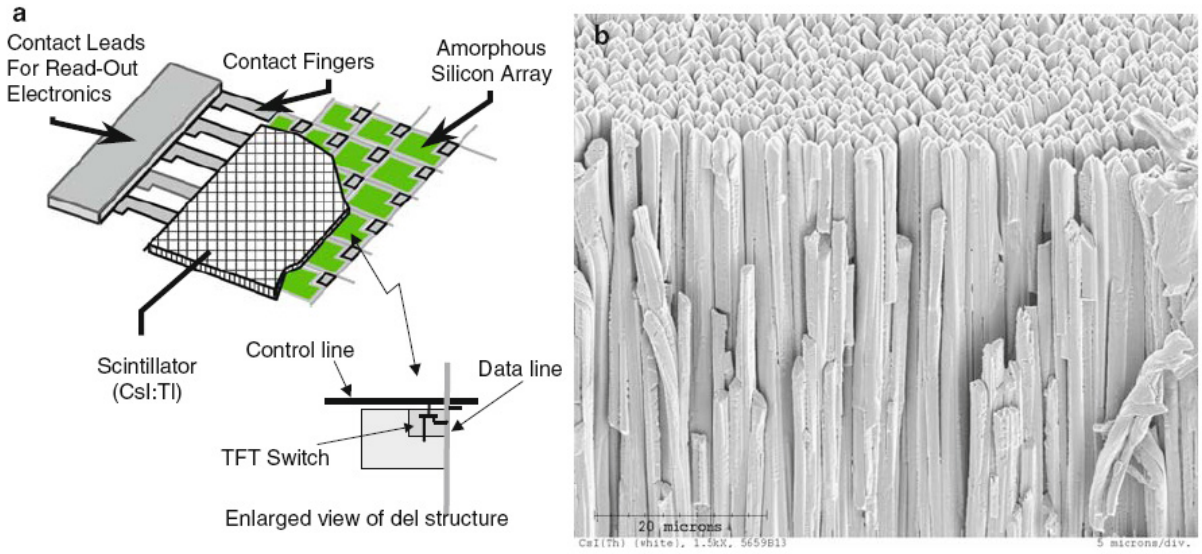


Figure 2.1. Flat panel detector with CsI(Tl) absorber. (a) Detector with photodiode array. TFT readout element is shown in inset. (b) Structure of CsI:Tl needle phosphor (Reprinted from Enhanced a-Si/CsI-based flat-panel X-ray detector for mammography [30, 32].

Phosphor-CCD System

In this detector, an X-ray absorbing CsI(Tl) phosphor is deposited on a fibre-optic coupling plate, which conducts light from the phosphor to several rectangular *charge-coupled device* (CCD) arrays, arranged end to end. The fibres transmit the optical image from the phosphor to the CCD with

minimal loss of spatial resolution. The CCD is an electronic chip containing rows and columns of light-sensitive elements. Light is converted in the CCD to electronic charge. The charge produced on each element in response to light exposure can be transferred down the columns of each CCD and read out by a single amplifier and analog-to-digital converter. In the commercial implementation of this type of detector, the detector is rectangular with approximate dimensions of 1 × 24 cm. The X-ray beam is collimated into a narrow slot to match this format. To acquire the image, the X-ray beam and detector are scanned in synchrony across the breast (figure 2.3). Charge created in the CCD is transferred down the columns from row to row at the same rate, but in the opposite direction to the physical motion of the detector across the breast so that bundles of charge are integrated, collected, and read out corresponding to the X-ray transmission incident on the detector for each X-ray path through the breast. This is referred to as time-delay integration (TDI).

Scanning systems usually require longer total image acquisition time than full-field detectors. The slot collimators only allow use of a small portion of the total emission from the X-ray tube so that the overall heat burden for the tube for a scan is generally considerably higher than for full-field collimation. Because only part of the breast is irradiated at one time in scanning systems, the scatter-to-primary ratio is reduced. Collimation occurs before the breast so that transmitted X-rays are not lost. Normally, an antiscatter grid is not required with scanning systems while grids are used with full-field detectors. This provides a significant dose advantage for the former.

A slot-beam CCD-based scanning digital mammography system was originally marketed by Fischer imaging Inc (Denver CO). It employs dels of 54 μm. Over a limited portion of the detector, data can be read out at 27 μm intervals to provide a high-resolution mode. Digitization is performed at 12 bits.

Figure 2.2. Photo of flat-panel detector. (Courtesy, GE Global Research Centre)

Photostimulable Phosphor (PSP) System

PSP systems, which are often referred to as, "computed radiography or CR," have been widely, used for many years in general radiographic applications. More recently, they were introduced for use in digital mammography. The operation of the detector in these systems is based on the principle of *photostimulable luminescence* (figure 2.4b). Energy from X-rays is absorbed in a screen composed of a phosphor material containing a high prevalence of electron trapping sites. The absorbed energy causes electrons in the phosphor crystal to be temporarily freed from the crystal matrix and then captured in "traps" within the crystal lattice where they can be stored with reasonable stability for times ranging from seconds to hours. The number of filled traps in a particular location is proportional to the amount of X-ray energy absorbed in that location of the screen.

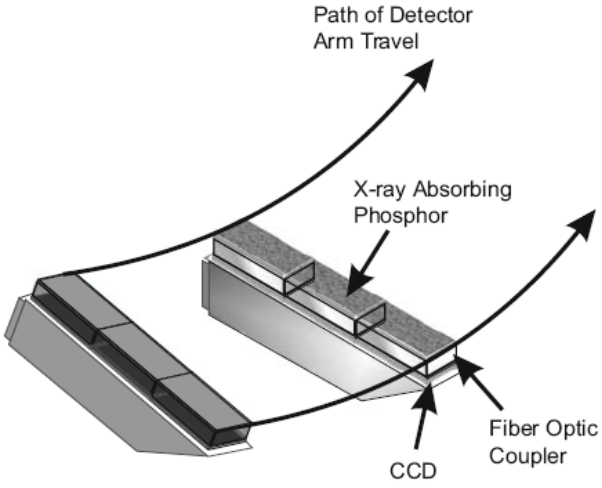


Figure 2.3. Slot-format scanned CCD detector with CsI:Tl phosphor [30].

This analog image is then read by placing the screen in a reading device where it is scanned with a red laser beam. This causes the electrons to be freed from the traps and to return to their original state in the crystal lattice. In doing so, they may pass between energy levels in the crystal structure. These energy levels are defined by small amounts of specific elements deliberately incorporated into the crystal. The choice of these materials thereby determines the colour of the light emitted (related to the difference in energy between the levels) as the electron makes its transition. A typical strategy is to design the crystal to emit blue light, so that this can be measured with an appropriate optical filter placed in a light-collecting system incorporating a sensitive photomultiplier tube (figure 2.5a), without interference from the red laser light. The amount of blue light measured is proportional to the energy of X-rays absorbed by the phosphor.

The phosphor plate is continuous and is not physically divided into dels. The laser beam is scanned across the plate along one dimension as the plate moves through the reader in the orthogonal direction and the location of the beam on the surface of the plate at each point in time is used to define the *x*–*y* coordinates of the image. The spatial sampling is determined by the size of the laser spot (aperture, d) and the distance between sample measurements (pitch, p). A photostimulable phosphor system for digital mammography was originally introduced commercially by Fuji Film. The dels are of a nominal size of 50 μm. The gain, affects both the sensitivity and noise of the imaging system. It is important that the light produced from the phosphor is collected efficiently. If an inadequate amount of light is measured from each interacting X-ray, then the image will contain additional noise above and beyond the quantum noise, causing the Signal-to-noise Ratio (SNR) and Detective Quantum Efficiency (DQE) to be reduced.

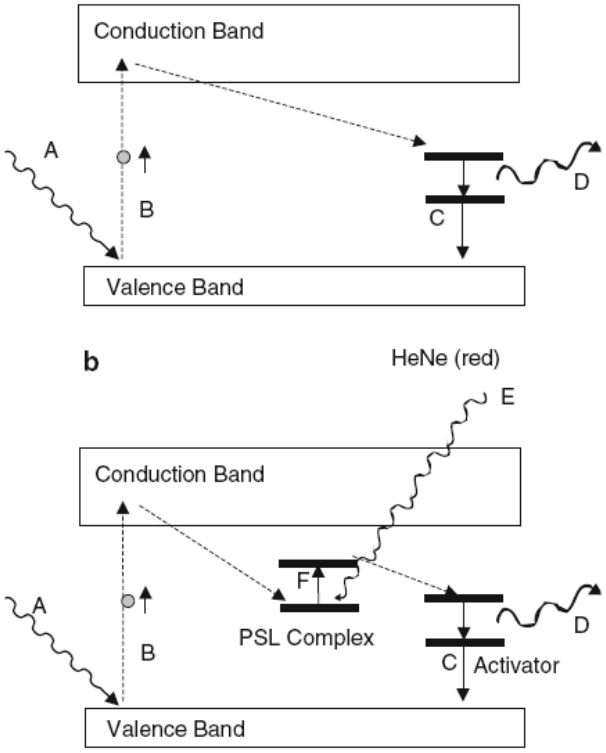


Figure 2.4. (a) Operating principle of a conventional phosphor X-ray detector, (b) a photostimulable phosphor [30].

To increase sensitivity and improve SNR, some photostimulable phosphor system manufacturers have refined their plate technology to reduce laser scattering and increased the efficiency of light collection by reading from both the top and bottom surfaces of the phosphor plate (figure 2.5b). Unlike the other systems, this system employs removable cassettes, which can be used in the Bucky tray of a standard mammography unit. While there are capital cost savings to this approach, it does require that phosphor plates be manually transported to the reader for processing. Because there are multiple detector plates, flat-field correction is normally not performed for the plates, but only for the plate reader. In principle, correction for nonuniformity of the individual plates could be done, but this would require precise registration within the reader and would be time-consuming.

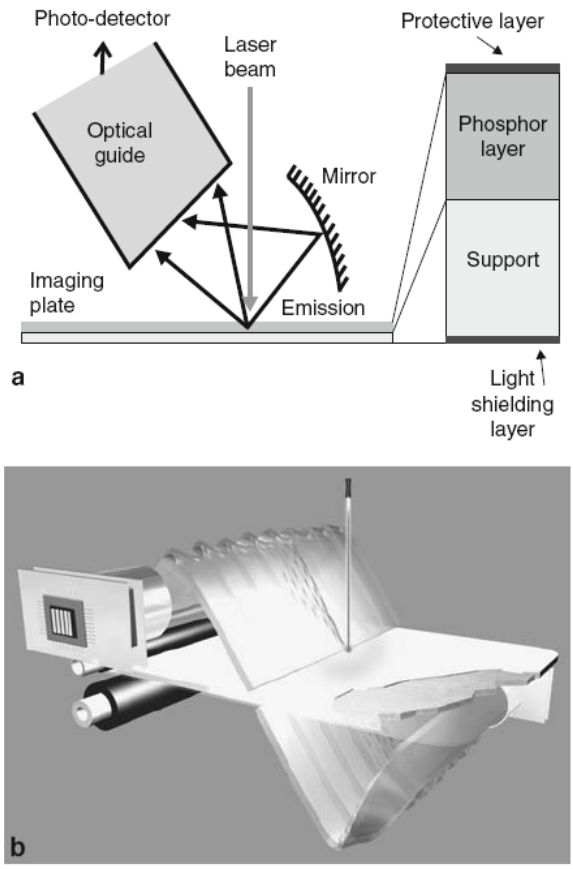


Figure 2.5. Photostimulable phosphor system (a) mechanism of the PSP, (b) double-sided readout to increase efficiency [30].

Selenium Flat Panel

In this type of detector, the X-ray absorber is a thin layer (100–200 µm) of amorphous selenium. When X-rays interact with the selenium and produce energetic photoelectrons; these lose their kinetic energy through multiple interactions with electrons in the outer orbitals of selenium atoms. The process causes some of these electrons to be liberated and the freed electron and the corresponding "hole" created by its departure, i.e. the electron–hole pair, form the signal. An electric field applied between electrodes deposited on the upper and lower surfaces of the selenium as in figure 2.6a sweeps the charges toward the electrodes. One of the electrodes is continuous while the opposing one is formed as a large matrix of dels on a glass plate [34]. The dels act as capacitors to store the charge. At the corner of each del is a TFT switch. Readout of charge from the dels is accomplished in the same manner as for the phosphor flat plate detector (figure 2.1), with control lines sequentially activating the TFTs for dels along individual rows.

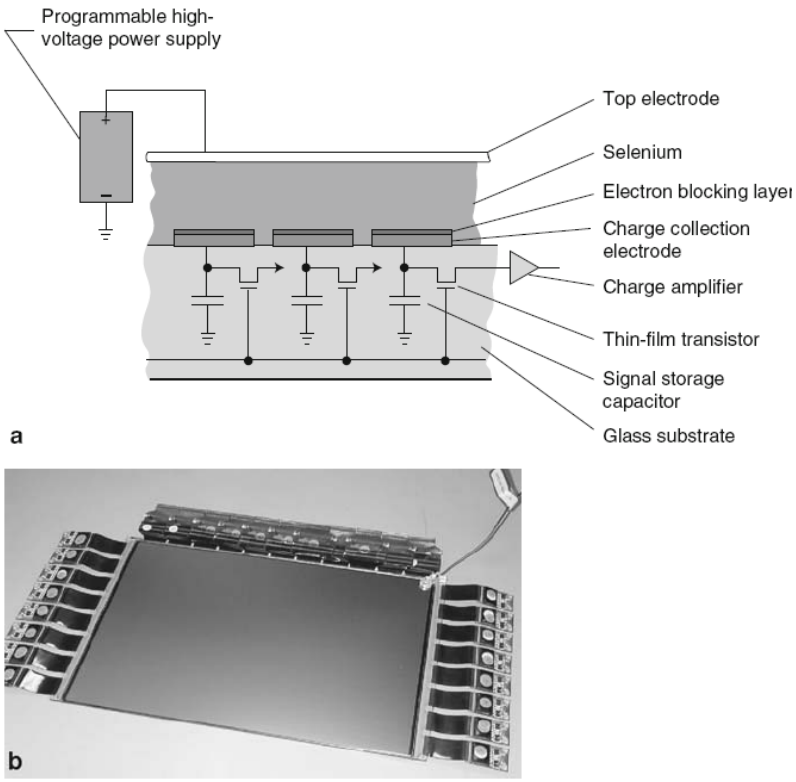


Figure 2.6. Selenium system (a) schematic of detector, (b) photo of Anrad detector [30].

The signals from all activated dels are then simultaneously transmitted along readout lines adjacent to the columns of the matrix to be amplified and digitized. A detector of this type is produced by Hologic (Danbury CT). Dels are 70 µm, with 14-bit digitization. A selenium flat-panel detector system is also being produced by Anrad (St Laurent Quebec, Canada) with 85 µm dels (figure 2.6b) and this detector is currently used on the Giotto, Planmed, and Siemens systems. Some of its performance characteristics have been described by Bissonnette et al [35]. Recently, an amorphous selenium detector that incorporates a different readout design (figure 2.7) has been introduced by Fuji. In this detector, there are two separate layers of selenium. The upper layer absorbs X-rays and produces electron–hole pairs similar to the operation of other selenium direct-conversion detectors. This charge is stored on the capacitance of each del. The lower selenium layer acts as an optically controlled switch that transfers the stored charge to a set of readout lines. This allows a del size of 50 µm to be achieved while avoiding the need for TFT switches, which would reduce the detector fill factor in the conventional design and so reduce the geometric efficiency of the detector.

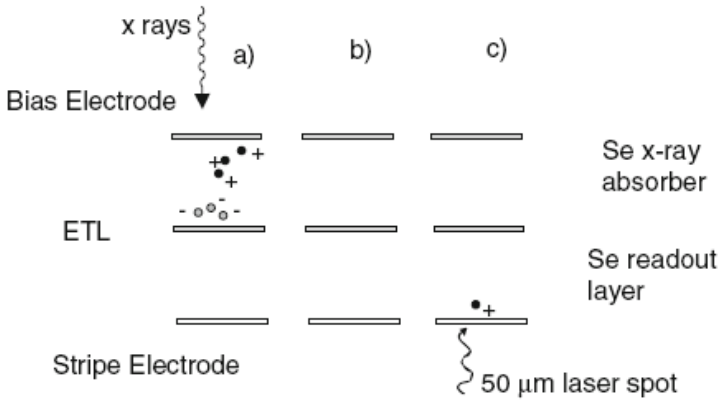


Figure 2.7. Selenium flat panel detector with optically switched readout (Fuji Medical) [30].

X-Ray Quantum Counting Systems

The detector systems described previously operate by absorbing the energy from X-rays interacting with each del in the detector and accumulating the electronic signal produced by all the X-rays received during that measurement. This signal is then digitized to create the information corresponding to a pixel of the image. One aspect of these types of detectors is that higher energy X-ray quanta produce more signals in the detector than those of lower energy and this tends to weight the image signal to higher energy quanta. These carry comparatively lower image contrast than lower energy quanta.

Alternatively, the detector can be designed so that each del produces an electronic pulse every time an X-ray quantum interacts with it. The pulses are then counted to create the signal for that pixel. Pulse counting has several desirable features. Each interacting X-ray registers exactly one count regardless of its energy, so that the secondary noise sources discussed earlier associated with fluctuation in gain are eliminated. In addition, the equal weighting shifts the emphasis in the image signal away from the higher energies. Counting systems do not require the traditional analog-to-digital converter; however, it is important that the counting electronics is properly designed to handle the high rate of incident quanta, which can exceed 10^6 per second.

Currently, two quantum-counting systems have been introduced. Both use a set of multiple linear detectors, which are scanned across the image field beyond the breast during image acquisition in synchrony with an appropriate set of collimator blades located on the X-ray entrance side of the breast. A precise mechanical scanning system is required in order to avoid image artifacts. The

detector in the SECTRA system (Stockholm, Sweden) absorbs the X-rays in crystalline silicon (figure 2.8a). The electron–hole pairs produced from each interacting X-ray are collected in an electric field and shaped into a pulse, which is counted [36]. The XCounter (Stockholm, Sweden) employs a pressurized gas as the X-ray absorber and pulses of ions created in the gas form the signal as illustrated in figure 2.8b [37].

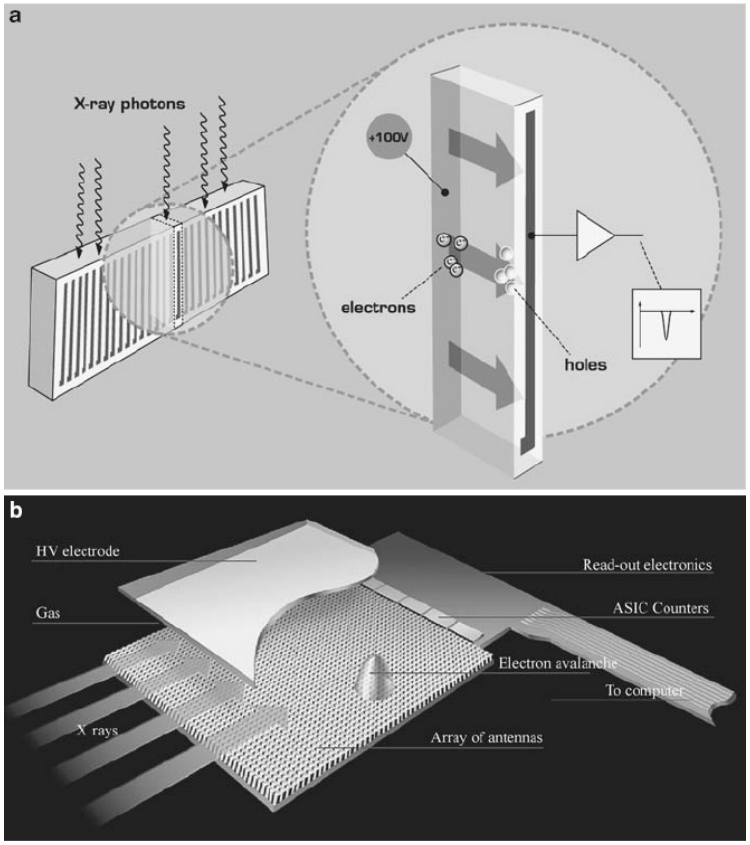


Figure 2.8. (a) X-ray counting detector based on silicon (photo courtesy SECTRA), (b) high pressure gas ionization detector (XCounter [30]).

In conclusion, it is important to differentiate between DR and CR mammography. True digital mammography takes and electronic image of the breast and stores directly into a computer or network. DR mammography is also known to employ less ionizing radiation than FS and CR mammography and facilitates work throughput as image transmission and storage is much easier and faster. For women under the age of 50 years and women with heterogeneously dense or extremely dense breasts, true digital, especially DR mammography, seems to be the modality of choice.

An anthropomorphic breast phantom study has shown that true digital mammography is superior to the FS for the detection and morphologic characterization of microcalcifications larger than 200 μm in diameter [38]. However, the resolution and dose performance of CR mammography is inferior when compared to FS and to DR detectors. In addition, CR offers no productivity advantage compared to FS mammography, which is enjoyed by true digital systems. Further still, CR is unable to perform dynamic imaging such as tomosynthesis, limiting its usefulness in future applications [39]. In fact, the NHS Breast Screening Programme (NHSBSP) states that [40],

> *"Adoption of direct digital technology*
>
> *On the 26th May 2010, the Department of Health Advisory Committee on Breast Cancer Screening decided that direct digital technology was the preferred option for the introduction of digital mammography into the NHS Breast Screening Programme.*
>
> *Any new mammography systems introduced by NHS Trusts providing breast screening within the NHSBSP should therefore now use direct digital technology rather than computerized radiography."*

Digital mammography performance testing

For the introduction of digital mammography in screening programmes the European Commission has published the European protocol for the quality control of the physical and technical aspects of mammography screening part B: digital mammography [2] and very recently a supplement to this document [3]. A very important change with respect to the protocol for conventional mammography is that the evaluation of the image quality is now based on a contrast-detail analysis of CDMAM phantom (Artinis, The Netherlands) images. Regarding image acquisition the protocol for digital mammography deals with image receptor aspects such as response, noise evaluation, detector homogeneity and memory effects. There are aspects specific for direct readout (DR) systems like detector element failures and for CR such as interplate sensitivity variations and image fading. Regarding image display, the quality assurance programme is based on monitor characteristics such as contrast visibility, resolution, luminance range and conformity with the DICOM greyscale standard display function [41].

Other relevant protocols and guidelines dedicated to digital mammography testing include the Institute of Physics and Engineering in Medicine (IPEM) Report 89 [15], the International Electro

technical Commission (IEC) standards in particular IEC 61223-3-2:2007 [22], NHBSP report 604 [24] and others.

Direct Radiography (DR) versus Computed Radiography (CR) performance

The limiting values introduced in the European guidelines [2] for a number of test parameters to be measured pose stringent requirements to digital equipment. This is especially the case for those tests related to image quality. To avoid the purchase of a digital mammography system with intrinsic characteristics not able to fulfil the quality requirements of the acceptance tests of the European guidance document, the Flemish government introduced type testing of digital equipment in the organization and legislation of their breast cancer screening programme [28]. The type testing protocol of digital equipment has two parts: a physical-technical part following the acceptance criteria of the European guidelines [2] and a radiological part with evaluation on screen of a set of clinical images by experienced radiologists. The radiological part brings additional value to this type testing since the European guidance document does not take into account objective quantitative measurements on image processing in the context of an acceptance test. Therefore, a type testing certificate guarantees that the system under consideration can pass the acceptance tests if tuned correctly and used in an X-ray setup of sufficient quality. The physical-technical part of type testing of a digital mammography system is performed independently by two teams recognized by the Flemish government for physical-technical control of digital mammography systems. On regular meetings of this working party the results obtained by the different teams in the type testing of digital equipment were compared and discussed [28]. From this study, DR systems passed the acceptance criteria more easily than CR systems. These results seem to be in agreement with the UK according to the Digital Steering Group of the NHSBSP [42].

Amongst others, the European guidelines [2] specifies quantitative methods for the evaluation of image quality of digital mammography systems i.e. the CNR and the SNR. The application of the European guidelines in the evaluation of CNR and MGD for two digital mammography systems has been reported [43]. This study consisted of an evaluation of image quality and dose characteristics of a CR and a DR system. The authors claim that the motivation behind this work was twofold; primarily that the image quality and the dose characteristics vary between digital systems and therefore requiring optimization techniques that are completely different than those needed in FS

systems. The second reason is the possible contribution through experience by the application of the European guidelines for better harmonization of mammography screening practices. Regarding the radiation dose to the breasts, early reports indicate a 30% lowering of the dose for equal image quality for full-field digital mammography (FFDM) compared to FS systems [44]. From the results achieved in this study [43] the MGD values achieved for both CR and DR systems were comparable to the achievable values set in the European guidelines [2]. The values achieved are well below the limiting values specified in the European guidelines. It was noted that the highest difference between the European limiting value and the measured MGD values was seen with the largest breast thickness at the 70 mm PMMA. This data indicates the existing potential for image quality improvement which may be achieved by increasing the dose. As a result it is necessary to analyze the dose and image quality balance in order to estimate reference values for FFDM systems [44]. The same results were achieved with the Flemish breast screening programme were the data obtained indicates that retuning of the AEC for DR systems is required if one is expecting a constant CNR for all breast thicknesses, particularly for the larger breasts. For CR systems this AEC retuning was less obvious and marked when compared to the DR systems [28]. In another study it was shown that in most cases, the AECs of FFDM systems successfully identified exposure parameters resulting in FOM values near the maximum ones, however, there were several examples where AEC performance could be improved [45].

The study by Schueller et al [46] was designed to compare the image quality, lesion detection and diagnostic efficiency of DR and CR mammography in the evaluation of breast lesions. In this prospective study, 150 patients with suspected lesions underwent imaging with both modalities. Nine aspects of image quality were evaluated i.e., brightness, contrast, sharpness, noise, artefacts and detection of anatomic structures (skin, retromammillary space, glandular tissue and calcifications). In addition the detection of breast lesions and the diagnostic efficiency, based on the BI-RADS classification (refer to page 37 and 41 for additional detail) [47], were evaluated with histological and follow-up correlation. For contrast, sharpness, and the detection of anatomic structures, DR was rated significantly better ($p<0.05$) than CR. Mass lesions were equally detected, whereas DR detected more lesions containing microcalcifications. The authors concluded that, based on image quality parameters, DR is, in part, significantly better than CR mammography. However, the diagnostic efficiency of DR and CR are equal. It was also concluded that the

perception and characterization of lesions is not defined solely by the mammography system employed but also influenced the interpreting radiologist.

The DR and CR systems employed in another study by Muhogora et al exhibited different characteristic behaviours in CNR and SNR with increased PMMA thickness and this may be attributed to the fact that in particular CNR based on different technologies is not directly compatible [43]. Further still, strong variation in SNR are known to be an inherent characteristic with CR systems [48]. Muhogora et al also stated that the MGD values of both modalities were comparable with the achievable levels and well below the limiting values specified in the European guidelines [2]. However, the performance of the DR unit exhibited higher detail visibility than the CR system when using the TOR MAX phantom (Leeds Test Objects, UK). The authors claim that this is probably due to the higher quantum efficiency of DR even at relatively low doses.

In a more recently presented study by Kalathaki et al [49], the performance of DR and CR mammography were compared in respect to dose and image quality for 52 mammography units participating in the Greek breast screening programme. K values were measured for all the participating units according to the European guidelines. Image quality was assessed by using the Gammex RMI 156 mammographic phantom and quantified by assigning a total score based on the identified phantom structures. Statistical analysis of the results indicated that DR systems provided a significantly better image quality than the CR systems. Comparison of the mean K values showed that DR systems provided slightly lower doses than CR however without indicating a statistical significance at the 5% confidence level.

Technical performance of Full-Field Digital Mammography (FFDM) versus Film-Screen (FS) mammography.

The results of acceptance testing of 18 FFDM systems for clinical use and of conducting annual physics surveys of 38 FS mammography systems were compared in terms of exposure times, MGD and image quality. These evaluations were made using the same test tools on all systems, with emphasis on assessing automatic exposure control performance and image quality on both digital and FS systems using clinical techniques. Survey results indicated that digital mammography

systems performed similarly to FS systems in terms of exposure times and MGD for thin to intermediate breasts, but that digital mammography systems selected shorter exposure times and lower MGD for thicker breasts. This is in agreement with previously mentioned studies [43, 44]. FFDM systems yielded mean contrast-detail scores higher than those for FS systems for all the breast thicknesses. For all breast thicknesses, the 18 digital mammography systems demonstrated less variance in terms of exposure times, MGD, and contrast-detail scores than did the 38 FS systems tested. These results indicate that the clinical use of digital mammography may generally improve image quality for equal or lower breast doses, while providing tighter control on exposures and image quality than FS mammography [50].

The results of an Italian study, evaluating patient dose showed that FFDM allows a significant clinical dose reduction compared with FS mammography [51]. From the four FFDM units evaluated in this study it was concluded that this dose reduction was provided mainly by the higher energy of the employed spectra. Detector linearity, wide dynamic range, and image processing make digital systems able to represent mammograms with optimized image contrast that is not totally dependent on the employed exposure parameters. The use of a Rh target as opposed to a Mo target, guarantees a lower MGD [51]. A more recent Italian study [52] comparing doses form a FFDM unit and a FS mammography having the same X-ray tube and hence the same spectral characteristics was performed. This study is considered important as it is difficult to perform a direct comparison in terms of dose between FS and digital mammography given the diversity in the available mammographic equipment, FS combinations and the different times in which the available research was performed. Results from this study [52] showed an overall reduction of MGD by 27% of digital over FS mammography. The dose saving was about 15% for thin and thick breasts while it was between 30% and 40% for intermediate thicknesses.

Microcalcification detection

Studies in mammography microcalcification detection can be grouped in two parts: studies with patients and studies with phantoms as follows.

Microcalcification detection – patients

Fischer et al [53] conducted a prospective study aimed at comparing FFDM and FS mammography in the detection and characterisation of microcalcifications. The image quality and the number of micro-calcification particles were evaluated and characterized using the BI-RADS method [47]. Image quality of FFDM was assessed as superior to FS mammography in more than 50% of the cases. FFDM depicted significantly higher number of microcalcifications than FS. The number of visible calcifications was the same for FFDM and FS in 59% of the cases while, FFDM showed more calcifications in 41% of all the cases. The authors concluded that FFDM (with a 100 µm system) provided better image quality than FS mammography with mammographic microcalcifications. The FFDM also has higher sensitivity and reliability in characterizing microcalcifications. FFDM demonstrated a higher diagnostic accuracy (deviation 0.86 BI-RADS steps) compared with FS (deviation 0.93 BI-RADS steps). In a more recent study to compare the microcalcifications detection rate and recall rate in FFDM and FS mammography [54]. It was concluded that screening using digital mammography leads to higher recall rates for assessment than FS mammography; however, similar rates of detection of microcalcifications occurred in both imaging modalities. In a previous study by Di Nubila et al cited by Young et al [55], it was stated that digital mammography was broadly similar to analogue imaging for the standard views, but appeared to have an advantage only when the magnification views were compared. Using a magnification technique allowed the detection of a larger number of microcalcifications than the standard technique with either system.

Microcalcification detection - phantoms

The results of a phantom study comparing different digital mammographic systems to analogue FS mammography suggested that direct flat-panel mammography was superior to FS mammography, to computed mammography and to indirect flat-panel radiography in detecting fibrous structures [56]. However, all methods detected microcalcifications and round densities equally well. In a later larger scale study using a breast anthropomorphic phantom it was shown that the detection rate and morphologic characterization of microcalcifications achieved with flat-panel DR mammography are comparably as effective as FS mammography [38]. The authors also stated that in digital mammography, monitor interpretation is diagnostically superior to film interpretation for the detection of large microcalcifications and that the experience of the interpreting radiologist is

of vital importance in the diagnostic accuracy using monitors. The authors conclude by stating that any future breast screening practices should give preference to digital mammography with monitor interpretation [38].

Image quality and lesion detection in clinical images

Obenauer et al [57] compared clinical FFDM and FS images from the same patients using subjective comparisons of image quality and lesion detectability. Digital and conventional mammograms were performed on 55 patients with cytologically or histologically proven tumours. Seventy-five digital mammograms of patients without tumours were also reviewed along with their FS mammograms taken previously. Aspects such as contrast, exposure and the presence of artefacts were evaluated. A three-point ranking scale was used to judge different details such as the skin and other structures. The detectability and characterisation of microcalcifications and lesions were also compared and correlated to histology. Artefacts were found in 78% of the conventional and in none of the digital mammograms. Some anatomical regions were better visualised by FFDM than by FS. The authors concluded that digital mammography offers better image quality without artefacts and equal lesion detection. Lesion characterisation was found to be slightly better using FFDM even though differences in the final diagnostic decision were not significant.

Digital mammography with soft-copy reading was also reported as superior to FS mammography in the detection and characterization of breast lesions in a retrospective study by Skaane et al [58]. Two-view mammograms were obtained with digital and FS systems at previous screening studies. Interpretation included the BI-RADS system [47] and a five-level probability-of-malignancy score. All cancers were confirmed histologically and images were interpreted by readers in two sessions that were five weeks apart; the same case was not seen twice in any session. Overall, however there was no statistically significant difference in the diagnostic performance rating between the two imaging modalities.

Fischmann et al [59] performed a study in which 200 women without visible or palpable breast lesions underwent digital mammography of one breast and FS of the other. For all women, one breast was imaged with the FFDM system and the other with the FS mammography to avoid

double radiation exposure of the breasts. The modalities were allocated randomly and the same compression force was applied to both breasts and by the same radiographer. The imaging parameters were set automatically by both systems. Three readers independently evaluated image quality, visualisation of calcifications and masses under the same viewing conditions. There was no difference in the diagnostic classification of the microcalcifications, and also there were no significant differences in the detection of masses. Readers A and B found better contrast with FFDM in parenchymal tissue, whereas reader C found a better contrast in fatty tissue. All three readers found the breast parenchyma to be less dense with FFDM. Finally, in contrast to earlier studies, there was a non-significant tendency for a higher MGD with FFDM. Digital mammography demonstrated improved image quality with significantly better depiction of the nipple, skin and pectoral muscle and better microcalcification detection [59].

Diagnostic Performance of digital and FS mammography – *Clinical Trials*

There is now general agreement that screening mammography reduces the rate of death from breast cancer. It is also apparent that younger women (up to 49 years of age) show a smaller benefit from screening mammography [60]. It is also true that younger women have a lower incidence for breast cancer, more rapidly growing tumours and greater radiographic density of breast tissue than women greater than 50 years of age [61]. Greater density reduces the sensitivity of mammography [62, 63] and increases the risk for breast cancer [64-66].

As stated in previous sections, FFDM employs low-noise detectors with higher detection efficiency than FS mammography. This may be expressed in terms of DQE and the Noise Power Spectrum (NPS). These parameters are now widely accepted to show the potential benefits of FFDM over FS mammography particularly with the limitations mentioned previously [60]. Image processing of the digitally acquired images whether CR or DR, allows the degree of contrast in the image to be manipulated and to improve the contrast resolution particularly in breasts having low subject contrast [38].

Despite the apparent advantages of digital mammography over FS, early clinical trials did not find digital mammography as being significantly more accurate than FS mammography in the diagnosis

of breast cancer. The first results of a trial, known as the *Colorado Screening Trial I* [67] were published in 2001. This study was performed on 4945 women aged 40 years or older. From the results of this study 20 cancers were detected using FS mammography and 21 cases were detected with FFDM. From the population investigated 4 cancers went undetected which eventually became clinically palpable within a year. These cases represent false negative findings in both modalities. In conclusion there was no significant difference in the cancer detection rate between FFDM and FS mammography, however the former showed a significantly lower breast recall rate (11.5%) as opposed to (13.8%) in FS mammography. However, in comment to this trial, the NHSPSP claim that the number of cancers detected is small to generalize any findings and that the sensitivity and the specificity of each modality on its own is low when compared to UK standards [55].

In 2002, the results of the *Colorado Screening Trial II* [68] were published. The main difference between the two trials is the increase in the screening population size. However, there was still no significant difference in cancer detection rates by FS mammography and FFDM. There was also no statistical difference between the sensitivities of the modalities despite the authors claiming such sensitivities are at or above the expected rate for their population. Nevertheless, a significant difference was again reported as in the first trial, in the recall rate which is lower for FFDM (11.8%) when compared to FS mammography (14.9%). However, the positive predictive value (PPV) for both modalities is effectively similar. An interesting observation in this trial was the difference in lesion conspicuity between FS mammography and FFDM. The authors claim, this may be attributed to various factors e.g. the view, the compression applied. Once again the NHSBSP claim that the performance parameters specified in this study (i.e. sensitivity, specificity, recall rate and PPV) are quite different when compared to the UK standards and caution is to be taken when comparing data between US an other countries [55].

In 2003, a comparative study known as the *Oslo I* study was performed comparing the diagnostic performance of FS to FFDM systems [69]. Two standard views of each breast were acquired with each modality. Images underwent independent double reading with use of a five-point scale indicating the probability of cancer. Recall rates and PPV were also calculated. The authors concluded that there is no significant difference in cancer detection rate between the two

modalities. The authors also performed a side-to-side analysis of cancer conspicuity by means of consensus as equal or slightly better and concluded that the two modalities were equal in some cases. FFDM is comparable to FS mammography in the detectability and characterization of microcalcifications and low-contrast objects. Interestingly, from this study it was observed that little additional information is achieved by applying post processing (zooming) beyond the suggested quadrant zooming. However, it must be emphasized that the mentioned conspicuity analysis was not a blinded study and sample size was small to be able to generalize findings. Nevertheless, authors of the Oslo I study concluded that missed cancers using FFDM were not due to a limitation in image quality, since they were easily visible in retrospect. Rather they believe that this may have been a result of a learning effect with the new digital environment in FFDM. Suboptimal viewing conditions e.g. extraneous light sources and reader training may also be considered as a factor causing cancers to be missed using FFDM [69].

The *Oslo II* study [70] was performed in 2004 and is not considered as a continuation of the Oslo I study. The women in this study were randomly selected to either FS mammography or FFDM. A total of 25 263 patients participated in this study as opposed to 3683 women in the previous one. The women age group consisted of ages between 45-49 and 50-69. Within these groups 70% of the women underwent FS mammography and the remaining 30% had their mammogram using FFDM. Again, two standard views of both breasts were acquired, and independent double reading was performed using a five point scale to indicate the probability of cancer. Recall rates, PPV and cancer detection rates were calculated and compared for the two age groups and modalities. The cancer detection rates for FS mammography and FFDM in all cases was 0.41% and 0.59% respectively. Of particular interest was the detection rate in the 50-69 age groups with a detection rate of 0.54% and 0.83% for FS and FFDM respectively, the difference between which approached statistical significance. As opposed to the Colorado I [67] and the Colorado II Trials [68], the recall rates in both age groups were significantly higher at FFDM than at FS mammography. However, PPV for both modalities was not significantly different. The authors concluded that FS mammography and FFDM with soft-copy reading are comparable techniques for population based screening mammography programs. However, it should be borne in mind that this conclusion may be safely applied only to this specific model, and after the training and viewing conditions issues raised in the Oslo I study [69] have been addressed [55].

The follow-up and the final results of the Oslo II study were discussed later in 2007 [71]. The purpose of this study was to prospectively compare the performance of FS mammography and FFDM in a population-based screening program. In this study, the higher cancer detection rate and the lower interval cancer rate in FFDM compared with FS mammography was of particular interest because this indicates that FFDM is superior to FS mammography screening. However, since the slightly larger mean and median size of detected cancers at FFDM and the small number of cancers do not justify this suggestion and the authors suggest further studies before making such conclusions. Of interest was the higher cancer detection rate noted in FFDM in the Oslo II study as compared to the Oslo I study. The authors in this follow-up state that possible reasons for this effect could be many however the most obvious being: a) the learning curve effect, i.e. the radiologist would have increased their knowledge and experience in viewing and interpreting soft-copy images from FFDM, b) the use of a dedicated screening/reporting room in the Oslo II study; c) the combination of inexperience with FFDM and too-fast soft copy batch reading might have contributed to the greater rate of false-negative findings in FFDM in the Oslo I study. In conclusion, this Oslo II follow-up study demonstrated a significant higher cancer detection-rate in FFDM with soft-copy image reading whereas PPV are still comparable for both modalities. FFDM with soft-copy image reading is well suited for breast cancer screening programmes [71].

The largest clinical trial performed to-date with a total of 49, 528 women, is the Digital Mammography in Screening Trial (DMIST) conducted by the American College of Radiology Information Network (ACRIN), better known as the ACRIN-DMIST Trial done in 2005 [60]. In this study digital mammography, included both FFDM and CR systems, were again compared to FS mammography in all the participants. Two standard views of each breast and any additional views required were taken using both modalities. The digital and the FS examinations were independently interpreted by two radiologists rating the mammograms on a 7-point malignancy scale (1 being definitely not malignant and 7 being definitely malignant) suitable for receiver-operating-characteristic (ROC) analysis and classification of the Breast Imaging Reporting and Data System (BI-RADS) [47] and recorded whether they recommended any additional tests. A BI-RADS score of 0 indicates incomplete data, 1 negative results, 2 benign findings, 3 probably benign, 4 suspicious abnormality and 5 findings highly suggestive of cancer. The breast density was also rated according to the standard BI-RADS scale (extremely dense, heterogeneously dense, scattered

fibro glandular densities and almost completely fat. The primary measurements of this study included the area under the ROC curves, sensitivity and specificity, PPV and NPV. Secondary aims of the study were to assess the effects of patient characteristics, including age, lesions type and breast density on diagnostic accuracy. Finally, a direct and long-term cost-effectiveness analysis is to be performed. It was concluded that there is no significant difference in diagnostic accuracy between digital and FS mammography for the entire study population. However, from the ROC analyses, the accuracy of digital mammography was significantly higher than FS among women under the age of 50 years, women with heterogeneously dense or extremely dense breasts on mammography and premenopausal and perimenopausal women. The recall rates for both modalities were 8.5% and there is no significant difference reported in sensitivity and specificity. There is also no evidence of a difference in diagnostic accuracy with the type of digital machine.

In 2008, the accuracy of digital versus FS mammography was retrospectively compared in the mentioned population subgroups within the ACRIN-DMIST Trial [72]. This study did not provide a definite answer as to why digital mammography performed better in women under the age of 50 years, women with heterogeneously dense or extremely dense breasts on mammography and premenopausal and perimenopausal women. There is also still no plausible explanation as to why FS mammography performs better in women aged 65 years or older with fatty breasts. In conclusion this study restates the conclusions made in the original ACRIN-DMIST Trial [60].

In 2009 a retrospective study was performed to review the performance of FFDM in a population-based screening program and to compare its performance with FS mammography [73]. A retrospective analysis was performed to determine if screening was performed on a digital system or a FS system, if the women were recalled for further assessment, if a biopsy was performed and if cancer was diagnosed. The recall rate, biopsy rate, cancer detection rate and the PPV for women who underwent FFDM were then determined and compared to those values for women who underwent standard FS mammography. The cancer detection rate for initial and subsequent screening, for invasive and ductal carcinoma in situ across all age groups, for cancers presenting as microcalcifications and recall rates were all significantly higher in FFDM when compared to film=screen mammography. However, there was no statistical significant difference in the PPV of recall assessment between the two modalities. In conclusion, FFDM resulted in higher cancer

detection and recall rates when compared to FS in women aged 50-64 years of age. The results of this study suggest FFDM can be safely implemented in breast cancer screening programs [73].

Table 2.1 summarizes the reviewed clinical studies comparing FFDM to FS Mammography.

Table 2.1. Previous Studies Comparing FFDM to FS Mammography.

	Study Design					Results		
Study	***Retrospective or Prospective***	***Population***	***Imaging Technique(s)***	**Patient Age (y)**	**Study Size (no. of examinations)**	***Recall Rate (p)***	***Cancer Detection Rate (p)***	***PPV_1 (p)***
Irish BSP [73]	Retrospective review	Population-based screening program; initial and subsequent screenings	FFDM or FS	50-64 y	188,823 total: 35,204 FFDM + 153,619 SFM	FFDM > SFM: 4.0% vs 3.1% (< 0.001)	FFDM > SFM: 0.63% vs 0.52% (0.01)	FFDM = 15.7%, SFM = 16.7% (0.383)
Oslo I [69]	Prospective case control study	Population-based screening program; subsequent screenings only	FFDM and FS	50-69 y	3,683 paired examinations: 3,683 FFDM and 3,683 SFM	FFDM > SFM: 4.6% vs 3.5% (NS)	SFM > FFDM: 0.76% vs 0.62% (NS)	SFM = 20%, FFDM = 12% (NA)
Oslo II [70] and follow up [71]	Prospective randomized control trial	Population-based screening program; initial and subsequent screenings	FFDM or FS	45-69 y	23,929 total: 16,985 SFM + 6,944 FFDM	FFDM > SFM: 4.2% vs 2.5% (< 0.001)	FFDM > SFM: 0.59% vs 0.38% (0.02)	SFM = 15.1%, FFDM = 13.9% (0.68)

DMIST [60]	Prospective study	Population recruited over 2 years at 33 sites in the United States and Canada	FFDM and FS	<50 y, 50-64 y, >65 y	42,760 paired examinations: 42,760 FFDM and 42,760 SFM	8.4% for both FFDM and SFM	Accuracy of FFDM was significantly higher in pre- or perimenopausal women < 50 y with dense breasts[a]	
Follow-up DMST [72]	Retrospective review	Population recruited over 2 years at 33 sites in the United States and Canada	FFDM and FS	<50 y, 50-64 y, >65 y	42,760 paired examinations: 42,760 FFDM and 42,760 SFM	FFDM performed significantly better than FS mammography for pre- and perimenopausal women < 50 y with dense breasts, but film tended non significantly to perform better for women aged >65 y with fatty breasts.		
Colorado I Study. [67]	Prospective study	Women presenting for screening at two institutions	FFDM and FS	40 y and older	4,945 paired examinations: 4,523 cases SFM followed by FFDM (same technologist). In 422 cases SFM followed by FFDM in an outlying screening centre on two separate visits.			
Colorado II Study. [68]	Prospective study	Women presenting for screening at two institutions	FFDM and FS	40 y and older	6,736 paired examinations: 6,736 FFDM and 6,736 SFM	SFM > FFDM: 14.9% vs 11.8% (< 0.001)	SFM > FFDM: 0.49% vs 0.4% (NS)	FFDM = 3.4%, SFM = 3.3% (NS)
Heddson et al [74]	Retrospective review	Population-based screening program; subsequent screening only	PCDR, CR, or FS	40-74 y	52,172 total: 25,901 SFM + 9,841 PC-DR + 16,430 CR	SFM > PCDR: 1.4% vs 1.0% (< 0.001)	PCDR > SFM 0.49% vs 0.31% (0.01)	PCDR = 47%, CR = 39%, SFM = 22% (< 0.001)

						SFM > CR: 1.4% vs 1% (< 0.001)	CR > SFM 0.38% vs 0.31% (0.22)	
Del Turco et al [75]	Retrospective review	Population-based screening program; initial and subsequent screenings	FFDM or FS	50-69 y	28,770 total: 14,385 FFDM + 14,385 SFM	FFDM > SFM: 4.56% vs 3.96% (0.01)	FFDM > SFM (NS) 0.72% vs 0.58% (0.14)	FFDM = 15.9%, SFM = 14.7% (0.65)
Vigeland et al [76]	Retrospective review	Population-based screening program; initial screenings only	FFDM only; compared with FS data from previous 9 y	50-69 y	343,002 total: 18,239 FFDM + 324,763 SFM	FFDM = SFM: 4.09% vs 4.16% (NS)	FFDM > SFM for all cancers: 0.77% vs 0.65% (0.058)	FFDM = 16.6%, SFM = 13.5% (0.014)
							DCIS: 0.21% vs 0.11% (< 0.001)	
Hambly et al [73]	Retrospective review	Population-based screening program; initial screenings only	FFDM or FS	50-64	188,823 total: 35,204 FFDM + 153,619 SFM	FFDM > SFM: 4.00% vs 3.10% (<0.001)	FFDM > SFM for all cancers: 0.63% vs 0.52% (0.01)	FFDM = 16.7%, SFM = 15.7% (0.383)

Note—PPV_1 = positive predictive value of recall to assessment, NS = non-significant, NA = not available, PCDR = photon-counting direct radiography, CR = computed radiography, DM = digital mammography, DMIST = Digital Mammographic Imaging Screening Trial, INBSP = Irish National Breast Screening Programme, DCIS = ductal carcinoma in situ.

[a] The end point of the study was diagnostic accuracy.

Diagnostic Reference Levels (DRLs) in Mammography

In mammography, the MGD within the breast is the used as a patient dose estimate because it is the glandular tissue that is believed to be the most sensitive to radiation induced carcinogenesis. Direct measurements on patients are not possible, and the MGD is calculated from measured X-ray tube output and exposure factors using the appropriate conversion factors [77]. In the UK, European and International Atomic Energy Agency (IAEA) dosimetry protocols, the MGD is estimated using:

$$D\,or\,MGD = Kcgs \qquad 7$$

In this expression K is the incident air kerma at the upper surface of the breast, measured without backscatter and g, c and s are Monte Carlo calculated conversion factors [78]. More information is given on page 114.

In order to improve dose optimization in radiological examinations the International Commission of Radiological Protection (ICRP) recommends the use of DRLs [79]. A DRL is a dose level for a typical X-ray examination of a group of patients with standard body sizes and for broadly defined types of equipment. These levels are expected not to be exceeded for standard procedures when good and normal practice regarding diagnostic and technical performance is applied [80, 81]. The justification of a breast screening programme requires that the radiation risks are outweighed by the benefits achieved. If the doses in different centres are regularly compared to the DRL, doses employed will be controlled and kept below the acceptable levels. Eventually, the DRL can be used to assess the risk induced in mammography screening [80]. The Medical Exposure Directive 97/43/Euratom and now EU Directive 2013/59/Euratom, mandate the use of DRL, stating that DRLs should be established in EC member states, in particular for screening [82, 83].

A proposal for a National DRL (NDRL) has been formulated by the NHSBSP Coordinating group for Mammography Physics and recommendations have been published by the Institute of Physics and Engineering in Medicine (IPEM) [84, 85]. In IPEM report 91 [84], MGD is categorized as either being to the 'standard breast' or to the 'patient'. The standard breast refers to a 4.5 cm PMMA

phantom which translates into a 5.3 cm thick breast. The MGD to the standard breast involves the determination of the mAs for the correct exposure to a 4.5 cm PMMA phantom to calculate the MGD having knowledge on the output/mAs, distance and HVL of the system used. The MGD may be calculated using well established methods having this data at hand [78]. The existing remedial level according to IPEM 91 is a MGD greater than 2.5 mGy per image. However, it is also stated that an MGD less than 1 mGy per image should be investigated to decide whether the image quality obtained is adequate.

With reference to MGD to patients, IPEM 91 recommends that a record of kVp, mAs, target, filter and the compressed breast thickness are kept for 50-100 patients. The MGD is calculated for each patient and the mean value and the range are assessed. Variations in the mean value may be attributed to changes or differences in machine performance, amount of breast compression and to variations in the patient population e.g. breast size and composition [84]. With respect to MGD to patients, the set remedial level of 2 mGy MGD set by the NHBSP Coordinating Group for Mammography Physics is translated to the equivalent requirement for a standard breast (5.5 cm compressed breast thickness) by multiplying with a conversion factor. A suitable conversion factor applicable to lateral oblique views has been derived [86]. The value of this factor is 1.73, hence the proposed NDRL for mammography in the NHSBSP is 3.5 mGy [84].

With reference to table 2.2, A recent study was performed in Iran aimed at evaluating the MGD and affecting factors during mammography examinations by FFDM [87]. The exposure factors including kVp, mAs, filter and target were recorded for 1145 mammograms including CC and MLO views. In addition to technical factors a set of data consisting of patient age, height and compressed breast thickness were also recorded. From the images achieved, the breast were classified according to their parenchymal pattern [88] by three expert radiologists as follow:

- Type 1: mostly composed of fat (<25% fibroglandular)
- Type 2: with scattered fibroglandular densities (25-50% fibroglandular)
- Type 3: heterogeneously dense (51-75% fibroglandular)
- Type 4: composed almost entirely of glandular tissue (>75% fibroglandular)

54% of the women in the Iranian study had breasts classified as type 2 followed by type 1 in frequency. The results showed that there exists a significant correlation between breast parenchymal pattern and MGD as follows: type 1 breasts yielded the highest mean MGD followed by type 3, 2 and 4. A significant correlation was also observed between MGD and the compressed breast thickness with mean compressed breast thicknesses of 4.9 cm and 5.8 cm and MGDs of 2 mGy and 2.4 mGy observed for CC and MLO views respectively. The authors also reported a significant correlation of MGD with the employed kVp and mAs [87].

In a similar study of the Irish breast screening programme in 2010, the impact of digital mammography screening on breast dose was analyzed [81]. As seen in table 2.2, the results from 12,110 mammograms in this study were used to determine DRLs for the screening programme. The average compressed breast thickness in this study was found to be 61.4±0.03 mm which is relatively high when compared to other studies. As a result, determination of the DRLs was based on the average MGD for the MLO views obtained for a compression thickness in the range of 60±5 mm. The 95th percentile of the distribution of the resultant MGD values was used as DRL. The data from each individual system was compared to the DRL. To compare the results of any particular system with the established DRL, the authors used the standard error of the mean (SEM) using the following equation calculated for all images of breast compressed thickness 55-65 mm:

$$SEM = \frac{\delta}{n^{0.5}} \qquad 8$$

Here n represents the number of patients in the thickness class and δ represents the standard deviation of the MGD. As proposed in another paper [89], the average MGD of any particular system is considered to be significantly greater than the DRL if the MGD plus twice the SEM exceeds the DRL. The authors concluded that the results in their survey demonstrated that the MGD for the Irish Breast screening programme was lower than that reported in other similar studies using digital mammography. The reported average MGD is 2.72±0.04 mGy which on face value appears larger when compared to other studies [44, 51, 52, 87, 90, 91] however, as stated earlier, the compressed breast thickness in this study is much larger when compared to the others (see table 2). The established DRL in this study was that of 1.75 mGy which is significantly lower than the UK value of 3.5 mGy [90] and 2.37 mGy of the Belgian programme [91]. The process of

setting up DRLs proves useful in the identification of systems that require optimization and the results clearly demonstrate the dose-benefits of digital mammography in breast cancer screening.

MGD to the standard breast using PMMA phantoms should be audited at least every 6 months. Doses to patients should be audited 1-3 yearly [84]. For the latter, the required dose audit measure can be obtained as the mean MGD from a sample of patients with a compressed breast thickness of 55±5 mm. Alternatively, to alleviate the data collection burden, a sample of ten patients, each with a compressed breast thickness lying within 50-60 mm can be used to derive the dose audit measure. The dose audit measure may be used to test compliance with local DRLs (LDRL) and if required to establish new DRLs for any particular organization [84].

In a large survey conducted in 2005 in the UK, it resulted that 97% of 290 X-ray screening mammography units in the NHSBSP complied with the standard for dose [90]. The average MGD was 2.23 mGy for MLO views and 1.96 mGy for CC views. It was concluded that with the increased use of sophisticated units with automatic beam quality selection have reduced the dose to large breasts. As a result it was concluded that this study confirms that the proposed NDRL of 3.5 mGy for 55 mm thick breast is appropriate to identify systems giving unusually high doses with just 3.5% of the participating systems exceeding this limit. However, it was noted that MLO views for average sized breasts yielded a MGD that is 42% higher than the standard breast. The authors stated that that this finding highlights the need to revise the definition of the standard breast in the UK to give values that better reflect doses received in clinical practice [90].

With reference to the European guidelines [1, 2], it is assumed that MGD in digital mammography systems should be equivalent to FS systems. To ensure this, the limiting dose values have been changed compared to the third edition of the European guidelines for quality assurance in mammography screening in three aspects [2]:

- In the present version of the protocol (4th Edition) the clinical spectrum is used for dose measurements instead of a standard spectrum, the dose limits have been made independent of optical density and a limiting dose value per PMMA thickness is introduced.

- The use of clinical spectra in dose measurements is introduced which is closer to clinical practice.
- In digital mammography the link between limiting dose values and OD is non-existent. Therefore a choice had to be made what limiting dose values would be appropriate for digital mammography. In the view of the authors, inspired by the *as low as reasonably achievable* (ALARA) principle, dose should not increase substantially when changing to digital mammography. Data from the Dutch, Swedish, Norwegian and UK (NHSBSP) screening programmes as cited in the European protocol [2] show that MGD in FS mammography systems is between 0.8 and 2.5 mGy for 4.5 cm PMMA in clinical settings (corrected for difference in standard PMMA thickness in the UK and the Netherlands). Therefore a MGD limit of 2.5 mGy at standard thickness in clinical settings has been chosen in the European protocol to ensure that dose levels in digital mammography will not exceed those of FS mammography. This limiting value is comparable to the objective of the NHSBSP in the UK to have MGD of 2 mGy or less (for 4.0 cm PMMA) and the limiting MGD value for the Dutch screening programme (3 mGy for 5.0 cm PMMA).

In the current version of the European protocol limiting dose values for a range of PMMA thickness have been introduced. This has been done because in some non-AEC systems it was noticed that manufacturers decreased dose at standard thicknesses to comply with the limiting value at standard thickness while dose levels at other thickness were found to be much higher than those found in FS mammography. Besides this it has been found that some systems did use much lower kVp than in FS mammography (thus increasing patient dose substantially). These very low kVp values proved unnecessary for image quality, hence not in-line with the ALARA principle. The limiting values for PMMA thicknesses other than standard thickness have been obtained by averaging all measured MGD levels per PMMA thickness from all X-ray units of the Dutch screening programme and some German screening trials. The resulting MGD values against PMMA thickness curve has been scaled to the limiting value at standard thickness. The results have been compared with the dose values per PMMA thickness found in the UK and some of the German screening projects. The limiting values were found to be reasonable [2].

A dose-reference level for certain diagnostic procedures means to identify practices with unusually high doses. If image quality is somehow quantified, the same 'worst quarter' principle can also be used to identify practices with less than optimal performance in terms of image quality. In Slovenia, the performance of mammographic centres was evaluated annually. Technical testing included MGD determination and evaluation of technical image quality using the image of a mammographic phantom. From the phantom image, simple image quality parameters are derived and for some of them reference levels can be established [77].

All DRL studies have been reviewed and summarised in table 2.2.

2.0 The use of a figure-of-merit for optimization in digital mammography

INTRODUCTION

This section was published [7]. The increased use of digital mammography in breast cancer screening and the introduction of new systems employing new and different technologies lead us to the question: how good are these systems and which are best? To date, a lot of research has focused on characterising the performance of digital systems in terms of detector performance using the MTF and the DQE. While these metrics are well suited to describe the intrinsic performance of the detector employed by the respective digital imaging system, they fall short of serving as a comprehensive image quality indicator [92, 93]. Since DQE and MTF are usually measured using a specific and strict protocol, their limitations include: lack of consideration for signal-specific and background-specific spectral changes, dose and risk to the patient, noise from anatomical structures and structures that usually contribute to or decrease scattered radiation e.g. grids [94].

It is generally acknowledged that image quality in digital mammography is best assessed using the CNR and SNR particularly at thicknesses other than the standard breast thickness [2]. Absolute values of SNR and CNR are usually system-dependent therefore limiting values need to be expressed in terms of variation in SNR over the whole range of simulated breast thicknesses and

the percentage of CNR at standard thickness respectively [2]. As a result CNR is usually a relative term and its acceptable value is a range which depends on the mammography unit being evaluated [95].

A relatively new concept in assessing the performance of digital mammography systems is that of the FOM. A FOM has been used to compare techniques and exposure factors in the optimization of mammography systems [45, 94, 96], it has been used to compare the performance of similar digital mammography systems employing different spectra [12] and has also been more rarely used to compare different digital mammography systems [43, 95]. The FOM is usually presented as a single number that reflects the status or the performance of any particular system under particular specified conditions. The FOM has been evaluated under different conditions including different target/filter combinations, different kVp and different simulated breast thicknesses. From the literature reviewed there still seems to be no common definition of the FOM in mammography however the one defined as the square of the CNR divided by the MGD seems to be the most promising for intra-system measurements. There is still no substantial literature on the FOM for use in inter-system comparisons. This section aims to review the various proposed roles of a FOM in digital mammography, its current definitions and discuss the benefits and drawbacks of its use.

SEARCH STRATEGY

A systematic review of the literature was performed using Pubmed. The search was limited to English language papers and the search words utilized were as follows: digital mammography, figure-of-merit, image quality, technique optimization, dose reduction, exposure parameters, techniques factors and beam optimization.

OPTIMIZATION AND THE FOM

A FOM is often defined as the end point of optimization [94]. In FS mammography optimization of mammographic techniques is totally dependent on the film OD. The image SNR in FS depends on the amount of radiation contributing to the image, and this is reflected by the OD. Although the

film OD will vary continuously through the image, a target value can be set or is desired when using an attenuating structure such a slab of 4.0 cm of PMMA. This target OD is eventually checked with a densitometer on the resultant processed images, this to ensure that the OD lies between the desired range [10].

In digital mammography, radiographic contrast can be manipulated to achieve any desired level of contrast. This means that in a digital system it is possible to produce an image of arbitrary brightness at virtually any radiation-exposure level [10]. This occurs because the capture and display elements of image formation are decoupled as opposed to FS systems. Samei et al [94] state that this decoupling causes greyscale appearance of an image, except for noise, to be no longer limited by the radiographic technique in terms of kVp and mAs but can be adjusted and manipulated to the desired level by the user.

Therefore, radiographic contrast is less important in digital mammography due to the decoupling process of the image capture and display elements mentioned previously. The contrast measure of choice in digital mammography should remain the CNR [10]. Consequently, optimization processes between FS and digital mammography differ. The FOM is an objective means of optimization in digital mammography whereby, a maximum CNR value is desired at the lowest dose possible to the patient.

From the literature reviewed there are various formats and definitions of various FOMs, all used for the optimization of digital mammography units. These are presented and discussed under separate headings in the following subsections.

Method 1

With the move from FS mammography to digital mammography, absolute contrast is no longer the dominant factor in image quality. Samei et al state that a more relevant quality parameter in digital mammography is the signal-difference-to-noise ratio (SDNR) [94] defined as:

$$SDNR\,or\,CNR = \frac{I_B - I_L}{\partial_B} \qquad 1$$

where, I_L and I_B refer to the intensity of the detector signal corresponding to a lesion and its background surrounding respectively and δ_B is the standard deviation of the detector noise in the background area. This definition of SDNR is congruent to the CNR [43]. Samei et al [94] state that assuming the consistent application of image processing to achieve comparable contrast, a digital image of higher CNR provides superior image quality. Assuming the system is *quantum limited*, for a given beam quality, SDNR or CNR is proportional to the square root of the incident exposure E to the breast, based on Poisson distribution and statistics. Hence, the authors define the following FOM, making it independent of E:

$$FOM = \frac{CNR^2}{E} \qquad 2$$

In their paper, Samei et al [94] conducted a study to evaluate the utility of a tungsten (W) target with an additional 50 µm Rhodium (Rh) filter compared with the conventional Molybdenum (Mo) filter and 30 µm Mo filtration using a prototype digital mammography unit. The background was represented by a variable thickness breast phantom (the 'Computerized Imaging Reference Systems (CIRS) phantom (CIRS, Inc., Norfolk, VA) between 2-8 cm in thickness, with 100% glandularity, 100% adipose and 50/50% glandular/adipose compositions. Anatomical details and simulated lesions were embedded in this phantom to enable the calculation of the CNR. The kVp was varied between 24 and 34 kVp using both tube targets with all the thickness combinations available. The breast entrance surface exposure E in conjunction with the measured HVL were used to calculate the MGD using Boone's program [97]. The results of this study show that the FOM is higher for thinner breast thicknesses and has relatively little dependence on the kVp employed except for the thin breast thicknesses. In most phantom combinations, the W target delivered a higher FOM when compared to the Mo beam. It was also indicated that up to 60% of MGD reduction is possible with the use of the W/Rh spectrum when compared to the Mo/Mo spectrum, particularly with the thicker breasts having more adipose content [94]. Samei et al concluded that the FOM defined in equation **2** is robust and accurate as means of optimization however it fails to incorporate the influence of anatomical noise, image post-processing and spatial frequency dependencies on CNR. The authors suggest that future research in this field should address and overcome these limitations, however that the simplicity of this model makes it easy to use in a clinical environment.

Method 2

The results of Samei et al [94] are in agreement with a later study by Delis et al [98] which evaluated a FOM in mammography utilizing Monte Carlo simulation to derive the energy deposition inside a breast phantom and the signal beneath it. Delis et al argue that the FOM in mammography depends on the specific use and demand of any particular system. For those FS units employed in breast cancer screening whose major objective is the detection of microcalcifications (μCs) and since the latter is a dominant radiographic sign in nearly 30-50% of breast cancers [99] the authors claim that the dosimetry index should be MGD rather than E as described by Samei et al and that the image quality indicator should be the subject contrast (SC) for μCs. E might be a more simplistic dose quantity since it is easier to measure and relatively more commonly used. Nevertheless, the MGD is the most appropriate quantity in mammography for the assessment of carcinogenic risk. Again assuming the system is quantum limited and that Poisson statistics are followed, the FOM suggested by Delis et al was defined as:

$$FOM = \frac{SC^2{}_{\mu c}}{MGD} \qquad 9$$

Delis et al state that in digital mammography systems the primary image quality index should be the CNR rather than the subject contrast (SC) for μCs.

In digital mammography taking into consideration system noise, expressed as the standard deviation of the output image, is of vital importance in the determination of image quality as this will definitely have a large impact in the resultant FOM calculation. The standard deviation in the digital image is assumed to follow the following relationship:

$$\delta = (\delta_e^2 + \delta_q^2 + \delta_s^2)^{0.5} \qquad 10$$

where δ_e, δ_q and δ_s represent the standard deviation terms representing electronic noise, quantum noise and structured noise respectively. The contribution of these noise terms are usually incorporated under the generic term 'quantum noise' given they have the same noise-dose

distribution according to Poisson statistics [100]. Two properties are expected for a well-designed detector: quantum noise should be the largest noise component for all kerma values used clinically and the structural and the electronic noise components should be low such that the quantum noise component remains the dominant fraction of the total noise.

The NPS is a scientific metric for describing the spatial frequency content of noise. It has been shown that NPS values has a very large dependence on exposure and decreases with increasing air kerma [101]. It has been also shown that noise expressed as the NPS is lower for harder spectra employing Rh/Rh combinations than for Mo/Mo. Garcia-Molla et al [101] also stated that by maintaining the same target/filter combination (Rh/Rh), NPS values are lower for higher voltages at all spatial frequencies. There is a twofold possible explanation for this decrease in NPS with increased beam hardness: (1) a higher photon fluence with increased kVp and the use of Rh/Rh over Mo/Rh over Mo/Mo at the same air kerma at the detector, minimizing the influence of electronic noise and (2) a higher mean energy spectrum lead to a larger transmission through the flat-panel cover [101].

This can justify the use of W as the anode material in digital rather than conventional mammographic systems which overall produced harder beams when compared to the conventional Mo anode material. Delis et al calculated the CNR for 2 mm air simulated masses and used equation **2** for the calculation of the FOM in digital mammography systems. The results of this study confirm that W target systems are best suited for digital mammographic systems while for FS mammography, Mo and Rh targets present an overall good performance [98]. In this study it was also generally shown that the FOM defined in equation 2 increased with kVp and also increased with increased beam hardening by altering the target filter-filter material. However, it still isn't clear why Delis et al used E rather than the MGD in their calculation of the FOM for digital mammography systems under simulation.

Method 3

With reference to figure 2.9, in another two studies [45, 96] a slightly different approach to the calculation of the FOM was used however still based on a methodology similar to that described by Samei et al [94]. In these studies, the SNR was calculated using a mass-equivalent step-wedge

with the application of specific regions of interest (ROI). With reference to equation **3**, the signal was defined as the difference in mean pixel value (MPV) in an ROI located on a particular step and another ROI of the same size located adjacently containing only background. The signal measured by Williams et al appears to be the measured contrast for the first step in the wedge in relation to corrected signal from the background.

For the measurement of the signal intensity, Williams et al made particular effort in correcting for background trends such as the anode heel effect, by applying extra ROIs in the background as seen in figure 2.9. The trend-corrected signal was then calculated using the following equation **3**:

$$signal = (ROI2 - ROI1) - (ROI4 - ROI3) \qquad 3$$

To quantify image noise Williams et al subtracted two images of the phantom in figure 2.9 with a common technique. Image subtraction was performed to remove correlated noise associated with phantom defects, grid-related noise, detector non-uniformity, and the heel effect. The root-mean-square (RMS) [102] uncorrelated noise in a single image was then given by the standard deviation of the pixel values in a large background ROI of the difference image (ROI 5 in figure 2.9), divided by the square root of 2 as follows:

$$Noise = \frac{STDEV(ROI5)}{\sqrt{2}} \qquad 4$$

In their study, Williams et al very importantly mention limitations that quantification of noise only by way of the RMS pixel-to-pixel fluctuations fails to include the effect of spatial frequency on noise. Although this wasn't within the scope of their study, the authors state that inter-system comparisons in terms of the FOM without a complete frequency-dependent description (in terms of MTF and NPS) of the system performance may be impossible.

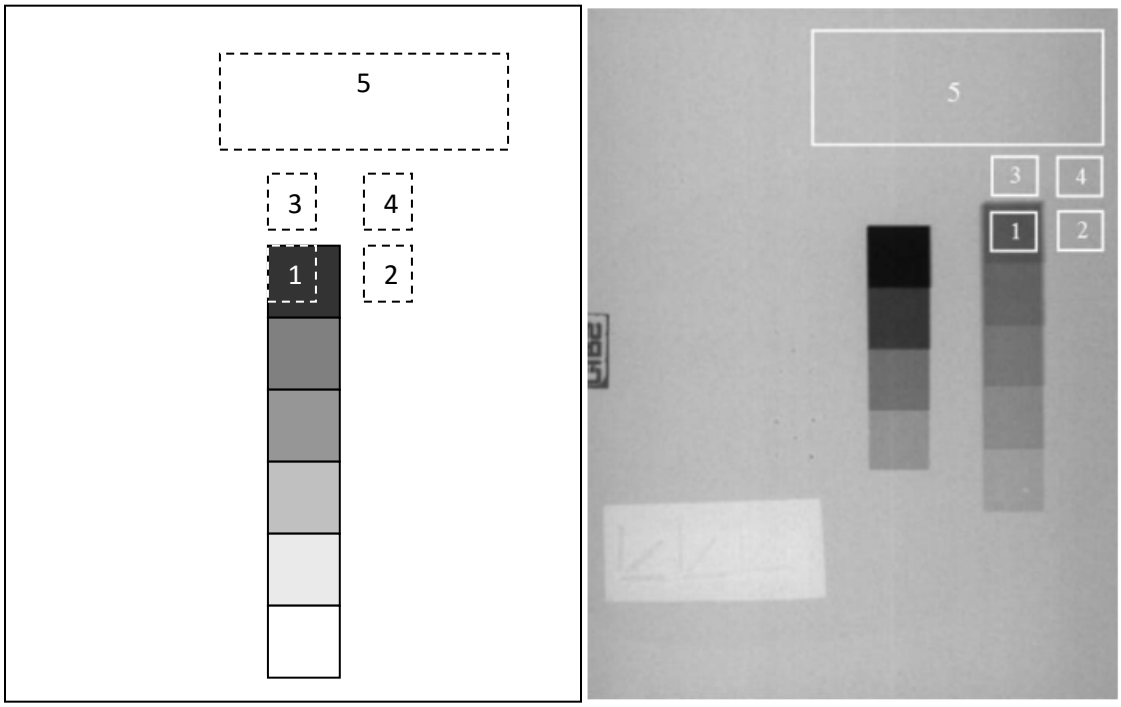


Figure 2.9. Pictorial representation and image of the phantom used by Williams et al [45, 96]). The ROIs shown are used for signal and noise calculations. Difference between ROI 1 and 2: difference between the average analogue-to-digital unit (ADU) value and the background; ROI 3 and 4: are of the same size as ROI 1 and 2, to correct for background trends such as the heel effect; ROI 5: larger ROI to quantify image noise. The chest wall edge of the phantom is the left edge of the image.

Williams et al in both studies [45, 96] define the SNR as the ratio between equation **3** and equation **4**. After the calculation of the MGD for all the spectral combinations including Mo/Mo, Mo/Rh, Rh/Rh, W/Rh and W/Al the authors defined the following FOM:

$$FOM = \frac{SNR^2}{MGD} \qquad 5$$

Williams et al claim that the FOM defined in equation **5** is independent of the exposure used to obtain the image always assuming the system is quantum limited. Higher values of FOM indicate the ability of a system to deliver better performance in terms of SNR at a lower dose to the patient, and thus it is useful for deciding amongst the various acquisition parameter options for a given digital mammography system. In agreement with Samei et al [94], Williams et al [45, 96] concluded that the FOM is not a strong function of the kVp. On the other hand, the FOM is more strongly determined by the choice of the target/filter combination. Interestingly, this

characteristic has also been reported by Berns et al [103] in which imaging performance was quantified with a 'CD score' (contrast x diameter) rather than a FOM. Their results showed that the CD score was a slowly changing function of kVp, similar to the FOM defined in equation **5**. Williams et al state that the FOM defined in equation **5** is useful for studies where the goal is to compare the relative trade-off between image quality and dose for a given imaging system when operational parameters are varied. However, absolute values of the FOM should be interpreted with caution for inter-system comparisons because of machine-specific attributes such as spatial resolution, pixel pitch and the presence of image processing algorithms.

Lo et al [104] utilized an approach and a FOM similar to those described by Williams et al [45, 96] to compare the image quality and dose performance for Mo/Mo and W/Rh target/filter combinations on an early prototype of the Mammomat Novation DR (Siemens). In agreement with the results presented here, the authors found that W/Rh consistently outperformed Mo/Mo for 4, 6, and 8 cm breasts of 0%, 50%, and 100% fibro glandular composition. As a result the FOM proved useful in the optimization process for this mammography system and help identify the spectra that are best suited for digital mammography for any particular breast thickness.

Method 4

Several papers were published suggesting that the square of the CNR divided by the MGD (equation **6**) is a suitable parameter for optimization in digital mammography systems [11, 12, 105, 106].

$$FOM = \frac{CNR^2}{MGD} \qquad 6$$

In a recent study by Kanaga et al [95] the FOM defined in equation **6** was used to compare three FFDM systems, namely Siemens Mammomat Novation DR, Hologic Lorad Selenia and General Electric Senographe Essential.

A breast equivalent phantom, namely the CIRS phantom, which is composed of different thickness and glandularity, was used. The MGD and the CNR were calculated using the methodology

described in the European guidelines [2]. The FOM was then calculated using equation **6**. The results showed that there was a statistical significance in the mean value of MGD and CNR between the three FFDM systems. The Hologic Lorad Selenia system contributed the highest MGD value while General Electric Senographe Essential had the highest CNR and FOM value. Kanaga et al state that their study may provide an objective criterion during the selection of a mammography unit by using the FOM for screening or diagnostic purposes although the authors also claim earlier on in their paper that a comparison of FOM values amongst the three FFDM systems is limited because the CNR values measured are closely associated to the different technologies. As a result, as pointed out by other researches in this review, CNR values are not directly comparable.

The same FOM described in equation **6** was used by Baldelli et al [12] in a study to investigate the effect of target/filter materials on the dose and image quality of a digital mammography system based on an amorphous selenium detector. In this study two FFDM units were compared from Lorad-Hologic. The original systems used the classic Mo/Mo and Mo/Rh target/filter combinations and the newer system uses W/Rh and W/Ag. Images of the CIRS phantom with simulated breast compressed thickness of 4, 5 and 6 cm and various glandular tissue equivalencies were acquired under kVps ranging from 24 to 34 in steps of 2 kVp. The mAs values were chosen manually in order to obtain a constant pixel value in the reference zone within the CIRS phantom (equivalent to using the automatic exposure control). The CIRS phantom has five details corresponding to different glandular composition of 0%, 20%, 50%, 70% and 100%. For each detail the CNR was calculated using the methodology described in the European guidelines [2]. For each setting employed the HVL and E were measured and the corresponding MGD was calculated using the methodology described by Dance et al [78, 107]. The optimization and the comparison of the two systems mentioned in this study were then done by comparing their performance based on the FOM. In agreement with Williams et al [45, 96], the authors state that assuming the two systems tested are both quantum limited, the FOM is independent of the dose level. Higher values of the FOM indicate a better ability of the system in terms of image quality at lower dose performance. This is very useful in establishing the best acquisition setting for a given mammographic system. Baldelli et al calculated the FOM for all acquisition conditions as a function of kVp for the three phantom thicknesses specified. An optimal kVp was determined by taking the maximum value of the FOM for each acquisition. The values achieved were compared with those chosen by the

system under AEC control. Results show that the W/Rh combination is the best choice in terms of dose and image quality for compressed breast thicknesses ranging from 4 to 6 cm with different glandularities. For thicknesses greater that 6 cm, the best target/filter combination reported was the W/Ag. In addition, the new system with a W target presents a better optimization of the AEC in comparison to its Mo target predecessor [12].

The FOM described in equation **6** was also used in another recent study by Ranger et al [11]. The objective of this study was to evaluate the dose savings possible by switching from analogue to digital mammography and to determine optimized radiographic techniques for a digital mammography system using an amorphous selenium detector, namely the Siemens Mammomat Novation DR. This FOM was evaluated for both Mo/Mo and W/Rh target/filter combinations for breast thicknesses of 2 to 8cm with various glandular compositions. Based on the FOM results, the dependency of system performance on spectral quality was assessed. The results from this study show that significant dose savings (ranging from 9% - 63%) could be achieved with the use of optimized W/Rh spectra in comparison with pre-existing clinical techniques using Mo/Mo, thereby demonstrating the importance of optimization. These dose-savings are in agreement with those specified by Samei et al [94]. Ranger et al also suggest that the optimization of the technique is independent of lesion type, i.e. there is no need to compromise the visualization of masses over calcifications, or vice versa. The authors also suggest that for dense breasts of thicknesses equal to or greater than 8 cm, there is still potential for further spectral optimization using novel target/filter combinations.

DISCUSSION

The ability of FFDM mammographic systems to detect early breast cancer is very desirable however this is heavily influenced by many factors that require optimization. These factors may include beam spectra, MGD, detector performance and detector characteristics, image processing, image display and radiographic technique. The objective of optimization is to establish standardized imaging protocols by determining the optimal trade-off between image quality and dose, which is especially important for screening mammography given the lifetime risk to women who undergo annual mammography examinations [11]. The use of a FOM presents a very

attractive way in performing such an optimization process in any available digital mammography unit. The general concept behind the FOM is that a number is calculated for any particular system, the magnitude of which reflects its performance. The larger the FOM for any particular system the better is its performance. The FOM should be independent of the dose level employed, given that the system used is quantum limited [12]. As a result, the image quality parameter used should be normalized to the patient dose such that we are left with an image quality value per unit dose making the FOM more objective in nature.

When it comes to the selection of image quality parameters, the CNR and SNR have been the parameters of choice in the assessment of mammographic image quality. From the literature reviewed the term SDNR has been used interchangeably with the term CNR. In digital mammography contrast can be adjusted but noise is also changing. Consequently, CNR is the best suited parameter to assess image quality objectively since it tells us how good the contrast is compared to the noise in a given image [108]. A digital image of higher CNR provides superior image quality. However, the problem with CNR (and SNR) is that it is a system-dependent measurement, not only in terms of image processing but also when considering detector-related characteristics such as pixel size [2]. Another pixel-related characteristic that can be overlooked is the pixel fill-factor i.e. the ratio of the radiation sensitive area of the pixel versus the total area of the pixel. It is only the radiation sensitive area that will contribute to image formation. Although manufacturers claim a particular pixel pitch as part of their equipment description, the fill-factor is of vital importance as this quantifies how much of the pixel is actually being used. As a result, as the pixel fill-factor decreases, the detective quantum efficiency of the detector will also below.

Noise is known to increase as the detector pixel size decreases, and different manufacturers are offering different pixels sizes associated with their units. It also relatively uncertain as to what kind or whether any noise corrective algorithms are being applied to help in noise reduction and these may obviously vary between one manufacturer and another. As a result a raw image associated to a particular mammography unit is not directly comparable to a raw image of another differently branded mammography unit creating a lot of uncertainty when measuring and calculating the SNR and the CNR from different units.

Assuming any quantum-limited system, for a given beam quality, the CNR is proportional to the square root of the incident exposure E or the MGD according to Poisson statistics. This explains the FOM defined in equations **9** and **6**. This FOM, in particular the one defined in equation **6**, is useful for investigations to compare the relative trade-off between image quality and dose for a given imaging system when operational parameters are varied. Various studies have been discussed in the previous section reporting the successful use of the FOM for optimization purposes and in identifying the ideal spectral conditions in digital mammography units. FOM analyses and optimization studies have proven essential in identifying the ideal spectral conditions in any particular mammographic unit. For instance, this process proved useful in the move from FS technology to direct digital mammography whereby it is generally becoming more accepted that W targets are more suited to digital mammography. The FOM has also been a very useful in the assessment of AEC devices in mammography systems to ensure that the right exposure parameters are being selected for a specified breast thickness [11, 12, 43, 45, 94, 96, 98].

However, the limitations in comparing absolute values of FOM between imaging systems because of system-specific attributes are still present and are the major limitation in using the FOM more vastly. This limitation is closely associated to the fact that CNR values achieved from different mammography units may not be directly comparable [108]. Apart from pixel size one needs to consider other various manufacturer-specific image attributes that cannot be controlled by the end-user. Amongst these are image post-processing algorithms that are applied to the raw image, for instance those that could be used for thickness compensation, contrast enhancement, MTF compensation, logarithmic transformations etc. [45]. As a result as it is, the FOM is useful for intra-system comparisons and for the optimization of any given mammographic unit. The comparison of mammographic systems using an absolute FOM is not recommended based on this literature review unless there is certainty that the images utilized from any two or more mammography brands are free from any form of manufacturer-specific post-processing or that detector-specific attributes that influence inherent image quality are made common to all.

Nevertheless, it is felt that there is still the scope for use of this parameter in routine quality assurance in digital mammography: as a constancy parameter. For any given mammography unit, the FOM described in equation **6** can be calculated at any particular instance e.g. at acceptance or

commissioning for a given breast thickness e.g. the standard breast, as a baseline value. Any future measure of the FOM will help in determining the overall performance for that particular unit. In this way, the FOM could serve as a quick and descriptive indicator that sums up performance in terms of dose and image quality in mammography. Mammography quality assurance protocols are usually long and time consuming and the FOM might prove itself useful as a 'synopsis', condensing the complex data measured into a simple figure. Saying this, the FOM should in no way replace existing image quality parameters but will help in complementing these parameters by getting a better picture of the performance of a mammography unit. However, the *forte* of the FOM as specified in equation **6** should lie in the optimization process since its sums dose and image quality simultaneously rather than seeing them as independent parameters.

The challenge lies in having a totally objective FOM that is totally independent of manufacturer or system-specific post processing thereby permitting inter-system comparisons. This would widen the scope of the FOM and not limit it to optimization only. For instance this would be useful in any breast screening programme wishing to invest in new equipment or wanting to audit the current equipment available in the programme. In this sense, the FOM could be seen as a handy tool to complement other test regimes in assessing the performance of different mammography units.

It might be worth considering redefining the FOM to permit inter-system comparison. The CNR should be replaced by another image quality parameter that is common to all digital mammographic units and that is not manufacturer or system-specific. It is also important to standardize the methodology for the calculation of the FOM. For instance in most of the literature data for CNR was taken using the CIRS phantom using the embedded step wedge [11, 12, 45, 81, 94, 96]. However, the European Protocol method i.e. using 0.2 mm of 99.9% purity aluminium [43] and also the Monte Carlo method based on micro-calcification attenuation have both been used [98]. It must be appreciated that different methodologies will lead to different results particularly if different materials are used for the calculation of CNR due to the different energy dependency and attenuation properties for each material.

With further investigation and research, it might also be worth considering using a different approach to image quality analysis rather than SDNR or CNR in defining the FOM such as threshold

contrast or the threshold contrast detectability index. However, with reference to equation **6**, the FOM stated can be confidently used in a particular digital mammographic system as an aid in technique optimization and as a simple assessment tool of the system performance for any particular mammography unit.

CONCLUSION

The use of a FOM in digital mammography for technique optimization appears promising. This is particularly important in digital mammography units employed in breast cancer screening programmes, given the lifetime risk to women who undergo repeated mammography examinations. From the literature reviewed it is evident that the FOM is an evolving quality indicator with many variations in its definition. In the author's opinion, the FOM defined as the ratio of the CNR^2 to the MGD seems to be the most appropriate and suitable for intra-system comparison. This FOM proved useful in identifying the potential for dose saving in digital mammography systems through the optimization of the spectra employed. Nevertheless, it is felt that further research is required to evaluate the use of a FOM for inter-system comparison because of system and manufacturer-dependent factors such as pixel size. However, one also needs to keep in mind that CNR itself may also be an unreliable metric because of image processing or corrective algorithms that are not apparent in the resultant raw images. As a result absolute values of the FOM are bound to be different between any one system and another. In the author's opinion mammography inter-system comparison could be made possible by replacing the CNR in the FOM definition by a different image quality indicator with less dependence on manufacturer and detector-specific attributes but rather gauges image quality holistically. Despite the evident need for further research in this matter, the FOM concept is promising and should be considered as a valid tool in the routine quality assurance test regime.

3.0 Mammography units included in this exploration

More detailed information on the detector types can be sought in the section

Digital Mammography page 27.

The mammography units that were included in this exploratory study comprised three detector types and described as follows: (1) **The General Electric Senographe Essential** is considered being an indirect flat-panel DR mammography unit, where x-rays are first converted into visible light photons which are eventually detected by photosensitive elements [109, 110]. The latter comprises a rectangular array of light-sensitive photodiodes deposited onto a plate. These are interconnected with an array of control and data lines as well as TFT adjacent to each photodiode. X-rays are absorbed by a layer of thallium-activated caesium iodide phosphor deposited on the photodiodes. The latter serve as dels that detect the light emitted by the phosphor and create and electrical charge signal that is stored in each del [111].

Another DR mammography unit included in this study is the (2) **Hologic Selenia Dimensions 2D**. This system is classified as a direct conversion flat-panel detector system, where X-rays are directly converted into electron-hole pairs [104]. In this type of detector, the X-ray absorber is a thin layer of amorphous selenium. When X-rays interact with the selenium and produce energetic photoelectrons, the latter interact with the electrons in the outer orbits of the selenium atoms. These interactions cause liberation of electrons and therefore the respective creations of 'holes'. Electrodes that create an electric field are applied on both surfaces of the selenium layer. One of these electrodes is continuous and the other acts as a del matrix which act as capacitors to store charge. Each del bares a TFT switch in each corner and signal readout occurs in the same way as with the previously mentioned indirect system [111].

Finally, the (3) **Fuji Profect CS**, is a CR system was employed. Energy from the X-rays is absorbed in a screen composed of a phosphor material containing a high prevalence for electron trapping sites. The absorbed energy causes electrons in the phosphor layer to be temporarily freed from the crystal matrix and then captured in 'traps' within the crystal lattice. The screen is then stimulated by a scanning laser beam to release the deposited energy in the form of visible light captured by a light-sensitive system to eventually define the acquired image [111, 112]. The Fuji system was later excluded from this study as it was removed from service.

4.0 The CDMAM phantom and contrast-detail scoring methods.

The contrast detail-mammography (CDMAM) phantom in figure 2.11 is typically referred to as a so-called 'Rose-type' phantom. Typically a Rose-type phantom has the shame shape objects with variations in thickness (contrast) and size (detail). The readers' task is to identify how many of the objects are detectable. The advantage of Rose phantoms is that they allow a determination of the limits the reader can detect, which is graphically illustrated with the construction of a contrast-detail diagram, which is a plot of the size against the diameter of the phantom details. The CDMAM phantom has a large number of details (410), making possible the evaluation of resolution and contrast properties of a digital mammography system [113].

The Artinis CDMAM 3.4 phantom [114] consists of an aluminium base with gold discs (99.99% pure gold) of various thickness and diameter. The gold discs are arranged in a matrix of 16 rows by 16 columns. Within a row the disc diameter is constant, with logarithmic increasing thickness and within a column the thickness of the discs is constant and the diameter increases logarithmic. The thickness is between 0.03 µm and 2.00 µm and the diameter is 0.06 mm between 2.0 mm. Each square contains two identical discs (same thickness, same diameter), one in the centre and one in a randomly chosen corner.

The aluminium base (0.5 mm thick Al, 99.5% pure aluminium) is attached to a PMMA cover (3 mm). Under normal mammography-radiation conditions (Mo/Mo, 28 kV) the aluminium base and PMMA cover together have an equivalent PMMA thickness of 10 mm. The phantom is delivered with four PMMA plates, each 10 mm thickness, which are used for the simulation of different breast thicknesses. The phantom and PMMA plates match the standard mammography film size (18 x 24 cm). The actual attenuation of the CDMAM phantom depends on the configuration of the phantom and PMMA plates. The effective energy of the phantom plane will be higher when more PMMA is added to the top and bottom of the phantom [114].

With reference to table 2.2, the European guidelines [2] published acceptable and achievable image quality levels in digital mammography based on contrast-detail measurements using the

CDMAM phantom. These minimum standards were set to ensure that digital standards are as good as or better than FS systems [30]. Although measurements of threshold contrast may be difficult to conduct reliably, advantage could be taken from the fact that on a given system threshold contrast is related to CNR as follows:

$$Threshold\ contrast = \frac{\lambda}{CNR} \qquad 11$$

where λ is a fitting variable. The advantages and disadvantages of using CNR as an image quality parameter have already been discussed. One potential source of confusion is that while a high CNR leads to a high image quality on any given system, it is ***not*** a valid measure for comparing systems that have different MTFs. This is because systems with the best MTF values need lower CNR for a given level of image quality [115]

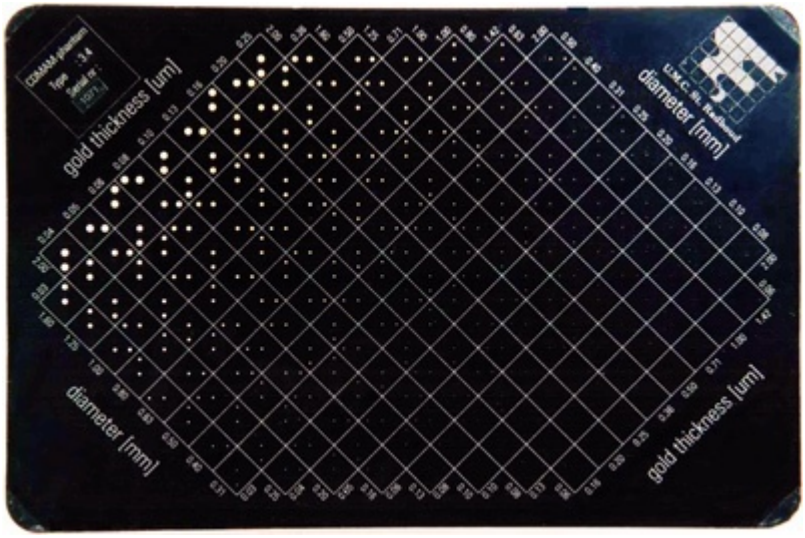


Figure 2.11. Photograph of the CDMAM Phantom (version 3.4, Artinis, St. Walburg 4, 6671 AS Zetten, The Netherlands.

There are basically five known methods of scoring the CDMAM phantom as follows. The first method 1) is the plotting of a contrast-detail (CD) curve representing a plot of the minimal detail diameters and respective thicknesses detected. The more the curve lays to the lower left of the plot the better the performance of the system in terms of image quality. The disadvantage with CD curves lies in the fact that there is no numerical expression given to quantify image quality [116, 117].

The CDMAM manual states that image quality can be expressed in a figure by calculating the ratio of the correctly identified disk positions to the total number of squares, known as the 2) correct observation ratio (COR) expressed as follows:

$$COR = \frac{Correct\ Observations}{Total\ Number\ of\ squares} * 100\% \qquad 12$$

Image quality may be defined and expressed as an 3) Image Quality Figure (IQF) or the IQF inverse (IQF_{INV}) defined as follows:

$$IQF = \sum_{i=1}^{16} C_i * D_{i,\min} \qquad 13$$

Where $D_{i,min}$ denotes the threshold diameter in contrast column i [118].

The 4) S parameter has been mentioned and described as the sum of all the phantom cells for which one or both disks are not detectable on the CDMAM image [118]. Zoetelief et al, also defined the 5) K value, as the square of the product of thickness and diameter of the smallest correctly identified discs. The K value can be determined separately for every row of a given CDMAM image, or its average value over all phantom rows can be used for overall characterization of image quality.

Thomas et al [113], also briefly mention a FOM to score CDMAM phantom. They define this parameter as the zero disk diameter value obtained from a linear extrapolation of the CD curves to the origin (e.g. zero disk diameter).

SCORING METHODS FOR THE CDMAM PHANTOM – MANUAL AND AUTOMATED APPROACHES.

Differences in the basic concept of the manual and computerized approach in the evaluation of contrast-detail performance should firstly be pointed out. In manual reading, an observer is not

obliged to score all the cells. He is forced to guess two more cells after the last visible one. This gives a matrix of correctly, falsely and not indicated cells. The scoring of one image results in a binary matrix, this matrix undergoes the nearest neighbour correction (NNC) analysis, and at the end the resulting threshold values of all analyzed scorings are averaged [119].

As mentioned earlier, the process of scoring CDMAM images by a human observer suffers from two main disadvantages. One is the significant presence of inter- and intra-observer error and the other is that the process is very time-consuming. A possible solution is the to use automated software to assess and score the phantom images [120]. In automated readout, all the cells of the image are scored. After correction, this results in a matrix of correctly and not correctly indicated disks. The CDCOM tool gives a binary matrix of the scored image and all matrices are averaged so that a probability matrix of correct detection is obtained. This probability matrix is further analyzed from which the threshold values are deduced. If commercially available applications such as the CDMAM Analyser Software are employed, as in this exploration, more than one image can be included in the analyses.

CDMAM 3.4 Analyser

The CDMAM 3.4 image analysis software [121] analyses the CDMAM 3.4 phantom images according to the supplement to the European guidelines [2]. The software analyses the images by:

- collection the results of the automatic readout software CDCOM (automated software by European Reference Organisation for Quality Assured Breast Screening and Diagnostic Services (EUREF)),
    - In CDCOM version 1.6 the position of the gold discs is estimated using the discs on the phantom which are clearly visible in combination with a template with the actual position of all discs.
    - This template is translated, rotated and skewed for an optimal fit to the phantom.
- determination the psychometric curves
- detection the threshold thickness of the different diameters,
- correction to human read values

- calculation the CD curves.

The software enables quick analysis and result overview of many sets of images for easy comparison of quality differences using different techniques, settings or mammography images and due to over time changes of your mammographic system. The results are shown for each image and each analysis step. The IQF_{INV} and percentage of detected gold discs are calculated for quick reviews of analysis results.

Hard threshold

It is easy to understand that averaging a range of binary (1 – 0) matrix cells results in values between 0 and 1. The value of a particular cell then represents the probability that it is correctly read. For the scored cells, correct or not, this is relatively straightforward. However, the *not* scored cells are debatable and open to further discussion [119].

From the perspective of probability, the most used scoring method would be as follows:

- Correctly indicated cell = 1
- Falsely indicated cell = 0
- Not indicated cell = 0.25

The cell has an actual probability of 25% to be correctly scored in a completely random guessing situation: there are 4 corners with equal probability in which the golden disc can be located. After correction of a single scoring, the matrix consists of 1's, 0's and 0.25's. Averaging results then in the final matrix which represents the probability that a cell is correctly indicated. However, an alternate reasoning could be applied: the cells that are not scored are not visible to the observer. If 'visibility' is the criterion, it makes more sense to have a binary matrix [119]:

- Correctly indicated cell = visible = 1
- Falsely or not indicated cell = invisible = 0

Averaging these matrices results in a probability matrix too: the probability that a particular disc is visible to the observer. This situation can be avoided by *forcing* the observer to score all the cells of the phantom, then the scoring is identical to the computerized scoring concept and the determination of the probability matrix is obvious. However, this would prolong significantly the time needed to read one single CDMAM image. A simple way to determine the threshold values, without any further analysis, is to set a limiting value to the probability (e.g. 50% or 62.5%). The cell with the lowest probability larger than the limiting value is then considered as the last visible cell. This analysis is repeated for each row in the phantom. All threshold diameters and thicknesses are deduced, they are subject to a second order polynomial fit which gives the final threshold values by which the system can be evaluated [119].

Gaussian filter

An ideal scoring of a CDMAM image would result in a perfect monotonous increasing detection probability along the line of diameter in the direction of increasing gold thickness and vice versa. In the border region of visibility, a disc is more likely to be detectable for a larger diameter or a thicker disc. However, in reality, jumps are often noted in the probabilities. A common tool used in image or data processing to smear out high frequency components is smoothing using the Gaussian filter as seen in figure 2.12.

$$\frac{1}{\sqrt{2\pi}\sigma}\begin{pmatrix} e^{-1/\sigma^2} & e^{-1/2\sigma^2} & e^{-1/\sigma^2} \\ e^{-1/2\sigma^2} & 1 & e^{-1/2\sigma^2} \\ e^{-1/\sigma^2} & e^{-1/2\sigma^2} & e^{-1/\sigma^2} \end{pmatrix}$$

Figure 2.12. Mathematical representations of the Gaussian filter.

Dealing with the borders of a matrix could be a challenge. In the CDMAM case, it is especially important that the low contrast/high diameter and the high contrast/low diameter corners have accurate results. *Zero padding, periodization* and *symmetrisation* are commonly used ways to deal with boundary conditions in image processing. However, none of these is appropriate for the CDMAM application [119].

A human, subjective interaction based on common sense is the only thinkable method to approach this situation. Figure 2.13 shows how the filter can be applied at the borders in the work by Verbrugge (2007) [119].

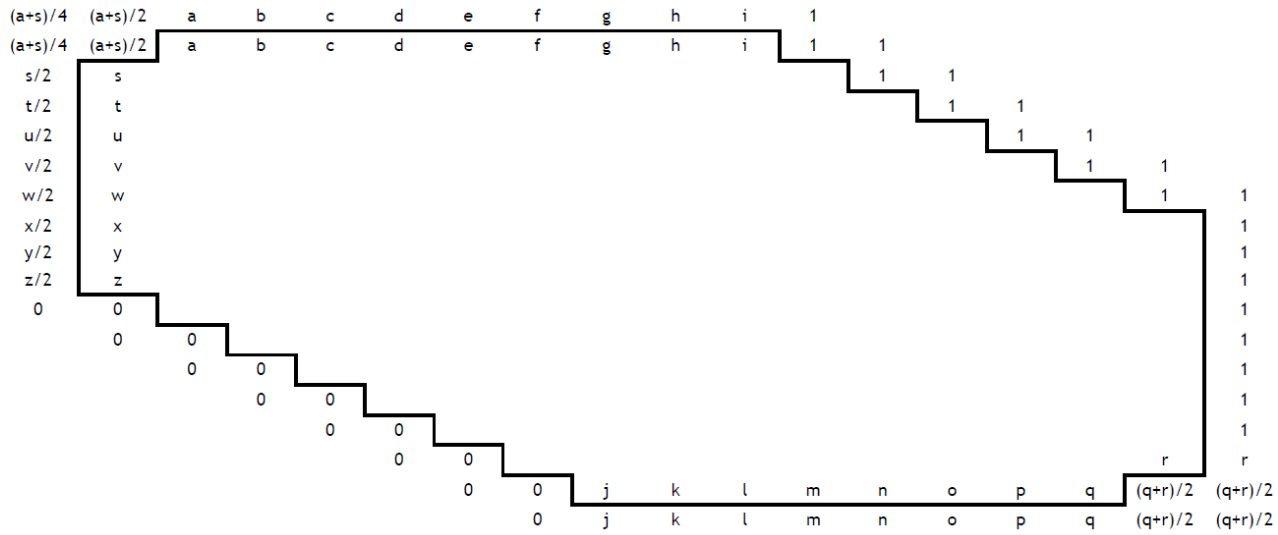


Figure 2.13. Border filter applied to CDMAM readout by Verbrugge [119].

Psychometric fit

A psychometric function relates a physical stimulus to an observer's response [122, 123]. Graphically, the correlation between the probability of correct responses and the physical parameter generally has a sigmoid shape. For a very large stimulus the observer will always be capable to respond correctly, on the other hand, the absence of the stimulus results in a percentage of correct responses at a random guessing level. In between, there is a transition zone where the observer does not respond always correctly, but has an above-chance rate of correct responses. *The threshold is often defined as the stimulus which produces a probability for correct responses halfway the psychometric function* [119].

Mathematically, this can be translated as follows [119, 122-124]:

$$P(x) = \gamma + (1 - \lambda - \gamma)\, p(x) \qquad 14$$

Where, γ= P(0) is the lower asymptote, 1 - λ is the upper asymptote, p(x) is the psychometric function going from 0% to 100% , P(x) is the psychometric function going from γ to 1 − λ and x is the stimulus strength.

If *lapses*, namely errors made to perceptible stimuli, are for simplicity reasons ignored, then λ = 0. The sigmoid character defines the psychometric function as a function of the following form:

$$P(x) \approx \frac{1}{1+e^{-x}} \qquad 15$$

with the threshold and deflection point at x = 0 . P(x) typically represents the percentage of correct observer's responses. In a CDMAM scoring statistically 25% of the responses are correct when the stimulus is not present. Therefore: P(0) = γ = 0.25. Bringing all these elements together, the proposed psychometric curve fitting procedure [117, 125] for the scored CDMAM data can be grasped in equation 16.

$$P(t) = 0.25 + \frac{0.75}{1+e^{-f(c(t)-C(tT))}} \qquad 16$$

P(t) is the probability to detect a disc of a certain diameter and thickness, C(t) $\log(1-e^{-\mu t})$ is the contrast of a disc with a thickness t, $C(t_T) = C_T$ is the threshold contrast which corresponds to a detectability of 62.5% and f is the free parameter to be fitted. Some minor calculations reform the contrast values in the function by gold thickness values. This gives a more practically and easy handling function in equation 17.

$$P(t) = 0.25 + \frac{0.75}{1+\left(\frac{t}{t_T}\right)^{-f}} \qquad 17$$

All the probability-data are fitted to the above presented function by means of the least mean square procedure. The resulting threshold values are deducted from the thicknesses that correspond to a 62.5% correct detection probability [119].

Conclusion

In this chapter the most recent and relevant literature available has been reviewed. The importance of quality assurance and quality control in mammography has been emphasized, particularly in systems employed in breast cancer screening. Most studies and literature reviewed make use of the recommendations of the European guidelines in this regard. The move from FS mammography to digital mammography has generally been accepted. It is also true that the move from FS to DR directly rather than CR mammography is recommendable in terms of image quality and work efficiency in breast cancer screening. From the literature reviewed it is evident that FFDM, using soft-copy image reporting performs better in terms of image quality and cancer detection capabilities when compared to FS and computed mammography. In general, the radiation doses for FS and FFDM mammography are similar whilst doses for CR mammography are usually reported as being slightly larger. Digital detectors are very efficient even with high kVps and filtration, which result in lower radiation doses. It must be noted with care that digital detectors, unlike FS have a wide dynamic range, in which noise improves significantly when higher doses are employed. As a result digital images look better with higher doses. Consequently, from the literature review the importance of DRLs has been highlighted to ensure that doses delivered in mammography, irrespective of the modality employed does not exceed the stipulated levels. The concept of FOM has been introduced and discussed with the aim of inter comparison of digital mammography systems. It is believed that the FOM could prove itself useful for the appropriate selection and optimization of mammography units in breast cancer screening. It is also felt that the FOM could be an essential contributor or performance parameter in the regime of constancy tests for digital mammography.

Table 2.2 DRL studies reviewed showing the study details, the quality assurance protocol employed, the method of dose calculation and the respective MGD calculated.

Study	Patient number	Protocol used	Conversion factors	System Type and info	CBT (mm)	MGD (mGy)		Anode-filter combos (%)			
						30% glan	50% glan	Mo/Mo	Mo/Rh	Rh/Rh	W/Rh
Gennaro et al [51]	800 CC images	EUREF	Dance	FFDM1	50.6 ± 11.73	1.25 ± 0.24	1.37 ± 0.26	11.9	34.6	53.5	NA
				FFDM2	48.6 ± 12.17	1.37 ± 0.55	1.49 ± 0.60	13.8	15.4	70.8	NA
				FFDM3	46.13 ± 13.52	1.57 ± 0.43	1.71 ± 0.47	39.3	23.1	37.6	NA
				FFDM4	46.48 ± 11.78	1.73 ± 0.45	1.89 ± 0.49	53.2	18.2	28.6	NA
Gennaro and di Maggio [52]	300 images	EUREF	Dance	FS	45.3	NA	1.92	41.3	39	19.7	NA
	296 images			FFDM	48.5	NA	1.40	13.9	28	58.1	NA
Moran et al [44]	20137 mammograms	EUREF	Dance	FFDM (CC)	49 ±12	1.80 ± 0.01		27.4	29	43.6	NA
				FFDM (MLO)	54 ± 13	1.95 ± 0.01					

Alizadeh Riabi et al [87]	1145 mammograms	ACR	Wu et al [126]	FFDM (CC)	49.7 ± 10	2 ± 0.70	78.5	21.5	NA	NA
				FFDM (MLO)	58.5 ± 12	2.4 ± 0.78				
Baldelli et al [81]	12110 mammograms	EUREF	Dance	FFDM (CC)	61.4 ± 0.03	2.72 ± 0.04 DRL set at 1.75mGy	NA	NA	NA	NA
Young et al [90]	53218 mammograms	NHSBSP	Dance	CC (FS and digital)	54.1 ± 0.2	1.96 ± 0.01 DRL set a 3.5mGy	58.5	40.0	1.2	0.2
				MLO (FS and digital)	56.8 ± 0.2	2.23 ± 0.01 DRL set at 3.5mGy				
Michielsen et al [91]	NA	EUREF	Dance	All (FFDM and CR)	55.0	2.50	NA	NA	NA	NA
Hendrick et al [127]	49528 patients	ACR	NA	FS (all)	53.0	2.37	NA	NA	NA	NA
				FFDM (all)	54.0	1.86				
Smans et al [80]		EUREF	Dance	All	48 – 58 (patients)	1.67	NA	NA	NA	NA

		DRL set at 2.46 mGy
	PMMA phantom	DRL set at 2.08 mGy

EUREF- European Reference Organisation for Quality Assured Breast Screening and Diagnostic Services; DRL- Diagnostic Reference Level; ACR-American College of Radiology; NHSBSP- National Health Service Breast Screening Programme; CBT- Compressed breast thickness; FFDM-Full-field Digital Mammography; FS-Film-Screen; MGD-Mean Glandular Dose.

CHAPTER 3

MAMMOGRAPHY EQUIPMENT PERFORMANCE, IMAGE QUALITY AND MEAN GLANDULAR DOSE IN MALTA

INTRODUCTION

This chapter was published [4]. The Maltese Islands, including the main island Malta and the smaller island Gozo, hereafter referred to as Malta, cover just over 300 km^2 (120 square miles) in area, with a population of approximately 400,000 people, making it one of the world's smallest and most densely populated countries. On the 1st of May 2004 Malta took its place as a Member State of the European Union (EU). Malta collaborates closely with the IAEA [128] and also with the EU in matters related to medical exposure.

The MNBSP provides free breast screening every three years for all women aged 50 to 60 years. Around 14,000 women are invited for screening each year [8]. Several other private clinics also offer breast cancer screening. The MNBSP includes two state-owned DR mammography units with a third DR unit planned. The private facilities mostly offer breast screening using CR systems.

The objective of this study was to determine the performance status of mammography in Malta and benchmark it against other countries. The main focus was on the evaluation of image quality and MGD using the European guidelines [1, 2]. Other international standards were also used [22, 23] including some manufacturer-specific quality control tests.

It is envisaged that the achieved results will contribute towards the establishment of formal national DRLs for mammography in Malta. The data collection for this survey was performed between September, 2010 and early 2012. It is acknowledged that some private clinics have now updated or replaced their mammography units and started using DR technology.

MATERIALS AND METHODS

Equipment

At the time of this study there were 12 mammography units in clinical use in Malta. Of these, 9 units were available for participation and inclusion in this survey, details of which are summarized in table 3.1. Clinics A, B and C comprise the state-owned mammographic facilities. Clinics A and B form the MNBSP and clinic C is present in the main state general hospital offering symptomatic mammography.

Table 3.1. The mammography units included in the survey.

Clinic Code	Model	Detector type	CR model/Film
A	Hologic Selenia Dimensions 2D	DR	Not applicable
B	GE Essential	DR	Not applicable
C	GE DMR+	CR	Fuji FCR
D	Philips Mammo Diagnost	CR	PCR Eleva
E	Philips Mammo Diagnost	CR	PCR Eleva
F	Siemens Mammomat 1000	CR	PCR Eleva
G	GE Senographe 800T	CR	Fuji FCR
H	Philips MD400	CR	Fuji FCR
I	Siemens Mammomat 1000	FS	Film: Agfa Mammoray HDR

DR – direct digital radiography; CR – computed radiography; FS – Film Screen; GE – General Electric

Tube Output, Tube Voltage, Beam Quality and Automatic Exposure Control (AEC) performance.

Tube output, voltage, beam quality and AEC performance were measured as specified in the European protocol [1, 2]. A Radcal ACCU-PRO radiation measurement system with a dedicated mammography chamber was used for dose measurements [129]. kV measurements were made using an Unfors Xi with a dedicated mammography solid-state detector specifically calibrated to accommodate W/Rh and W/Ag spectra [130]. Where it was possible, repeated readings were taken to reduce experimental uncertainties. The figures presented are mean values including the standard deviation.

Mean Glandular Dose

A parametric method of incident K calculation was used and adapted from a previous study by Robson [6, 131]. The MGD was then calculated using Dance's method through the appropriate conversion factors in equation 7 for each clinical exposure using PMMA phantoms [78, 107].

Image Quality

In this survey image quality tests were divided in two parts: 1) low contrast-detail detectability using the CDMAM phantom and 2) technical image quality using other image quality phantoms. All the phantom images were scored on a dedicated workstation using two 5 megapixel monitors by Barco in suitable ambient light conditions. Prior to any image quality scoring, evaluation of the monitors was performed according to the European protocol with the use of the integrated Barco Medical QAWeb software. The ambient light and monitor brightness were tested using a PTW lux/candela meter. Clinic I was the only FS clinic included in this study, meaning that all image quality assessments were performed on-site on the available view boxes.

Contrast-detail (CD) measurement - CDMAM phantom

Contrast-detail measurements on both CR and DR units were made using the CDMAM phantom (version 3.4, UMC St. Radboud, Nijmegen University, Netherlands). The phantom was positioned with a 20 mm thickness of PMMA above and below the phantom, to give a total attenuation approximately equivalent to 50 mm PMMA or 60 mm thickness of typical breast tissue. This procedure was repeated with small adjustments to the phantom position to obtain a representative sample of 16 images using the AEC. All images were acquired using standard

AEC exposures and with the least processing possible i.e. in their 'raw' format when possible and available. The images acquired were scored automatically using the CDMAM 3.4 analyzer software based on the CDCOM analysis core from the European Reference Organisation for Quality Assured Breast Screening and Diagnostic Services (EUREF) [121].

Technical Evaluation of Image Quality

Technical evaluation of image quality of both digital and FS mammography systems was performed by imaging the TOR MAM and the TOR MAX test objects (Leeds Test Objects, Leeds, UK). The TOR MAX has embedded a high resolution contrast pattern for the measurement of limiting spatial resolution and low contrast circular objects for the measurement low-contrast sensitivity. The TOR MAM is a two-fold test object: one side includes micro-filaments, circular details and simulated microcalcification groups and the other side simulates the appearance of breast tissue with calcification clusters. The TOR MAM test object is useful because the details identified are counted and the respective image is assigned an image quality score [23]. The TOR MAM was imaged under clinical conditions using the standard AEC mode and processed under typical clinical conditions. [25]

Image quality assessment in FS mammography

A sensitometer was used to expose the mammographic film used in clinic I and the optical density of the film step wedge was read to generate a sensitometric curve. This curve was then compared to the manufacturer recommendations to ensure that film and processing are of adequate performance. Measurements of temperature and processing time were performed to establish baseline performance. The image quality in Clinic I was assessed by using the TOR MAM and the TOR MAX phantoms. The image quality parameters assessed included the spatial resolution, the image contrast and threshold-contrast detail detectability. As with the digital systems, an overall image quality score was then generated [23].

Patient Dose Survey

The dose and exposure factors to a total of 759 patients (2812 images) in clinics A, B and C of mean breast compression thickness (BCT) between 5.0 cm and 7.0 cm were collected and analyzed. The data from the two DR systems was extracted using the Digital Image and Communication in Medicine (DICOM) headers and metadata retrieved from the Picture Archiving and Communication System (PACS). The data from clinic C was collected manually with the help of the resident radiographers in the mammography room over a period of four

weeks. Most of the information required for calculating the MGD such as K, tube output and HVL was retrieved from the most recent system survey.

RESULTS

Tube Output, Voltage and Beam Quality.

In general, all the mammography units in this survey had tube outputs, generators and AEC devices that were European protocol compliant. The mean reproducibility of the kV for the broad focus (BF) and fine focus (FF) was 0.07±0.001% and 0.08±0.001% respectively. The mean output reproducibility for the BF and the FF was 0.11±0.001% and 0.17±0.001% respectively. Mean output linearity over a wide range of exposures for the BF and the FF was 3.30±0.04% and 2.04±0.03% respectively. The mean output rate measured for the reference exposure as defined by the European protocol [1, 2], excluding clinic A due to the different anode and filter materials employed, was 14.43±3.50 mGy/s and 3.59±1.50 mGy/s for the BF and FF respectively. Given there is no specification or recommended tolerance for the output rate in the European protocol, reference was therefore made to IPEM 89, [23] recommending an output rate better than 7.5 mGy/s for the BF. There is no reference to FF specifications in either of the mentioned documents.

The output rate measured for system A for a reference exposure using W/Rh as target/filter was 6.73 mGy/s. The output rate measured using the Hologic-recommended method (i.e. 28 kV, 320 mAs W/Rh) was 5.71 mGy/s which is well above the recommended acceptable level of 2 mGy/s or more [132].

AEC performance

With reference to figure 3.1a the DR units (Clinics A and B) in this study resulted in the largest CNR values for the standard breast. In figure 3.1b the average CNR was compared between the CR and DR systems surveyed. In agreement with the literature reviewed [44, 91, 127, 133-136], it is clearly shown that DR mammography yields better overall image quality in terms of CNR at all PMMA thicknesses. In figure 3.1c, the CNR in these DR units was then compared to the similar units in the UK with reference to published technical reports by the NHSBSP [133, 134]. Although the performance of both DR units was very good when based on criteria set by the European protocol some differences in CNR performance can still be observed when

compared to the UK systems. The Hologic Selenia Dimensions 2D (clinic A) performed relatively close to the respective NHSBSP technical report. There is however, a larger difference noted between the local GE Essential and the CNR values published by Young and Oduko, 2012 [133]. Looking at figure 3.1c, it is evident that CNR values for the latter are considerably larger than the values retrieved locally from clinic B, with CNR values becoming closer as PMMA thickness increases. Despite these differences both local and published values tally in the fact that CNR values for the GE Essential are larger than CNR values from the Hologic Selenia Dimensions 2D for all PMMA thicknesses.

The CR systems in Clinics F and G resulted in unexpectedly high CNR values. With reference to figure 3.2, the large CNR value in clinic F may be attributed to the large MGD values when compared to the other systems surveyed. However, clinic G still had low MGD values despite the high CNR at 45 mm and 50 mm PMMA. AEC tests in the European protocol, are based on the evaluation of image quality for one thickness (at the equivalent of 50 mm PMMA) using contrast threshold measurements. At other PMMA thicknesses the CNR limiting value is related to the CNR limiting value at 50 mm PMMA to ensure image quality at other thicknesses. This high CNR achieved in clinic G for the medium PMMA thicknesses caused poor thickness AEC compensation at other PMMA thicknesses in the AEC tests, especially at PMMA thicknesses greater than 60 mm of PMMA. These results are a symptom of an unoptimized AEC device.

Equally notable was the low CNR level in clinic C especially when compared to the other CR systems in this study despite it being appropriately optimized and calibrated. After further investigation it resulted that the users had modified the existing CR reader processing algorithm. The radiologists appeared to have been dissatisfied with the previous image quality claiming that images lacked sufficient contrast. The new optimized algorithm by Fuji Medical resulted in far better image contrast. However, the CNR decreased by approximately 30% due to the large amount noise introduced in the images. This shows that CNR *alone* may not be an appropriate image quality indicator especially when comparing different digital systems.

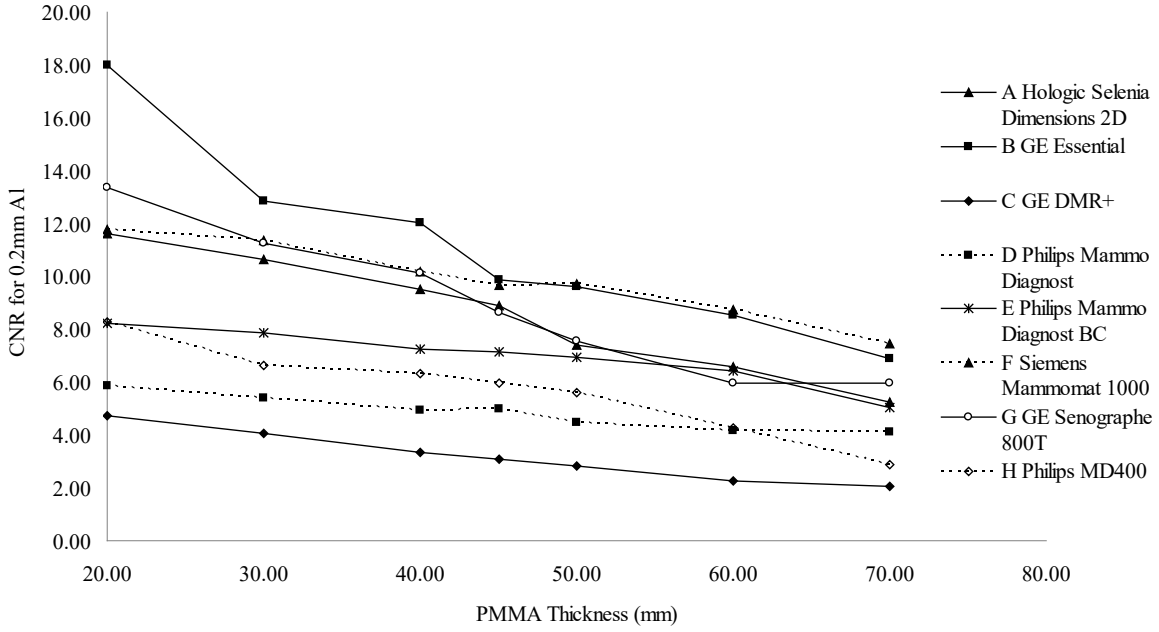


Figure 3.1a. Plot of CNR versus PMMA

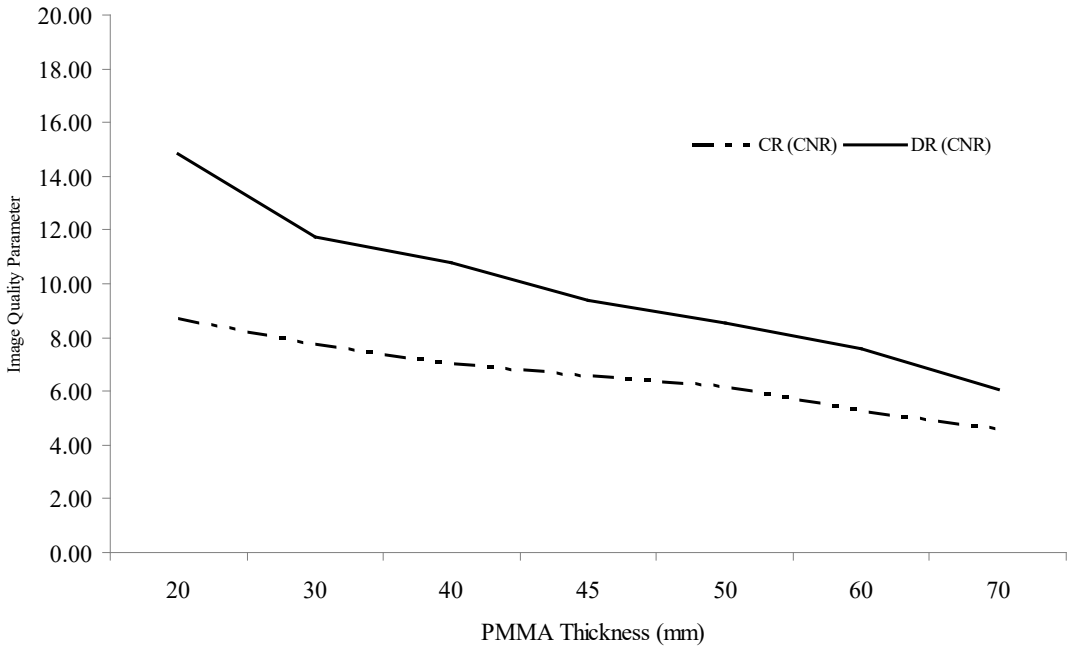


Figure 3.1b. Average CNR versus PMMA for CR and DR.

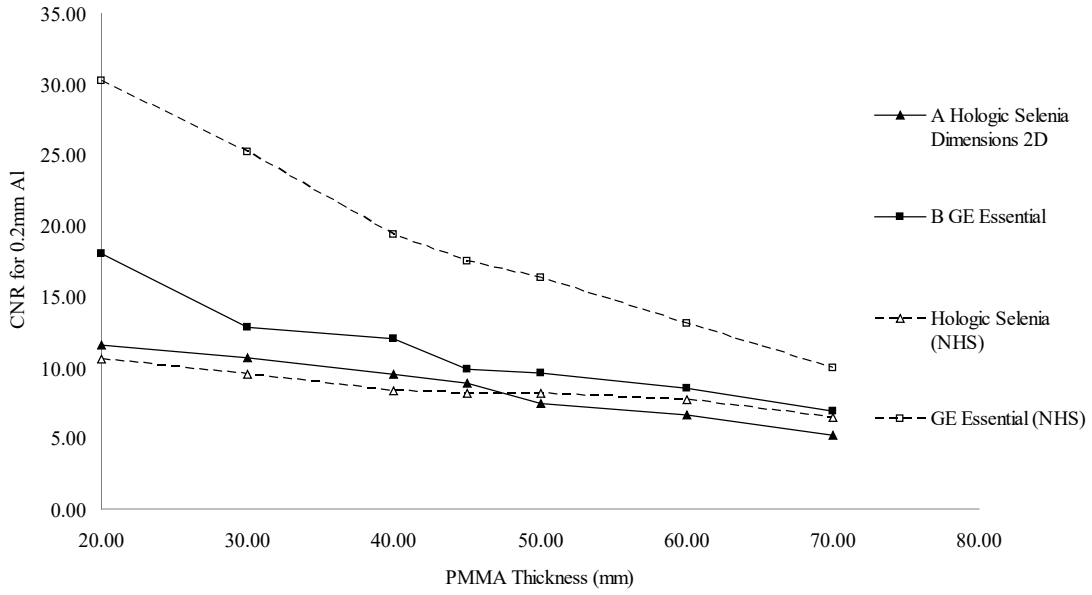


Figure 3.1c. CNR values for DR units in clinic A and B compared to the same units from data published in the UK NHSBS [133, 134].

Mean Glandular Dose

The mean MGD to the standard breast was 1.53±0.65 mGy with values ranging from 0.93 mGy to 2.87 mGy with kV values ranging from 27 to 29. Clinics having a semi-automatic AEC device mostly employed 28 kV with Molybdenum/Molybdenum (Mo/Mo) target/filter when using the AEC. Figure 3.2a shows the MGD as a function of the PMMA thickness under AEC control. This figure also shows the acceptable and the achievable MGD levels set by the European protocol. With reference to figure 3.2b, the average doses between all three technologies are compared showing that the DR units yielded the lowest relative MGD when compared to CR and FS mammography. The recorded overall dose difference between CR and DR at the standard breast was 34% in MGD and increasing to a 62% reduction in MGD at 70 mm PMMA.

It appears that all the mammography units showed MGD levels below the acceptable levels in European protocol except for clinic E which resulted in unacceptably high doses for PMMA thicknesses above 40 mm. The dose to the standard breast in clinics F (1.91 mGy) and H (2.09 mGy) was below the 2.5 mGy acceptable limit and above the achievable 2.0 mGy European protocol limit. With increased PMMA thickness, doses in clinic F were close to the acceptable limit set by the European protocol. The remaining clinics all yielded relatively low MGD values.

It is important to note that when CR mammography systems are properly calibrated, the expected MGD vis-à-vis DR mammography is usually in the range of 50 to 100% larger [23, 24]. Doses that are too low may compromise image quality and conversely high doses may result in unnecessary image quality at the detriment of the patient. Since most CR systems had poorly set up AEC devices, reassessment is recommended once further consideration has been given to optimization of the AEC before any further recommendations can be made.

The DR units in clinics A and B stand out in this survey which yielded low relative MGD at all PMMA thicknesses, particularly with larger breasts. The doses from these DR mammography units were compared to the technical evaluation of the same units by the NHSBSP in the UK. The dose in clinic A i.e. the Hologic Selenia Dimensions 2D with software version 1.4.2 yielded overall MGD values lower than those reported by Young and Oduko, 2012 (Report 1201)[133] except for PMMA thicknesses at 20 mm and 30 mm. The MGD at the standard breast for clinic A in this study was 1.28 mGy as compared to 1.40 mGy in the mentioned report making the local MGD approximately 8.5% lower. With reference to NHSBSP report 0803 by Young et al, 2008 [134] similar findings can be reported for the GE Essential in clinic B. MGD at the standard breast in clinic B was 0.93 mGy, compared 1.01 mGy by Young et al, making the former approximately 8.0% lower. Therefore, the overall trend is that doses reported in the Maltese survey are generally lower than those published for the same units in the NHSBSP. Although both DR systems in this study report doses lower than those in the UK, the difference between the two values is statistically insignificant ($p = 0.538$ and $p = 0.506$ for the Hologic Selenia Dimensions 2D and the GE Essential respectively). With reference to the 'p' value, when you perform a hypothesis test in statistics, a p-value helps you determine the significance of your results. Hypothesis tests are used to test the validity of a claim that is made about a population. This claim that's on trial, in essence, is called the *null hypothesis.* A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis (statistical significance), so you reject the null hypothesis. A large p-value (>0.05) indicates weak evidence against the null hypothesis (statistical insignificance), so you fail to reject the null hypothesis.

Both local and UK MGD values for these two DR units are low when compared to acceptable and achievable levels in the European protocol. However, it is important to point out that in general, the Hologic Selenia Dimensions 2D yielded larger MGD than the GE Essential. From

the NHS reports evaluated and from the current study it appears that the percentage difference in MGD at the standard breast between these two DR units is approximately 20% in favour of the GE Essential.

Contrast-Detail (CD) Analysis

With reference to figures 3.3 and 3.4, and with the knowledge that the best performance lies in system having CD curves towards the left lower end of the graph, it can be seen that the DR units in clinics A and B showed a relatively good performance in the CD analysis, particularly for small diameter object details. The difference in CD performance becomes less evident as the object detail diameter increases. Also notable is that there were CR systems which also demonstrated very high CD capabilities at small object diameters, particularly clinic E which is associated with the already mentioned elevated MGD in this system.

The European protocol includes minimum standards for the image quality of digital mammography based on CD measurements. The method involves the determination of the threshold object thickness visibility of the gold details within the CDMAM phantom. The minimum standards are set to ensure that digital systems are as good as or better than FS mammography. These achievable and acceptable levels have been plotted in figure 3.4. These survey results suggest that although diverse, most mammography units performed within the recommendations made by the European Protocol in terms of CD performance. However all systems exhibited improved CD detectability with increasing detail diameter. Below, the 0.25 mm object diameter the performance of all systems approached the acceptable levels. From figure 3.4 it is seen that system G failed to perform below the acceptable levels requiring further investigation. Consequently, doubt still lies in the high CNR values achieved for clinic G particularly with such poor relative threshold contrast detail detectability. The latter may be totally attributable or rather, may be more significantly related to the low MGD in this clinic. These issues ought to be revaluated once all relevant AEC devices are adequately optimized ensuring that the appropriate thickness compensation is happening in terms of CNR.

From the CD assessment of the CDMAM phantom the image quality figure (IQF_{INV}) was used to present the overall final results. With reference to figure 3.5, the mean IQF_{INV} for the standard AEC exposure for all the units was 106.1$\pm$20.89 with values ranging from 70.49 to 129.72. In agreement with the results of the CD analyses, the highest IQF_{INV} score was observed in the DR

mammography units. Also in agreement with the CD analyses was the fact that clinic E once again scored a very high IQF_{INV} related to the elevated MGD values.

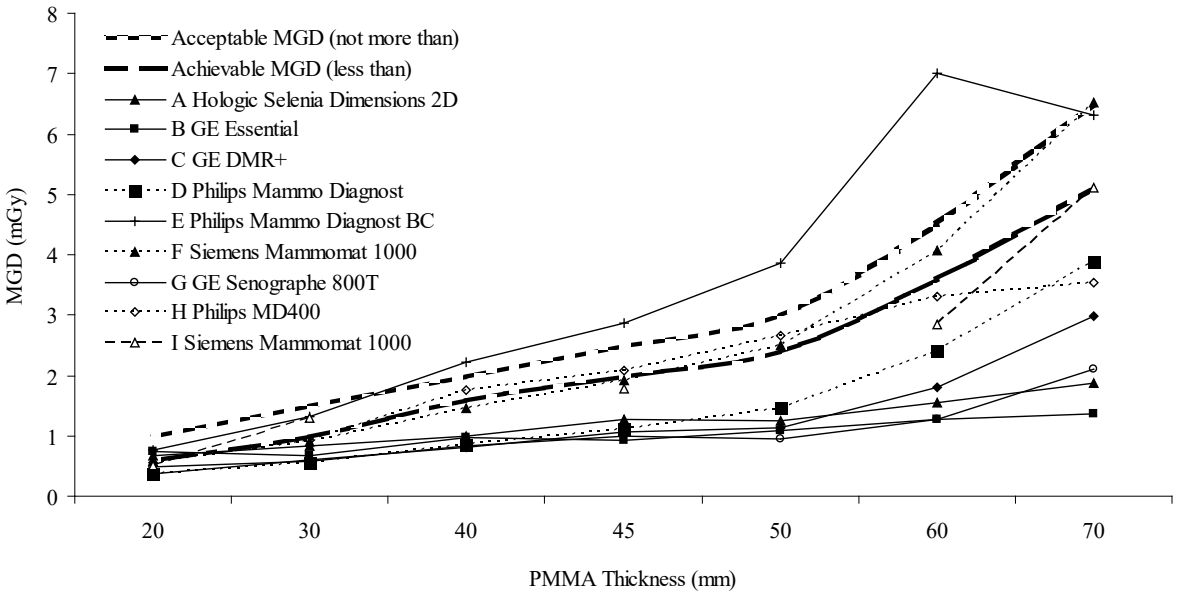


Figure 3.2a. Plot of MGD (mGy) versus breast thickness

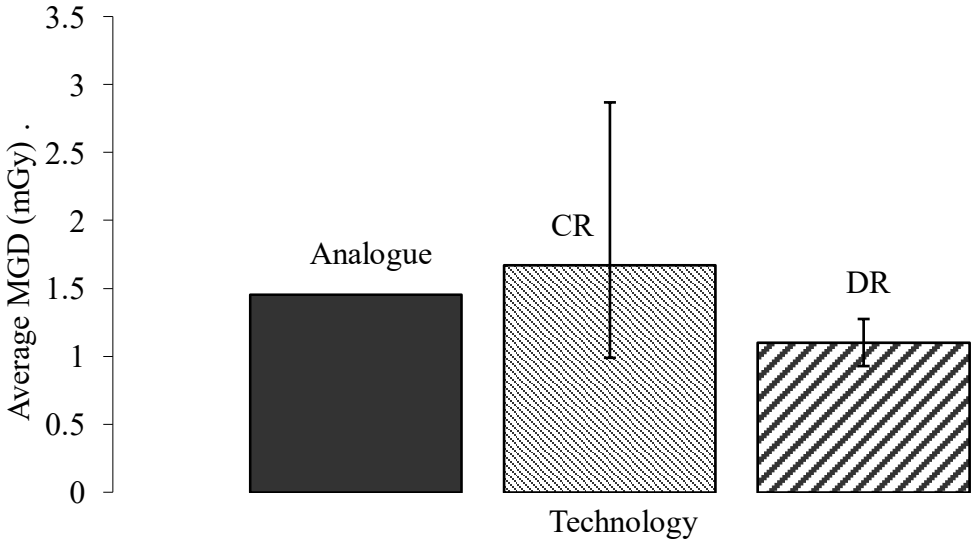


Figure 3.2b. Average MGD for analogue (FS), CR and DR mammography units at the standard breast.

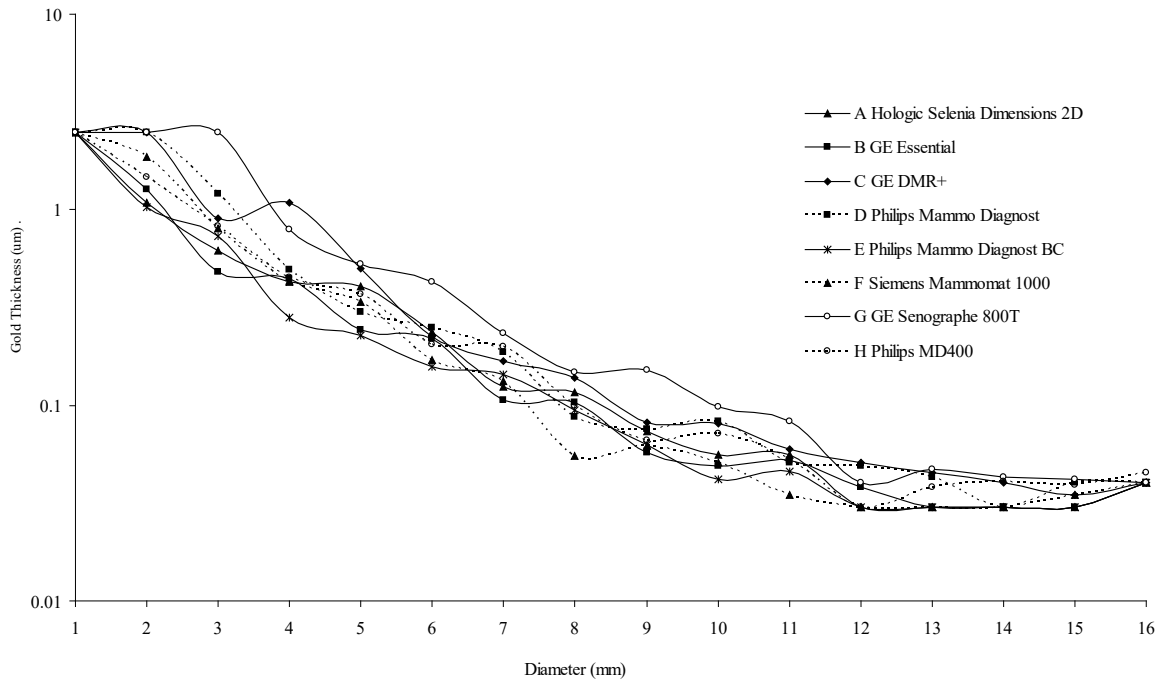


Figure 3.3. CDMAM plot of Gold Thickness versus diameter.

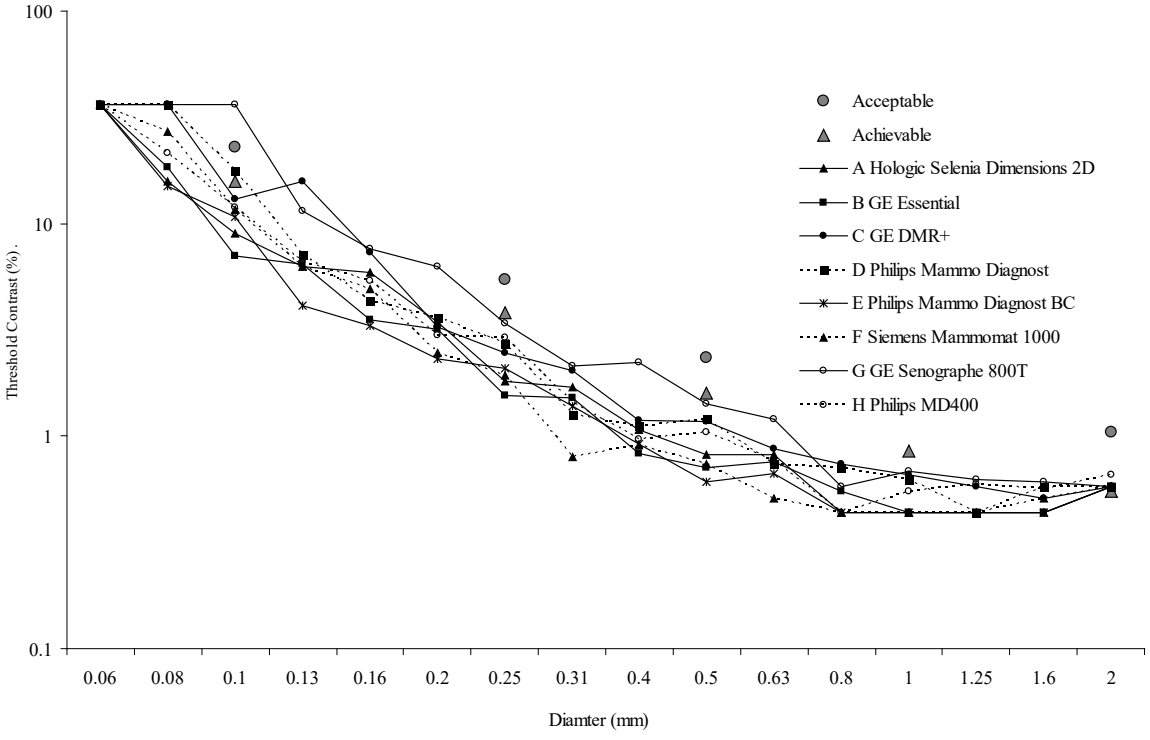


Figure 3.4. Threshold contrast versus diameter

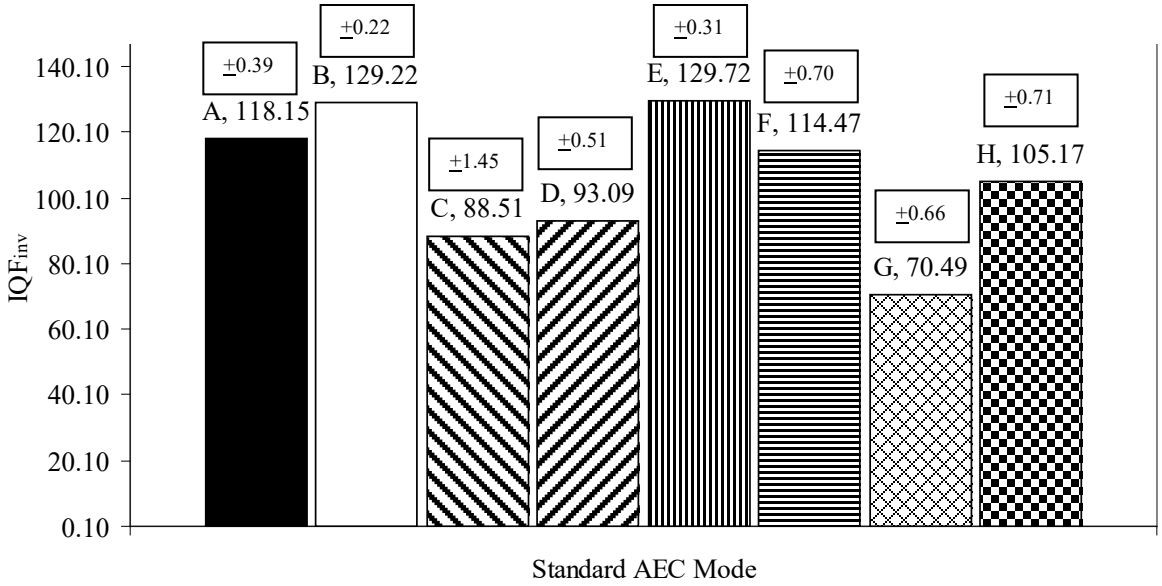


Figure 3.5. Column plot of the IQF_{INV} calculated from the automated CDMAM evaluation software using the standard AEC mode (data label includes the mean value and the text box includes the standard deviation).

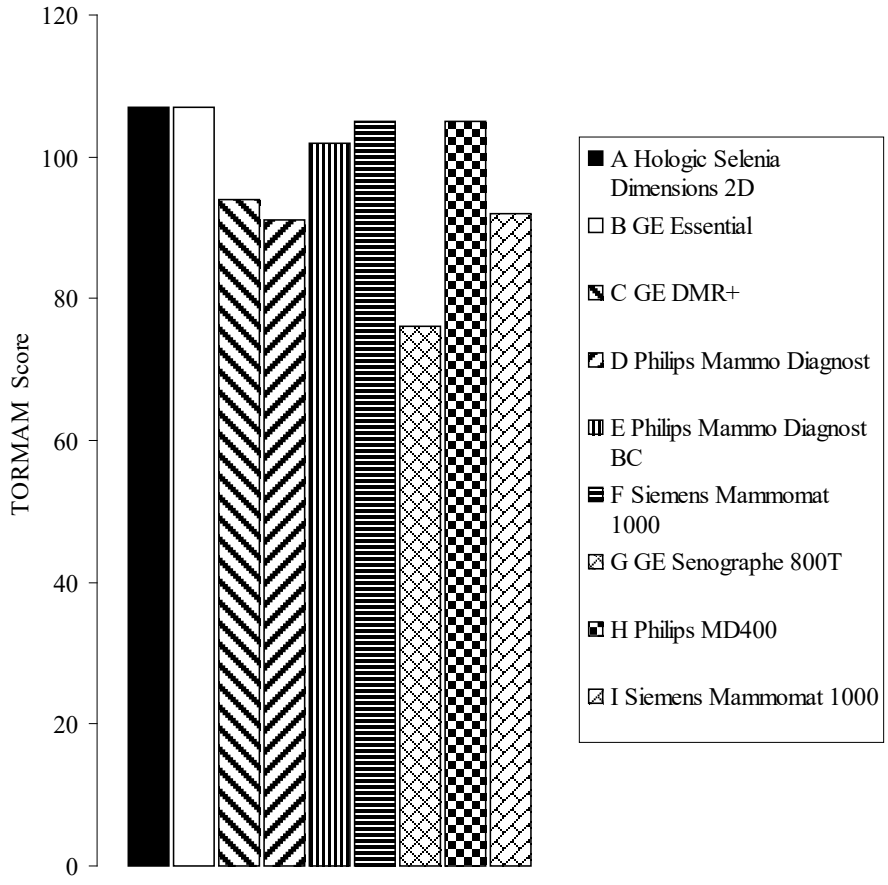


Figure 3.6. Column plot of the TOR MAM Image quality figure using the standard AEC mode.

TOR MAM, TOR MAX and Spatial Resolution

From figure 3.6, the mean TOR MAM image quality score was 98.4±10.85 with scores ranging from 76 to 107. Overall, there seems to be agreement between these results and those achieved by the automated CDMAM scores including the CD analyses and the IQF_{INV}. The most notable conclusion from the TOR MAM analysis was that the DR mammography units yielded the highest score. The unit in clinic G scored the lowest TOR MAM score which as stated earlier is believed to be a consequence of the excessively low MGD.

It is important to note that, unlike FS systems, there is no specification in terms of high contrast spatial resolution in the European Protocol for digital mammography units. However, from the results in this study the spatial resolution using the TOR MAX phantom ranged from 7.1 till 8.0 lp/mm with both DR units scoring 8.0 lp/mm. The spatial resolution score for clinic I was 14.3 lp/mm, which is a typical value in FS systems.[1]

Patient Dose Survey

From the patient cohort, the mean patient age was 57±3.8 years ranging from 42 to 87 years. The screening age for the Maltese population is currently between 50 and 60 years of age and patients of this age are examined exclusively in clinics A and B. Patients undergoing clinical mammography including those falling outside the screening age are examined in clinic C. The distribution of the BCT was normal with an average of 5.75±1.4 cm, with 5.38±1.30 cm for the CC view and 6.34±1.35 cm for the MLO view ranging from 2.0 cm to 8.9 cm BCT for both views combined. The collective mean and the third quartile MGD values were 1.17±0.4 mGy (range 0.47 - 4.29 mGy) and 1.19 mGy respectively.

Results in table 3.2 include patients with BCT between 5.0 cm and 7.0 cm. The data presented in this table and in figure 3.7 showing the distribution of doses, may be considered as the proposed national DRL for mammography for the Maltese population. The 95th percentile MGD value [85, 89, 135] for the MLO views of the selected patients was considered as the DRL since the 75th percentile value for the patient data would yield a too low DRL making it unachievable by most clinics. Consequently, the proposed DRL for the MLO view is *1.87 mGy.*

Table 3.2 Results of MLO views of Maltese Women with BCT between 5.0 and 7.0 cm.

	Number of images	kV ± SD	K (mGy)			MGD (mGy)			
			Mean +CI (Range)	75th Percentile	95th Percentile	Mean +CI (Range)		75th Percentile	95th Percentile
Total	1517	29.28±0.81	4.83±0.06 (2.55 -12.03)	**4.67**	**6.98**	1.07±0.02		**1.11**	**1.68**
CC	935	29.26±0.79	4.79±0.06 (2.55 -9.27)	**4.67**	**6.41**	1.06±0.02		**1.11**	**1.65**
MLO	580	29.30±0.84	4.90±0.10 (2.55 –12.03)	**4.68**	**8.23**	1.07±0.02		**1.11**	**1.87**

CC, cranio-caudal view; MLO, medio-lateral-oblique view; SD, standard deviation; CI, confidence interval for the mean of 95 %.

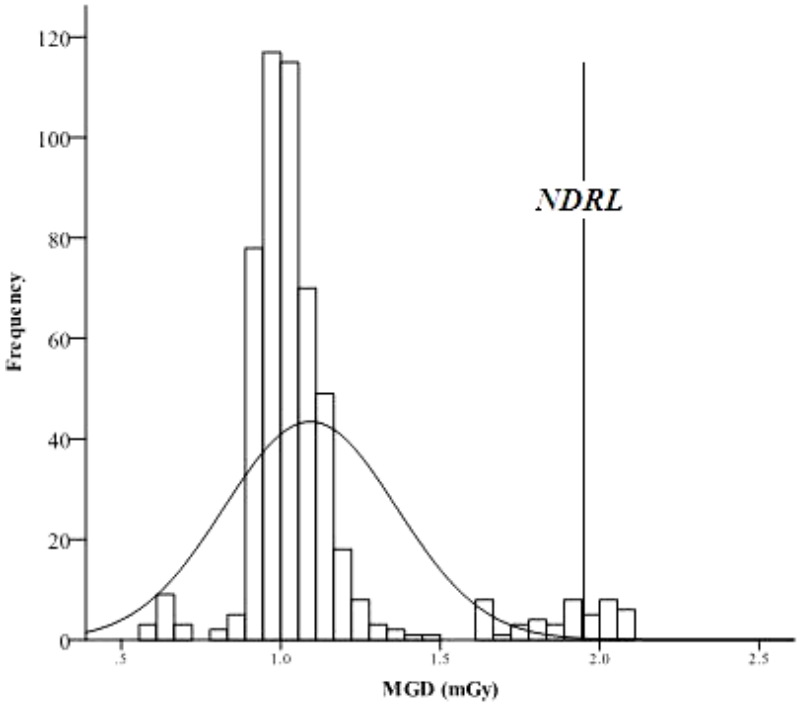


Figure 3.7. Histogram showing the distribution of MGD for the MLO view of patients with BCT between 5.0 cm and 7.0 cm.

Table 3.3 Summary of the K and MGD – Patient data

		Clinic		
		A (DR)	**B (DR)**	**C (CR)**
Mean (mGy)	**K ± SD**	7.94±2.02	4.92±1.05	4.6±3.20
	MGD ± SD	1.9±0.39	1.12±0.33	0.93±0.48
75th Percentile (mGy)	**K**	8.82	5.1	5.3
	MGD	2.05	1.15	1.18
95th Percentile (mGy)	**K**	9.26	5.23	11.83
	MGD	2.11	1.16	2.04

Minimum standards for image quality in Malta.

Based on a similar study in Slovenia [77] and Croatia [25], the survey data from all the mammography units in this study was used to set minimum standards for image quality for mammography in Malta. The same 'worst quarter' principle was adopted to identify those practices with less than optimal performance in terms of image quality. The data for these reference levels hereafter referred to as DRL_{IQ}, was extracted from the tests done using the TOR MAM and the TOR MAX phantoms on all systems and also from the CDMAM phantom results done on the digital units.

The results for the image quality parameters are summarized in tables 3.4 and 3.5. Based on these results the proposed DRL_{IQ} for image quality based on the worst quartile values is: 92 for the TOR MAM image quality score including all mammography units, 91.45 for the CDMAM IQF_{INV} and 5.72 for the CNR at 45 mm PMMA for all units except FS. The third quartile value for the spatial resolution using the TOR MAX phantom was 7.1 lp/mm for digital units and the only FS unit scored a value of 14.3 lp/mm. Using the same worst quartile principle, a DRL is being proposed for the gold thickness detected for the 0.1, 0.25, 0.5, 1.0 and 2.0 mm diameter details on the CDMAM phantom. The respective worst quartile values achieved were 0.71, 0.13, 0.05, 0.03 and 0.04 µm respectively. This equates to the respective following contrast levels: 10.53%, 2.04%, 0.79%, 0.47% and 0.63%.

Table 3.4 Equipment Survey Results

Quantity	Average	75th Percentile	95th Percentile	Standard Deviation	Worst Quartile	Range	Reference Value
K (mGy) (Without backscatter)	7.82	9.21	13.3	3.36	9.21	5.05 - 14.01	10mGy [136, 137]
MGD (mGy) (45mm PMMA)	1.53	1.91	2.56	0.65	1.91	0.93-2.87	2.5mGy [1, 2]
High-Contrast Spatial Resolution (lp/mm)	Digital (CR and DR) : 7.55			Digital (CR and DR) : 0.48	Digital (CR and DR) : 7.1	Digital (CR and DR) : 7.1 – 8.0	No specification available for digital mammography.
	FS (FS) : 14.3			NA	14.3	NA	12 lp/mm[1, 2]
TOR MAM Score	97.66	92	107	10.84	92	76 - 107	
CDMAM Score (IQF_{INV}) (excluding FS)	106.1	91.95	76.8	20.89	91.45	70.49 – 129.72	
CNR to 45 mm PMMA (excluding FS)	7.28	5.72	3.74	2.44	5.72	3.07 – 9.89	

Table 3.5 Descriptive Statistics for Image Quality Parameters

				Mean		Std. Deviation	Variance
	N	**Minimum**	**Maximum**	**Statistic**	**Std. Error**	**Statistic**	**Statistic**
TOR MAM	9	76.00	107.00	97.67	3.46	10.37	107.50
CNR	8	3.07	9.89	7.28	0.86	2.44	5.94
IQF_{INV}	8	70.49	129.72	106.10	7.39	20.89	436.47
Threshold Contrast	9	0.01	0.01	0.01	0.01	0.00	0.00
Spatial Resolution (lp/mm)	9	7.10	14.30	8.30	0.76	2.29	5.27

DISCUSSION

The mammography units

The mammography units investigated demonstrated substantial differences in terms of dose and image quality. These differences are related to the different technologies, age and technical condition of the equipment surveyed.

A good start would be to compare the mean and quartile values calculated to the European protocol reference values of 2.0 mGy and 2.5 mGy as achievable and acceptable values respectively [1, 2]. The resulting MGD DRL from this survey may be considered comparable to the achievable and considerably lower than the acceptable values. With reference to table 3.6, the present MGD values were compared to the literature reviewed. However, direct comparison of doses with other published studies using a phantom approach, was at times difficult since different phantoms were used namely, 4.0 cm PMMA [86, 90, 138, 139], RMI 156 phantom[140] and the ACR phantom[136] as opposed to the 4.5 cm PMMA employed in this survey and other reviewed studies[20, 27, 141]. Nevertheless, in table 3.6 it is clear that the overall mean MGD of 1.53 mGy in this survey is larger than the dose reported in studies listed in table 3.6 except for Serbia, which reported a mean MGD of 1.9 mGy using the standard phantom. The probable reason for the high dose to the standard breast in this survey is related to the use of unoptimized CR systems locally. In fact, the mean MGD from the DR units alone was only 1.10 mGy, making this value conversely lower than the mean dose in most reviewed countries.

Besides the optimization and calibration issues related to AEC devices used with newly installed CR systems, other various problems were noted during the equipment surveys. For instance, Clinic E was equipped with a Rh filter which needed to be inserted manually by the radiographer. However, this was jammed during the mentioned tests; meaning that it's use was inhibited even though its use is particularly important for the thicker and denser breasts. This system also made use of the usual two detector sizes i.e. 18 x 24 cm and 24 x 30 cm. Surprisingly, the larger plate holder did not incorporate an anti-scatter grid. It also became apparent that most radiologists were dissatisfied with the image quality from mammograms of larger breasts at this clinic; something that is attributable to this fact. Other notable problems in this unit were related to the tube and generator. The variation of the output with mAs

exceeded 10% which may be attributed to problems in tube current calibration or timer inconsistencies. The guard timer/security cut-out of this unit did not function properly which may consequently damage the X-ray tube and may also indirectly contribute to an increased patient dose.

The other common problems found in most mammography units were related to light field – x-ray field inaccuracies and also CR-related artifacts. In one particular unit a replacement of the CR plates was recommended due to the quantity of artifacts identified and also due to severe ghosting on the same CR plates. Other less common problems included: out of specification HVL for W/Rh; AEC dose modes calibrated on the same output curve; ambient light >10 lux in reporting rooms; reflections on monitors due to stray sources of light, faulty brakes and inconsistent compression devices.

In agreement with other studies [136], the most valuable conclusion from the survey on the mammography units in Malta using the European protocol is that DR mammography yields lower MGD when compared to CR and FS mammography. DR mammography is also superior to CR based on CNR and threshold contrast detail detectability results, the latter being scored by both human observers and by automated methods.

Patient Data

It is acknowledged that the selected bandwidth for BCT in this study is very wide however these ranges were chosen because of the diversity in BCT between centres. Future evaluations of patient doses in Maltese women should therefore take into consideration clinics in the private sector to give a more representative DRL. Further still, larger population sizes from each respective clinic would be recommendable to make similar national results more significant.

On comparing the total mean BCT values in this study to the standard breast as defined by the European protocol i.e. 5.3 cm it is evident that the latter may not be representative or ideal for the Maltese population. The standard breast is close to the local CC view mean of 5.38 cm but if the DRLs are based on the MLO views then redefinition of the standard breast is highly recommendable since the Maltese MLO view mean is 6.34 cm. Young et al [90] in their national survey on radiation doses received in the UK Breast Screening Programme between

2001 and 2002 found that the average BCT for the women in the UK was approximately 5.7 cm for the MLO views and 5.4 cm for the CC views. Given these results, Young et al also suggest that the current definition of the standard breast in the UK may not be appropriate and needs to be revised. Consequently, the UK's definition of the standard breast with the inclusion of patients with BCT between 5.0 cm and 6.0 cm is also not appropriate to the Maltese population if the MLO view is to be considered for DRL purposes.

With reference to table 3.6 and similar to the limitation in the previous section, direct comparison of the MGD results achieved in this survey with other countries was difficult due to the diversity in BCT and methods of DRL presentation. Nevertheless, the overall interpretation of the results in this survey demonstrated that the mean MGD to the patient population having their mammogram in state-owned facilities was lower than that reported by other similar surveys. Using a percentile approach, both the 75th and the 95th percentile values in this study were lower than the 2.23mGy and 1.96 mGy for the MLO and the CC views respectively in the UK [90] even though the average BCT was larger amongst Maltese women. The doses in this study were also relatively low when compared to more recent studies. Baldelli et al [135] surveyed Irish breast screening patients setting up a DRL at 1.75 mGy. The reported Irish CC and MLO views combined mean MGD was 2.72 mGy and the combined mean BCT was 6.14 cm. To facilitate comparison with the latter, the combined mean MGD and the combined mean BCT in the present study was calculated, resulting in 1.07 mGy and 5.75 cm respectively for Malta.

Baldelli et al [12], made use of mammography units similar to those present in clinics A and B. In agreement with the present study the average BCT recorded from the Hologic system was larger than that for the GE Essential. Nevertheless, MGD values for the Hologic were conversely lower than those from the Essential. In the present survey the Hologic system always appeared to yield the highest patient dose when compared to the Essential and it is believed that this might be attributed to the larger average BCT of patients attending clinic A. However, the need for further optimization in clinic A cannot be excluded. In their survey, Baldelli et al state that the differences between the Essential and the Hologic systems are insignificant since the overall mean MGD were close: 3.03 mGy and 2.91 mGy per examination respectively.

Smans et al [80] reported a mean MGD of 1.69 mGy from 80 Belgian centres in 2005 also combining both CC and MLO views. The DRL set by Smans et al for the 95th percentile value for the combined CC and MLO views was significantly higher than the Maltese value in table 3.2, reported being 2.46 mGy compared to 1.68 mGy in the present survey. In contrast, the proposed DRL for 45 mm PMMA in the present study exceeded the DRL presented by Smans et al. This discrepancy may be indicative that failure to include the private clinics in the patient survey might be underestimating the patient DRL value in terms of MGD.

Another Belgian study made in 2008 by Michielsen et al [91] reporting a 95th percentile value of 2.37 mGy and an overall average MGD of 1.69 mGy. These values are relatively close to the Maltese values. In relation to the study by Michelsen et al, current proposed DRLs exceed the Belgian values at the 95th percentile value, however the overall mean Maltese MGD values are lower. The most recent reviewed publication by Hendrick et al [127] in a screening trial using data from the American College of Radiology screening centres reported a total mean MGD, irrespective of the view, of 1.86 mGy for a BCT of 5.38 cm. In comparison to this Maltese survey, MGD values are significantly lower locally although BCT are larger in relation the study by Hendrick et al.

CONCLUSION

The results of this survey provide valuable information on the current status of mammography units in Malta. The results also indicate the success of the newly established MNBSP in terms of dose and image quality to the patients. Most of the success of the programme is clearly attributable to the use of DR mammography and to regular quality assurance procedures. Some of the systems surveyed lacked regular quality assurance and appropriate optimization making their performance at times inadequate. This is concerning and worrying especially if these units are employed for screening mammography. The results of this study also challenge the effectiveness of CR mammography especially with the increased use of DR in breast cancer screening. Although initial costs associated with DR mammography are high, long-term benefits in terms of dose and image quality are obvious in view of the results achieved in this study. Consequently, it is understandable why any new mammography system introduced by the NHSBSP in the UK providing breast screening should now be DR rather than CR.

Table 3.6 A comparison of this study with other studies.

Data Source	Technique Used	MGD calculation Method	No. of Women	Mean BCT (mm)	MGD standard breast MGD (mGy)	MGD per view (mGy)	MGD per woman (mGy)
Heggie, Australia [142]	estimation	Wu et al	490	52		median: 2.17	median: 4.42
Thiele et al, Australia [136]	estimation	Wu et al		42	1.16[d]		mean:4.6
Eklund et al, Sweden [143]	estimation	Rosenstein et al	1350	50		mean: 1.25	
			1596	54			mean: 0.68
			1496	62			mean: 0.76
Moran et al, Spain [144]	estimation, TLD	Dance	350	52	1.3[a]	mean: 1.6	
Moran et al, Spain [44]	estimation	Dance	5034	52		Total mean: 1.88	
				CC: 49		mean CC: 1.8	
				MLO: 54		mean MLO: 1.95	
Klein et al, Germany [145]	estimation	Klein at al	1678	55.9		mean: 1.59	
			945	50.8		mean: 2.07	
Gentry and Deward, USA	TLD	Wu et al	4400	CC: 45		mean 1.49	

[146]							
Young et al, UK [90]	estimation	Dance	16, 505	CC: 54.1 MLO: 56.8	1.42[a]	mean CC: 1.96 mean MLO: 2.23	
Jamal et al, Malaysia [140]	estimation	Wu et al, Dance	300	CC: 37.5 MLO: 44.5	1.23[b]	median CC: 1.44 median MLO: 1.65	median CC: 3.37 median MLO: 3.21
Koustic et al, Serbia [27]	estimation	Dance			mean 3 centres: 1.68[c]		
Ciraj-Bjelac et al, Serbia [141]	estimation	Dance	53	CC: 49.5 MLO: 56	1.9[c]	mean CC: 2.8 mean MLO: 4.3	
Smans et al, Belgium [80]	estimation	Dance				mean: 1.67	
Michielsen et al, Belgium [91]	estimation	Dance				mean: 1.69 median: 1.48	
Faj et al, Croatia [25]	estimation	Dance			1.1[c]		
Vassileva et al, Bulgaria [20]	estimation	Dance			range: 0.35 - 3.47c		
Baldelli, Ireland [135]	estimation	Dance	2910	61.4 CC: 60.5 MLO: 63		mean: 2.72 mean CC: 1.27 mean MLO: 1.34	

Tsapaki et al, Greece [147]	estimation	Dance	250	42		mean: 1.4 mean CC: 1,2 mean MLO: 1.5
Alizadeh et al, Iran [87]	estimation	Wu et al	298	CC: 49.7 MLO: 58.5		mean CC: 2.0 mean MLO: 2.4
Present study, Malta	estimation	Dance	759	57.5 CC: 53.8 MLO: 63.4	1.53[c]	mean: 1.07 mean CC: 1.06 mean MLO: 1.07

a 4cm PMMA; b RMI 156 phantom; c 4.5cm PMMA; d ACR phantom 4.2cm PMMA

CHAPTER 4

TO DETERMINE THE DIFFERENCES BETWEEN THE DISPLAYED DOSE VALUES FOR TWO FULL-FIELD DIGITAL MAMMOGRAPHY UNITS AND VALUES CALCULATED USING A RANGE OF MONTE CARLO-BASED TECHNIQUES:

A PHANTOM STUDY

Introduction

This chapter was published [6]. Modern DR units now display the MGD and the exposure to the breast following each examination. This information is usually displayed with each image and is included in the DICOM data for each patient. Of particular interest is how well the displayed dose values compare to the values calculated using well-established Monte Carlo-based methods, for instance those that have been recommended in the European protocol [1, 2] or other reputable publications [78, 97, 107, 126].

The purpose of this work was to compare the displayed doses to those measured and calculated using these well-established methods by Wu, Dance and Boone [78, 97, 107, 126, 148, 149] for a range of exposures to PMMA. This study was also useful to determine how the MGD calculated using the three Monte Carlo-based techniques compared to each other. This work also attempts to identify correlations between the values determined from the three methods and devise models from which users can use the displayed MGD to predict the calculated values from Wu's, Dance's or Boone's methods [78, 97, 107, 126, 148-150].

Materials and Methods

The DR units evaluated for the purpose of this study were the General Electric Senographe Essential (hereafter referred to as the Essential) and the Hologic Selenia Dimensions 2D (hereafter referred to as the Hologic). The performance of these mammography units was evaluated by applying the European protocol [1, 2].

MGD and K

The MGD values calculated and the K values measured were compared and correlated to the system-displayed values for both the Essential and the Hologic under different specified conditions. K as defined in the European protocol [1, 2] was measured using a Radcal [129] ACCU-PRO radiation measurement system with a 10 x 6 dedicated mammography chamber [129]. The ion chamber was positioned at a distance of 4 cm from the chest wall edge and 4 cm high from the breast support with the compression device inserted halfway between the focal spot and the detector [1]. The former position is close enough to the chest wall edge of the breast support in order to avoid variation due to the anode heel effect but still ensuring full chamber coverage. The elevation of the chamber ensures that no backscattered radiation is being included in the measurements.

A Parametric method for the determining output and HVL for W/Rh and W/Ag system.

A parametric method of establishing the mammography unit tube output and HVL was used and based on a previous study by Robson [131]. This method was necessary for the estimation of K for any possible tube voltage, mAs, target/filter combination and breast compression thickness. To date, this method has never been adapted for systems with a W tube using Ag filtration.

A mammography ionization chamber was placed at a tube-chamber distance of 500mm on the Hologic system. K was measured at tube voltage values ranging from 25 to 32 using a fixed mAs value, with the compression paddle inserted. For each tube voltage setting the HVL was calculated using 99.9% purity Aluminum placed directly on the compression paddle. The K measured was expressed as the output i.e. by normalizing the dose measured per unit mAs. Based on the theoretical relationship in equation 18 that:

$$Output = A(tube\ voltage)^{n} \qquad 18$$

where A and n are constants, the output was plotted against the tube voltage and a power trendline was applied to the plot using Microsoft Excel as shown in figure 4.1a. The same program was then used to plot tube voltage with the HVL as shown in figure 4.1b. A second order polynomial trendline was then applied with reference to equation 19.

$$HVL = a(tube\ voltage)^2 + b(tube\ voltage) + c \qquad 19$$

where a, b and c are constants. In figure 4.1c, the logarithm of the calculated air kerma was plotted against the logarithm of the tube voltage for the W/Ag target/filter selection.

Figure 4.1a. Plots for the W/Ag target/filter combination on the Hologic: Output (mGy/mAs) against tube voltage.

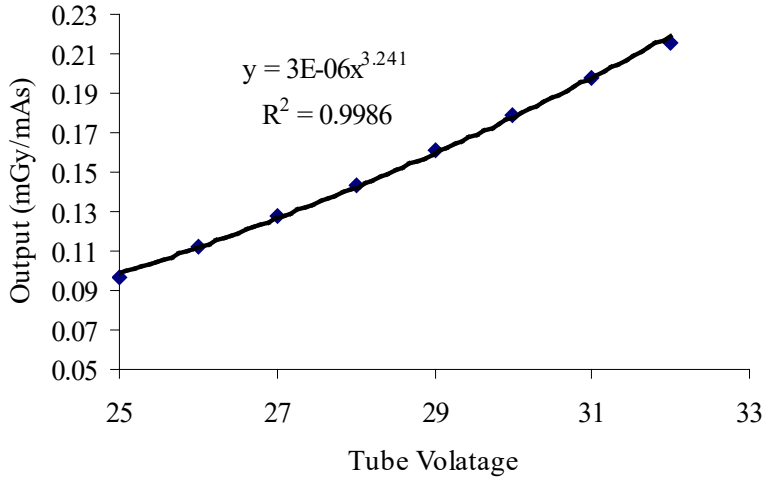


Figure 4.1b. HVL (mm Al) against tube voltage

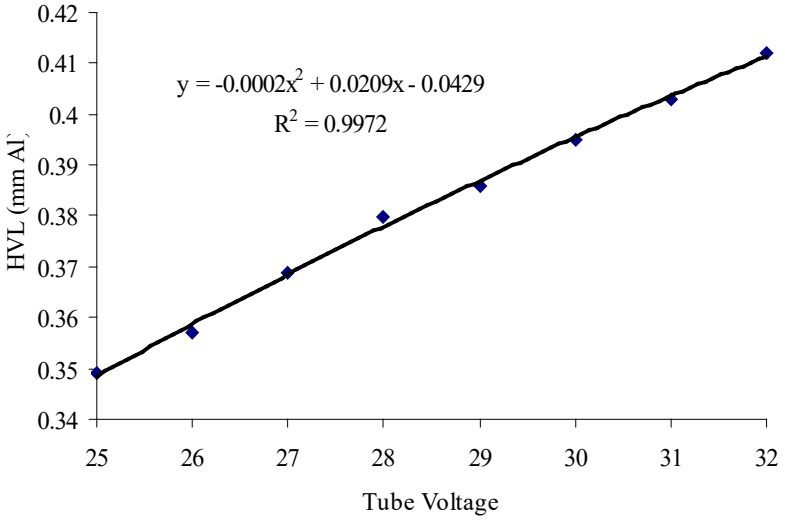
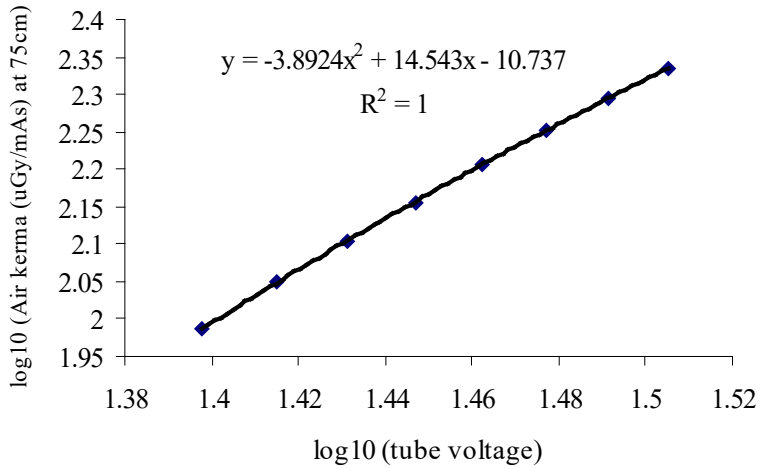


Figure 4.1c. $\log_{10}$ (Air kerma (µGy/mAs) at 75 cm) against $\log_{10}$ (tube voltage).



The parameters for the output and the HVL predictions for different tube voltage settings for the W/Ag target filter combination are listed in table 4.1.

Since all MGD Monte Carlo-based methods used in this study require knowledge of K, using various tube voltage, mAs, target/filter combinations and breast thicknesses, Robson's method [131] and the described adaptation was used. With reference equations 18 and 19, the required *n, a* and *b* factors for the Hologic system were eventually calculated for the W/Ag spectrum. Using equations 18 and 19, and knowing the tube output and HVL at 28kV, the values for *A* and *c* can be determined using the parameters in table 4.1. With all the parameters now known, the tube output and the HVL can be calculated for any other measured value of tube voltage in the range 26 to 32 kV for the Hologic system using W/Ag. The objective of this process was to facilitate the calculation of K for all the exposures required in this study. The adaptation of Robson's method for K estimation will be useful for any future reference, especially in quality assurance procedures for system that make use of similar spectra.

Table 4.1. Calculated values of the constants n, a and b for the W/Ag

Target	Filter	n	a	b
Filter	thickness			
W/Ag	35µmAg	3.1521	-0.0009	0.0733

MGD calculation dose models

A number of different algorithms have been published to estimate the MGD for mammography. Dance et al [78, 107] , Wu et al [126, 148]) and Boone [149]calculated their conversion factors from Monte Carlo simulations on a semi-circular or a cylindrical breast phantom. Each of these algorithms embodies slightly different assumptions about the anatomy and image formation process in the associated Mote Carlo simulations, for instance the subcutaneous fat surrounding the breast. The conversion factor of Dance et al. were developed to take into account the various breast thicknesses (g-factor), mass densities (c-factor), target/filter materials (s-factor) and recently the projection angles in breast tomosynthesis (t-factor). These factors are provided as a function of thickness and beam quality expressed in terms of HVL. Wu et al tabulated the normalised glandular dose, DgN, for different breast densities and added tables for different target/filter combinations.

Method 1 (Dance's method)

In the UK, the most well-established method of MGD calculation is the method by Dance et al [78, 107]. In this method the MGD is calculated using the appropriate conversion factors in equation 7 for each clinical exposure using PMMA phantoms. This method in association with Robson's parametric method was implemented by using a Microsoft Excel sheet developed in-house.

Method 2 (Wu's method)

Wu's method [126, 148] includes the application of normalized glandular dose values per unit incident air kerma (DgN) and the K associated with any particular mammogram. The DgN values are essentially the roentgen-to-rad conversion values, calculated for the "at risk" glandular component of the breast [149]. Wu's method of MGD calculation is based on equation 20 as follows:

$$MGD = DgN * K \qquad 20$$

Wu's method was implemented by using the parameterization algorithm present in the paper by Sobol and Wu [150] in a Microsoft Excel sheet. The algorithm consists of three separate functions one for each target/filter combination namely: Mo/Mo, Mo/Rh and Rh/Rh. There are no data provided for other spectra, namely W/Rh and W/Ag in any of the original mentioned papers by Wu et al [126, 148, 150]. Consequently, this method could not be used to calculate the MGD for the Hologic unit included in this study.

Method 3 (Boone's method)

Boone's method [149] includes the use of Monte Carlo simulated and interpolated lookup tables of DgN values with extended utility including values for much thicker breasts and including spectra like W/Rh and W/Ag which are essential in MGD calculations for the Hologic system. The data tables provided by Boone [149] were used to produce a comprehensive dataset for all the glandular fractions employed in this study. Based on Boone's suggestion, DgN values may be linearly interpolated from 0% to 100% from the glandular tables. As with Wu's method, the data from the processed lookup tables were parameterized and incorporated into a Microsoft Excel sheet.

Relevant parameters were kept constant for all methods of data calculation and data collection. Breast thickness and glandularity were based on the assumptions listed in table 4.2 adapted from Dance et al [78, 107] as follows:

Table 4.2. Simulation of typical breasts for women aged 50-64 years using blocks of PMMA. Surface layers of 100% adipose 0.5 cm thick are assumed.

Perspex Phantom thickness (mm)	Equivalent breast thickness (mm)	Glandularity for age 50 to 64 (%)
20	21	97
30	32	67
40	45	40
45	53	29
50	60	20
60	75	9
70	90	4
80	103	3
85	109	3

Displayed doses compared to the calculated MGD and measured K values.

For the purpose of this study, the displayed doses and the calculated MGD and measured K values were compared for PMMA thicknesses including 20, 30, 40, 45, 50, 60 and 70 mm. For each PMMA thickness the addition of expanded polystyrene blocks as 'spacers' in between

each block were used to reach thicknesses equivalent to normal breasts as specified in table 4.2. This is particularly important since the AEC exposure parameters on both systems are dependent on the breast compression thickness detected. For each specified thickness the displayed and the calculated MGD and the measured K were recorded.

Exposures with varying PMMA thickness were done using the standard (STD) automatic exposure control mode on the Essential while the *'autofilter'* option was employed on the Hologic since this offers fully automated exposure parameter selection. Manual measurements using predetermined fixed mAs values were also taken for tube voltages between 26 and 32 kV in steps of 1 kV for all target/filter combinations on both systems using a fixed PMMA thickness of 45 mm (also referred to as the standard breast). Dose values were also recorded and compared for the standard breast, keeping tube voltage constant for all the target/filter combinations available.

MGD calculation Models (Regression Models)

In statistics, linear regression is an approach for modeling the relationship between a scalar dependent variable *y* and one or more explanatory variables (or independent variables) denoted *X*. The case of one explanatory variable is called *simple linear regression*. For more than one explanatory variable, the process is called *multiple linear regression* [151]. A multiple linear regression analysis was performed on the data collected with the aim of defining models by which the displayed MGD values could be converted with ease to Dance's, Wu's and Boone's calculated values. The predicted values were also compared to the actual or calculated values to confirm the reliability and validity of the models presented.

Results and discussion

Data analyses was done using IBM SPSS Statistics version 19. The data were initially analyzed for their normal distribution using the Kolmogorov Smirnov test; this to determine the best suited statistical test. It was concluded that there is a very good normal distribution of the collected data including the displayed dose values, the calculated MGD and measured K values for both systems. Consequently a correlation analysis using the Pearson correlation coefficient r method was used to determine the relationship between the displayed and calculated MGD and the displayed and the measured K values. In statistics the Pearson correlation coefficient is a measure of the linear correlation (dependence) between two variables *X* and *Y*,

giving a value between +1 and −1 inclusive, where 1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation. It is widely used in the sciences as a measure of the degree of linear dependence between two variables [152].

In general, the results indicate a very strong positive correlation between all dose values ($r>0.98$) for both the Essential and the Hologic. Further still, all the results achieved can be considered statistically significant since the p values for all the readings in both systems are much less than the 0.05 level of significance.

Albeit the strong correlation, with reference to figures 4.2-4.5, the results also show that in most cases there is a noticeable difference between the displayed and the calculated dose values. There are also notable differences in values between the Monte Carlo methods of MGD calculation employed which is probably more significant than the difference between the displayed and each Monte Carlo method for both systems. Although the results achieved from this study strongly suggest that the displayed values from both the Essential and the Hologic may be confidently used for dosimetric purposes, an algorithm that would enable medical physicists to easily convert the displayed MGD to values calculated using these other methods [78, 97, 107, 126, 148, 149] may be useful and very convenient when and if needed.

The Essential

In figure 4.2a the displayed MGD on the Essential appears greater than that calculated trio of MGD methods for all PMMA thicknesses. The difference between all dose values seems to increase with increased PMMA thickness. Figure 4.2a indicates a similar trend in the dose curves, particularly noticeable at the standard breast showing a plateau in MGD probably related to a change in the target/filter selection i.e. from Mo/Rh to Rh/Rh. The diversity in doses is also observable when looking at the percentage differences in figure 4.2c. The latter represents a graph comparing the percentage difference between each method and the displayed MGD value. The collective mean represents the arithmetic mean of all the Monte Carlo methods used and may be considered as representing the 'true MGD value'. From this figure it can be seen that for PMMA thicknesses above 40 mm, Wu's method appears being the closest to the displayed MGD given the least percentage difference. For thinner PMMA thicknesses, Dance's method resulted in the least percentage difference. When the Monte Carlo methods are inter-compared it appeared that all methods yielded similar results from

thin to medium sized-breasts. However, for PMMA thicknesses above 45 mm, the difference between each method appeared being more pronounced.

Based on correlation analysis, the displayed MGD correlated very strongly with all the three calculation methods across the whole range of PMMA thickness ($r>0.98$). The best correlation resulted with Boone's method with a Pearson r value of 0.995. This strong relationship is reflected in the scatter plot in figure 4.6, showing a line of linear regression between the displayed MGD and Boone's method with an R^2 value of 0.99.

This system displayed dose *overestimation* is also visible on comparing the displayed MGD with the calculated values with increasing tube voltage in figure 4.3a. On face value, the curves are relatively identical in shape also having a strong positive correlation with Pearson r values greater than 0.98 with all three Monte Carlo methods. The percentage difference between the displayed and the Monte Carlo methods with changes in tube voltage may be seen in figure 4.3c. For the range of tube voltages assessed it appeared that the largest percentage difference noted was with Dance's method (>18%) at the highest tube voltage. When looking at the collective mean it is very clear that Monte Carlo methods are closer to the displayed MGD values at lower tube voltages. As the latter increased, the percentage difference with the displayed values increased.

With reference to figures 4.2b and 4.3b it also appears that the system displayed K values are larger in magnitude to those measured both with changes in PMMA thickness and with change in tube voltage. It must be noted that the K values in this study do not account for any backscattered or forward scattered radiation contributions to the dose readings as recommended in the European protocol. Backscatter factors are dependent on the mammographic spectra and the phantom thicknesses employed and may contribute to anywhere between 8% to 15% increase in the dose [153]. Forward scattered radiation for typical dose measurements from the compression paddle is also often neglected. Forward scattered radiation may contribute to anywhere between 2% and 10% in dose depending on the thickness of the paddle and the HVL [154]. As a result it is not clear whether the displayed K values are taking these factors into account. From the results in figure 4.3c, the percentage difference calculated between the displayed and the measured K ranged from 2% till 9.5% with measured values being constantly smaller in magnitude when compared to the displayed

values. This difference may be due to the lack of inclusion of scatter contributions in the measured values.

In agreement with the MGD results, in figure 4.2b the difference between the displayed and the measured K values appears to increase with increased PMMA thickness, with the maximum difference recorded at 70mm PMMA. Nevertheless, both values still correlate remarkably well (r >0.99). Further still, the scatter plot in figure 4.8 shows an R^2 value for the line of linear regression of 0.999, showing the strong predictability between these two variables.

Unfortunately, there seemed to be no obvious correlation or statistical differences between the data with changes in target/filter. All dose methods appear to display a trend in the data where the move from Mo/Mo to Mo/Rh to Rh/Rh corresponds to decrease in MGD. The shift between target/filter combinations contributed to an expected reduction in MGD, anywhere between 4 to 10%. However, this was not the case in the displayed values. When shifting from Mo/Rh to Rh/Rh the MGD displayed remained constant at 0.83 mGy. Wu's and Boone's method appear to be closer to the displayed MGD, particularly, Wu's method at all three target/filter combinations. Dance's method appears to yield the lowest MGD estimate when compared to all other methods.

The Hologic

The MGD results for the Hologic system with increased PMMA thickness are shown in figure 4.4a. The displayed values lie between Dance's and Boone's values. Similarly to the Essential, it appears that all dose values have a strong positive correlation (r>0.99) and that the disparity between all three readings increases with increased PMMA thickness. Also in agreement with the Essential, the best correlation resulted between the displayed and Boone's method with a Pearson r value of 0.99 also yielding a line of linear regression having an R^2 value >0.99 in figure 4.7. With reference to figure 4.4c and relatively similar to figure 4.2c, the percentage difference between the Monte Carlo methods and the displayed MGD and also between the Monte Carlo methods themselves seems to increase with increased PMMA thickness. Also notable, and also in common to both figures 4.2c and 4.4c, is the fact that Dance's method rendered the largest MGD figures as the PMMA was increased. Nevertheless, using the collective mean approach, it is clear that both Monte Carlo methods employed resulted in

dose values with an overall percentage difference below 10% for the Hologic and below 18% for the Essential when compared to the displayed values.

With reference to figure 4.5a, the variation in MGD values with tube voltage using fixed mAs values seemed to be more complex to interpret. The displayed MGD was lower than Dance's method from 25 to 28 kV but eventually exceeded Dance's value above 29 kV; this difference appeared to increase with increasing tube voltage. Boone's method always appeared to yield the largest MGD estimate when compared to Dance's and displayed values. Also from figure 4.5a the displayed MGD values approach Boone's values as the tube voltage increases. Nevertheless, from the Pearson correlation analysis both Boone's and Dance's methods correlated similarly well with the displayed value ($r > 0.99$).

From figure 4.5c, the percentage difference between the displayed and the Monte Carlo methods appears to be at its largest (>15%) at low tube potentials. From the collective mean curve in figure 4.5c, it appears that in the Hologic system, Dance and Boone MGD values appear closer to the displayed values at medium to high tube potential values as opposed to the Essential were Monte Carlo values were closer to the displayed MGD with lower tube potentials.

With reference to figure 4.4b and in contrast with figure 4.2b the displayed K, despite being close, appears to be smaller than the measured K for the Hologic with all the range of PMMA thicknesses; however the trends in the dose curve appear identical. Once again it is unknown whether Hologic have included any backscatter or forward scatter contributions in their K displayed values. With reference to figure 4.5b and in agreement with figure 4.4b, it is evident that the measured K values are once again larger than the displayed K values with change in tube voltage. It is important to note that the opposite was noted in the Essential, whereby measured values were lower in magnitude than the displayed. Despite the mentioned differences, the measured and the displayed K values again demonstrated a strong positive correlation ($r > 0.99$) as seen in figure 4.9.

As with the Essential, no conclusive or interesting results can be drawn or made with changes in target/filter at the standard breast for the Hologic system. The largest notable difference was actually between the Monte Carlo methods themselves i.e. between the Dance and the Boone methods rather than between the displayed and the calculated values. This was

particularly notable for the W/Ag spectrum with an MGD percentage difference of 13% between the two methods with the largest estimate for Boone's method. However, such discrepancies may be attributed to uncertainties or multiple causation effects between the MGD-related measurements and calculations.

Figure 4.2a. Calculated and Displayed MGD for the STD mode on the GE Essential.

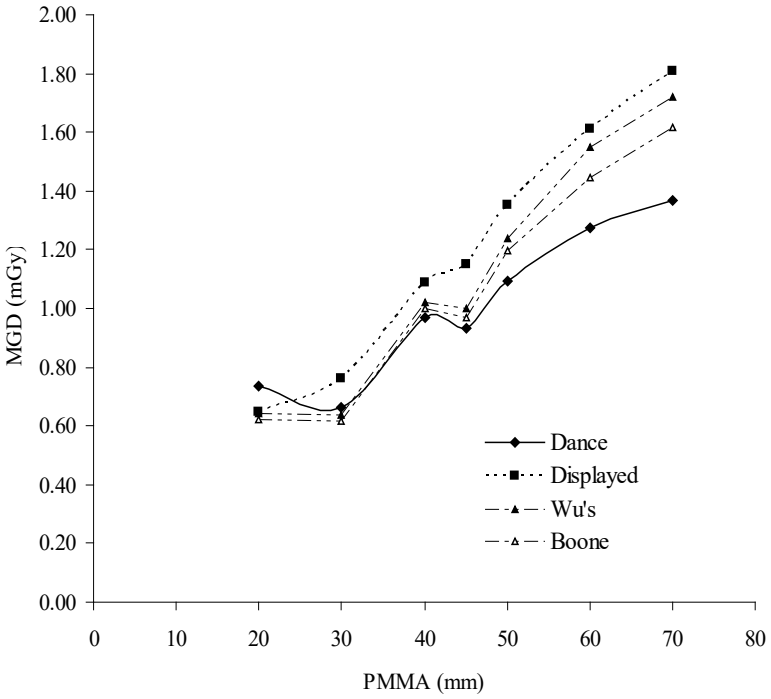


Figure 4.2b. Measured and Displayed K for the STD mode on the GE Essential.

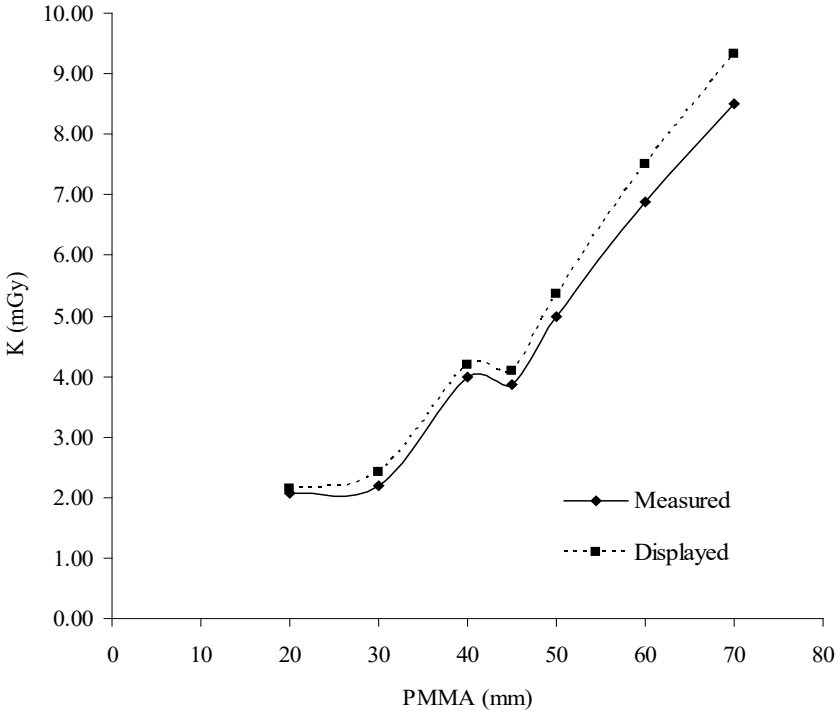


Figure 4.2c. Percentage difference between the displayed and the MC calculated MGD on the GE Essential using different PMMA thicknesses.

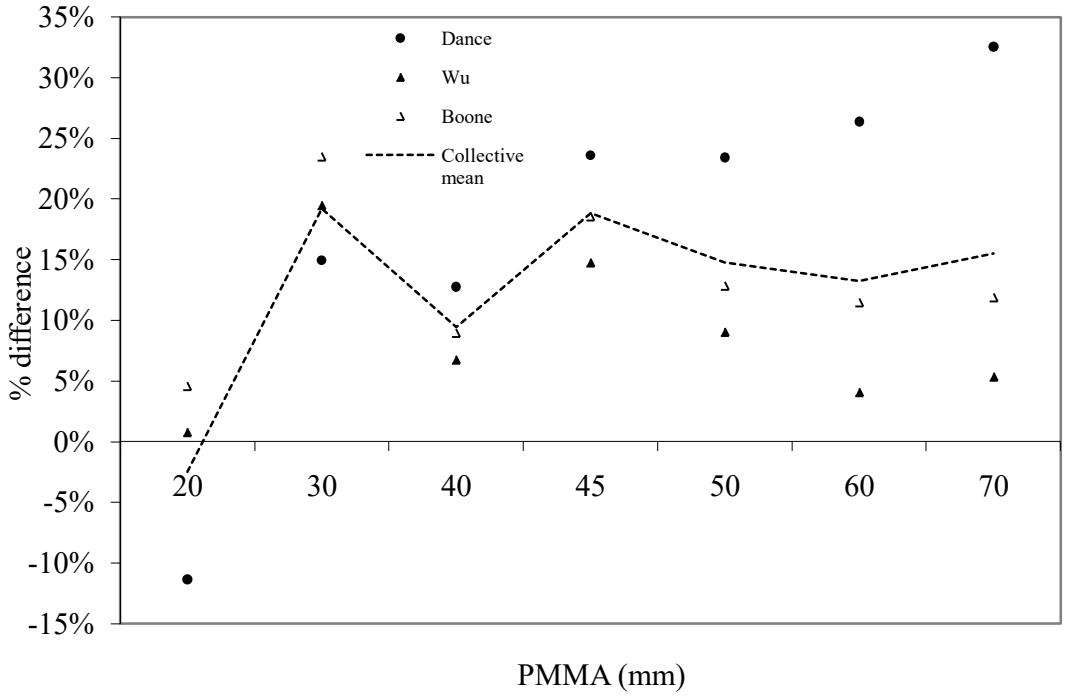


Figure 4.3a. Calculated and Displayed MGD with tube voltage at fixed mAs on the GE Essential.

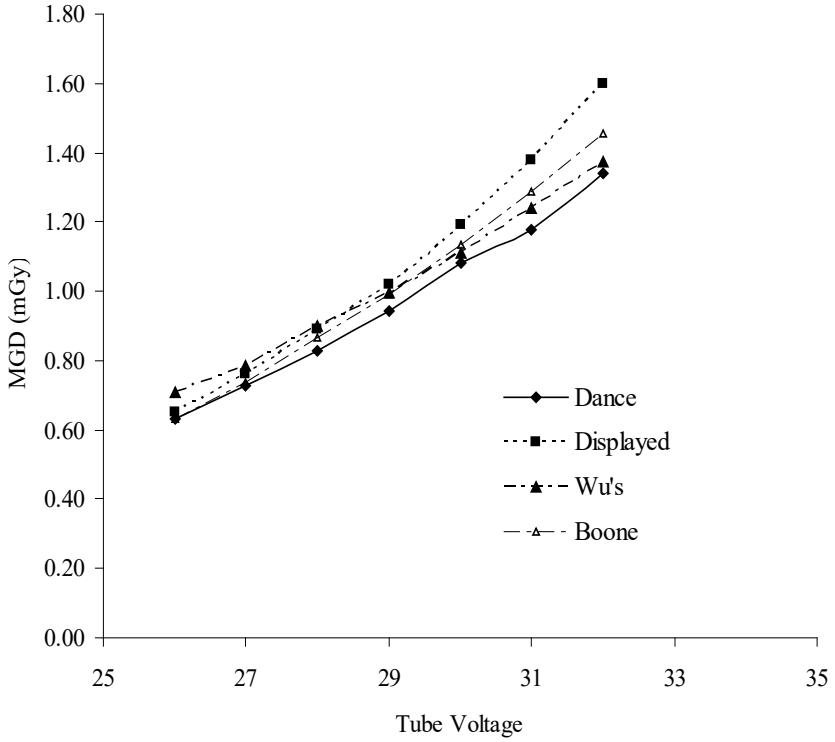


Figure 4.3b. Measured and Displayed K with tube voltage at fixed mAs on the GE Essential.

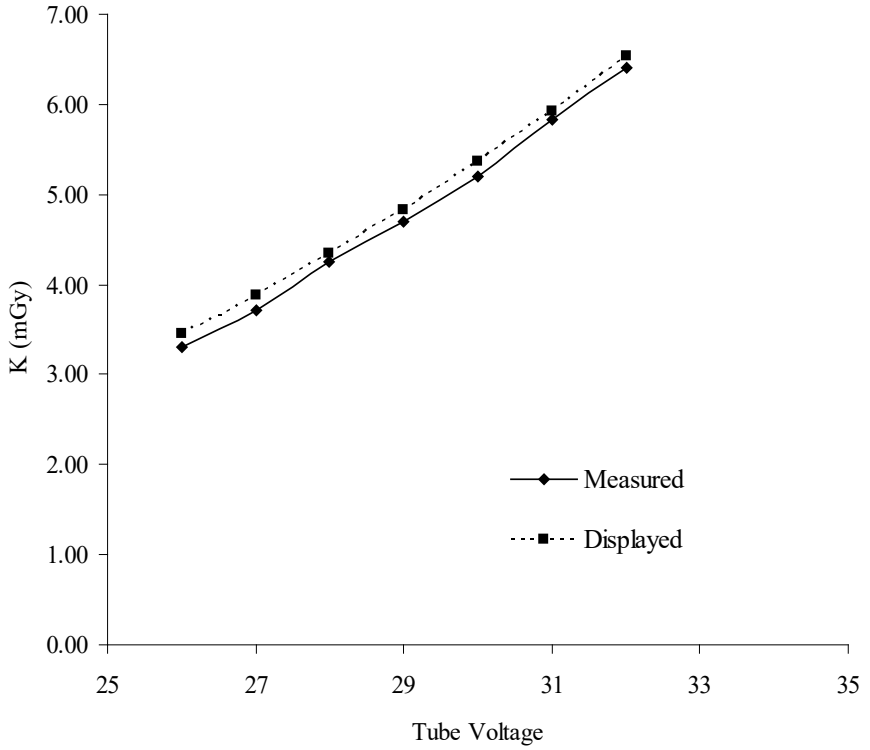


Figure 4.3c. Percentage difference between the displayed and the MC calculated MGD on the GE Essential using different tube voltages.

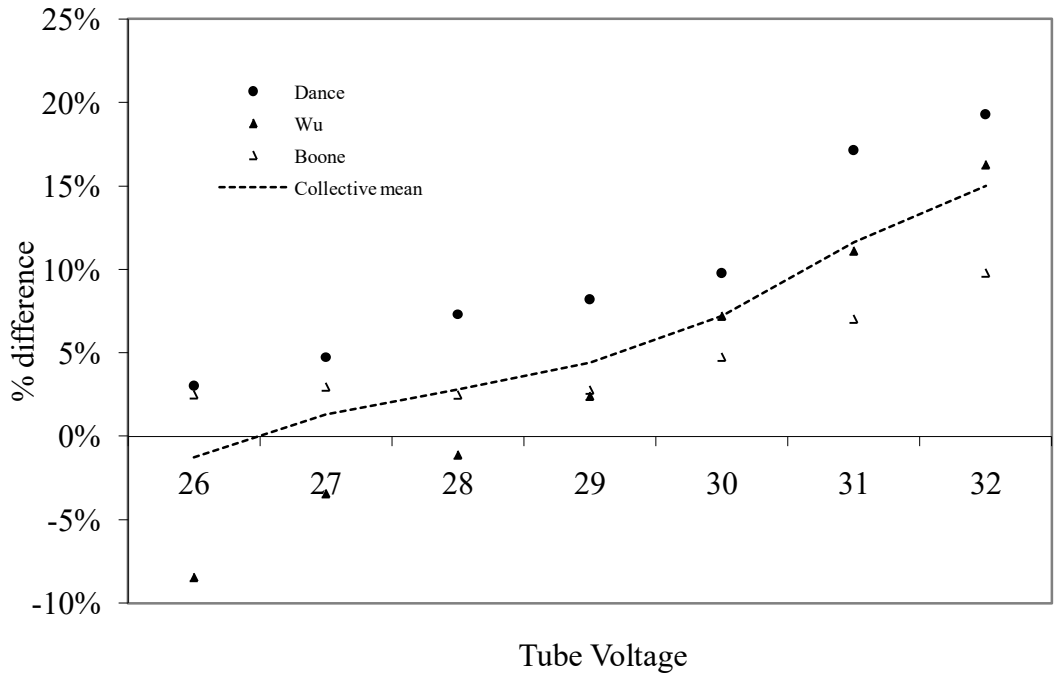


Figure 4.4a. Calculated and Displayed MGD for the Autofilter mode on the Hologic.

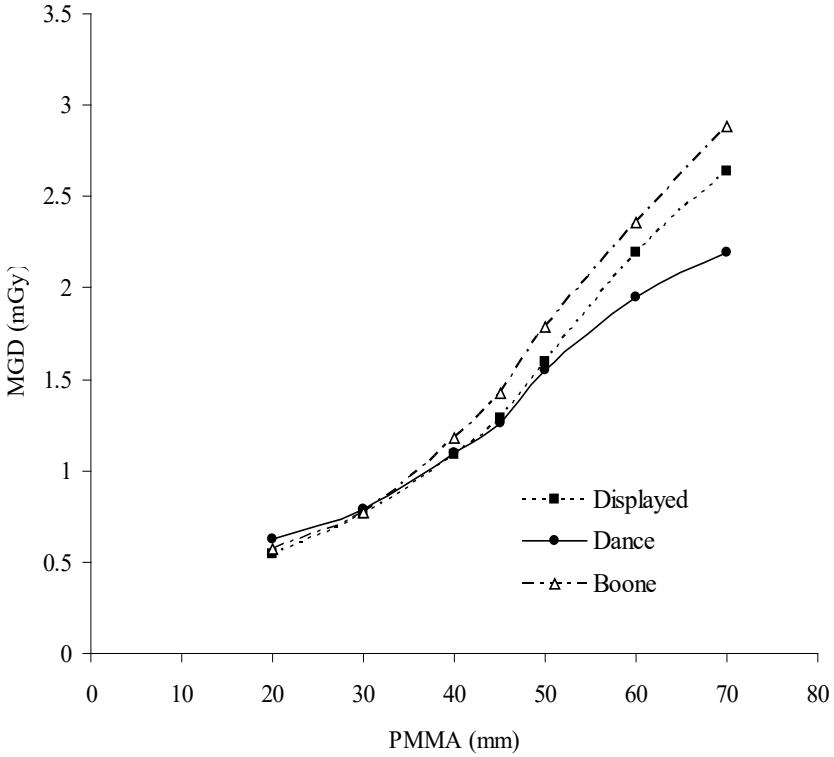


Figure 4.4b. Measured and Displayed K for the Autofilter mode on the Hologic.

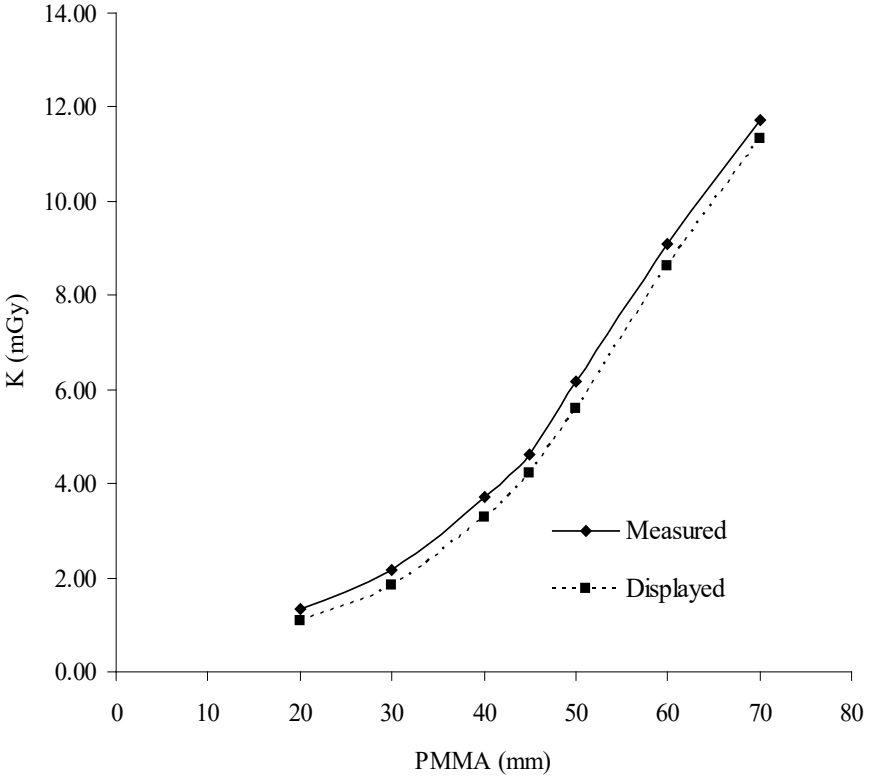


Figure 4.4c. Percentage difference between the displayed and the MC calculated MGD on the Hologic using different PMMA thicknesses.

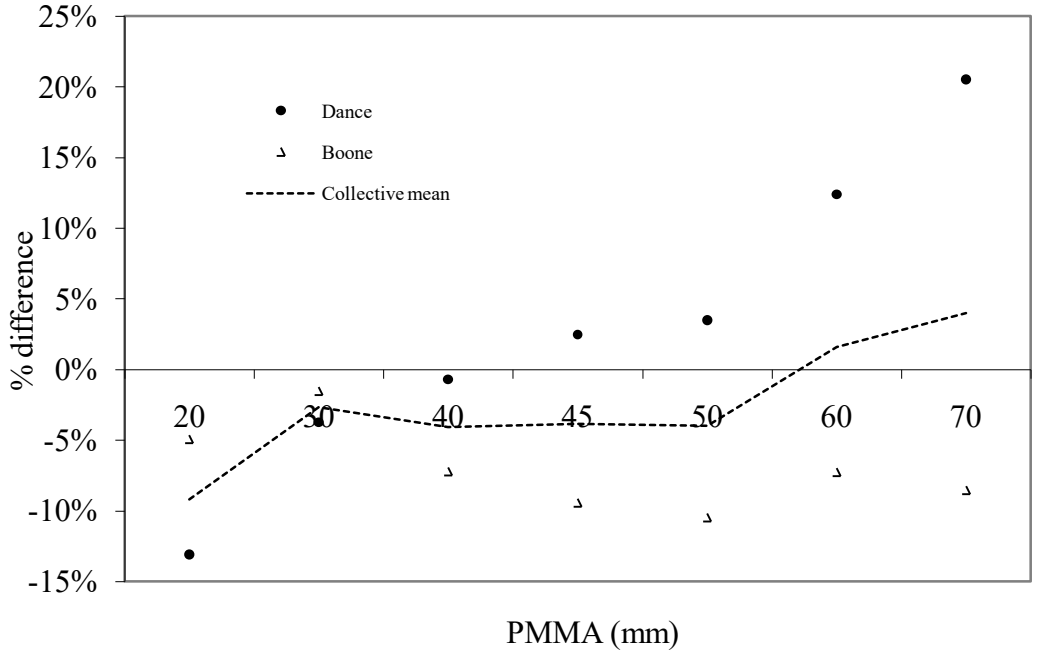


Figure 4.5a. Calculated and Displayed MGD with tube voltage on the Hologic.

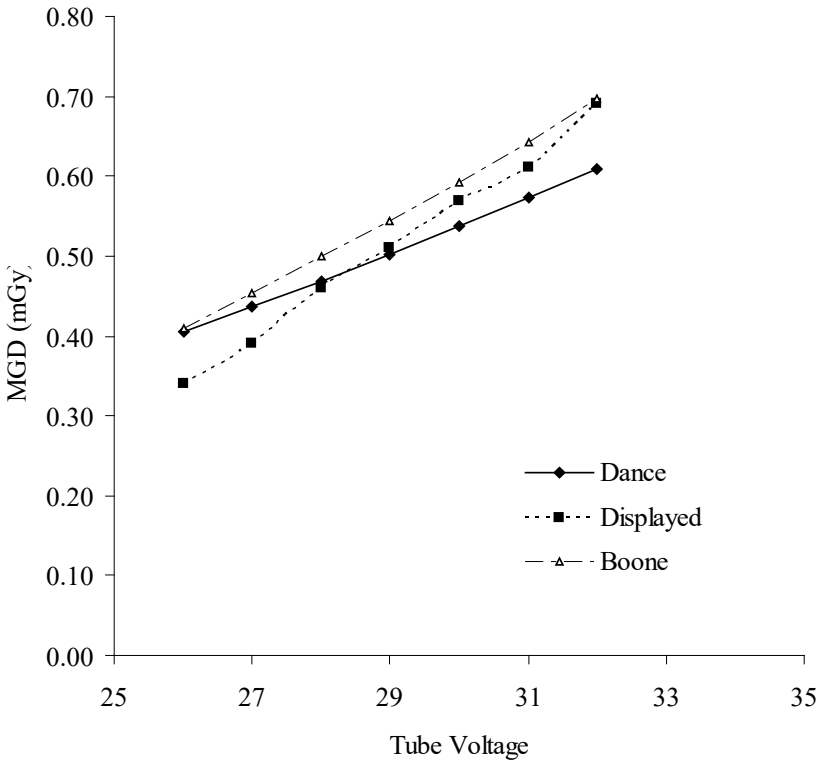


Figure 4.5b. Measured and Displayed K with tube voltage on the Hologic.

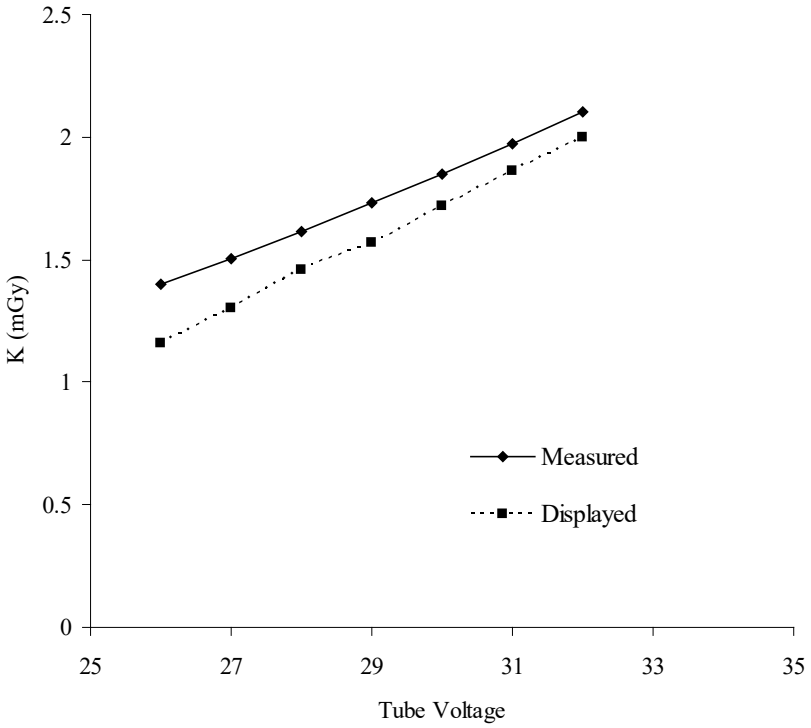


Figure 4.5c. Percentage difference between the displayed and the MC calculated MGD on the Hologic using different tube voltages.

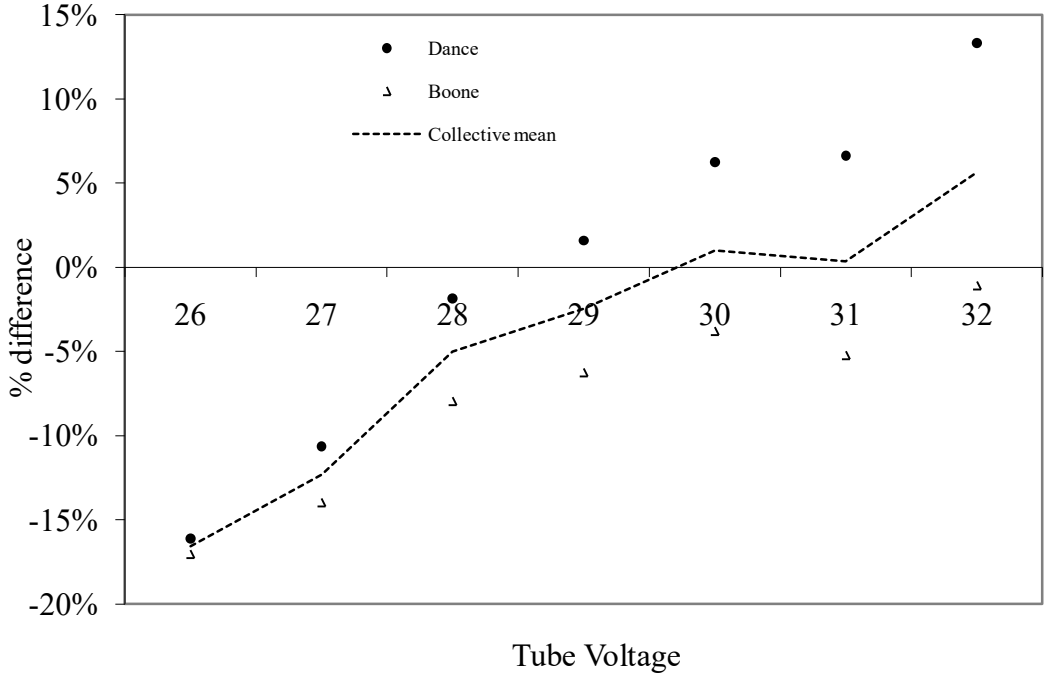


Figure 4.6. Scatter plot of the displayed against the different MGD calculated in mGy for the GE Essential.

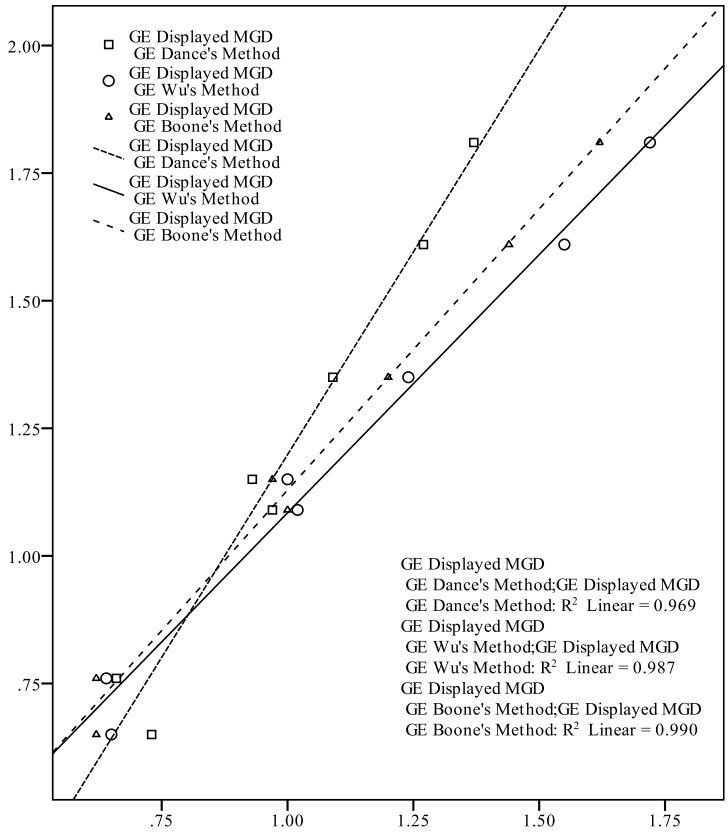


Figure 4.7. Scatter plot of the displayed against the different MGD calculated in mGy for the Hologic Selenia.

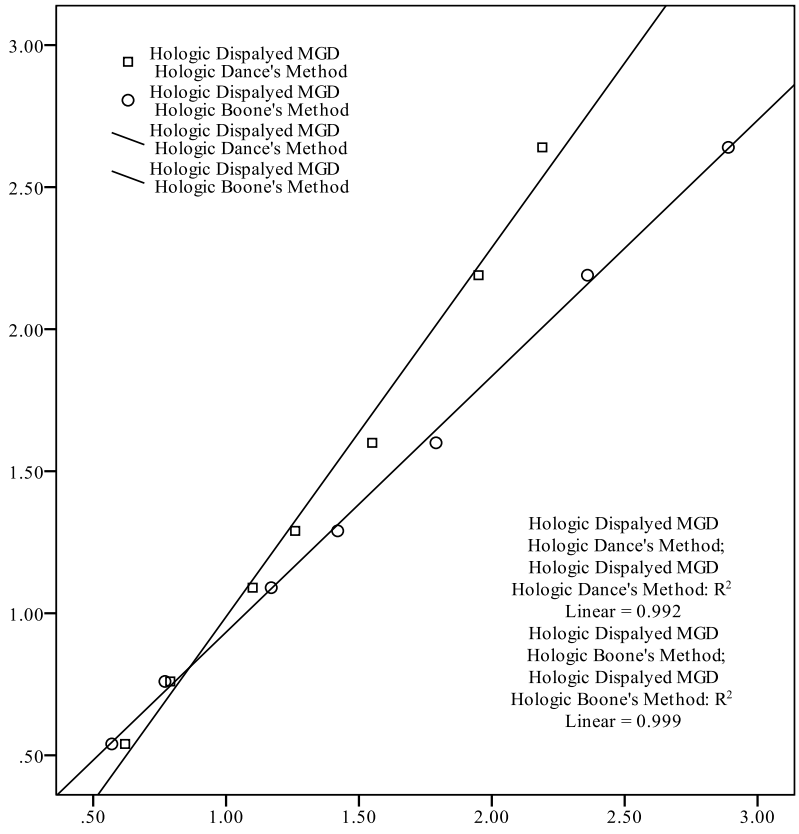


Figure 4.8. Scatter plot of the displayed against the measured K in mGy for the GE Essential.

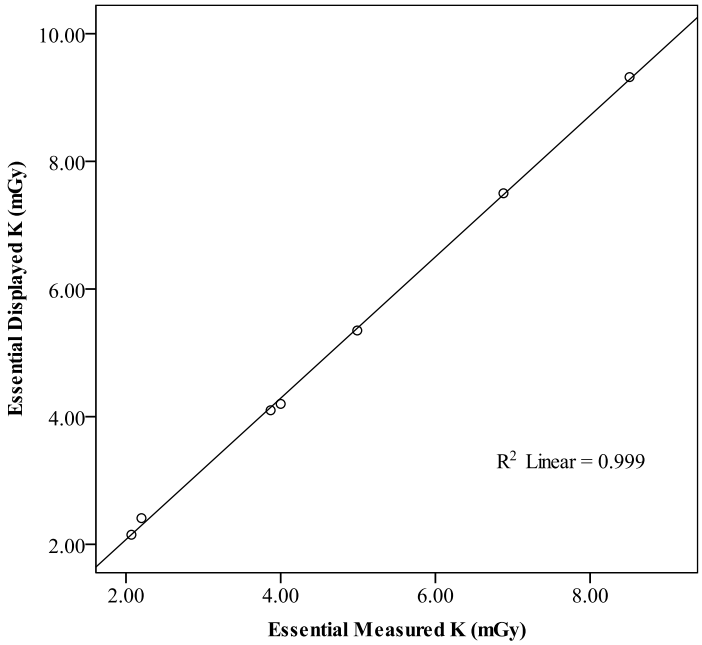
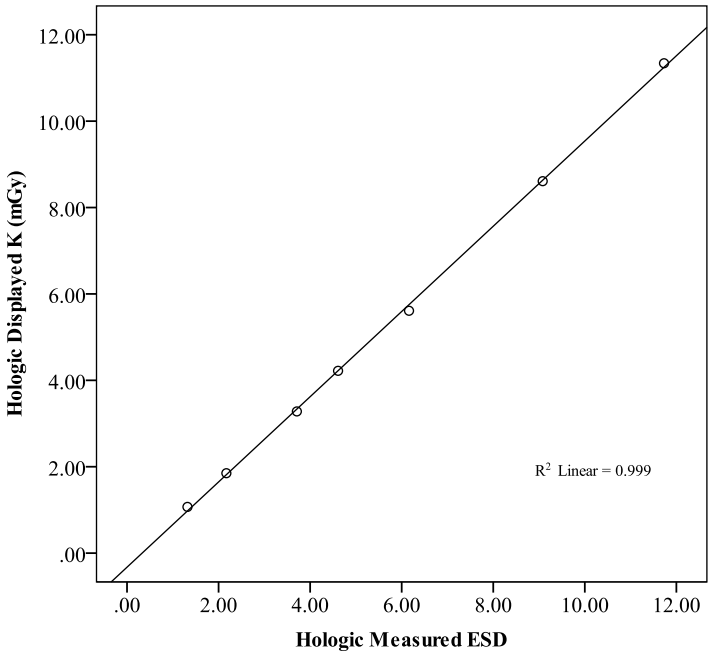


Figure 4.9. Scatter plot of the displayed against the measured K in mGy for the Hologic Selenia.



Regression Analysis and MGD Prediction Models

The strong Pearson correlation between variables implies that there is a predictable relationship between the displayed doses and those calculated using the Wu, Dance or Boone methods as applicable. Given the dependent variable in this study (displayed MGD) is highly correlated with each predictor (Dance, Wu and Boone methods) for both the Essential and the Hologic systems and also given that the data are normally distributed, a multiple linear

regression analysis is indicated [155]. Consequently, models were generated using this method to define coefficients to convert the displayed MGD value on each system to values using the different Monte Carlo-based methods.

From the regression analysis results on the Essential it was concluded that with a single predictor approach, Dance's method was the best predictor for displayed MGD, with an R^2 value of 0.969. This model was followed by a two predictor approach including Dance's and Wu's method with an R^2 value of 0.990. A three predictor model i.e. using all three dose calculation methods resulted in an R^2 value of 0.999. With a closer look at the single predictor approach as a starting point to this algorithm, it was observed that the 30 mm PMMA thickness data point on the Essential was considered an influential point on evaluation of the Cook's distance using SPSS. In statistics, Cook's distance or Cook's *D* is a commonly used as an estimate of the influence of a data point when performing least squares regression analysis. In a practical ordinary least squares analysis, Cook's distance can be used in several ways: to indicate data points that are particularly worth checking for validity; to indicate regions of the design space where it would be good to be able to obtain more data points. By removing this data point from the regression analysis the new R^2 value for the Dance's method improved to 0.982. As a result, for the Essential, Dance's method could be predicted from displayed MGD as follows using equation 21:

$$Dance_{predicted} = \frac{(Displayed + 0.564)}{1.737} \qquad 21$$

Wu and Boone's methods for the Essential were eventually predicted using equations 22 and 23 based on the regression analysis using SPSS.

$$Wu_{predicted} = 0.9623 Displayed - 0.0323 \qquad 22$$

$$Boone_{predicted} = 0.8788 Displayed + 0.0192 \qquad 23$$

Dance's method was also the best displayed MGD predictor for the Hologic system with an R^2 value of 0.992. Using a two predictor model i.e. including Dance's and Boone's methods, the R^2 value was 0.999. There were no significant outliers or influential points in any of the models

presented for the Hologic system. The following models are presented based on the regression analysis for the Hologic in equations 24 and 25:

$$Dance_{predicted} = \frac{(Displayed + 0.312)}{1.3} \quad 24$$

$$Boone_{predicted} = 1.1099 Displayed - 0.0367 \quad 25$$

The reliability and the validity of the presented models/algorithm was verified and assessed using Q-Q plots in figures 4.10, 4.11 and 4.12 showing plots of the predicted values versus the calculated MGD Monte Carlo-based methods on both systems. The Q-Q plot is a graphical technique for determining if two data sets come from populations with a common distribution. The Q-Q plot is used to answer the following questions [156]:

- Do two data sets come from populations with a common distribution?
- Do two data sets have common location and scale?
- Do two data sets have similar distributional shapes?
- Do two data sets have similar tail behaviour?

It is easily observable that there is a very strong linear relationship between the predicted values and the actual calculated values using the established Monte Carlo methods with R^2 values greater than 0.960 on both systems. It must be noted that the R^2 values for plots of the predicted against the actual MGD values were marginally higher for the Hologic system, implying that the models hereby presented show a marginally stronger prediction power for the Hologic when compared to the Essential.

However, in general the presented empirical models can be confidently used to predict the MGD from Dance's, Wu, and Boone's MGD values from displayed MGD values on both the Essential and the Hologic (except for the Wu's method in the Hologic system).

Figure 4.10. Q-Q plot: Dance's predicted MGD values using equations 21 and 24 against the Dance's calculated MGD for both systems.

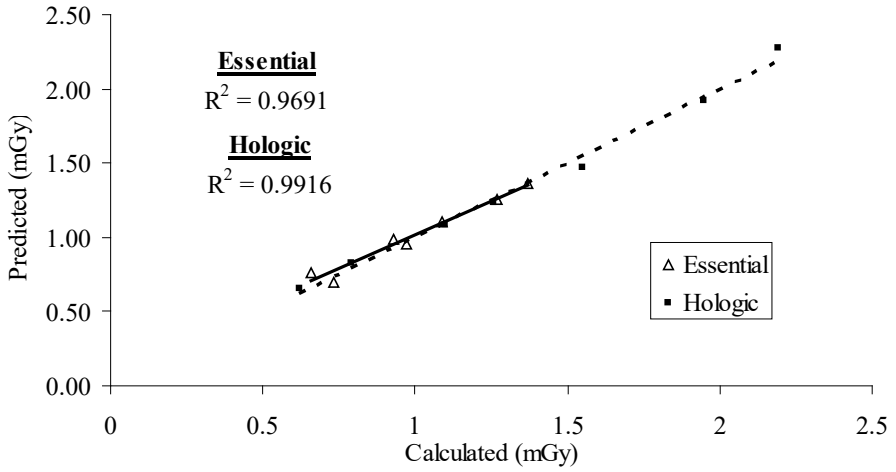


Figure 4.11. Q-Q plot: Wu's predicted MGD values using equation 22 against the Wu's calculated MGD for the Essential.

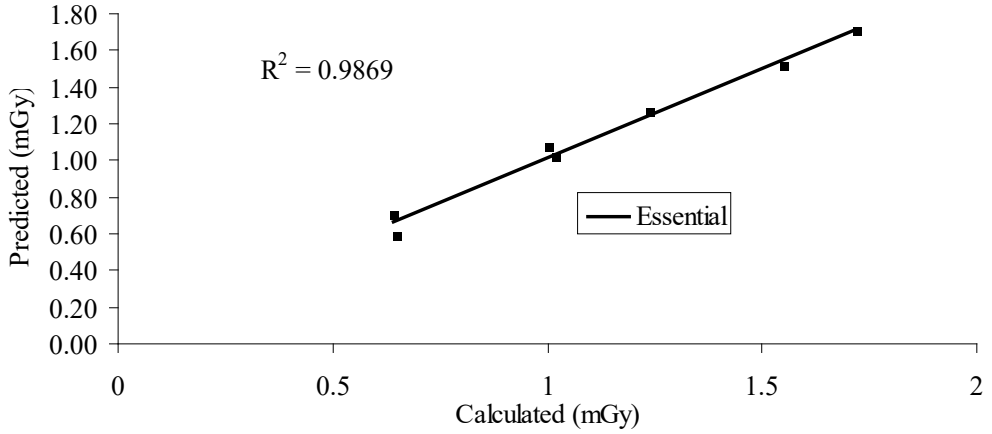
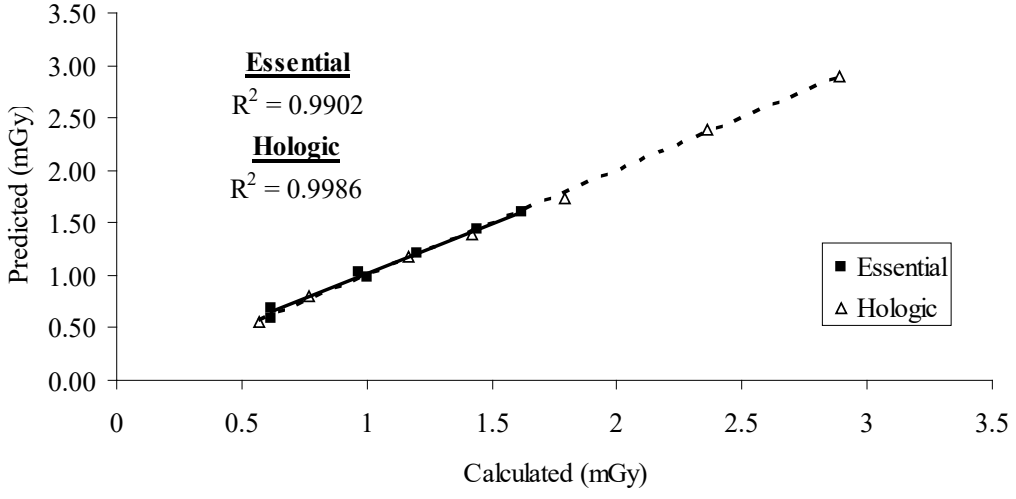


Figure 4.12. Q-Q plot: Boone's predicted MGD values using equations 23 and 25 against the Boone's calculated MGD for both systems.



CONCLUSION

From this study it appears that the displayed MGD compares well to all Monte Carlo-calculated values for small and medium PMMA thicknesses using fully automatic exposures. The displayed MGD on both the Essential and the Hologic may be confidently used for routine dosimetric purposes for small and medium PMMA thicknesses since their overall values under various conditions in this study related well to all well- established Monte Carlo methods. However, with increased PMMA thickness the difference between displayed and calculated MGD values appears to be more pronounced. Displayed K value measurements can be relied upon for all PMMA thicknesses.

Despite the differences shown for larger PMMA thicknesses, the displayed MGD correlate extremely well with the calculated values. Given the strong correlation, predictive models have been devised to easily convert system-displayed MGD values into conventional Monte Carlo-calculated MGD values for larger PMMA thicknesses.

The models generated are promising predictor tools since their reliability and validity has been proven in the various Q-Q plots within this study. Easy inter-conversion between Monte Carlo-based MGD methods for larger PMMA thicknesses should be a very handy tool for medical physicists in the estimation and calculation of doses during routine quality assurance checks and should reduce equipment survey time.

CHAPTER 5

SHOULD WE USE PROCESSED OR RAW IMAGE DATA IN MAMMOGRAPHIC IMAGE QUALITY ANALYSES? A COMPARATIVE STUDY IN THREE FULL-FIELD DIGITAL MAMMOGRAPHY UNITS.

Introduction

This chapter was published [5]. A feature that is a characteristic to many DR mammography units is the option to 'choose' between processed and raw or unprocessed images of the acquired DICOM images. DICOM is the international standard for medical images and related information (ISO 12052). It defines the formats for medical images that can be exchanged with the data and quality necessary for clinical use. The European protocol [2] requires that raw images must be used (when available) for the evaluation of particular quality control tests, such as the threshold contrast detail detectability, using the CDMAM phantom and CNR. Some mammography units do not have raw images available and in CR systems the acquired images should be linearized from the determined response function. For systems in which raw images are unattainable the European protocol suggests using processed images for TCDD. However, this should be avoided since image processing may introduce artefacts on phantom images and may be different from image processing for mammograms due to histogram or local texture based processing techniques. Therefore, care needs to be taken in the interpretation of processed images for quality assurance purposes [157, 158]. Certainly, and without any doubt, detector characterisation tests including: response function, MTF, NNPS and DQE as specified in IEC 62220-1-2 [9] have to be performed using raw and/or linearized data in all mammography units. The latter also applies for the estimation of the NEQ.

On the other hand, using unprocessed or raw images implies that only the image acquisition part of the system is used for evaluation. This means that more weight is given on detector performance that in turn governs the resultant image quality. However, it is also true that image processing must not be overlooked as this may have a substantial contribution on the final image and its presentation. In a study by Zanca et al [157], it has been shown that

CDMAM images did not show any effects, neither benefits nor disadvantages from image processing algorithms using one DR mammography unit. However, if processing algorithms change through system upgrades, then image quality results of any digital systems are bound to change, which again suggests the use of raw data for any analyses.

The purpose of this study was to compare processed and unprocessed phantom images for image quality tests using two DR units and a CR mammography system and quantitatively analyse differences between the two image sets. This study also aims to confirm the need to use raw images in image quality analyses in digital mammography.

Materials and Methods

Mammography Units

The mammography units evaluated in this study were 1) the General Electric Senographe Essential (hereafter referred to as the Essential), 2) the Hologic Selenia Dimensions 2D (hereafter referred to as the Hologic) and an analogue unit with a CR system 3) the General Electric DMR+ with a Fuji Profect FCR system (hereafter referred to as the Fuji).

Image acquisition

CDMAM and CNR images were acquired using AEC [159]. In the Essential and the DMR+ (used for the Fuji) this was done using the 'AOP standard' mode. Unprocessed images in the Essential are readily available with each exposure by default and in the Hologic raw data is acquired using a separate protocol: the *'flat field'* protocol. Processed images from the Hologic were acquired using the autofilter mode. With the Fuji CR system, processed data were collected using the standard 'breast' protocol on the CR reader. Based on European protocol recommendations a fixed exposure data recognizer (EDR) mode with level (L) 2 and sensitivity (S) 120 were used to minimize image processing and considered to produce unprocessed images in such systems. Separate descriptions are given for image acquisitions used for detector characterization tests under the respective headings in the coming sections.

Image Quality

Contrast-Detail (CD) phantom

Threshold contrast detail detectability (TCDD) measurements were made using the CDMAM phantom. More details on the CDMAM phantom can be found in pages 74 and 88. The phantom was positioned with a 20 mm thickness of PMMA above and below the phantom, to give a total attenuation of approximately equivalent to 50 mm of PMMA or 60 mm thickness of typical breast tissue. This procedure was repeated with small adjustments to the phantom position to obtain a representative sample of 16 images using automatic exposure control [159] using image acquisition parameters described in the previous section.

Automated Scoring of the CDMAM images

All CDMAM images acquired were scored using the CDMAM 3.4 analyzer software [121, 160, 161]. See also pages 76 and 88. This program has the ability to score digital images obtained with the CDMAM phantom automatically, using the CDCOM v1.6 analysis core from EUREF [121, 162].

CNR

CNR or SDNR region of interest [163] measurements were linearized using the data from the detector response function. The specific linearization method is described elsewhere [24]. The CNR was calculated using equation 1, for a range of PMMA thicknesses (20 mm, 30 mm, 40 mm, 45 mm, 50 mm, 60 mm and 70 mm) using a 0.2 mm, 99.9% pure Aluminum square, as recommended in the European protocol.

The limiting value for CNR (using 50 mm PMMA) was eventually determined according to equation 26. This equation determines the CNR value (*CNR limiting value*) that is necessary to achieve the minimum threshold gold thickness for the 0.1 mm detail in the CDMAM phantom (i.e. *threshold gold limiting value* = 1.68 µm), which is equivalent to *threshold contrast limiting value* = 23.0% using 28 kV Mo/Mo. Threshold contrasts were then calculated as described in the European protocol and used in equation 26.

$$\text{Threshold}\quad \text{Contrast}_{\text{measured}} * \text{CNR}_{\text{measured}} = \text{Threshold}\quad \text{Contrast}_{\text{limiting}} * \text{CNR}_{\text{limiting}} \qquad 26$$

The relative CNR was then calculated according to equation 27 for each PMMA thickness and compared to the limiting values determined by the European Protocol. The minimum CNR required to meet this criterion was then established.

$$\text{Relative CNR} = \frac{\text{CNR}_{\text{measured}}}{\text{CNR}_{\text{limiting value}}} \qquad 27$$

The CNR value was also calculated for a range of gold details within the CDMAM phantom at 50 mm PMMA thickness for both the processed and unprocessed images acquired. This exercise was done to determine whether the CDMAM phantom can be used effectively for CNR measurements and whether any changes in the results between this method and the conventional method of CNR analyses can be identified.

Detector Characterization tests

Objective X-ray performance evaluation of an X-ray detector includes calculation of MTF, NNPS and DQE. There is growing consensus in the scientific world that the DQE is the most suitable parameter for describing the image performance of an X-ray imaging detector. The DQE describes the ability of an imaging detector to preserve the SNR from the radiation field to the resulting digital image area. Since in X-ray imaging the noise in the radiation field is associated with the K, DQE values can be considered to describe the dose efficiency of any given X-ray detector [9]. A system with a higher DQE will reach the same contrast-detail resolution at lower dose or higher contrast-detail resolution at the same dose level compared to another system with lower DQE [30, 92-94, 164].

Detector Response Function

The detector response function was measured using a 45 mm PMMA block placed directly on the detector at 28 kVp using Mo/Mo for the Essential and the Fuji (GE DMR+) and W/Rh for the Hologic. Different tube loadings were employed ranging from 10 mAs to a maximum of 400 mAs yielding a range of K values to the detectors. An area of 10 x 10 cm^2, 60 mm from the chest wall, was used to evaluate the MPV and the SD on both the processed and the raw data. The detector response function was compared to what was specified by the manufacturer or the available literature. This test was also important in determining any pixel offset related to the detector device and for the linearization of data, particularly the Fuji system.

MTF

The pre-sampling MTF was measured using the edge technique as described in IEC 62220–1-2 [9, 23]. The test object used consisted of a stainless steel edge angled at 2°–3° with respect to lines/columns of pixels and forms part of the DMAM kit (Leeds Test Objects). The edge was put in contact with the detector surface excluding the antiscatter grid in the middle of the field of view and adjacent to the chest wall side. The MTF was evaluated using a detector K of 120 µGy for all the units employed along the two orthogonal axes with a collimated x-ray beam. The spectra and the detector K employed in this study were selected and based on other studies [100, 101]. The spectrum used for the Essential and the Fuji comprised 28 kVp with a Mo/Mo target/filter combination with 2 mm of 99.9% purity Al added at the tube head also described as the RQA-M2 spectrum [9, 101]. There are many possible spectral and detector dose combinations; however the dependence of detector performance on K or the spectrum chosen wasn't within the scope of this study. On the linearized images a ROI of 50 mm across and 50 mm along the edge was used for analysis [9].

The RQA-M2 spectrum is not applicable to the Hologic system given it has a W tube system with either Rh or Ag as filtration options. Consequently, a similar setup as with the RQA-M2 spectrum was adopted however using W/Rh. Reference was also made to manufacturer technical recommendations as required [132].

With reference to IEC 62220-1-2, the angle of the edge image was determined by using a linear least square fit and the distances of each pixel to the edge was computed to obtain an oversampled edge spread function (ESF). The ESF was smoothed with a median filter to reduce high frequency noise and then the ESF was differentiated to obtain an oversampled line spread function (LSF). Finally, the LSF was Fourier transformed to give the pre-sampling MTF [9].

Noise Assessment

The noise was evaluated with the method prescribed in the European protocol [2]. The mean pixel value and standard deviation were measured on the acquired images of the detector response function measurement. For systems with a linear response such as the Hologic and the Essential, the SNR^2 was plotted against K. The linearity was determined by plotting a best fit through all measured points. The square of the correlation coefficient (R^2) was determined using Microsoft Excel. Non-linearity is an indication for the presence of additional noise

sources besides quantum noise. This value served as a baseline value for any future noise evaluations of any given system within this study.

For CR systems with a logarithmic response, as is the Fuji CR, the standard deviation squared was plotted against the reciprocal of K. Once again, the linearity was determined by plotting a best fit through all measured points and the R^2 value was determined. The offset is an indication for the presence of additional noise sources besides quantum noise.

NNPS

The methodology employed to estimate the NNPS was done according to IEC 62220-1-2 [9] by using uniform images with spectral conditions and K values to the detectors similar to the MTF measurements. The X-ray field was collimated to an area 13 x 18 cm^2. A region of 10 x 10 cm^2 was extracted 60 mm from the chest wall and it was divided into squares with areas of 256 x 256 pixels. These areas were overlapping by 128 pixels in both the horizontal and the vertical directions. A two-dimensional second-order polynomial function was used to correct each region of interest [163] for low frequency background effects including the Heel effect. The two-dimensional noise NPS was calculated by using equation 28 [23]:

$$\mathrm{NPS}(u_n, v_k) = \frac{\Delta x \cdot \Delta y}{M \cdot 256 \cdot 256} \sum_{m=1}^{M} \left| \sum_{i=1}^{256} \sum_{j=1}^{256} (I(x_i, y_j) - - S(x_i, y_j)) \exp - 2(\pi i (u_n x_i) + (v_k y_j)) \right|^2 \qquad 28$$

where *u* and *v* are the spatial frequency variables, Δx and Δy are the *x* and *y* pixel size (assumed to be the same), *Nx* and *Ny* are the number of elements in the *x* and *y* directions, respectively (*Nx* = 256, *Ny* = 256), *M* is the number of ROIs used, I (xi , yi) is the linearized data, and S (xi , yi) is the 2D polynomial function. The ROIs were divided by the square of the large area signal, to obtain the NNPS. The large area signal was calculated as the mean pixel value in each ROI of the linearized image. In order to evaluate the one-dimensional NNPS in the horizontal and vertical direction, ten central frequency bands on both sides of the u and v axes were averaged. The central axes were excluded to avoid a residual trend in signal background. In addition, image data in the absence of X-ray irradiation were acquired and analyzed in a similar manner to obtain a measure of dark NNPS, which describes additive noise and can show the presence of fixed pattern noise [9].

DQE

The DQE for all the units in this study was eventually calculated using the MTF and NNPS using equation 29.

$$\mathrm{D}QE(u) = \frac{\mathrm{MTF}^2(u)}{\mathrm{q.E.NNPS}(u)} \qquad 29$$

where *q* is the X-ray quantum fluence per unit exposure, MTF is the modulation transfer function, NNPS is the normalized noise power spectrum and *E* is the air kerma on the detector. In order to evaluate *q*, the X-ray spectrum was calculated using an ImageJ plugin [130, 165] and a pre-established Matlab routine [101]. The achieved DQE values were eventually compared to any values specified by the manufacturer were available and other reputable publications. [100, 101]

NEQ

NEQ expresses the quality of the image data by the photon fluence that the image is worth at each spatial frequency. The NEQ was calculated from the DQE using equation 30 where q is the number photons per unit area and was verified using a pre-established Matlab routine (MIQuALa) [101] .

$$NEQ(u) = DQE(u) * q \qquad 30$$

Results

CNR

The exposure parameters selected by the AEC for each exposure is presented in table 5.1. With reference to figure 5.1, the three systems investigated showed different relationships between the raw and processed data. The data from the Essential in figure 5.1a shows little difference between the raw and processed CNR images. In contrast, data comparing the raw and processed images from the Hologic and the Fuji differ. All results were statistically analyzed using the independent samples T-test. This test compares the mean scores of the two groups, with CNR being the dependent variable. If the null hypothesis (H0) holds, this implies that there is no significant difference between the raw and the processed CNR mean values. If the null hypothesis does not hold (H1), then the two groups have significantly different means. The p value for the Essential was 0.785 and the p value for the Hologic and

Fuji was less than 0.05. This confirms observations made in figure 5.1a that there is no statistical difference between the mean scores or values between raw and processed CNR data in the Essential while the opposite holds for the Hologic and the Fuji systems.

However, having different mean values does not necessarily mean the datasets do not correlate. In fact, a Pearson correlation analysis followed between the datasets for the three systems. All three systems resulted with r values greater than 0.9, meaning that raw CNR data correlates very well with processed CNR data. However, it was worth noting that the Essential scored the largest r value (~1.0), showing a remarkable correlation between the raw and the processed images. In figure 5.1b the strong relationship between the raw and processed data in the Essential is confirmed by yielding the largest R^2 value, followed by the Hologic and the Fuji system.

Table 5.1 AEC exposure parameters for the CNR exposures

	kVp	Target / Filter	Phantom Thickness (mm)	Recorded mAs	Mean Glandular Dose - MGD (mGy)
Essential	26	Mo/Mo	20	27	0.77
	26	Mo/Rh	30	36	0.64
	28	Mo/Rh	40	46	0.83
	28	Mo/Rh	45	55	0.91
	29	Rh/Rh	50	60	0.97
	30	Rh/Rh	60	70	1.10
	31	Rh/Rh	70	73	1.11
Hologic	25	W/Rh	20	44	0.63
	26	W/Rh	30	69	0.86
	28	W/Rh	40	89	1.11
	28	W/Rh	45	120	1.38
	29	W/Rh	50	136	1.60
	31	W/Rh	60	210	2.36
	30	W/Ag	70	273	3.19
Fuji	25	Mo/Mo	20	17	0.36
	26	Mo/Mo	30	35	0.65
	27	Mo/Mo	40	51	0.89
	27	Mo/Mo	45	79	1.26
	25	Mo/Rh	50	124	1.27
	26	Mo/Rh	60	194	2.04
	27	Mo/Rh	70	318	3.37

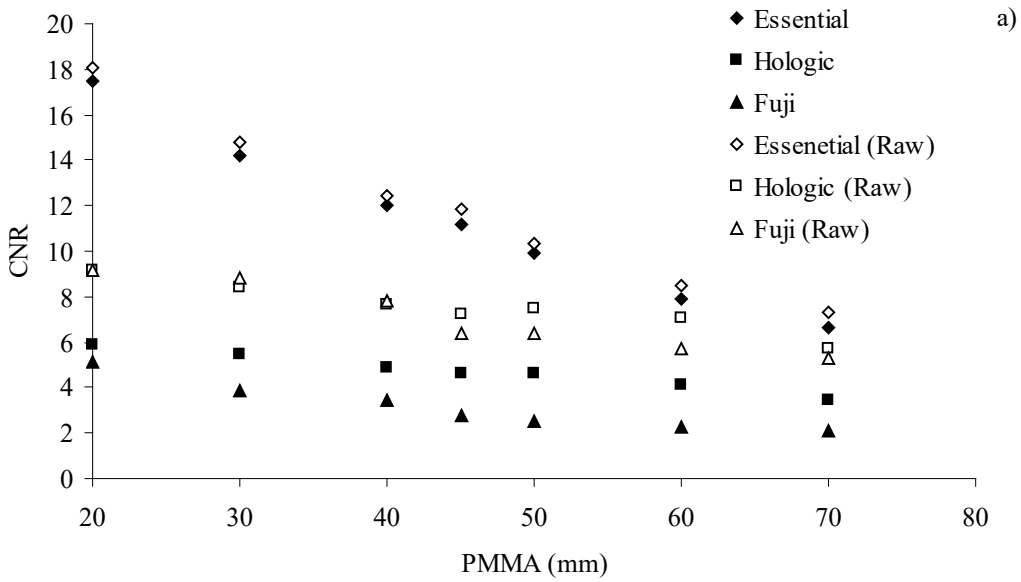


Figure 5.1a. Plot of CNR data (raw and processed) with PMMA thickness

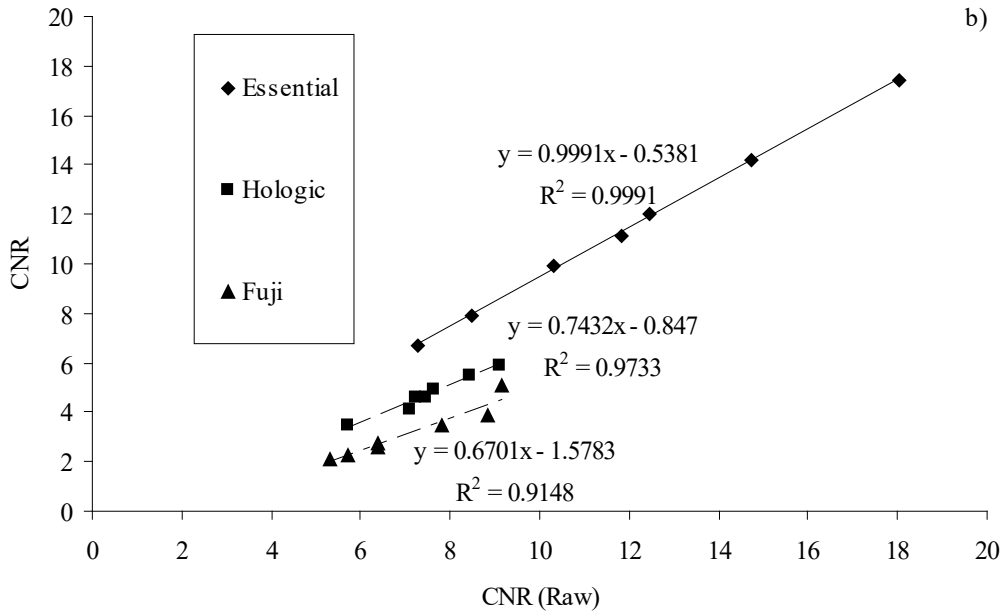


Figure 5.1b. Scatter plot of the CNR from processed versus the CNR from raw data.

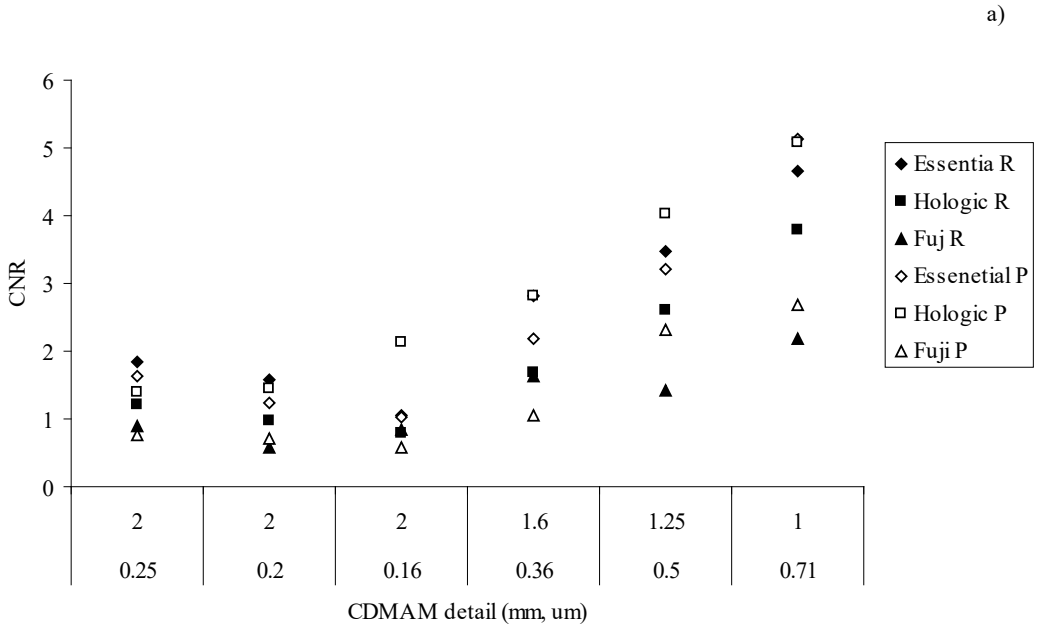


Figure 5.2a. Plot of CNR data (raw and processed) using the CDMAM phantom at 50 mm PMMA (x axis label: top value - gold detail diameter; bottom value - gold detail thickness).

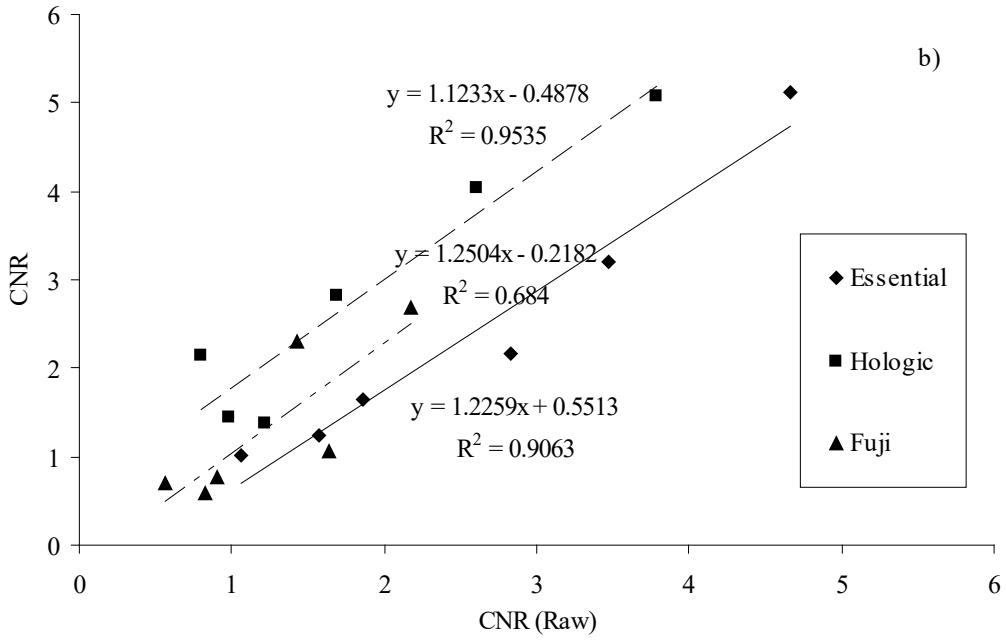


Figure 5.2b. Scatter plot of the CNR from processed versus the CNR from raw data.

From figure 5.2, no strong findings can be drawn on CNR calculations using the CDMAM phantom. Nevertheless, the results are somewhat different from those achieved in the previous section. In figure 5.2a, it appears that as the detail diameter decreases the difference between the raw and the processed images increases. Moreover, CNR from the raw and

processed data converges i.e. appears closest in magnitude, for all three systems at 2 mm; 0.25 µm and the 2 mm and 0.2 µm CDMAM detail groups. The CNR using these CDMAM details could be investigated in further detail with changes in PMMA thicknesses in future similar evaluations. With reference to figure 5.2b, the R^2 values achieved are less remarkable for the Essential and the Hologic. In the case of the Fuji, the R^2 value is also notably lower using this methodology.

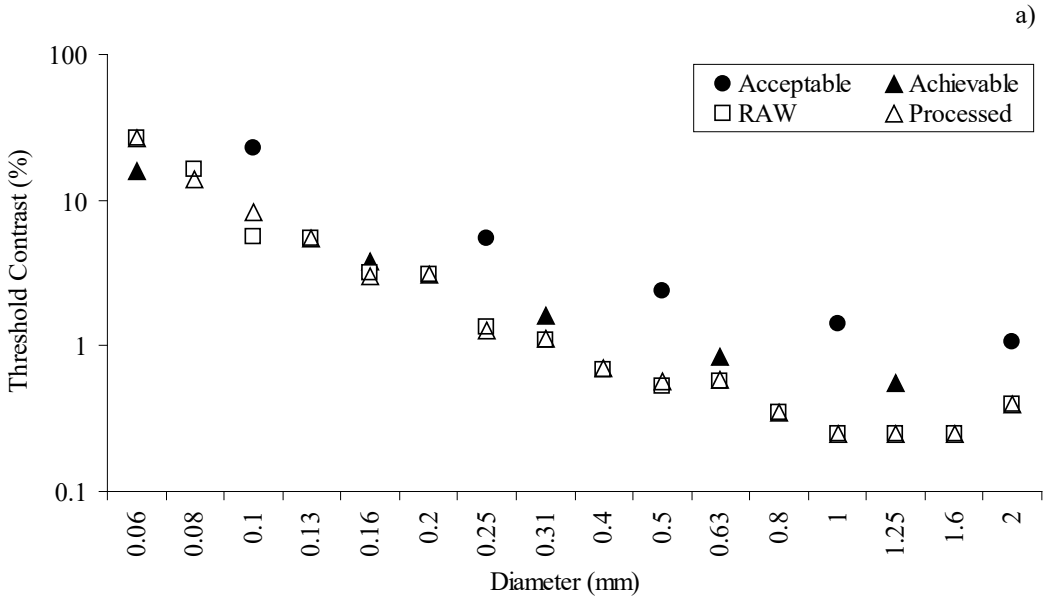
Despite the apparent differences, the T-test results on this data group show that there is no statistical difference between the means of the raw and processed groups when using the CDMAM phantom for CNR estimation. This actually contrasts the results from the T-test in the previous section which showed a significant difference in the Hologic and the Fuji mean values. In agreement with the European protocol method, there is also a very good correlation between the two groups in all 3 systems with Pearson r values greater than 0.8. Also in agreement is the fact that the best correlation was present in the Essential with an r value of 0.98.

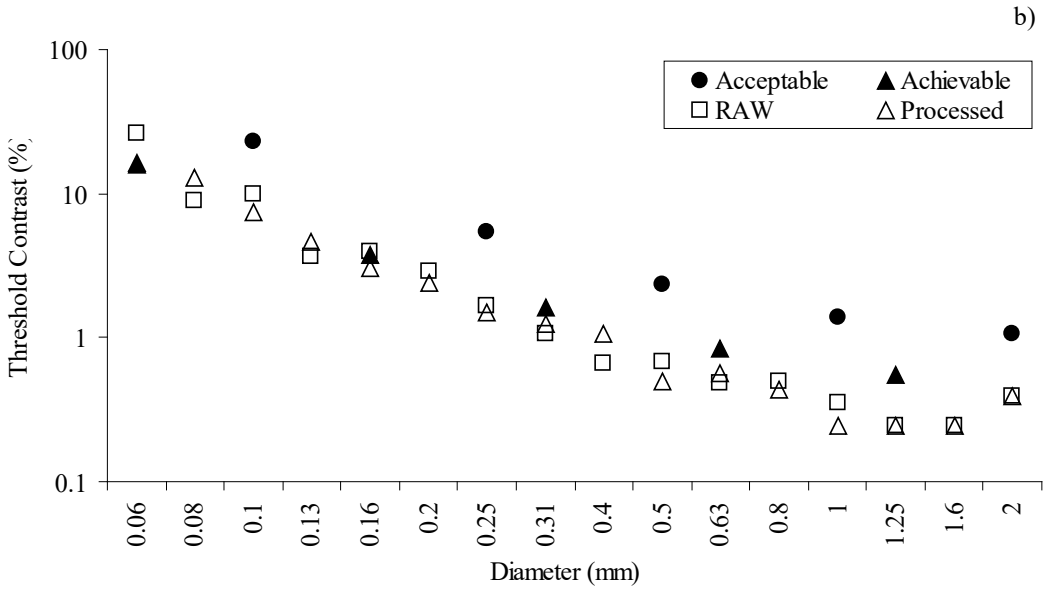
Threshold contrast, H_T and the IQF_{INV}

In figures 5.3 and 5.4 it is seen that the TCDD results suggest that the performance of the Essential and the Hologic was equivalent. Further still, the image quality of the DR systems is in general superior to the Fuji system. An independent sample T-test performed on all the data, showed that there is no statistical difference between the threshold contrast and the H_T data between the processed and the raw images in all three systems. Moreover, with reference to table 5.2 and figures 5.5a and b, there is an exceptionally good correlation between the same processed and raw threshold contrast and H_T data in the three systems investigated. The correlation between processed and raw images appears better in the threshold contrast data which is less obvious with H_T. This is particularly noticeable with the Hologic system which showed an R^2 value of 1.00 for threshold contrast which then decreased to 0.76 for the H_T data. Nevertheless, the percentage difference between the raw and processed images was less than 5% for both threshold contrast and H_T data in all three systems.

Table 5.3 shows the automatically generated IQF_{INV} using the CDMAM Analyzer software [162]. It is interesting to note that the Essential and Fuji systems have relatively similar IQF_{INV} for their

raw and processed images with 0.69% and 5.11% difference respectively between the two groups. However, the Hologic yielded a considerable difference (9.41%) in the IQF_{INV} score between the two datasets. It is also worth noting that there is less than 1% difference in IQF_{INV} between the two DR units themselves however there is approximately a 35% reduction in IQF_{INV} between the two DR units and the Fuji CR system confirming the superior image quality of the former.





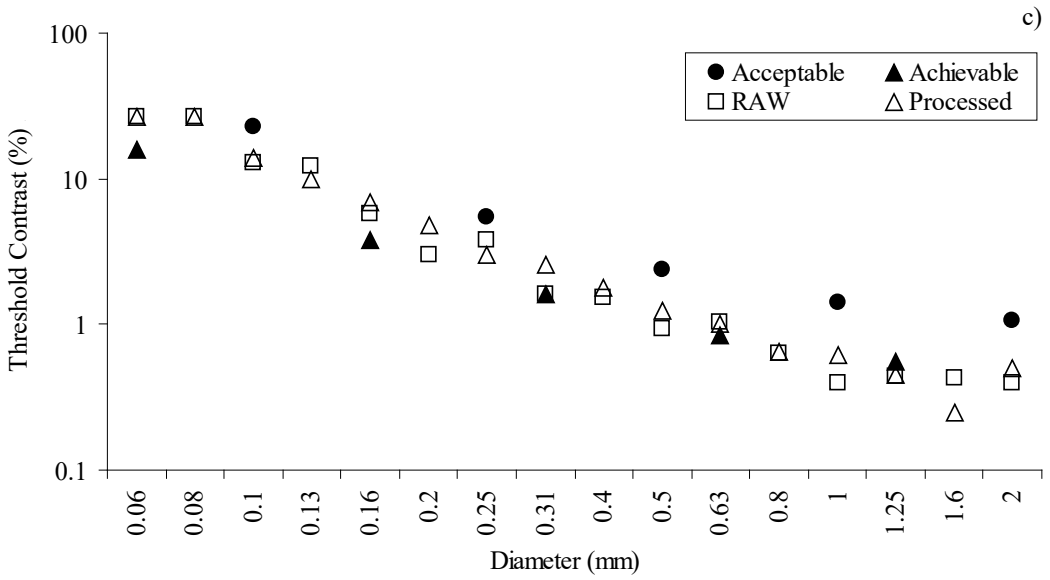


Figure 5.3. Plot of the threshold contrast (%) against the CDMAM detail diameter for a) Essential, b) Hologic and c) Fuji. Also, showings are the acceptable and achievable threshold contrast levels specified by the European Protocol.

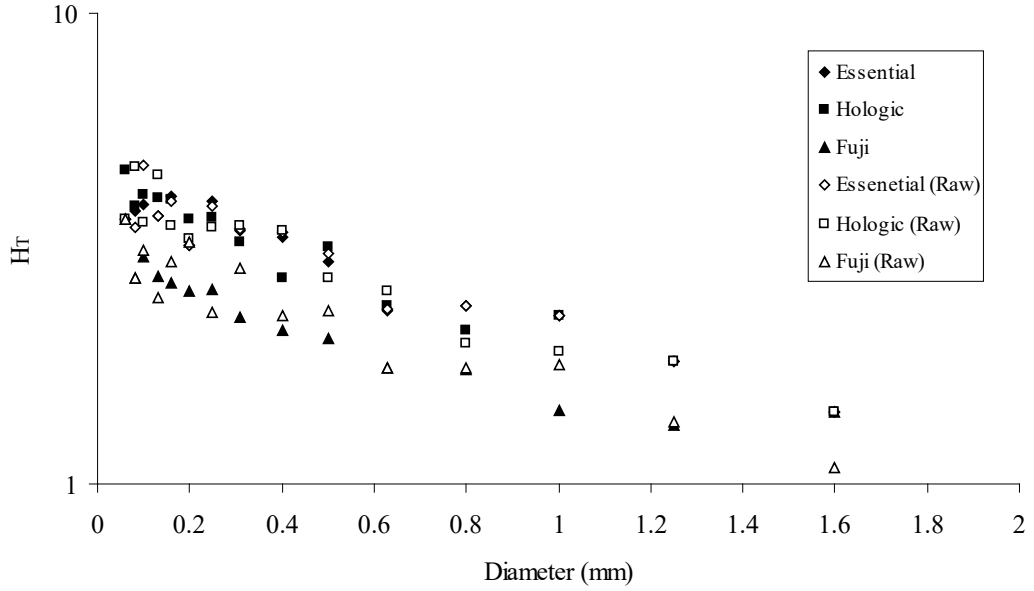


Figure 5.4. Plot of the threshold detection index (H_T) (mm^{-1}) against the CDMAM detail diameter.

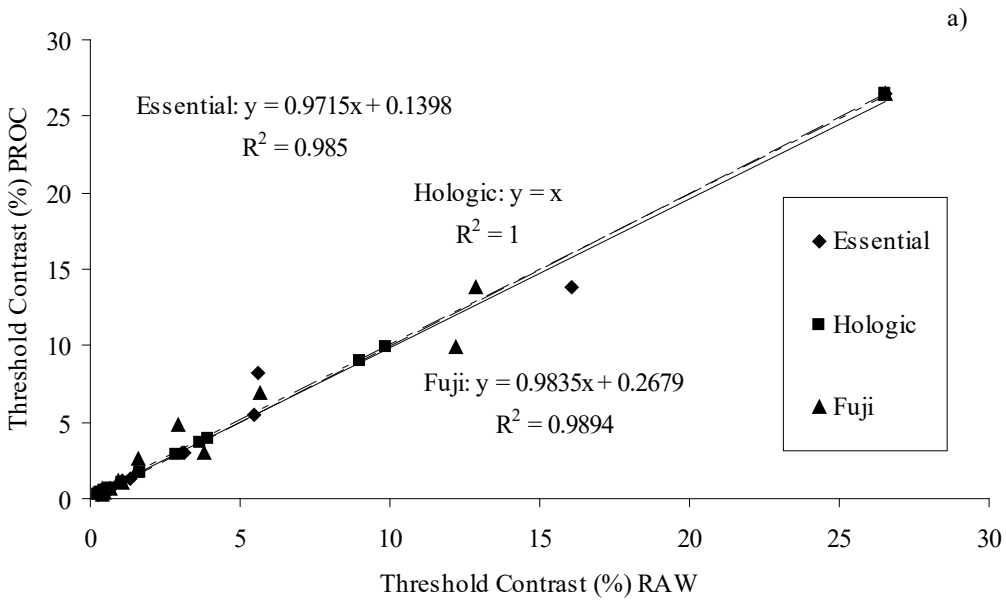


Figure 5.5a. Plots of raw against processed data for threshold contrast

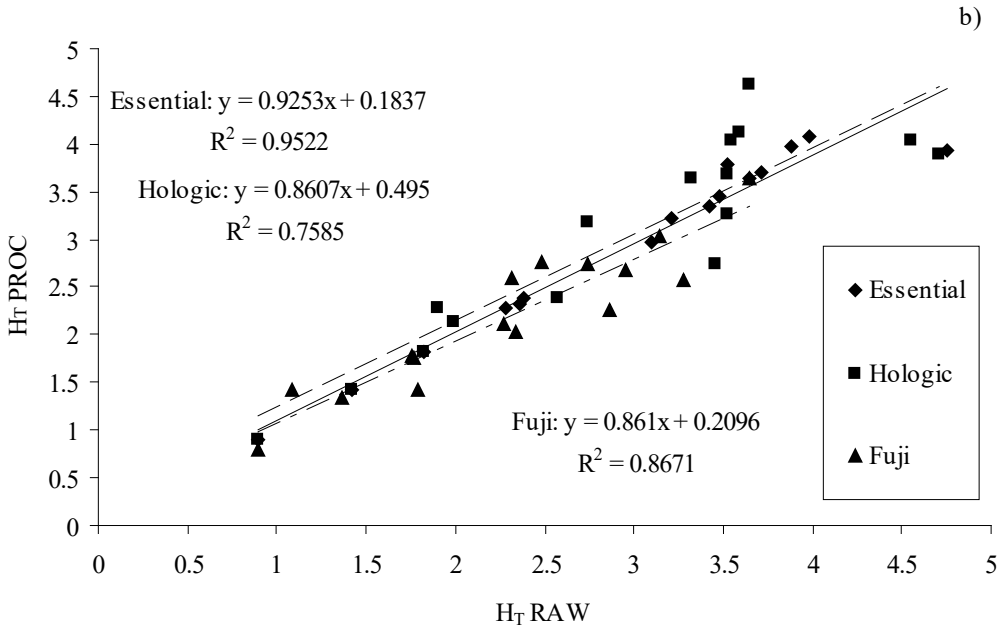


Figure 5.5b. Threshold contrast detail detectability index H_T

Table 5.2. Pearson r values of correlation between the raw and processed images for threshold contrast and H_T.

Pearson r values for threshold contrast			
	GE Processed	**HOL Processed**	**FUJI Processed**
GE Raw	0.992		
HOL Raw		1.000	
FUJI Raw			0.995
Pearson r values H_T			
	GE Processed	**HOL Processed**	**FUJI Processed**
GE Raw	0.976		
HOL Raw		0.904	
FUJI Raw			0.931

Table 5.3. Mean Image Quality Figure (IQF_{INV}) score and the standard deviation for the three systems using the CDMAM Analyzer application by Artinis.

Image Type	Essential	Hologic	Fuji
Raw	127.85±1.68	128.86±1.75	83.06±2.28
Processed	126.98±2.15	142.25±1.77	79.02±1.37

Detector Response Function

Raw images from both the Essential and the Hologic systems resulted in linear response functions both with R^2 values being very close to 1.0. With reference to the Hologic system, the manufacturer specified a pixel offset of 50 which was very easily replicated in this study using the raw data. The Fuji CR system using the processing methodology specified in the European protocol (L2 and S120) resulted in the expected logarithmic detector response with an R^2 greater than 0.90.

When the same tests were repeated using the typical clinical mode, the results achieved were completely different from those achieved using raw data. The response function achieved for the Essential using the processed data is strongly logarithmic, where the MPV decreased with increasing dose. In the case of the Hologic it appeared that on analysis of the processed images, MPV did not change with increasing dose (coefficient of variation 0.54%) but the SD decreased with a power relationship with increased K (R^2>0.99). Despite the mentioned differences, the relationship of the SNR^2 and K for both the processed and the raw data in the Hologic system was positively linear with R^2 values greater than 0.99. The only difference in the latter relationship was the magnitude of the SNR, being much larger in the raw data.

As with the Hologic, the MPV values hardly changed (coefficient of variation 0.98%) in the processed data of the Fuji system with increasing K. Consequently, the detector response function couldn't be calculated by plotting MPV against K. Similar to the Hologic, the SD also decreased with increasing K with a power relationship (R^2>0.98). When the SNR^2 was plotted with K both raw and processed data from the Fuji showed a strong logarithmic response function (R^2>0.91) with data from the raw images being again larger in magnitude when compared to the processed images.

A correlation analysis was performed between raw and the processed MPV values for the three systems. The best correlation was observed for the Essential (r>-0.97). In contrast, the correlation for the MPV between the raw and the processed images for the Hologic and the Fuji was poor (r = 0.48 and 0.14 respectively) given this did not change with increased K as stated previously. Consequently, a correlation analysis was performed between the SNR of the processed and the raw data for all the systems. It resulted that the best SNR correlation identified was for the Hologic (r = 0.98) followed by the Fuji system (r = 0.97) and worst SNR correlation was reported for the Essential (r = 0.78).

From these correlation analyses and results, the raw MPV values for the Essential (equation 31) and the raw SNR values for the Hologic (equation 32) and the Fuji (equation **33**) systems could be easily predicted from the processed values by using a nonlinear regression model using SPSS V20 as follows:

$$MPV(Raw) = 16765.897e^{-0.00121*MPV(\text{Processed})} \qquad 31$$

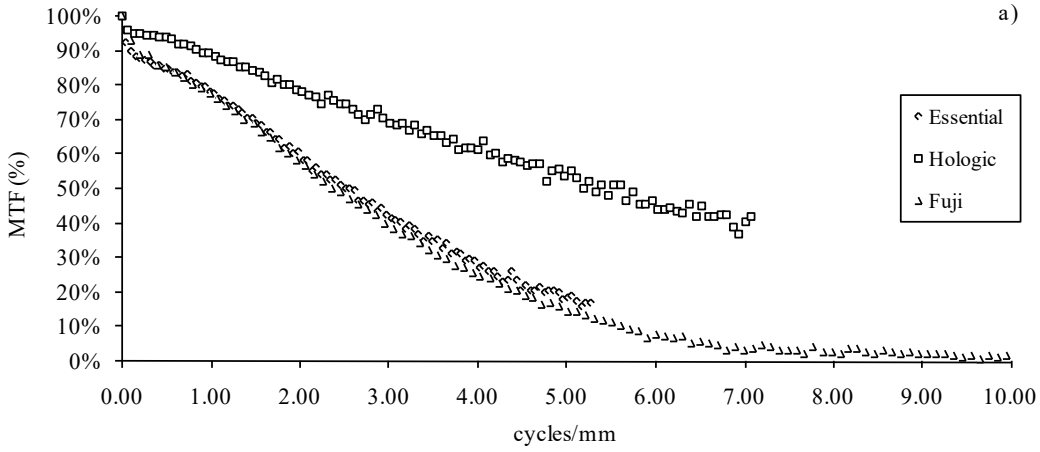
$$SNR(Raw) = 16.165 + (6.349 * SNR(\text{Processed})) \quad 32$$

$$SNR(Raw) = -47.453 + (13.343 * SNR(\text{Processed})) \quad 33$$

Any significant outliers were removed from the data to improve the regression models generated. The predicted MPV and SNR values as applicable for the raw images were plotted against the actual data from the 'real' raw images. The R^2 values achieved for all three system was better than 0.98, meaning that the above equations may be confidently used to predict raw data information from any processed image in the relevant system.

MTF, NNPS, NEQ and DQE

Plots of the MTF, NNPS, NEQ and DQE data have been plotted in figures 5.6, 5.7, 5.8 and 5.9. From these figures differences between the raw and processed images are more obvious. In fact, replication of specifications made by the manufacturers or comparison with other studies [112, 166] could only be done using the raw data. Nevertheless, for the scope of this study a correlation analysis was performed between the mentioned parameters for the three systems simply to accentuate the difference or identify any similarities between these two datasets.



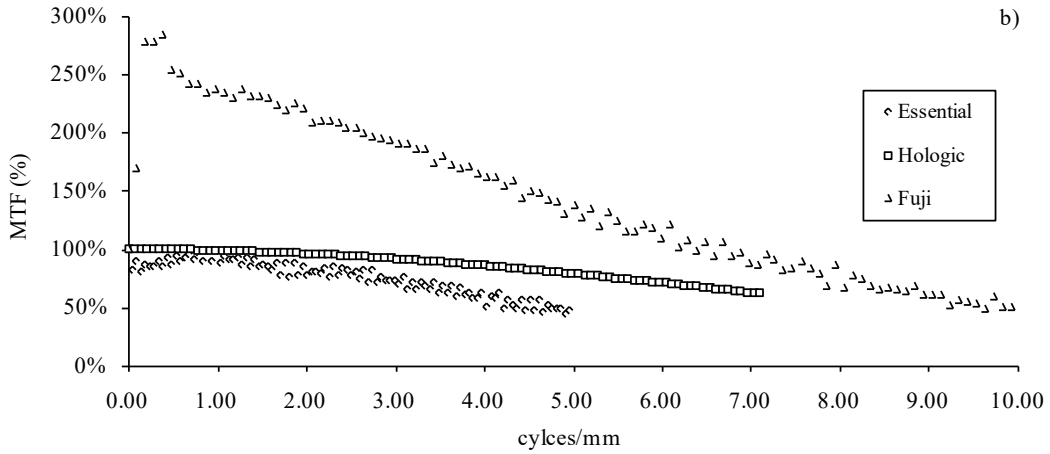


Figure 5.6. MTF of the three systems a) raw b) processed data.

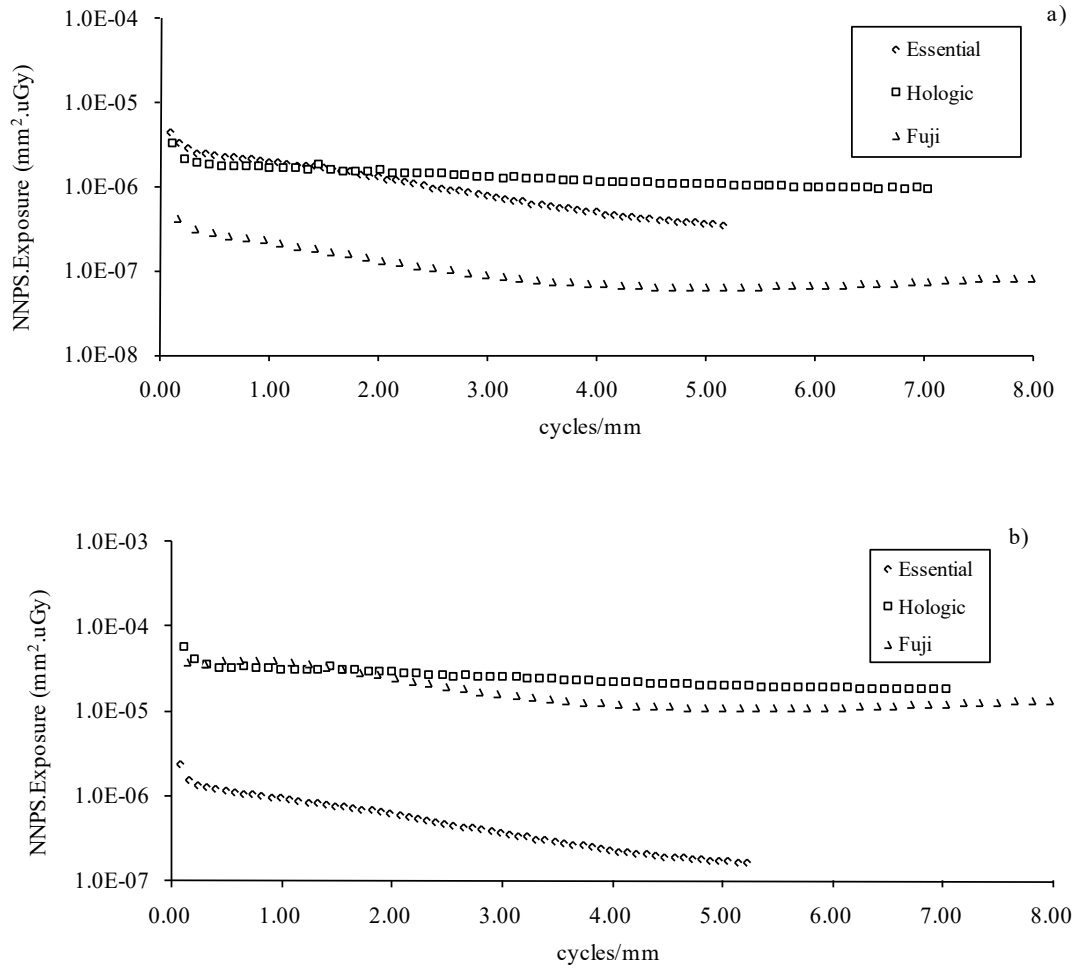


Figure 5.7. NNPS of the three systems a) raw b) processed data.

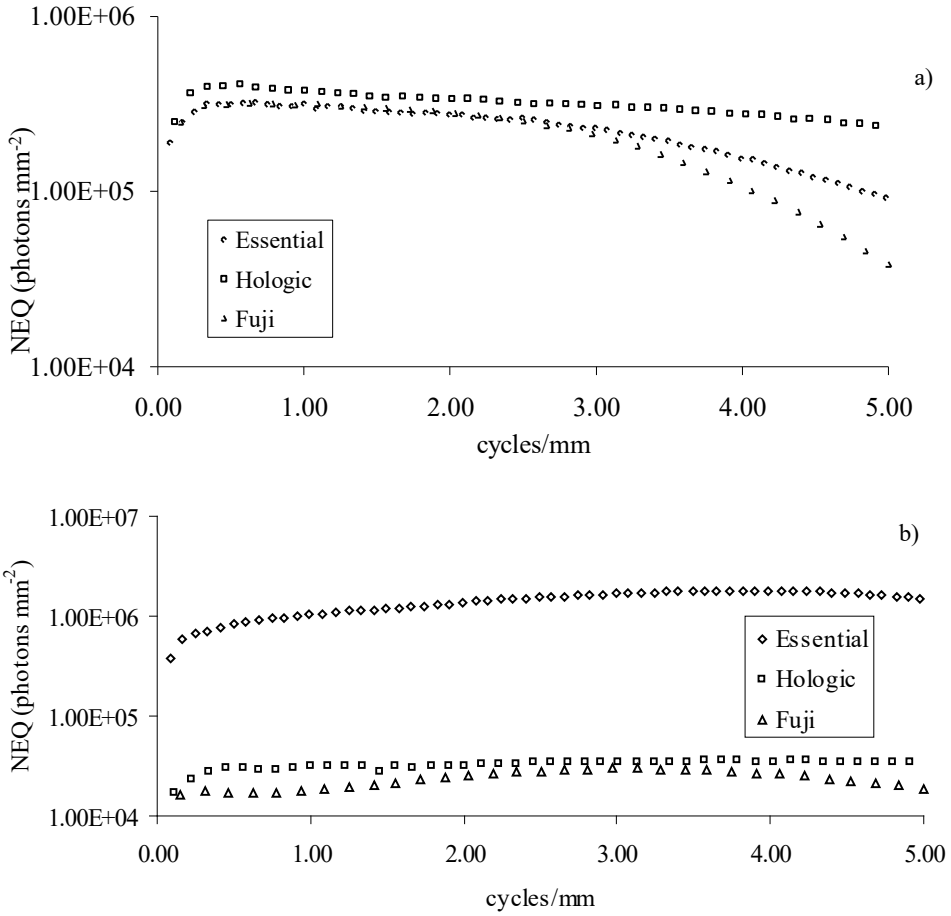
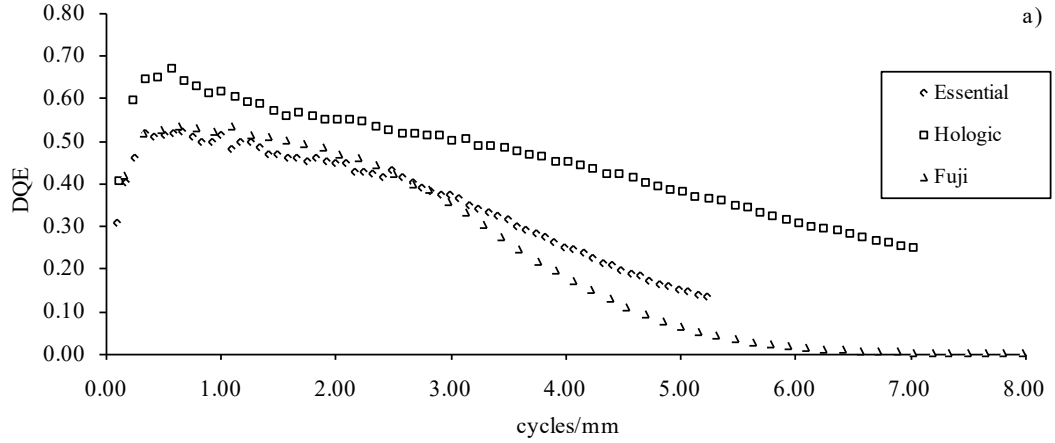


Figure 5.8. NEQ of the three systems a) raw b) processed data.



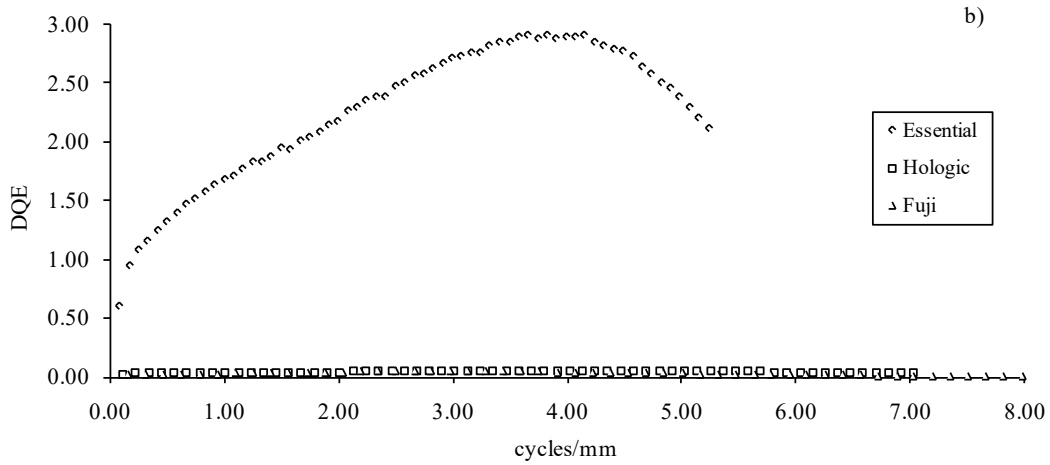


Figure 5.9. DQE of the three systems a) raw b) processed data.

A statistically significant positive correlation (r>0.85) was reported between the raw and processed data on MTF and NNPS for all three units (also refer to figures 5.10a and b). The best correlation for all parameters resulted in the Hologic system (r>0.96). In contrast, weak correlation results were observed in all three systems for the NEQ (r<0.69) and DQE data (r<0.60) being particularly poor in the Hologic. However, the results for both NEQ and DQE were not statistically significant (p = 0.33; 0.55).

Despite the poor results in DQE, the good correlation results with the MTF and the NNPS led to further regression analyses to predicted raw data from any available processed data. These prediction models are summarized in table 5.4. Any significant outliers were omitted from the regression analyses to improve the significance of the models generated. The Q-Q plots in figure 5.11 prove the reliability of the models in table 5.4. R^2 values for the MTF were all greater than 0.89 while for the NNPS the lowest R^2 value was scored by the Fuji system (0.77). However, the NNPS R^2 values for the Essential and the Hologic were both greater than 0.99.

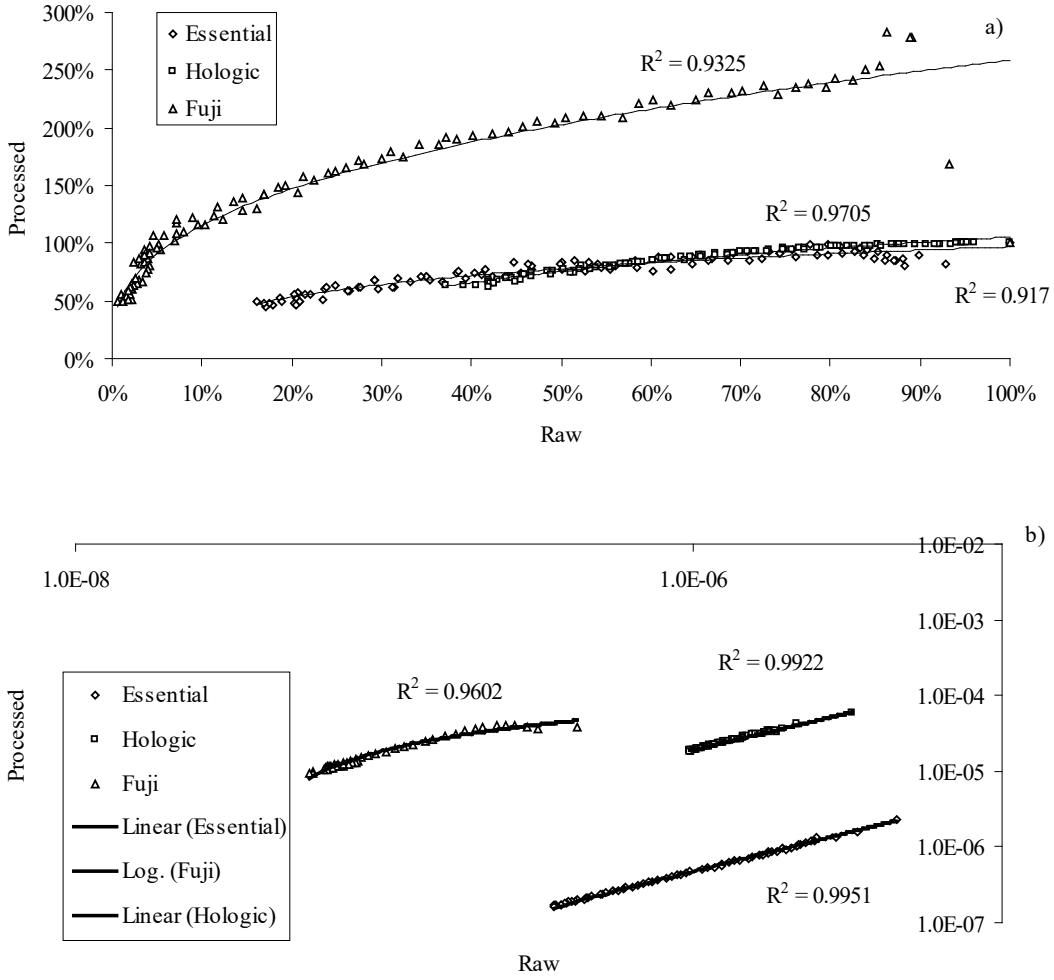


Figure 5.10.. Plots of a) MTF and b) NNPS between the processed and the raw data.

Table 5.4 Prediction MTF and NNPS regression models to derive raw data from the processed images.

	MTF		NNPS	
System	**Model**	**R^2**	**Model**	**R^2**
Essential	$MTF(raw) = 0.0012 * MTF(proc)^{2.449}$	0.89	$NNPS(raw) = 1.326E-8 + (2.087 * NNPS(proc))$	0.99
Hologic	$MTF(raw) = 0.00796 * MTF(proc)^{2.022}$	0.95	$NNPS(raw) = 1.331E-9 + (0.53 * NNPS(proc))$	0.99
Fuji	$MTF(raw) = 8.096E-5 * MTF(proc)^{2.496}$	0.98	$NNPS(raw) = 3.299E-8 + (0.005 * NNPS(proc))$	0.77

MTF – modulation transfer function; NNPS – normalised noise power spectrum; raw – data from the raw images; proc – data from the processed images.

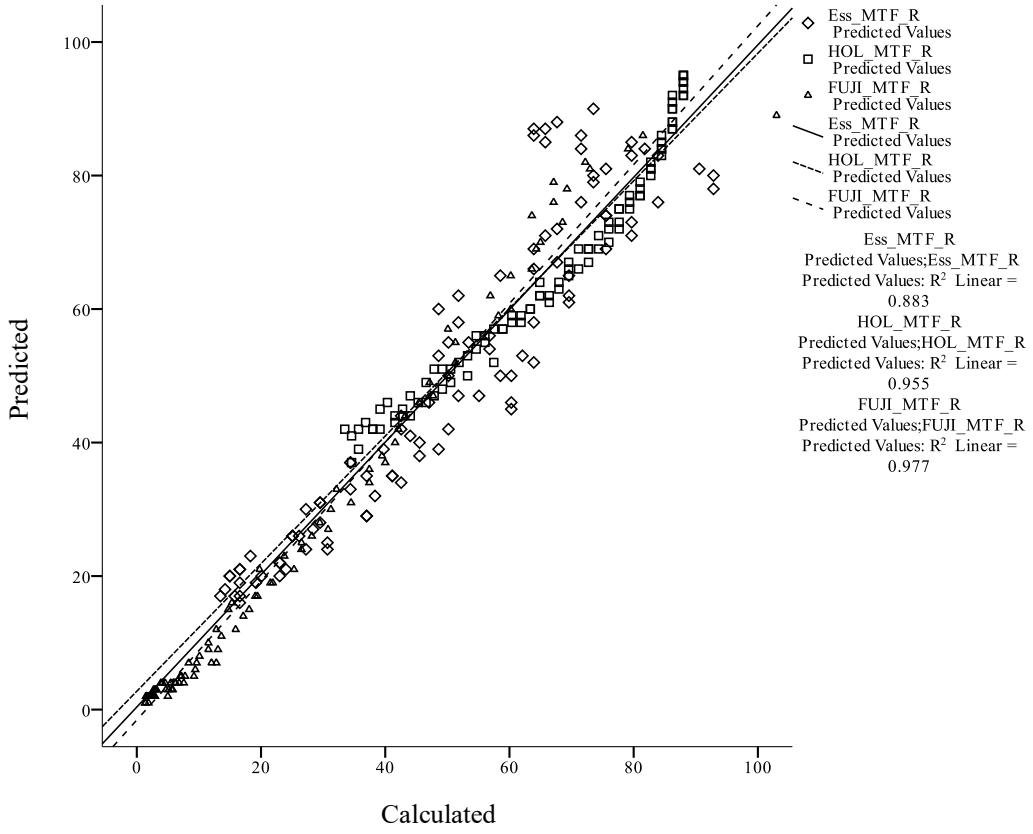


Figure 5.11a. Q-Q plots of Raw MTF predicted against calculated values

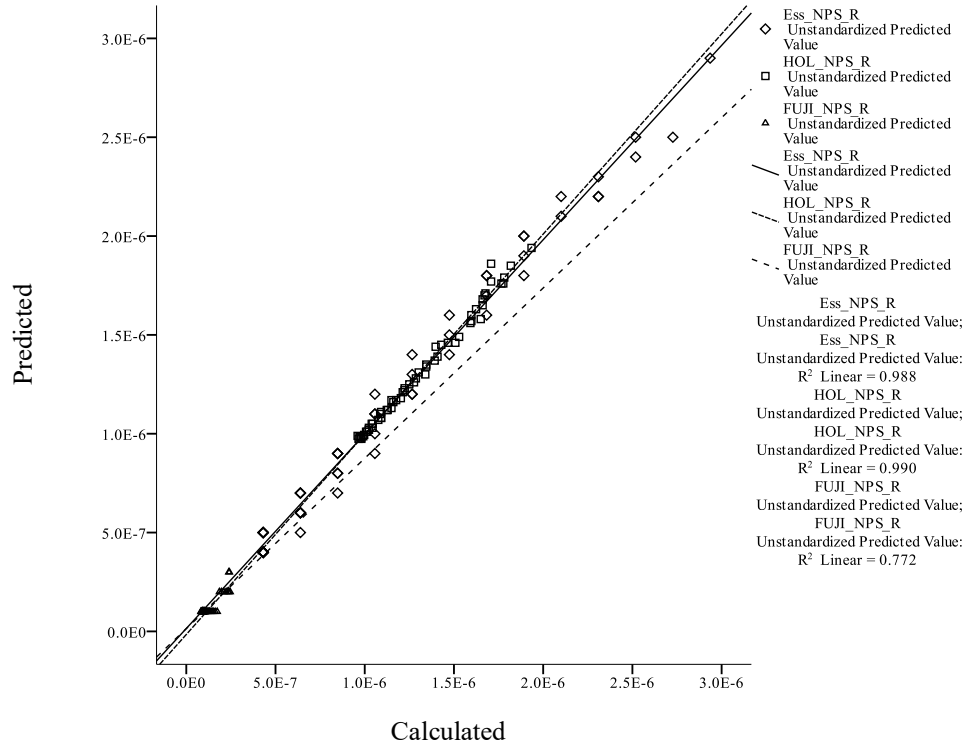


Figure 5.11b. Q-Q plots of raw NNPS predicted against calculated values.

NEQ and CDMAM Results

NEQ results were correlated with all the results achieved using the CDMAM phantom including threshold contrast, H_T and IQF_{INV}. Weak correlation results were observed in all three systems for the NEQ ($r<0.60$) again being particularly poor in the Hologic. However, the results for the NEQ correlations and CDMAM results were all not statistically significant.

Discussion

The European protocol recommends the use of unprocessed images for threshold contrast tests using the CDMAM phantom. Recommendations on image processing related to CNR and CNR-related tests such as thickness compensation are unclear. It is generally understood, and quite rightly so, that detector-specific image quality tests such as MTF, NEQ, NNPS and DQE are performed using exclusively raw image data. It is also true that these tests are optional in the European protocol. Further still, with reference to IEC 62220-1-2 [9], that is a better suited standard to describe detector characterization tests, mandates the use of raw image data. Also, other studies aimed to physically characterize digital mammographic detectors employed raw images [100, 112, 167].

In this work the Essential showed absolutely no difference between the processed and the raw images for CNR measurements. Apart from yielding a good correlation between the raw and processed images there was also no difference in the mean values between these two datasets. In the Hologic and the Fuji systems, there is also a notable correlation between the two datasets but, in contrast to the Essential, the measured values are completely different. Having said this, these findings can be only applied on these two systems and in no way should these results be generalized to all digital mammography units. Further still, any future software upgrades of these systems may influence to processing algorithms, which in turn are bound to influence the CNR results.

No particular conclusions can be drawn from the CNR measured using the CDMAM phantom. However, an interesting pattern or feature was noted in this dataset where CNR appeared closest for all kinds of data for all the three systems for two CDMAM detail groups. This is worth further investigation particularly when one considers changes in PMMA thickness. Such

future investigation may prove the CDMAM useful not just for TCDD but also for the calculation of CNR and consequently in the investigation of thickness compensation tests in any given digital mammography unit.

From the results, there is no statistical difference between the mean values of threshold contrast and H_T of raw and processed images of the three systems evaluated. There is also a very good correlation between raw and processed threshold contrast and H_T data. Of particular interest were the IQF_{INV} results in table 5.3. Although there was no particular difference in threshold contrast results there was notable difference in the IQF_{INV} score of the Hologic system between the raw and the processed data evaluations. Conversely, there was no significant difference in IQF_{INV} between the raw and processed images in both the Fuji and the Essential. As with the CNR results, there is still not enough conclusive information to use raw and processed images interchangeably for threshold contrast and H_T analysis. Nevertheless, such results give an insight on the magnitude of differences, if any, between raw and processed data.

Detector characterisation test results specified by manufacturers were met and successfully replicated using image raw data. The same can be stated for results from other studies [112, 166]. It is very well-known and accepted by medical physicists that such tests require the use of raw data but the effect of image processing has been very rarely shown. As expected detector response function, MTF, NEQ, NNPS and DQE did present diverse results when comparing the raw and processed data. However, surprisingly, the two datasets correlated and it has been shown that when knowing one set of data the other can be predicted with confidence for the systems in this study. The only lack of correlation was observed in the DQE and NEQ values and consequently, a regression analysis is not recommended.

Conclusion

This study shows that differences between raw and processed image data is system-specific and cannot be generalized to all digital mammography units. In detector characterization tests, differences between the raw and processed data are more dramatic and significant. IEC standards and manufacturer recommendations, if present, in general recommend the use of raw data for such evaluations. In fact, manufacturer recommended specifications could be

only replicated by using raw images. Mathematical conversion models between the raw and processed data have been proposed in tests including detector response function, MTF and NNPS, owing to the extremely good correlations. The predicted values are very similar to the actual raw values from the various regression analyses presented in the study. However, the calculation of DQE and NEQ resulted in very poor correlation results. Although, various results throughout this study showed similarities between the processed and the raw image data, it is still strongly recommended to use raw images for image quality analyses. In conclusion, quantitative image quality analyses in digital mammography should be done using raw image data only, especially when performing detector characterization tests. The use of processed images may results in completely wrong or unpredictable values and are inconsistent due to possible future software upgrades in any particular system.

CHAPTER 6

THE FIGURE-OF-MERIT ANALYZED FOR TWO FULL FIELD DIRECT DIGITAL MAMMOGRAPHY UNITS IN MALTA

Introduction

Mammography is the gold standard used for the early detection of breast cancer although there are complementary modalities such as breast-self examination, ultrasound and clinical examination, since it is able to detect abnormalities that are asymptomatic. The introduction of digital mammography has been relatively slow when compared to other imaging modalities due to the high demands on image quality and low dose. It is generally assumed that the glandular tissue of the breast is most vulnerable to the induction of cancer by ionization radiation [168, 169]. As stated in the European protocol [1, 2], the MGD is the quantity of radiation risk related to induced carcinogenesis [124]. Periodical quality control of the system is essential for obtaining high quality mammograms to detect any lesions or microcalcifications that suggest potentially malignancy.

Therefore optimization of the exposure parameters with respect to dose and image quality using the AEC has become one of the current issues in digital mammography as the AEC is one the key components of FFDM. For qualitative assessment of image quality, several types of parameters and scoring systems using a breast phantom have been employed [68, 103, 170]. However, variance in the evaluation of the images may exist depending on the evaluation procedures or viewing conditions, even when the phantom images are viewed by the same person [171, 172]. Besides, it is generally acknowledged that image quality can be assessed quantitatively by CNR or SDNR as said in previous sections, however CNR as commonly defined for mammography equipment is not an absolute quantity, but its value is a range and is manufacturer and system-dependent. Several papers were published showing that square of the CNR divided by MGD in equation 6, termed the *figure- of-merit*, is a good parameter for optimization of FFDM systems [94, 95, 98, 173, 174].

Background

Digital mammography has the advantages of high detection efficiency, wide dynamic range, and the separation of image acquisition from display. Unlike FS mammography, digital mammography permits the use of a higher energy spectrum for optimal results so patient dose reduction is possible [175]. Although the contrast can be further adjusted on the display stage, the detectable contrast is limited by the SNR. Meanwhile, the SNR cannot be arbitrarily increased due to the constraint of patient dose [176]. The selection of the energy spectrum is therefore important in order to balance image quality and dose.

There were numerous studies on finding the optimal spectrum for digital mammography [11, 12, 45, 94, 106, 177]. In a phantom study, Samei et al proposed a FOM using the squared CNR normalized to K to compare the systems with different target/filter combinations [94]. Bernhardt et al utilized a different FOM by replacing the entrance exposure with the MGD in a Monte Carlo simulation and phantom study [177]. They concluded that the W/Rh combination works the best when compared with Mo/Mo and Mo/Rh. Ranger et al applied the FOM normalized to the MGD to the slab phantoms with embedded masses and microcalcifications [11]. It was found that W/Rh performed better than Mo/Mo. Baldelli et al concluded that for each phantom thickness and target/filter combination, there is an optimum voltage [12]. It is believed that these investigations might have helped design better digital mammography systems. However, it is difficult to implement the FOM-based approaches in the current AEC devices because the FOM quantification for the lesions are not available in advance. Instead, the objective of the AEC is to achieve a preset digital signal level in a certain ROI with different kVp values and filter choices [111]. Typically, several AEC look-up tables reaching the same signal level are stored in the commercial system and a specific look-up table is preloaded by the manufacturer. The challenge for the radiographer is that it is not clear whether the preloaded table optimizes the SDNR and dose efficiency [111].

Materials and methods

The exposure regime for Essential included a spectrum of 29 kVp and Rh/Rh combination. While for the Hologic a spectrum of 30 kVp and W/Rh combination was used. These spectra where also selected by the systems when using AEC for the imaging of 50 mm of PMMA. The

standard AEC exposure was labelled 'EN'. With reference to figure 6.1, subsequent exposures involving image quality, detector characterization and the FOMs where taken at EN and from EN divided by 4 to EN multiplied by 3 for each system. Radiation dose and exposure was measured with an Unfors Xi meter with dedicated mammography probe [130]. From the calibration certificate, the relevant uncertainties for kVp measurements were 0.7%. For dose, the uncertainties were 2.4% (W/Rh) and 2.5% (W/Ag). The calibration certificate states that the reported expanded uncertainty is based on a standard uncertainty multiplied by a coverage factor k=2, which is for normal distribution, provides a level of approximately 95% confidence (ISO/IEC Guide 98-3:2008). kV meters are calibrated by the Swedish National Testing and dose meters are calibrated by PTB and are traceable to NIST where applicable.

The mentioned exposures were coded and used to expose the CDMAM phantom sandwiched between 40 mm of PMMA, yielding a total equivalent thickness of 50 mm of PMMA. The same equivalent thickness of PMMA was used to calculate the CNR using the conventional method as described in the European protocol [2]. This procedure was repeated with small adjustments to the phantom position to obtain a representative sample of 8 images for each setup. The CNR was calculated using PMMA slabs similar to those employed by the CDMAM phantom to avoid the influence of difference in the material compositions. In addition, the CDMAM phantom itself was used to calculate the CNR using the 0.36; 1.6 mm gold detail (figure 6.2). This was done to verify whether the CDMAM phantom could be used reliably within the scope of this study instead of the conventional method specified in the European protocol and eventually whether this information can contribute towards further research in terms of the defined FOM in table 6.1b. These FOMs were based on the established relationship between MGD and the respective image quality parameters summarized in table 6.1a derived from plots and regression analyses between dose and image quality parameters.

All known CDMAM-related image quality indicators were calculated from the mentioned exposures to the phantom. These included IQF_{INV}, threshold detection index (H_T), described in equation **3**, and the threshold contrast for each gold detail. The mentioned quality indicators were estimated using 8 images for each setup. This was done to ensure repeatability and minimize experimental uncertainty. The relevant values presented represent the mean values and the standard deviation. In order to minimize human interaction and subjectivity, the CDMAM Analyzer software [121] was employed for the scoring of the acquired phantom images which makes use of the CDCOM analysis core from EUREF.

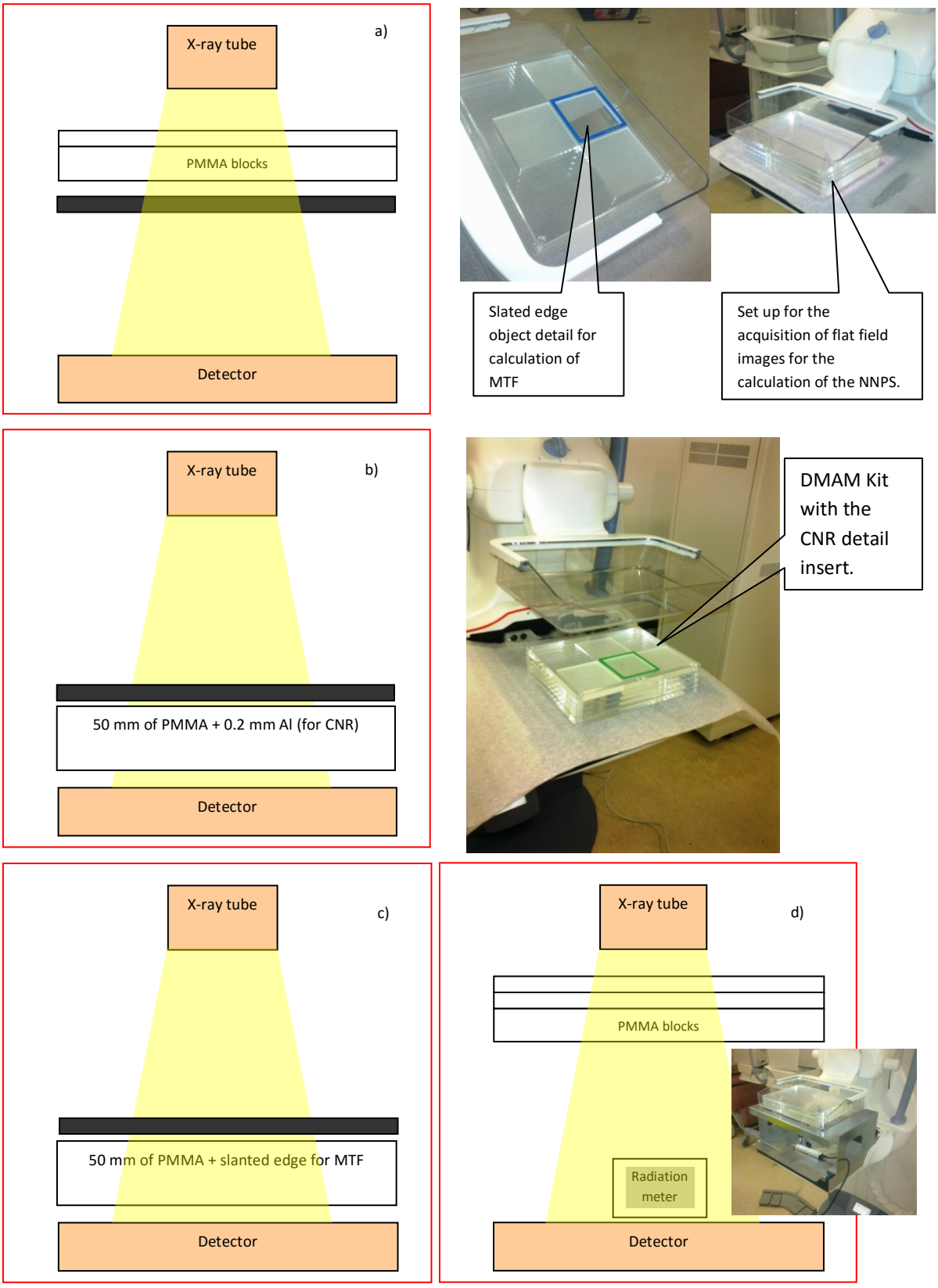


Figure 6.1. Experimental setup a) exposure determination b) CNR c) MTF d) detector dose measurement

Detector characterisation parameters such as MTF, NNPS and DQE were also calculated using the same experimental setup described in figure 6.1.

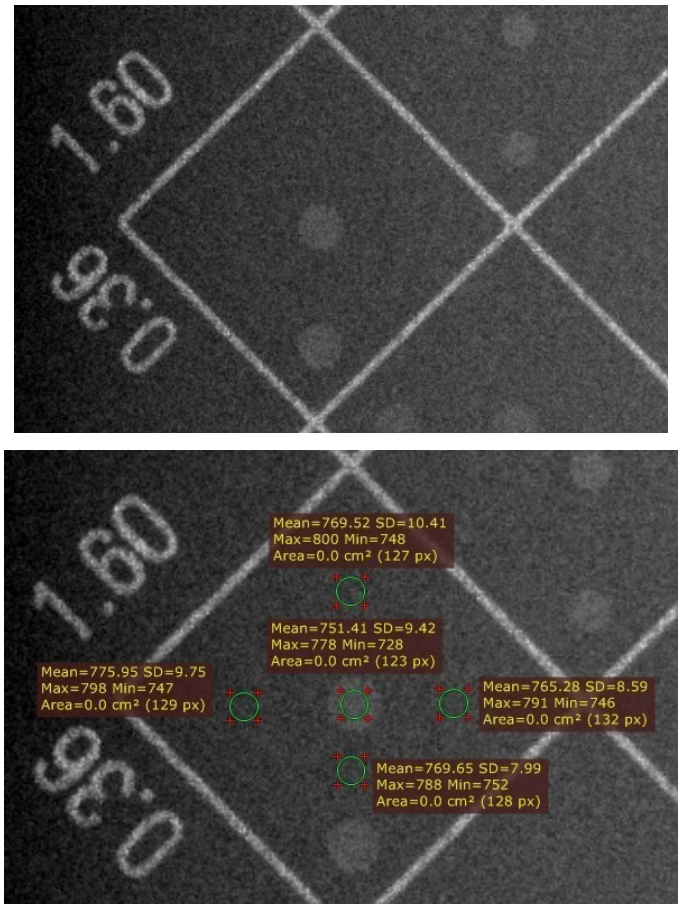


Figure 6.2. CDMAM phantom showing the 0.36; 1.6 mm gold detail for CNR calculation

The image quality indicators calculated using the CDMAM phantom show how good the image quality is for a particular exposure. Although there are no manufacturer-defined tolerance values for image quality, the values achieved are relative and may still be useful in indicating the image quality status of any particular system. This is very useful in constancy testing and in routine quality assurance. Further still, the European protocol provides acceptable and achievable limits for threshold contrast using the CDMAM phantom, assuming an exposure of 28 kVp Mo/Mo. Contrast resolution, including threshold contrast detail detectability and CNR are closely related to dose, whereby higher dose values to the detector and hence higher MGD to the patient yield better image quality.

Consequently, MGD and image quality are antagonistic factors whereby any one factor should be never seen in isolation of the other. A good performing system is one which yields good image quality at the lowest possible dose, achieved through optimization.

Table 6.1a. Relationship between the MGD and Image Quality Parameters

Fitted Equation	Proportionality
$MGD = 0.0068 * CNR^{2.33}$	$MGD \propto CNR^{2}$
$MGD = 1E-7 * H_T{}^{4.22}$	$MGD \propto H_T{}^{4}$
$MGD = 1E-5 * IQF_{INV}{}^{2.39}$	$MGD \propto IQF_{INV}{}^{2}$
$MGD = 12661 * Qfactor_1{}^{4.22}$	$MGD \propto Qfactor_1{}^{4}$
$MGD = 0.192 * Qfactor_2{}^{2.39}$	$MGD \propto Qfactor_2{}^{2}$

The various FOM defined in table 6.1b, are seen as simple quantitative means or methods of potentially comparing one system to another and also in reaching an optimized state of any given mammography unit. There are no rigid rules governing the construction of a FOM, but in general it is useful to distinguish desirable attributes from undesirable attributes. It is often most intuitive and convenient to express the FOM as a ratio. Placing desirable attributes in the numerator and undesirable attributes in the denominator forms a simple FOM for which the highest value is best, however other designs could be valid too. The philosophy behind this work was to identify, 'guess' and propose possible FOM in this area, possibly to elicit further research and evaluation. Groundwork on the relationship between dose and the various image quality indicators using the CDMAM phantom are summarised in table 6.1a. Consequently, this helped in formulating the equations in table 6.1b, particularly in defining FOM 1, 2, 10, Qnorm1 and Qnorm2. At this stage, with reference to chapter 5, it was decided that raw or unprocessed images in DICOM format will be employed for this FOM exploration. Nevertheless, processed images may be considered in the future. Since, users have limited control on inherent image processing established by the manufacturer and also due to the fact that image processing has a fundamental role in the image quality of the final image, it may be worth investigating this matter further by including processed images rather than raw images in the future.

SPECTRUM-SPECIFIC CONTRAST LEVELS

The European protocol [2] and the relevant NHSBSP documents [24] make reference to acceptable and achievable values for the threshold contrast of particular gold details for the CDMAM phantom. However, AEC exposures and the resultant spectra at the specified thickness of 50 mm of PMMA (and equivalent to the CDMAM phantom with 40 mm of PMMA) may differ from system to system when compared to 28 kVp Mo/Mo as specified in these

documents. A case in point is the exposure used (EN) in the Essential using 29 kVp Rh/Rh and the Hologic employing 30 kVp W/Rh.

Table 6.1b. Figures of merit included in this study.

FOM	Definition
1	CNR^2/MGD
2	IQF^2/MGD
3	CNR/MGD
4	IQF/MGD
5	CNR/ESAK
6	IQF/ESAK
7	CNR^2*/MGD
8	CNR*/MGD
9	CNR*/ESAK
10	$1*10^7(H_T^4/MGD)$
Qnorm1	$1*10^5(Qfactor_1^4/MGD)$
Qnorm2	$Qfactor_2^2*100/MGD$

CNR – contrast to noise ratio; MGD – mean glandular dose; ESAK – entrance surface air kerma; IQF – image quality figure acquired from CDMAM analysis; CNR* - contrast to noise ratio acquired from the 0.36;1.6 mm CDMAM detail; H_T – Detection Index.

Consequently, it is envisaged that the other mammography units will make use of different spectra under AEC control. As a result the acceptable and achievable limits hereby defined are not spectrum-specific but merely an indication of adequate performance. It is believed that by calculating spectrum-specific contrast levels for any given gold detail within the CDMAM phantom, acceptable and achievable levels may be redefined and made specific for any given spectrum and hence for any given system. The tube angle and the inherent filtration also influence the resulting contrast levels between one system and another and will be considered in these calculations. The spectrum-specific gold detail contrast levels calculated are believed to be useful for the calculation of FOM10, Qfactor1 and consequently Qnorm1.

The contrast levels for the different gold detail thicknesses and diameters for both spectra are summarised in table 6.2. The spectra and the related contrast levels were estimated using IPEM report 78. The same process should be repeated for any system that needs to be evaluated [100]. The calculation of these contrast levels may be also used by physics staff in related quality assurance tests mentioned in the European protocol e.g. thickness

compensation, enhancing the specificity and accuracy of the results using the CDMAM phantom. Any future spectral calculations may be performed using other Monte Carlo-based applications or methods

Table 6.2. a) Calculated contrast levels for the GE Essential (29 kVp Rh/Rh) using IPEM 78

CDMAM Detail Thickness (μm)	mm	Calculated Contrast (%) from IPEM 78
0.1	0.0001	1.33
0.5	0.0005	6.67
1	0.001	13.64
1.5	0.0015	18.06
2	0.002	23.08

b) Calculated contrast levels for the Hologic (30 kVp W/Rh) using IPEM 78

CDMAM Detail Thickness (μm)	mm	Calculated Contrast (%) from IPEM 78
0.1	0.0001	1.32
0.5	0.0005	6.74
1	0.001	13.84
1.5	0.0015	21.35
2	0.002	29.23

.

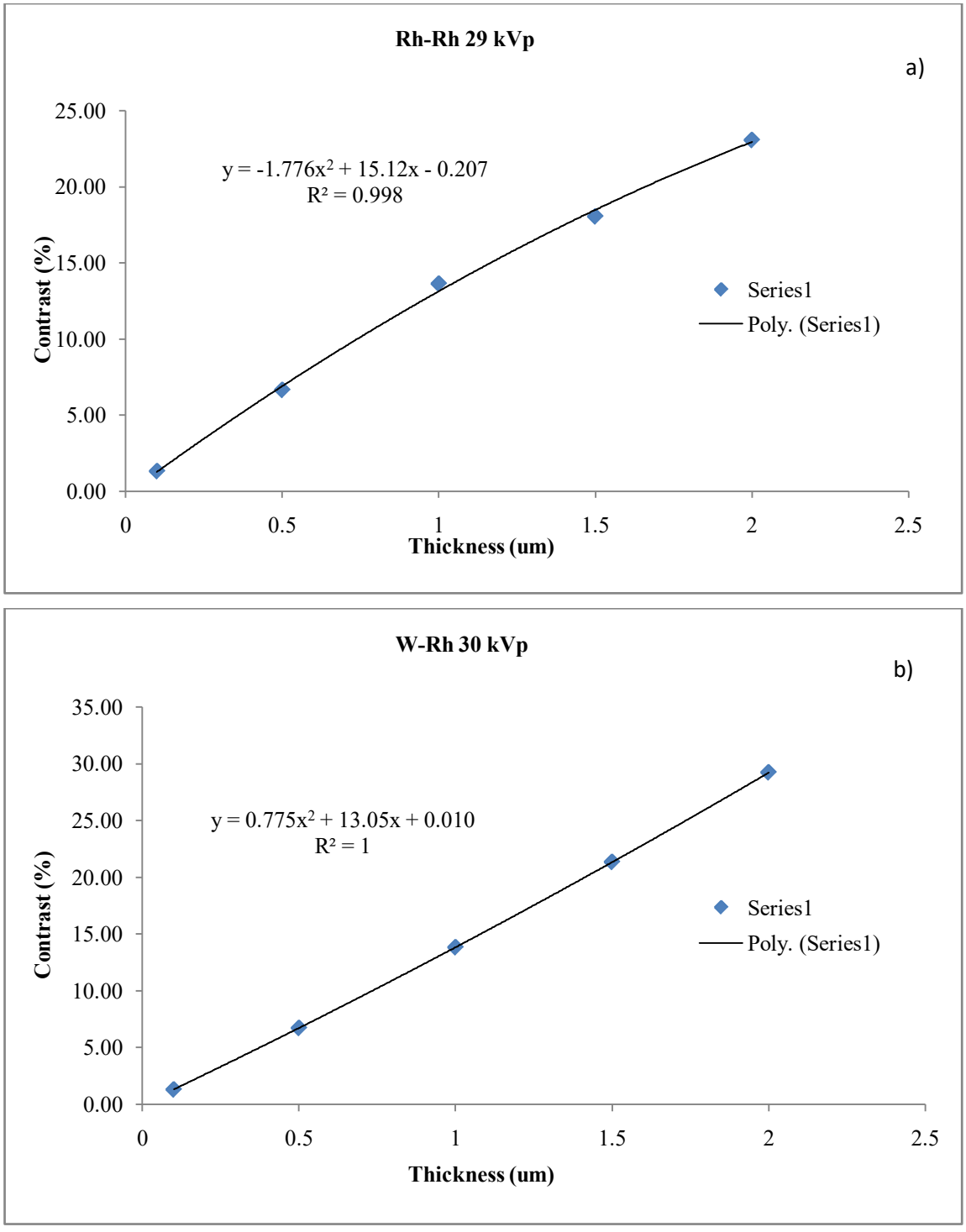


Figure 6.5 Calculated Contrast (%) for each CDMAM gold detail using IPEM 78 a) Essential b) Hologic

Summarised method for the calculation and estimation of spectrum-specific contrast levels using IPEM report 78

- Radiation contrast (%) levels for each CDMAM gold detail were re-established for the Essential and the Hologic and summarised table 6.2.
- The anode angles and the filter thicknesses were confirmed by the field engineers.
- To the above materials, 0.65 µm of Beryllium were added as attenuation material to account for the inherent filtration.
- It was assumed that the CDMAM was totally composed of 50 mm of PMMA as recommended in the literature reviewed.
- The specified ***I*** (incident air kerma) values were simulated using IPEM 78 i.e. as if the CDMAM had no details incorporated. The ***I(D)*** (transmitted air kerma from detail D) values were then simulated with the specified CDMAM gold detail thicknesses. Consequently, the contrast of each detail was estimated.
- The data calculated was plotted in figure 6.5.
- A second order polynomial fit was applied. The resulting R^2 value was very high for both systems, implying that the equation for these regression lines can be used confidently in predicting the contrast values for other detail thicknesses.

DETECTOR RESPONSE

The detector response curves are required for linearization of the DICOM greyscale images prior to calculation of the image quality parameters such as CNR. Although these two systems have a linear response, linearization removes any offset that may be present that would confound measurements of DQE and effective DQE (eDQE). Detector response was measured for the AEC spectra selected by the system at 50 mm PMMA. A range of mAs values were then selected manually such that the detector dose is ranged from low values to extremely high values. The antiscatter grid was removed for the detector response measurements and the K at the detector was measured with an Unfors Xi dosimeter. This is a solid state detector that has been calibrated explicitly for Mo/Mo, Mo/Rh, W/Rh, and W/Ag target/filter combinations; calibration is traceable to a national standard. MPV were measured using a 5 x 5 mm ROI placed 6 cm from the chest wall edge and centred left-right, from DICOM raw images. The MPV was then plotted against K at the detector to give the detector response curve for the

given combination for each system, namely the Essential and the Hologic. The response function was then used to linearize the MPV readings.

MTF, NNPS and DQE

The MTF, NNPS and DQE were measured as stated earlier in the relevant sections on page 137. The main difference in methodology used was that instead of 2 mm of Aluminium filter as mandated by IEC [9] these detector characterisation tests were measured using scatter conditions i.e., 50 mm of PMMA, as for the CDMAM and SDNR measurements. The presampling MTF (pMTF) was measured using the slanted-edge method [94, 178]. The phantom used forms part of the DMAM kit, Leeds Test Objects. First, the edge was placed on top of the 40 mm PMMA blocks and the detail itself was placed within a 10 mm PMMA slab to give a total thickness of 50 mm of PMMA (figure 6.1c). The NNPS and subsequently the DQE data were then collected using the experimental setup explained in figure 6.1. The exposures described on page 159 were then used. From this setup and data, pMTF, NNPS and DQE were calculated using a Matlab routine designed by Konstantinidis [13].

EFFEFECTIVE DQE AND EFFECTIVE NEQ BACKGROUND

The quantities of eDQE and eNEQ (e meaning 'effective') can be seen as a development of the work in previous studies [173] [179] with application to digital radiographic systems. The eDQE is an attempt to broaden the notion of the detector metric of DQE to include system parameters such as focus blurring and system scatter rejection methods. This chapter and work applies the eDQE and eNEQ methodology to a digital mammographic system. The formulation of eDQE attempts to compare a system against an ideal imaging system with perfect X-ray detection and scatter rejection, however a patient dose marker is not included at first. Eventually, the eNEQ is normalized for MGD and was included in this work namely, the $eNEQ_{MGD}$ and described in equation **40**. The system configurations that produce high eDQE may not be the best clinical choice since the patient dose associated with parameters that give a high value eDQE could be higher than the patient dose of configuration that is *suboptimal* in terms of eDQE. The $eNEQ_{MGD}$ describes the impact of the patient dose on the chosen configuration but this term does not give any indication on object detectability. As an example, lower patient dose configurations can lead to higher $eNEQ_{MGD}$ without leading to higher object

detectability in images. For this reason, it was decided to extend $eNEQ_{MGD}$ with an object contrast (C%) that is characteristic for the studied configurations, and named this quantity $eNEQ_{DC}$, described in equation **41**. This was done using both the 0.2 mm Al detail as described in the European protocol and the CDMAM phantom as described in figure 6.2.

Effective DQE

The eDQE was evaluated for both systems for 50 mm of PMMA and evaluated for the exposures mentioned on page 159. The equation used was equation **46** where *u'* is the spatial frequency scaled to the object plane, MTF(u') is the presampling modulation transfer function, SF is the scatter fraction, NNPS(u') is the normalized noise power spectrum, TF is the narrow beam transmission factor of PMMA, *E* is the prephantom exposure corrected to the detector plane, and *q* is the number of photons expressed as Gy^{-1} mm^{-2}.

Equation **46** is written such that it is analogous to the standard equation for detector DQE in equation **29** but with some important differences. First, the MTF in the numerator is measured at the object position to include X-ray source geometric unsharpness. The term $(1\text{-}SF)^2$ corrects for contrast loss due to scatter and hence the numerator can be thought of as the total system MTF. In the standard equation for DQE, the term K x q gives the total number of X-ray photons per unit area at the detector entrance plane. In the equation for eDQE, this is replaced by the term E x TF x q and thus the detector-grid-cover combination is considered as the "image receptor plane" and correction is only made for object transmission (and not grid transmission) when estimating the number of photons at the image receptor.

Scatter Fraction (SF) and Transmission Factor (TF)

The SF was estimated using the beam stop method [174] using six Pb squares of dimensions ranging from 4 to 30 mm. The primary radiation component was estimated by placing the PMMA at the tube using narrow beam geometry to produce a low scatter image. Using the same configuration, the Pb squares were positioned in the beam at the detector and the pixel values behind the disk were used to estimate the detector glare component. Finally, full field images were acquired with the PMMA placed at the detector to give the total signal (primary, glare and scatter components). The total signal was calculated as the MPV in the same ROI where primary radiation and glare were estimated. Consequently the MPV of the scatter free configuration images were subtracted from and the MPV representing detector glare from the

total signal to give the scattered radiation component. The narrow beam TF was then measured by placing the Unfors Xi on the detector cover and collimating the X-ray beam to be slightly larger than the radiation meter/detector. The PMMA blocks were placed at the X-ray tube. K was than acquired for all the exposures mentioned on page 159 with and without the PMMA in the beam. The ratio of the measured air kerma with and without PMMA gave the narrow beam TF.

Pre-phantom exposure and number of photons

The term *pre-phantom exposure* refers to the E in equations **29** and **46** and was estimated as follows. First, X-ray tube output per unit mAs was measured for the spectrum using EN (the spectrum given using the AEC). The setup is explained in figure 6.1d. The desired mAs value was then predicted for each system for the wanted K value. This was then corrected by the inverse square law used to give the pre-phantom exposure corrected to the detector entrance surface, E.

Finally, q, the number of photons in Gy^{-1} mm^{-2} was calculated using the data of Boone et al [180] and IPEM report 78. The photon spectrum of interest (kV, target/filter setting, including added system inherent filtration and the added PMMA were simulated and then the corresponding air kerma was calculated for this X-ray beam using the photon-to-air kerma conversion factors [181, 182]. Normalizing for air kerma gave the number of photons per unit air kerma (in units of Gy^{-1} mm^{-2}) for the X-ray spectra and exposures being investigated.

FIGURES-OF-MERIT

THE QUALITY (Q) FACTOR AS A FOM

The Qfactor1 and Qnorm1

The gold detail thicknesses detected using the CDMAM Analyzer software for each CDMAM detail diameter, for each exposure, were converted to contrast using the equation in figure 6.5 as follows:

$$Contrast\ (\%) = -1.78x^2 + 15.13x - 0.21 \qquad 35a$$

$$Contrast\ (\%) = 0.78x^2 + 13.05x + 0.01 \qquad 35b$$

where *x*, is the gold detail thickness. With reference to page 163, the new spectrum-specific contrast values for the Essential and the Hologic for each gold detail were used for the calculation of H_T for each particular exposure defined in equation 34.

$$H_T = \frac{1}{Contrast * \sqrt{Area\,(\text{detail})}} \qquad 34$$

The calculation of the H_T is useful in comparing exposures between any particular system for the optimization of techniques for instance between the AEC exposure and a range of manual exposures. Eventually, the H_T values for each CDMAM exposure were summed and averaged for their use described later on in this section. It is believed that by using the H_T summed or averaged, image quality evaluation for any particular exposure is more complete as both detail thickness (contrast) and diameter (detail) are being considered. A higher H_T, implies that a detail is more easily visible, consequently a larger total or average H_T value means that more details are visible for a particular CDMAM exposure. Typically, with increased dose, the resulting total H_T values increase.

This relationship is strongly logarithmic as seen in figure 6.6. The summed H_T value for each CDMAM exposure was eventually used as ***numerator ($H_T(Ai)$)*** in the calculation of the Q factor defined in equation 37, adapted from the study by Dragusin et al [14] based on cardiology systems employing fluoroscopy.

With reference to table 6.3, the acceptable and achievable levels specified in the European protocol were used to define target reference values for the calculation of the Q factor in equation 37. As previously stated these reference values are set using 28 kVp Mo/Mo, hence it is worth revising the mentioned acceptable and achievable levels to take into account spectral changes in new mammography units such as the Hologic systems that make use of W tubes with Rh and Ag filtration.

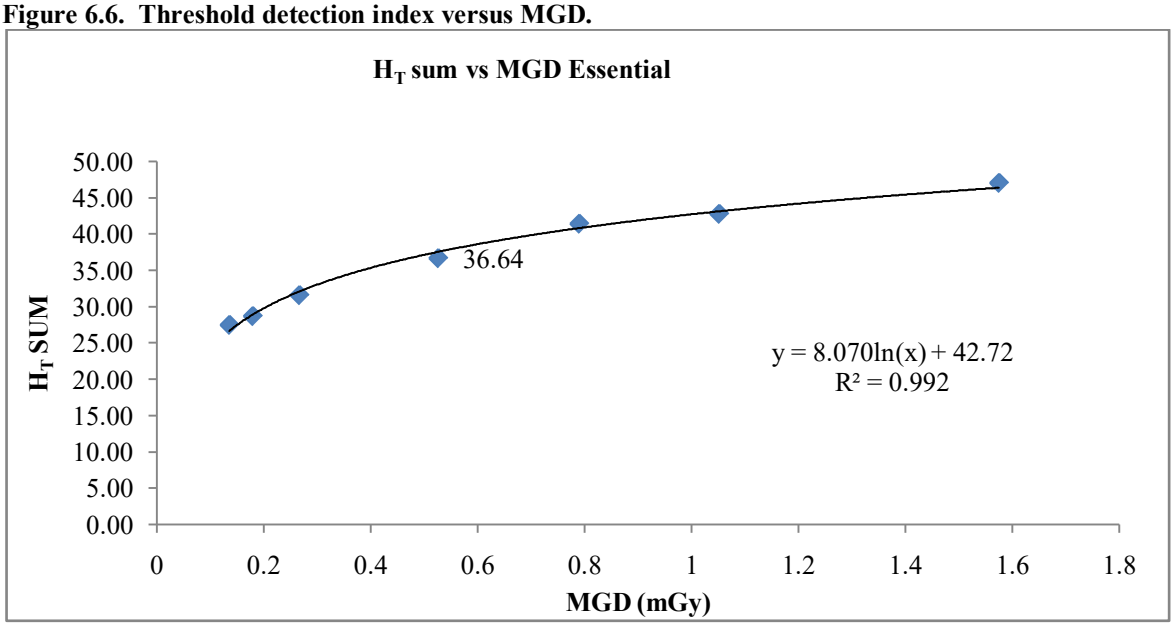
The total and average H_T values for the acceptable and the achievable threshold contrasts was computed using all the CDMAM phantom detail diameters investigated including: 0.06, 0.08,

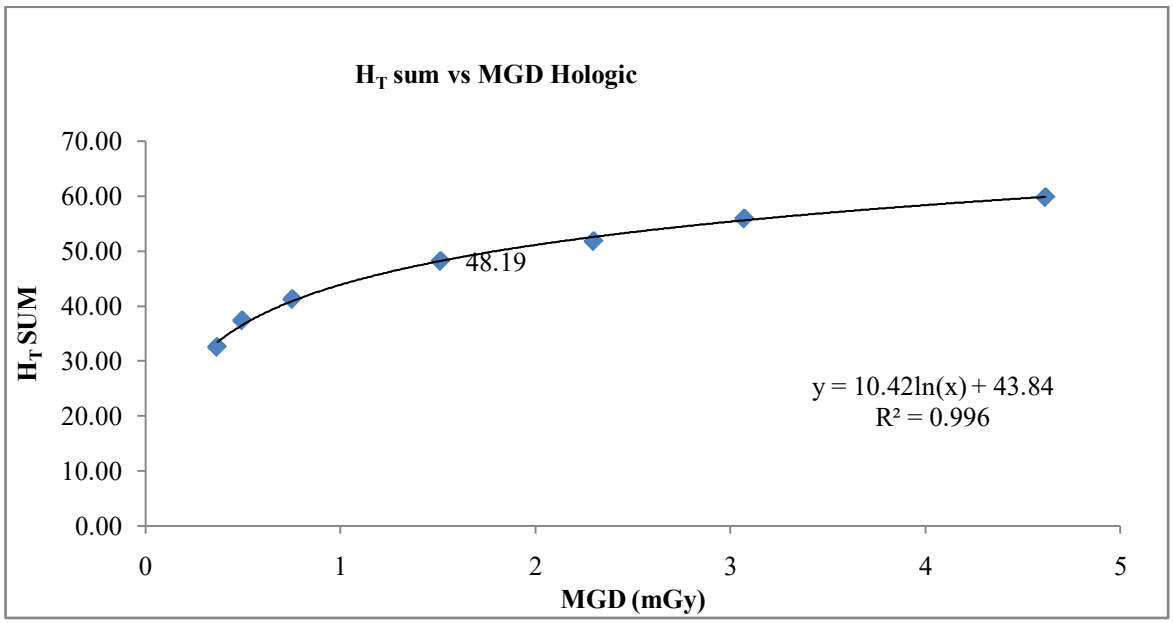
0.1, 0.13, 0.16, 0.2, 0.25, 0.31, 0.4, 0.5, 0.63, 0.8, 1, 1.25, 1.6 and 2.0 mm. Using the data in table 6.3, a nonlinear regression analysis using IBM SPSS v20 was performed using the least squares model as follows:

$$Tc = a + b.x^{-1} + c.x^{-2} + d.x^{-3} \qquad 36$$

where a, b, c and d are coefficients adjusted to achieve the least squares fit, x is the detail diameter and *Tc* is the given threshold contrast value in table 6.3. From the regression analysis the resultant parameters are presented in table 6.4.

Figure 6.6. Threshold detection index versus MGD.





Consequently, by using equation 36 and the parameters generated in table 6.4, the threshold contrast for all the detail diameters and not just those specified in table 6.3 can be confidently predicted in the entire CDMAM phantom. In fact, the predicted values were verified using Q-Q plots against the given values provided by European protocol. The resultant R^2 values for both the acceptable and the achievable were close to 1.0. The predicted acceptable and achievable threshold contrast for all the CDMAM detail diameters and the calculated H_T values at 28kVp using Mo/Mo are summarised in table 6.7.

Table 6.3. Acceptable and Achievable Threshold Contrast values in The European Protocol.

	Threshold contrast			
	Acceptable value		Achievable value	
Diameter of detail (mm)	Gold thickness (µm)	Radiation contrast using Mo/Mo 28 kV (%)	Gold thickness (µm)	Radiation contrast using Mo/Mo 28 kV (%)
5*	0.056	<0.85	0.032	<0.45
2	0.069	<1.05	0.038	<0.55
1	0.091	<1.40	0.056	<0.85
0.5	0.150	<2.35	0.103	<1.60
0.25	0.352	<5.45	0.244	<3.80
0.1	1.68	<23.0	1.10	<15.8

*This diameter size is optional and is not included in version 3.4 of the CDMAM test object.

The calculated **H_TSum** in table 6.7 was used as the ***denominator (H_{ref} (Ai))*** in equation 37, representing the reference curve for any mammography system in terms of H_T.

$$Q(Qfactor1) = \frac{1}{n}.\sum_{i=1}^{n}\frac{H_T(Ai)}{Href(Ai)} \qquad 37$$

where Q is the quality number, n is the number of detail diameter groups (16 in the CDMAM), $H_T(Ai)$ are the threshold detection indexes for each particular exposure as defined earlier. For the purpose of this study, the calculated acceptable H_T values only were considered as reference values (denominator in equation 37).

Notable is the fact that the H_T value for the CDMAM exposures in figure 6.6 and the Q factor in figure 6.20 both yield a strong logarithmic response with MGD. Nevertheless, the Q factor alone cannot be considered a FOM since it only takes image quality into account without due consideration to the related dose or MGD.

Table 6.4. Regression analysis parameter estimates – acceptable and achievable threshold contrast

Parameter Estimates (Acceptable)

Parameter	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
a	.813	.018	.586	1.039
b	.386	.024	.078	.693
c	.195	.007	.103	.287
d	-.001	.001	-.008	.005

Parameter Estimates (Achievable)

Parameter	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
a	.317	.010	.183	.450
b	.414	.014	.233	.595
c	.115	.004	.061	.169
d	.000	.000	-.004	.004

As a result equation 38 addresses this issue by *normalizing* the calculated Q-factor to the calculated MGD based on the relationship between the Qfactor1 and MGD listed in table 6.1a. The new FOM is termed Qnorm1. The relationship between MGD and the Q factor may be observed in figure 6.20 suggesting a strong logarithmic relationship between the two variables.

$$Q_{norm}1 = \frac{Qfactor_1^{\,4}}{MGD} \qquad 38$$

*The Qfactor2 (Qfactor2*100) and the Qnorm2*

The second approach towards the Q factor was taken through Qfactor2 (Qfactor2*100). The principle is similar to the definition in equation 37 however the philosophy and the approach is slightly different. The $H_{ref}(Ai)$ in the Qfactor2 is considered being the maximum achievable IQF_{INV} using the CDMAM phantom. By using equation **13**, the maximum achievable IQF_{INV} using the CDMAM phantom is 351.99. Then by using equation 37, the achieved IQF_{INV} for a given

exposure is placed as numerator and compared to the $H_{ref}(Ai)$. Consequently, this means that a system that scores the CDMAM phantom perfectly i.e. all the details located correctly, would score a Qfactor2 or rather a *Qfactor*100* of 6.25.

The Qfactor2*100 was also *normalized* to the MGD and may be referred to as the Qnorm2 (table 6.1a and b). The Qnorm2 is another FOM and its relationship with MGD can be explored.

Approach 3 – $eNEQ_{MGD}$ and $eNEQ_{DC}$

These approaches have been discussed in considerable detail on page 168. Reference is also made to equations 43 and 44. These methods have been used previously by Salvagnini et al [183-185].

Approach 4 – FOM X and FOM X1

An attempt was made to explore two FOMs based on the principle that desirable attributes are placed as numerator and the undesirable as denominator namely the FOM X and FOM X1 in equations **47** and **48** respectively.

$$FOM\ X = \frac{Qfactor_{\ 1}.DQE.CNR}{MGD} \qquad 47$$

$$FOM\ X1 = \frac{Qfactor_{\ 2}.IQF_{INV}.CNR\ (using\ \ CDMAM\ )}{MGD} \qquad 48$$

In FOM X, the numerator is composed of the Qfactor1 multiplied by the CNR using the method described in the European protocol and the DQE. The value of these parameters multiplied together is then normalised by the MGD for any particular exposure.

FOM X1 follows the same principle but the image quality parameters are replaced by the Qfactor2, the IQF_{INV} and the CNR used was the one described in figure 6.2 using the CDMAM phantom.

Results

For characterizing the quality of the image in terms of the potential detectability of structures in the breast, it is useful to introduce the SDNR or CNR. The SDNR is described in equation 1. The SDNR is related to the SNR and the radiation contrast as follows [186]:

$$SDNR = 2C_{RAD}.SNR \quad 49$$

The performance parameters of an X-ray detector are described in both the spatial and the frequency domain. Usually the signal transfer property (STP), the MTF, the NNPS and DQE are sufficient to describe the signal and noise properties of the X-ray detector. The STP shows the response of the detector to the input signal. The MTF describes the contrast reduction at different spatial frequencies that compose the image and the NPS describes the frequency components of the noise. Finally, the DQE shows the ability of the detector to transfer the squared SNR from the input stage to the output stage [187].

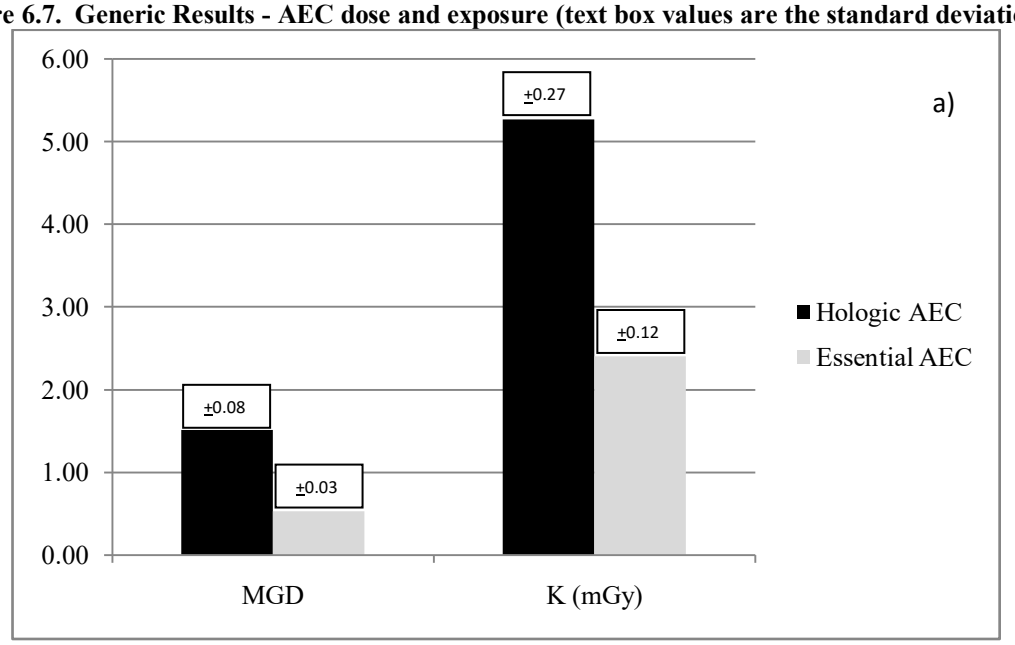
The following sections will discuss the results achieved for the Essential and the Hologic.

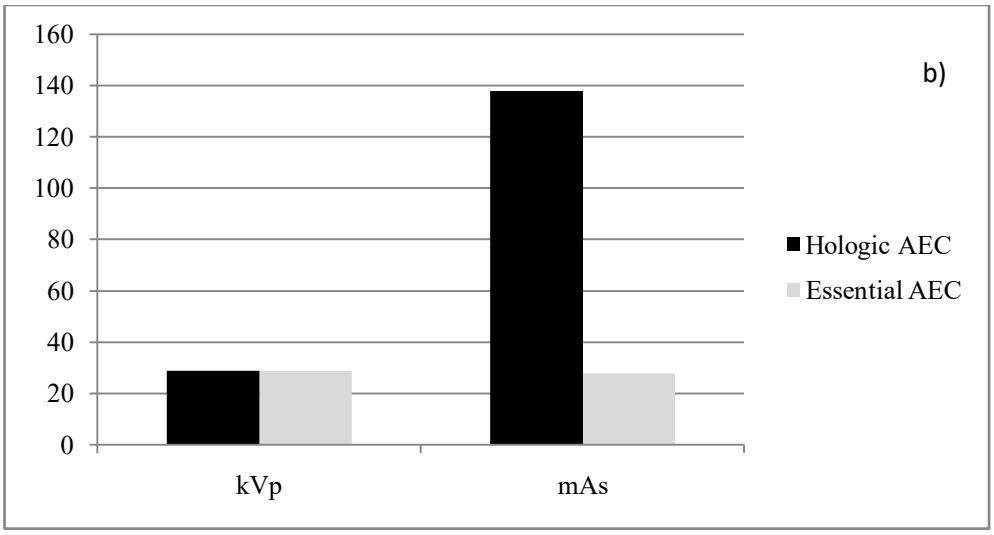
GENERIC RESULTS

With reference to figure 6.7, it can be observed that both systems analyzed demonstrated different AEC behaviours in terms of exposure and dose when applying the methods on page 159.

From figure 6.7 it can be seen that the Hologic resulted in AEC exposures with higher mAs values consequently resulting in AEC exposures with larger detector doses (DAK), MGD and K. These values were rechecked, in the presence of the field engineers and achieved through the same methodology for both systems.

Figure 6.7. Generic Results - AEC dose and exposure (text box values are the standard deviations)





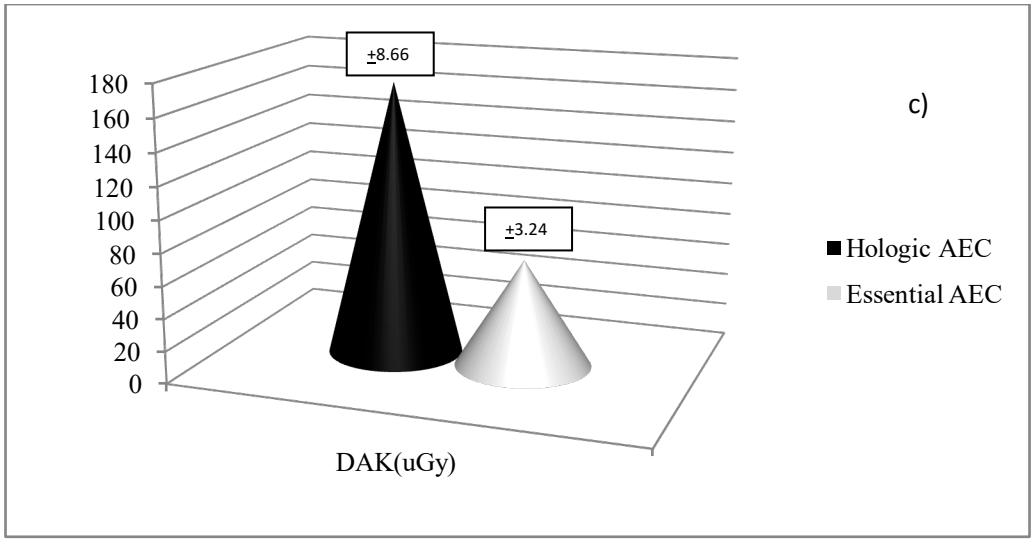
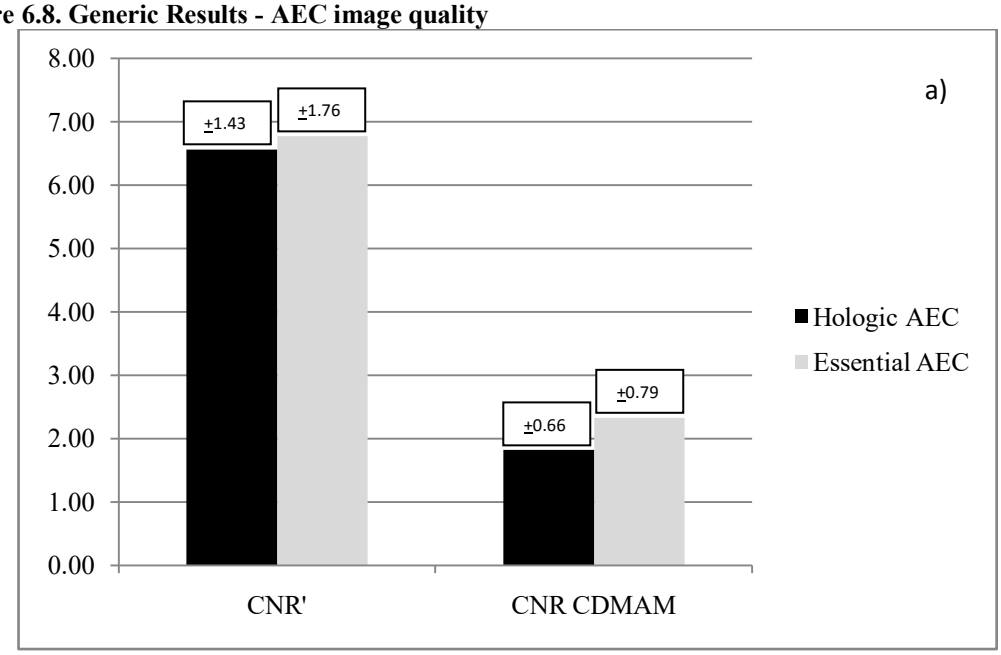
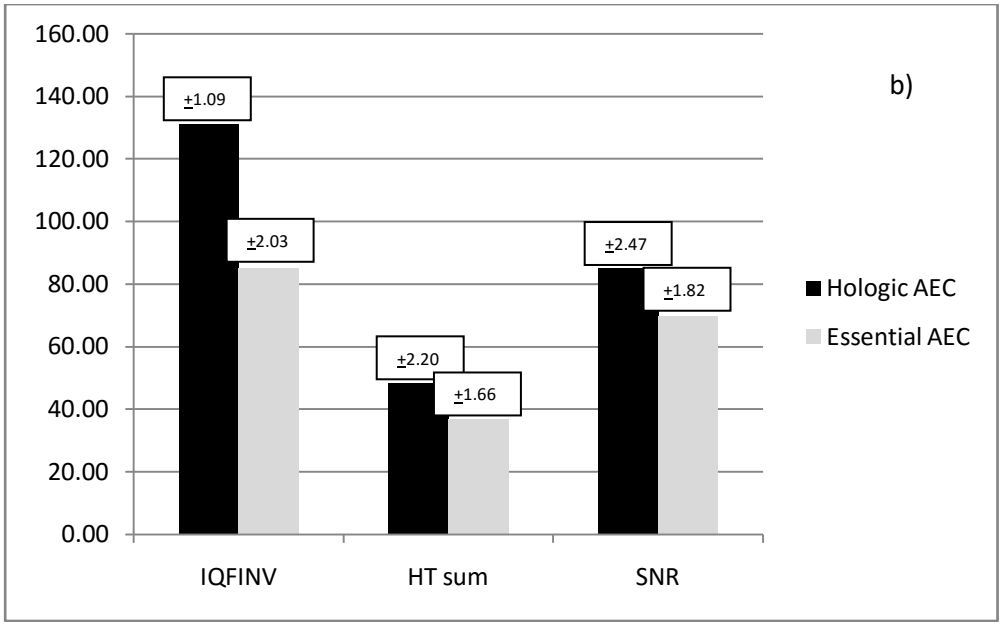


Figure 6.8. Generic Results - AEC image quality





With reference to figure 6.8b, this increase dose was reflected in CDMAM-related image quality parameters but surprisingly this was not reflected in CNR measurement in figure 6.8a. It seemed that although doses were significantly higher for the AEC exposure in the Hologic, the CNR values, using both methods, was lower when compared to the Essential. With reference to figure 6.9 the Hologic revealed a better MTF value than the Essential. However, the DQE values of the former system were not significantly higher than those of the Essential. More details on the DQE for all the available DAK values are given in figure 6.10 and the next section.

Figure 6.9. Generic Results - AEC detector characterization

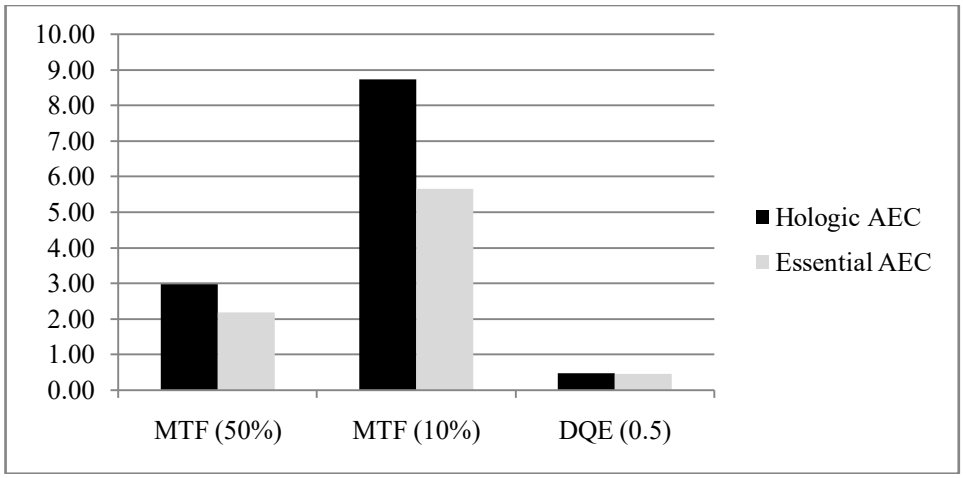
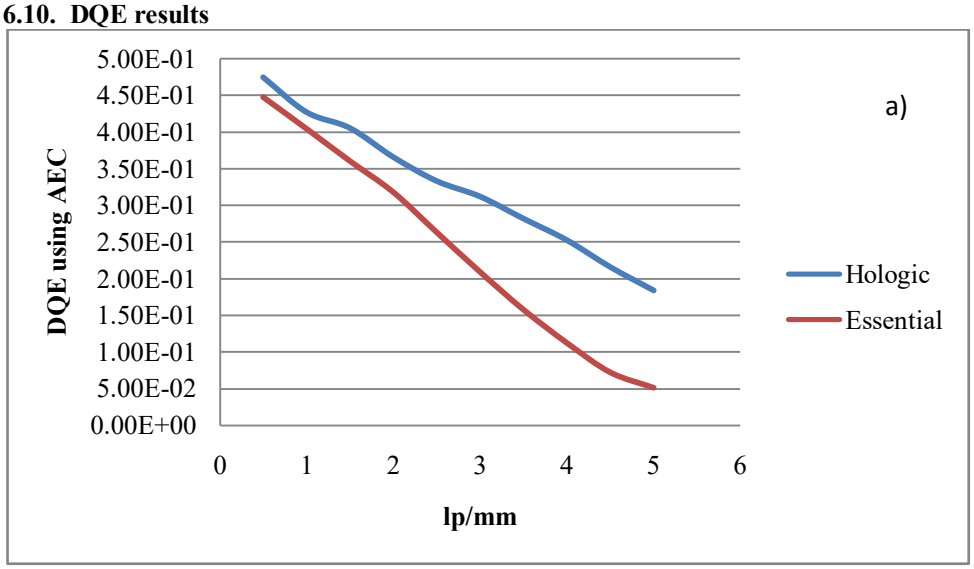
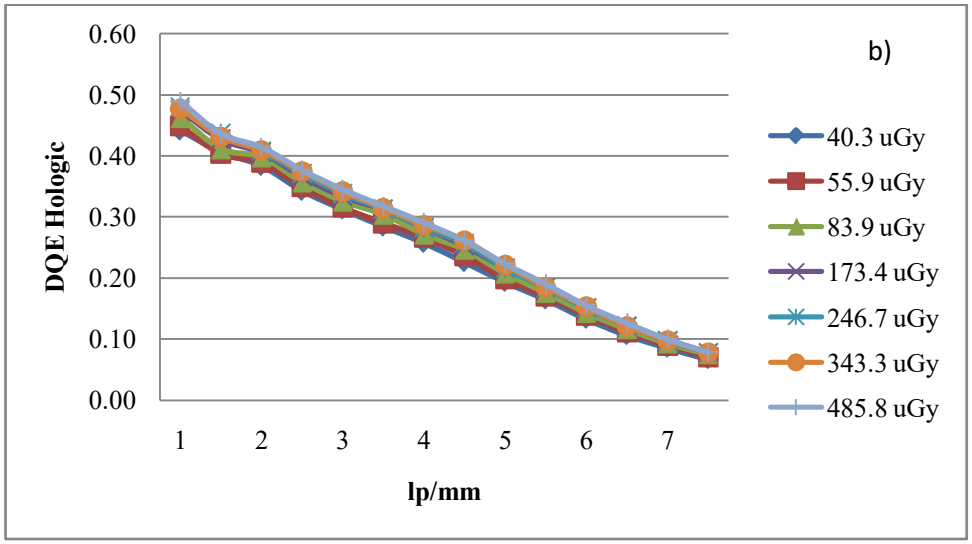
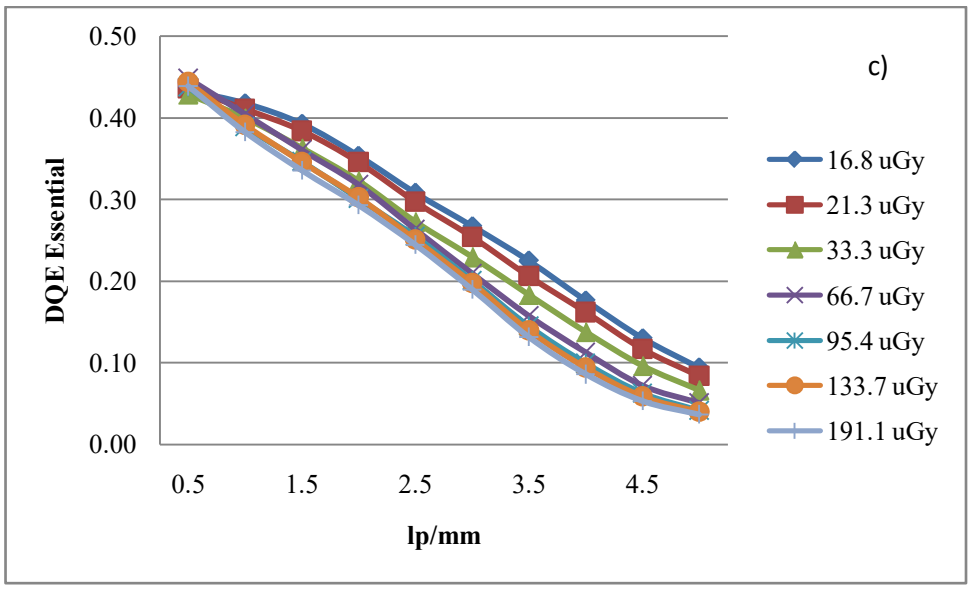


Figure 6.10. DQE results







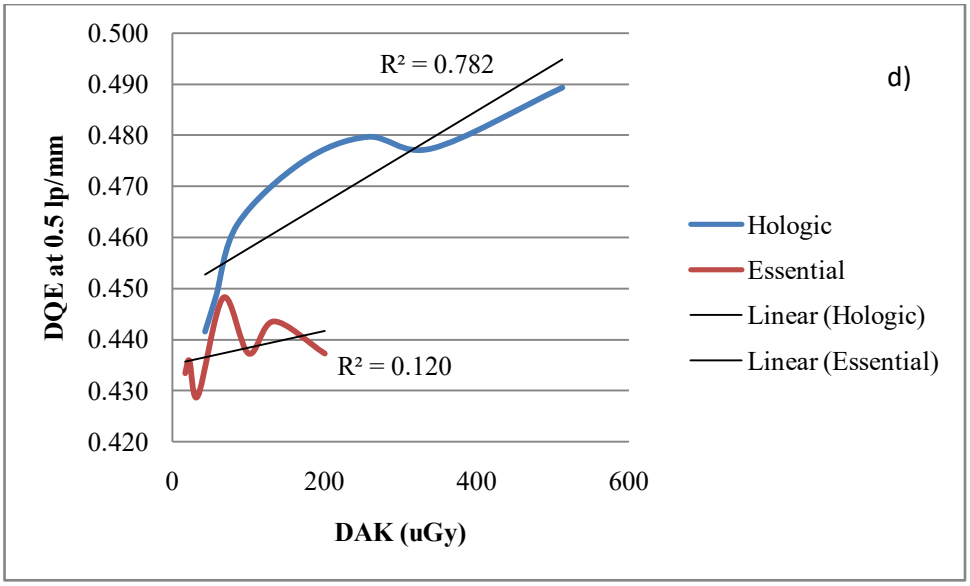
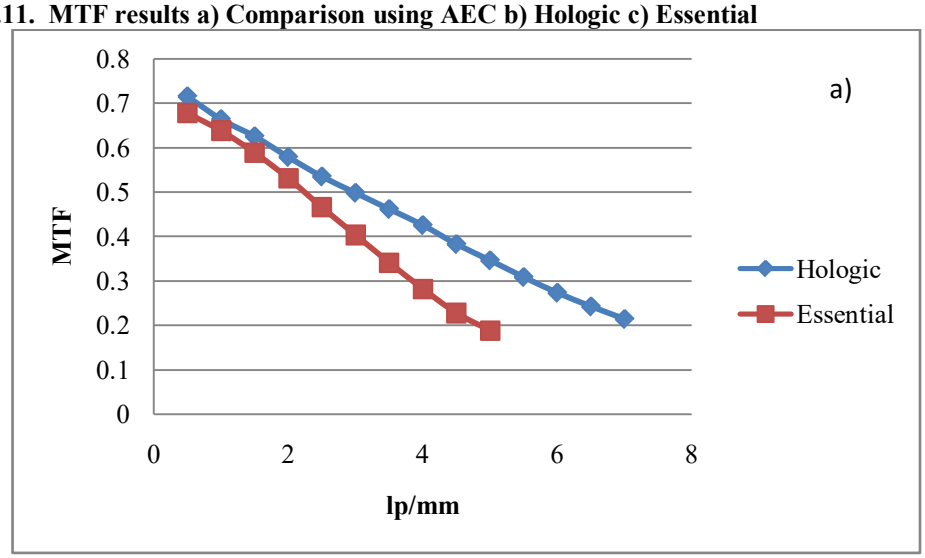
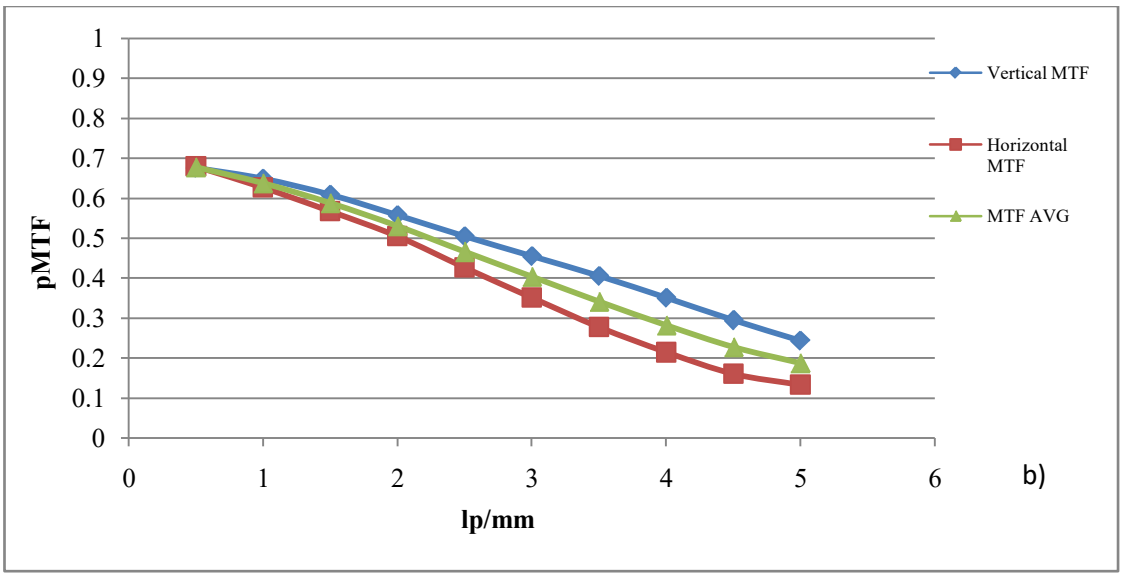
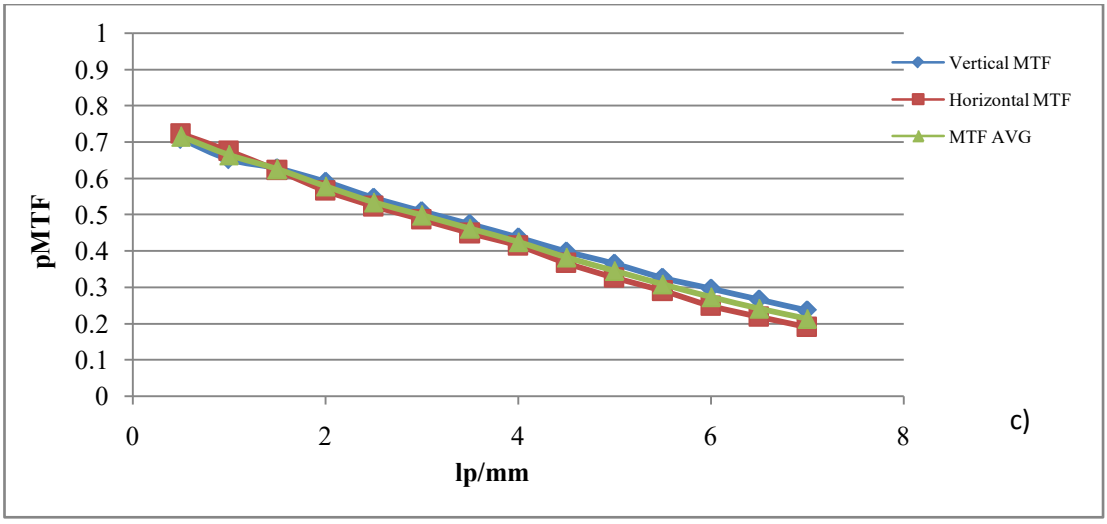


Figure 6.11. MTF results a) Comparison using AEC b) Hologic c) Essential







DQE AND NEQ

The SDNR is the effective quantitative descriptor of the *quality* of the information carried by the radiological image. The larger the signal or signal difference is, when compared with the random fluctuation, the better the image is. The SDNR increases with increasing exposure and with higher values of quantum detective efficiency (ɳ) [186]. This wasn't the case with the Essential as seen in figure 6.8. Although the DAK was much lower than that of the Hologic the resulting SDNR or CNR was 3% higher. In theory, the SDNR decreases when there are sources of noise other than quantum noise contributing to the image. The highest SDNR occurs in the pattern of X-rays transmitted by the breast. This can be considered as the input signal. If the number of these X-ray quanta in a specified area was n_0, its value would be:

$$SNR_{IN} = \sqrt{n_0} \qquad 50$$

For a system that is perfect (i.e., having no additional noise sources) except for incomplete absorption of all incident X-rays by the detector, the signal would be $n(n_0)$and the noise $\sqrt{n}(n_0)$ giving a reduced signal of $\sqrt{n}(n_0)$. The performance of the imaging system can be characterised by determining how efficiently it transfers the input SNR to the system output (i.e., the observer). The DQE computes the ratio:

$$DQE = \frac{SNR_{OUT}^2}{SNR_{IN}^2} \qquad 51$$

Therefore, for a perfect system the DQE would equal 1.0. Considering just the efficiency of X-ray interaction described above, DQE would be: $n(n_0)/ n_0 = \eta$, the fraction of incident X-rays used by the detector. If there are other sources of noise, SDNR will decrease below the value predicted by the number of interacting quanta so that DQE will fall below η. From the measurement of SNR_{OUT} it will appear that fewer X-rays have been used to form the image than has actually been the case and DQE is a measure of that apparent lack of efficiency. In fact, the quantity, SNR^2_{out} which, in the absence of additional noise sources is just the number of X-rays detected, is known as the number of *noise-equivalent quanta* or NEQ [186].

It is common to present DQE and NEQ values as a function of spatial frequency were DQE(f) specifies how well the system transfers the SNR information present at its input at each level of detail. This has been demonstrated in figure 6.10 and 6.13.

DQE variation with air kerma at the AEC spectrum is shown in figure 6.10a-d. The maximum DQE values reach 0.49 at 458.8 µGy for the Hologic and 0.45 at 66.7 µGy for the Essential but falls to 0.08 and 0.05, respectively, at the Nyquist frequency. The rapid decrease of DQE with increasing spatial frequency in flat panel detectors is probably related to the shape of the MTF curve [101].

In 6.10a, DQE values were plotted at the AEC exposure. The Hologic demonstrated a 5.9% higher DQE than the Essential at the AEC exposure probably attributed to higher DAK values. The Essential showed also a more marked decrease in DQE with spatial frequency when compared to the Hologic. At a matched DAK value of 120 µGy at 0.5 lp/mm, the Hologic still resulted in higher DQE of 4.37%. At higher spatial frequencies the difference in DQE between the two systems becomes more substantial as shown in table 6.5 in favour of the Hologic.

Figure 6.10d shows DQE variation with DAK at a matched spatial frequency of 0.5 lp/mm. It can be seen that DQE rapidly increases with increased DAK in both systems however is much more marked and significant in the Hologic system. In the latter unit there appears to be a steady increase in DQE with DAK at 0.5 lp/mm, manifested by the higher R^2 value when compared to the Essential.

NEQ

Measurement of the NEQ may provide a practical and robust way to assess the image quality of a system objectively. The combined effects of signal and noise, in terms of spatial frequency, can be expressed by the NEQ. The latter provides an index of the SNR associated with the diagnostic value of a medical image. NEQ can be written as follows [188]:

$$NEQ(f) = \frac{SPS(f)}{NPS(f)} \qquad 52$$

where SPS is the product of the squared MTF, the squared signal, or detector response (G), where G is the gain factor (in digital units per μGy)[188].

The shapes of the NEQ curves are affected by both NPS and MTF. As shown in figure 6.13, in the low frequency range image data has higher NEQ values for both systems. This is attributed to the low NPS and high MTF values of the image data in this spatial frequency range. Accordingly, low contrast objects of large dimensions may be better visualized. From figure 6.13 it can be seen that the Hologic system showed significantly better NEQ values due to higher DAK values resulting in lower noise images. Apart from this, although dose-independent, the Hologic system showed better MTF values at all spatial frequencies which will affect the NEQ values too.

In the study by Michail et al [188] it was shown that in the medium to high spatial frequency range, NEQ of the images is clearly higher than the corresponding values of the raw image data. This may be explained by the higher MTF values in this spatial frequency range. The 'sharpen' and 'sharpen more' filtered images have reduced NEQ values due to their high NPS caused by the noise that the filtering process introduces to the images.

It is therefore important, as was done in this exploration study, that raw images are used for both systems.

MTF

Spatial resolution can be characterized quantitatively and more usefully through MTF. The MTF describes how well the imaging system or one of its components performs, for instance how well the detector transfers the contrast of sinusoidal patterns from the incident X-ray

pattern to the output. A sinusoid is a repetitive function, characterized as having a frequency (in this case a *spatial frequency* specified in line pairs (lp)/mm) and amplitude. Low spatial frequencies represent coarse structures and high spatial frequencies (short wavelengths) describe fine detail.

Table 6.5 Percentage difference in DQE at 120 µGy

lp/mm	% Difference in DQE
0.5	4.37%
1	5.48%
1.5	11.66%
2	13.84%
2.5	20.62%
3	31.52%
3.5	44.63%
4	56.99%
4.5	67.00%
5	73.10%

Any pattern can be represented as a combination of sinusoidal shapes, each spatial frequency having specific amplitude. The MTF simply describes how well each spatial frequency is transferred through a system, i.e., is simply the ratio of the amplitude of the sinusoid at the output of the system or a component to that at the input. The MTF of an imaging system is often 1.0 at very low spatial frequencies and falls with increasing spatial frequency. In a system containing several elements that affect the spatial resolution, the overall MTF is determined as the product of the MTFs of the individual components. For example, the MTF of a radiographic system is the product of that due to the focal spot, the detector and any motion of the patient during the exposure. This is helpful in determining what part of the system is responsible for limiting its performance [186].

Several factors affect the spatial resolution in digital mammography. Some of them are identical to those that apply to film mammography e.g. the focal spot must be sufficiently small, prevent excessive image unsharpness by reduction motion of the patient, the detector size must be minimized etc. Another factor affecting resolution is related to the lateral spread of signal (light photons or electronic charges) in the detector from the point where an X-ray is

absorbed to the point where the signal is recorded. Generally, as the detector is made thicker to increase h, more blurring of this sort will occur and therefore there is a compromise between spatial resolution and η.

In digital mammography systems, an additional factor, the spatial sampling, affects resolution. The X-ray signal from each del is averaged over the aperture, d. This causes the MTF of the detector to decrease. For practical purposes, detectors for digital mammography are not designed to provide as high a spatial resolution as obtainable from a FS system and its MTF is typically considerably lower. Other factors such as contrast and noise characteristics provide offsetting improvements in imaging performance. Typically, the size of the del ranges from 50 to 100 µm. If d is expressed in mm and the del is a square, then the MTF falls to 0 at a spatial frequency of 1/d cycles/mm in both the x and y directions. For example, a del with d = 100 µm, has an MTF that falls to 0 at 10 cycles/mm [186].

The spacing between dels or pitch, p is also an important factor. The larger that p is, the more information that will be lost through the sampling process and this restricts the maximum value of the spatial frequency of information in the image that can be represented accurately. The Nyquist theorem stipulates that for a pitch p, the highest spatial frequency that can be accurately represented is:

$$f_N = \frac{1}{2p} \qquad 53$$

Higher spatial frequencies in the X-ray pattern will be misrepresented, a phenomenon known as *aliasing*. Coarse sampling causes the erroneous creation of sinusoids of lower spatial frequency. So, undersampling both prevents reliable depiction of high spatial frequency information in the image, also because of interference due to these aliased low frequencies impairs the representation of the true low spatial frequency information in the image. A detector with d = 100 µm and p = 100 µm will be susceptible to aliasing if there are spatial frequencies in the image above 5 cycles/ mm [186].

Figure 6.11 shows the system presampling MTF that is essentially the same at all exposure levels and beam qualities [101]. Presampling MTF values at 2 lp/mm and 4 lp/mm are 0.58 and 0.43 for the Hologic and 0.53 and 0.28 for the Essential. The Hologic system showed better MTF values at all spatial frequencies and this difference became more pronounced as the spatial frequency increased.

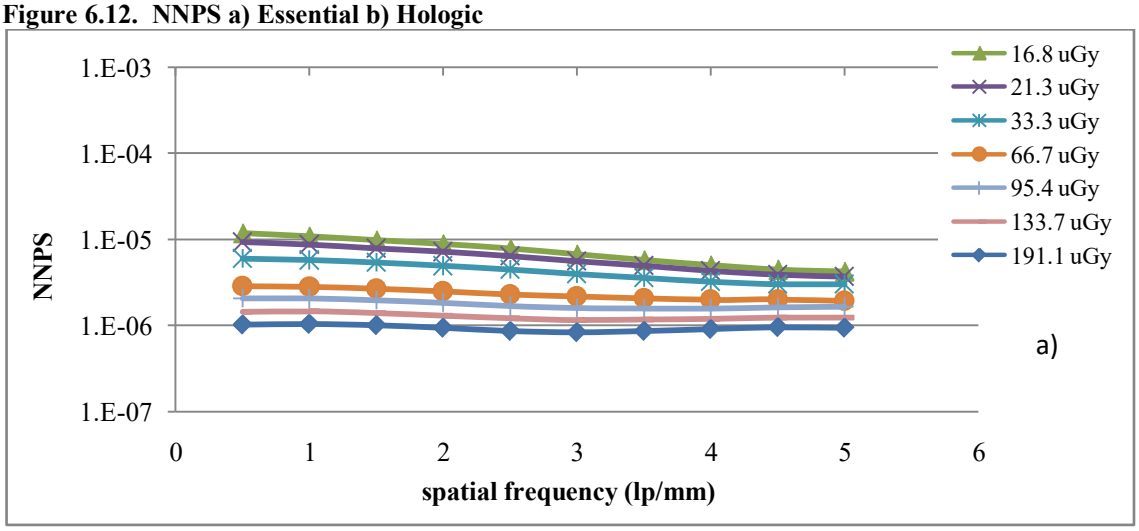
Table 6.6 Percentage difference in MTF

lp/mm	% Difference in MTF
0.5	5.31%
1	3.95%
1.5	5.93%
2	8.28%
2.5	13.00%
3	19.06%
3.5	26.08%
4	33.67%
4.5	40.53%
5	45.60%

NNNPS

Figure 6.12 shows the 1 dimension NNPS at several values of DAK for both systems. It is clear that NNPS values exhibit a dependence on exposure, and decrease with increasing DAK. The Hologic exhibited lower NNPS values because of the higher DAK values associated with the AEC exposures. From figure 6.12 it can be seen that the Hologic requires a lower DAK value at a matched NNPS level. This could be attributed to the fact that (1) there is a higher photon fluence when using a harder beam (W/Rh) instead of Rh/Rh for the same kV, which implies a smaller influence of electronic noise, and (2) the higher mean energy of the spectrum that leads to a higher transmission of the flat-panel cover [101].

Figure 6.12. NNPS a) Essential b) Hologic



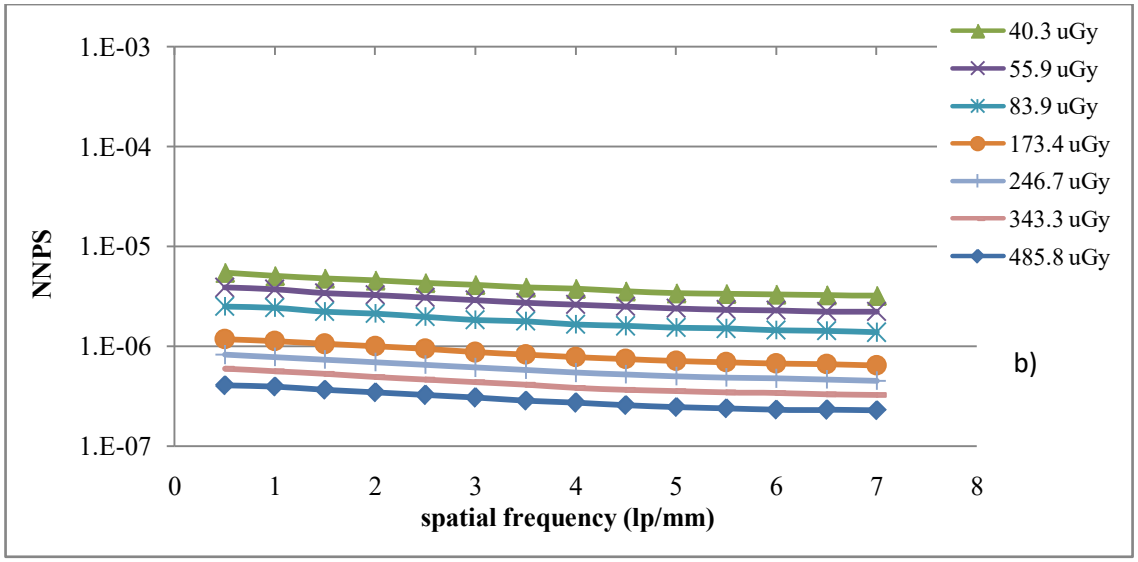
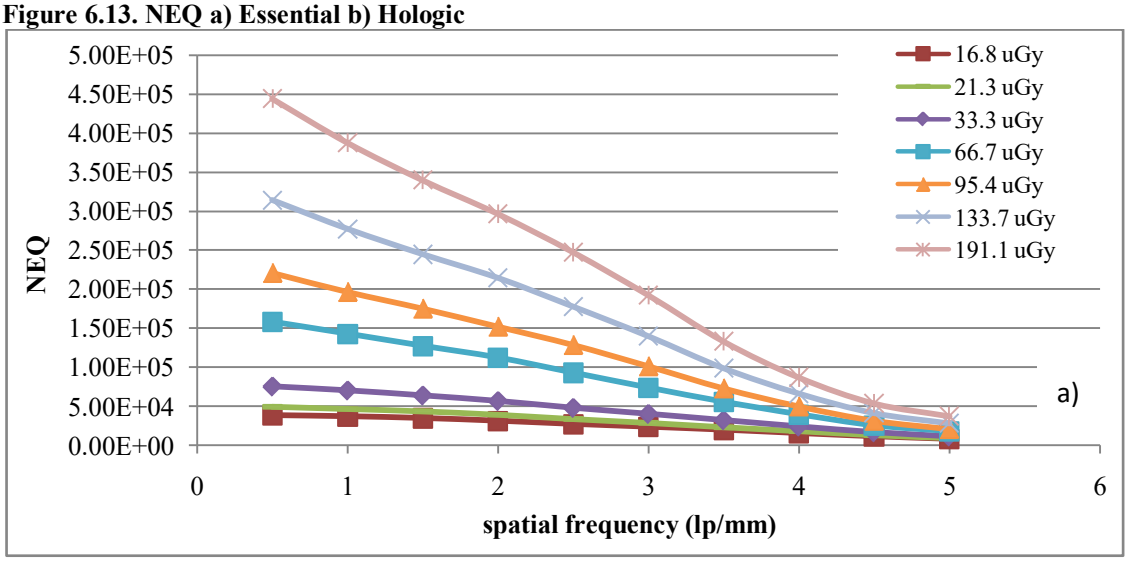
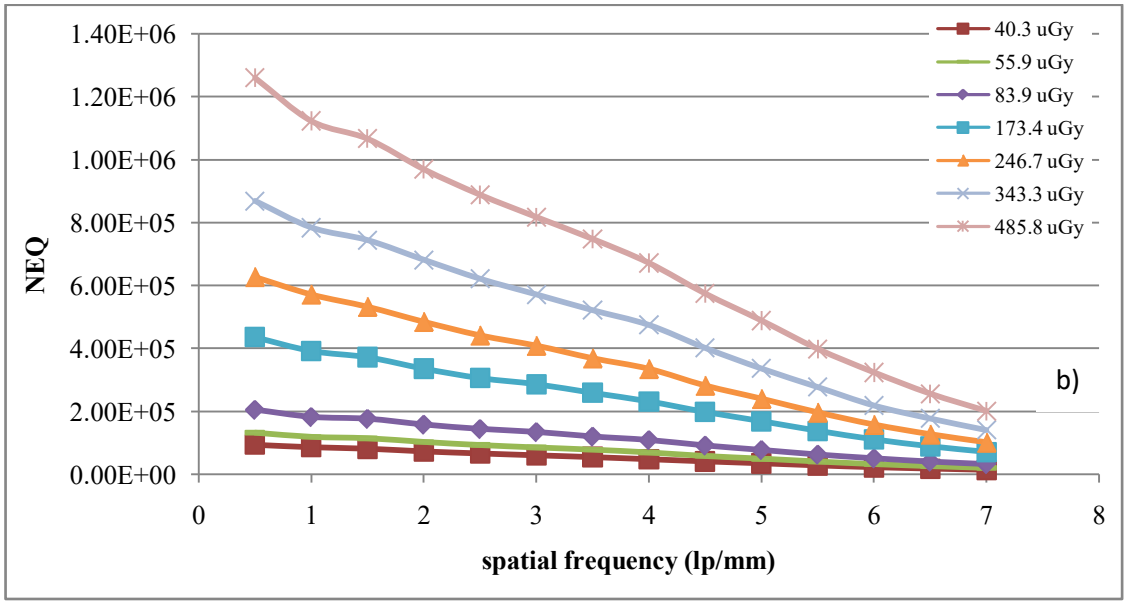


Figure 6.13. NEQ a) Essential b) Hologic



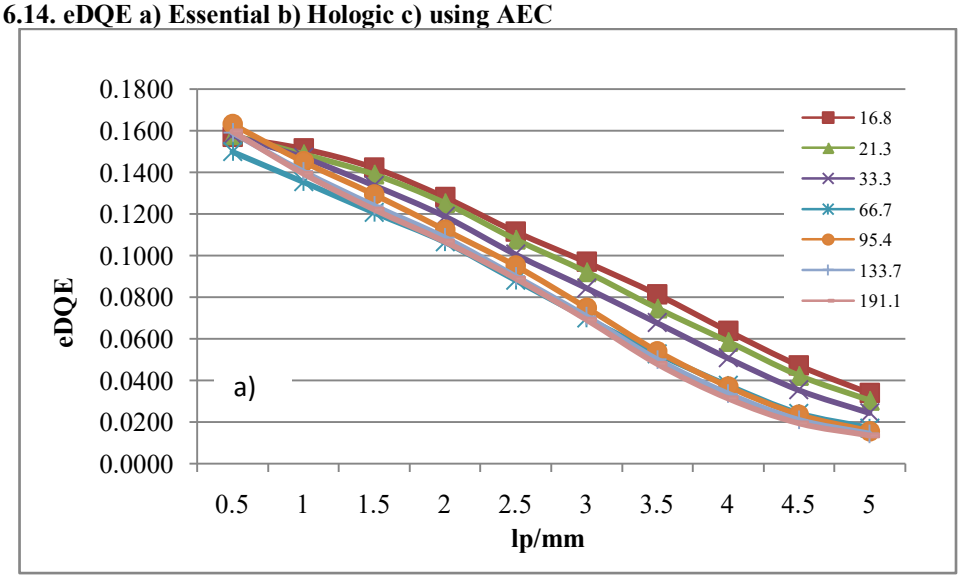


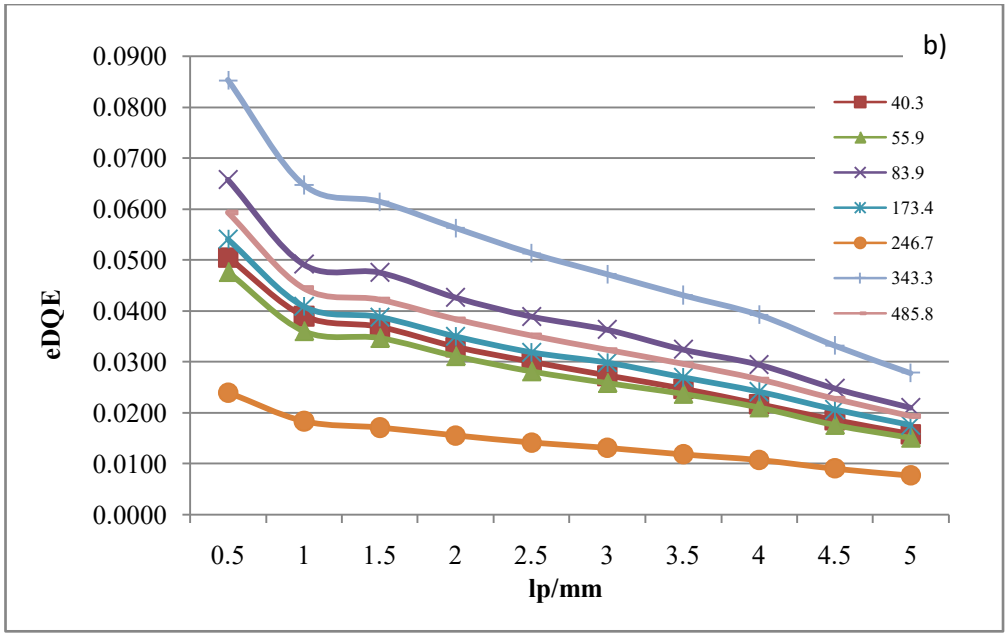
EFFECTIVE DQE, EFFECTIVE NEQ, EFFECTIVE NEQ_{MGD} AND EFFECTIVE NEQ_{DC}

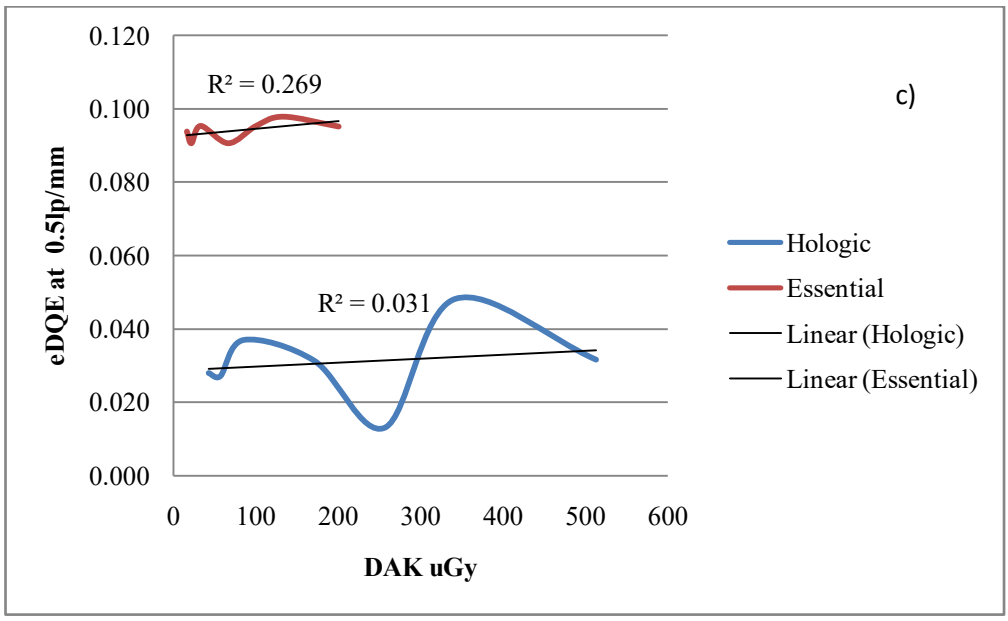
As already stated, the most commonly used metric of detector performance is the DQE, a parameter that describes how effectively a detector captures the incident SNR relative to an ideal detector. DQE places detector efficiency on an absolute scale and gives an idea of the imaging potential of the detector. eDQE and eNEQ were introduced to broaden the notion of DQE and NEQ by including system parameters such as focus blurring, magnification and system scatter rejection methods [183, 185]. The formulation of eDQE attempts to compare a system against an ideal imaging system with perfect X-ray detection and scatter rejection.

Since the patient is not included and the objective is to compare systems with the patient as the focus, eNEQ normalized for the MGD was included ($eNEQ_{MGD}$), as system configurations that produce high eDQE may not be the best clinical choice. It is possible that the patient dose associated with parameters that give a high value eDQE could be higher than the patient dose of configuration that is suboptimal in terms of eDQE. The $eNEQ_{MGD}$ describes the impact of the patient dose on the chosen configuration but this term does not give any indication on detail detectability. Consequently, the $eNEQ_{MGD}$ was redefined with object contrast in mind and named $eNEQ_{DC}$[185]. The purpose of this section was (1) to perform a characterization of the digital mammography system efficiency of the Essential and the Hologic using eDQE as opposed to a DQE approach that focuses on the detector, (2) to investigate the influence of the MGD using $eNEQ_{MGD}$, and (3) to investigate $eNEQ_{DC}$ in relation to contrast detectability.

Figure 6.14. eDQE a) Essential b) Hologic c) using AEC







From figure 6.14 it is immediately appreciated that the resulting eDQE levels of the Essential are far better that those of the Hologic. Further still, it seems that the maximum eDQE was achieved by a DAK value of 343.3 µGy and the least eDQE values were achieved by 246.7 µGy at all spatial frequencies for the Hologic. This shows that the Hologic, unlike the Essential demonstrated eDQE values that are specific to particular DAK values. In the Essential the highest eDQE values resulted from the lowest DAK. However, this wasn't very clear at all the spatial frequencies. It seemed that at lower spatial frequencies (0.5 to 1 lp/mm) the highest eDQE value resulted from a DAK value of 95.4 µGy and lowest eDQE value, at all spatial frequencies resulted from the AEC exposure, 66.7 µGy.

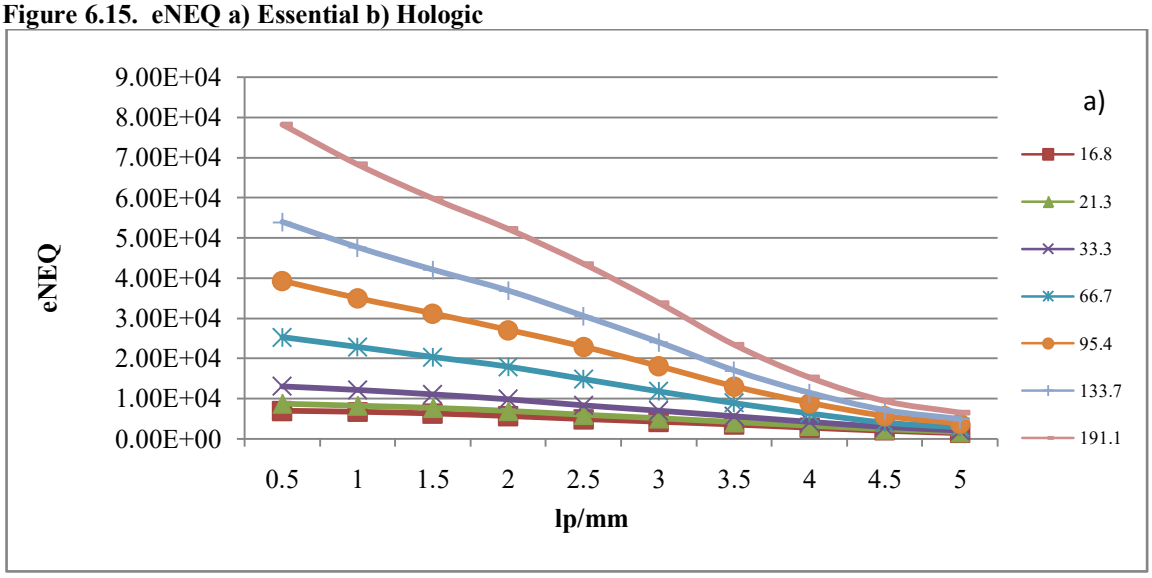
With reference to 6.10d and 6.14c, both systems showed an overall mild and insignificant increase in eDQE with DAK at 0.5 lp/mm being more pronounced in the Essential. In figure 6.14c, the eDQE seemed to reveal a somewhat different behaviour than its counterpart. There seems to be a less influence of DAK on eDQE at 0.5 lp/mm on the overall eDQE, especially for the Hologic unit which now appears to have a lower eDQE than that of the Essential (refer to figure 6.10d for comparison).

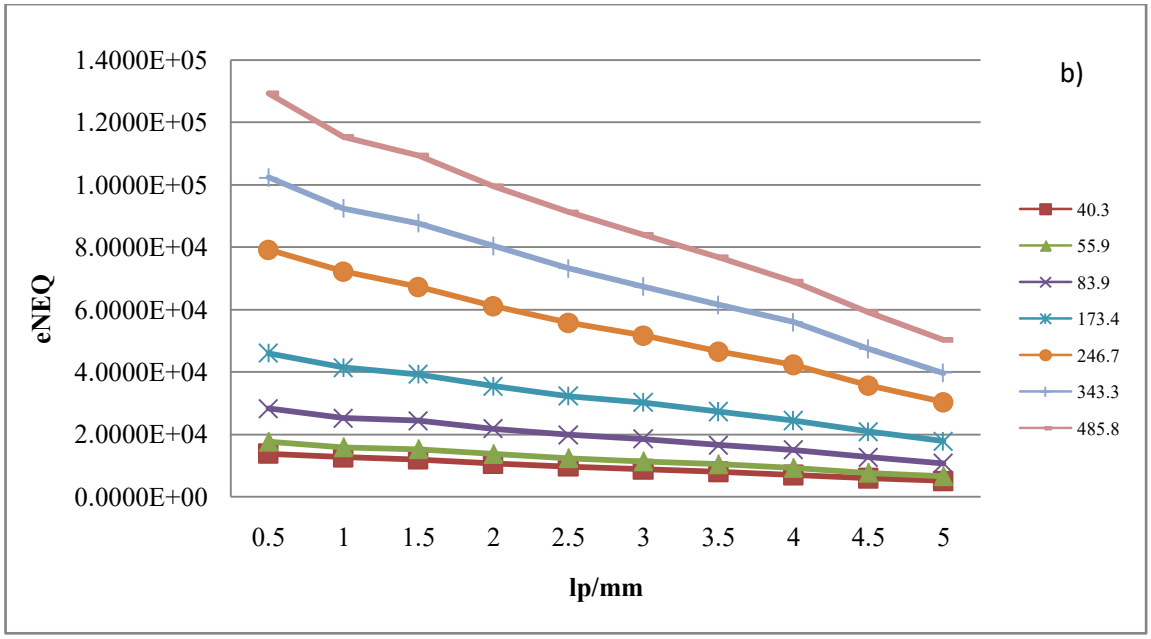
The eNEQ results in figure 6.15 appear to more systematic for both systems being larger for the Hologic than the Essential at all the spatial frequencies. The eNEQ seems to be affected by the DAK values whereby the larger DAK values yield the highest eNEQ values. With reference to figure 6.15, the Essential revealed a larger drop-off eNEQ value with spatial frequency when compared to the Hologic.

Effective NEQ_{MGD} and Effective NEQ_{DC}

As stated earlier to introduce a dose metric, eNEQ normalized for the MGD was included in this section ($eNEQ_{MGD}$). The $eNEQ_{MGD}$ describes the impact of the patient dose on the chosen configuration and the eNEQ with object contrast that is characteristic for the studied configurations is referred to as $eNEQ_{DC}$.

Figure 6.15. eNEQ a) Essential b) Hologic

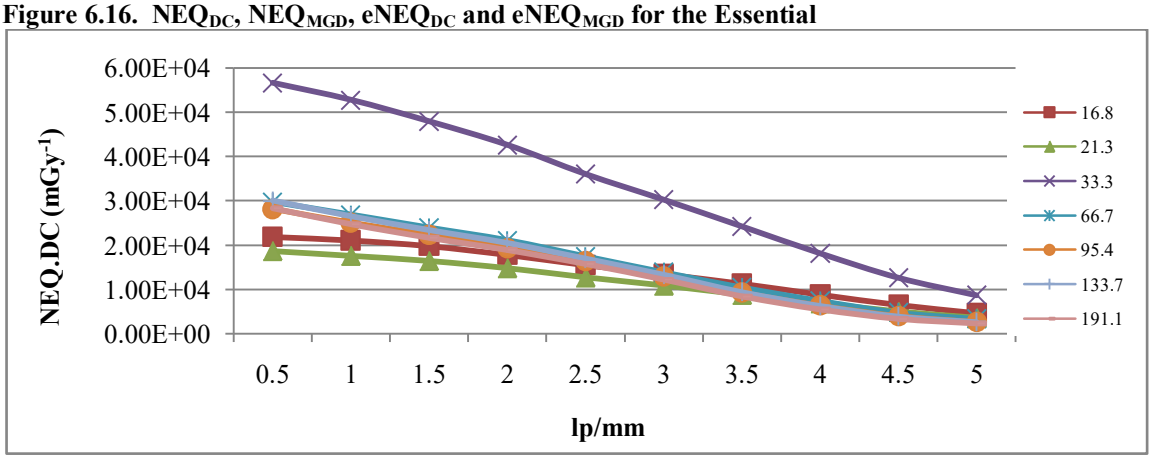


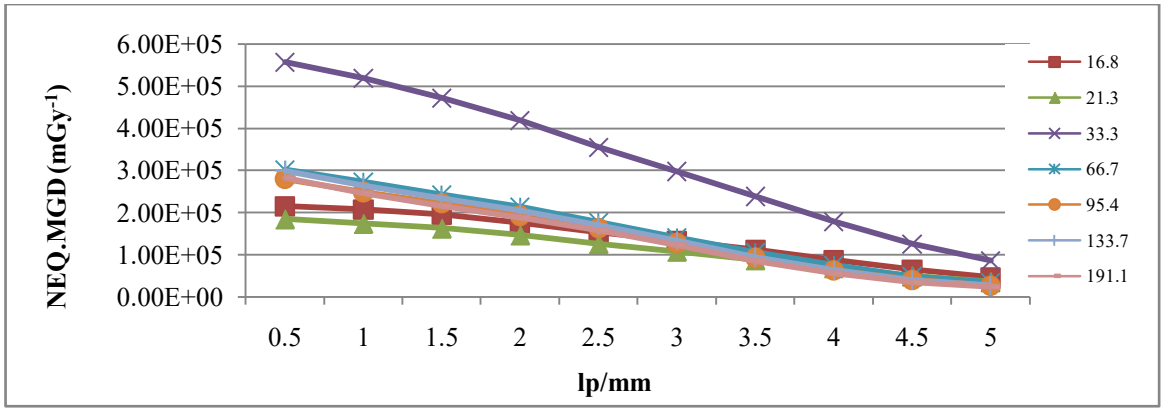


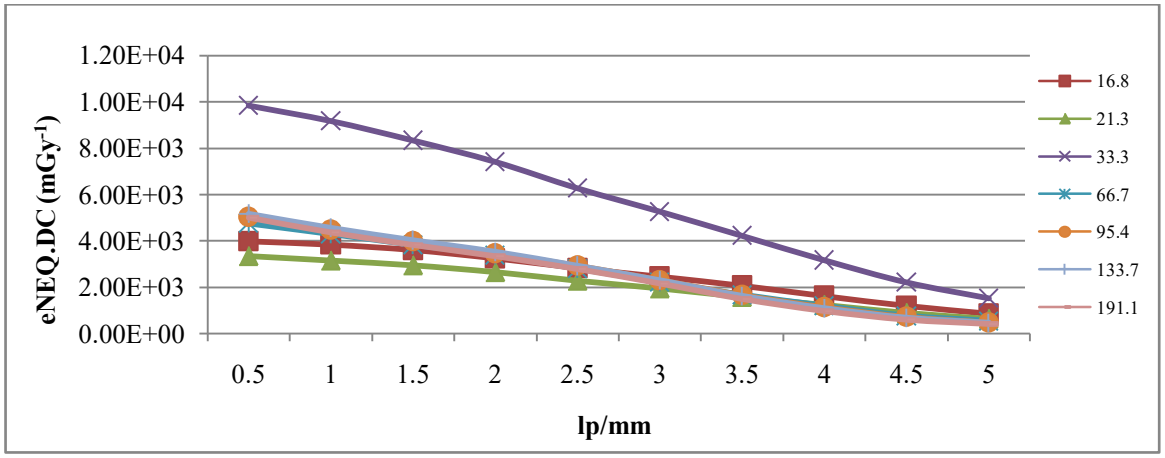
With reference to figures 6.16 and 6.17 both the non-effective and effective values were calculated. From these graphs it can be seen that the effective values were lower than non-effective values for both systems. The Essential showed lower values overall in NEQ_{MGD}, NEQ_{DC}, $eNEQ_{MGD}$ and $eNEQ_{DC}$ probably attributed to the lower DAK values employed which affect the MGD but also a lower DAK will affect contrast detail detectability.

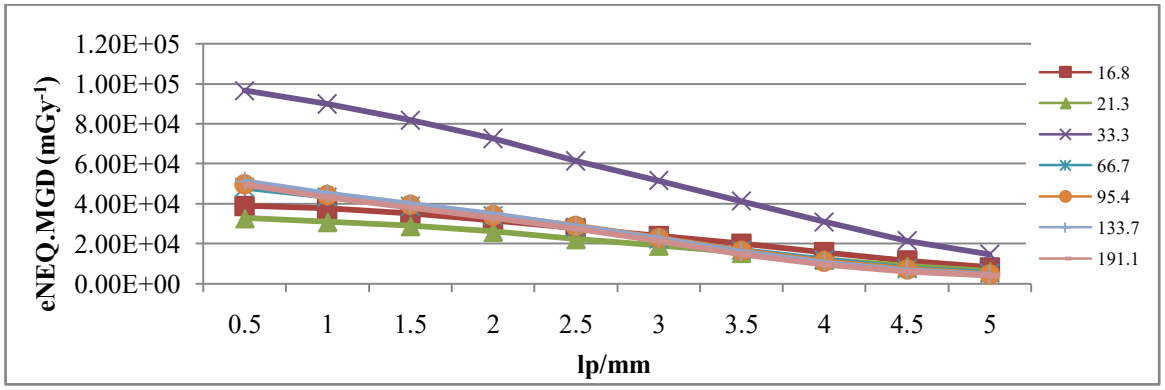
The Essential showed similar patterns for all the datasets calculated. Interesting to note was that the largest NEQ_{MGD}, NEQ_{DC}, $eNEQ_{MGD}$ and $eNEQ_{DC}$ were recorded at 33.3 µGy and the lowest values were observed at 21.3 µGy. The former value seemed to exhibit a large drop-off in NEQ_{MGD}, NEQ_{DC}, $eNEQ_{MGD}$ and $eNEQ_{DC}$ for all the spatial frequencies. Nevertheless, as the spatial frequency increased (4.5 – 5 lp/mm) all the values seem to converge and making differences less obvious.

Figure 6.16. NEQ_{DC}, NEQ_{MGD}, $eNEQ_{DC}$ and $eNEQ_{MGD}$ for the Essential



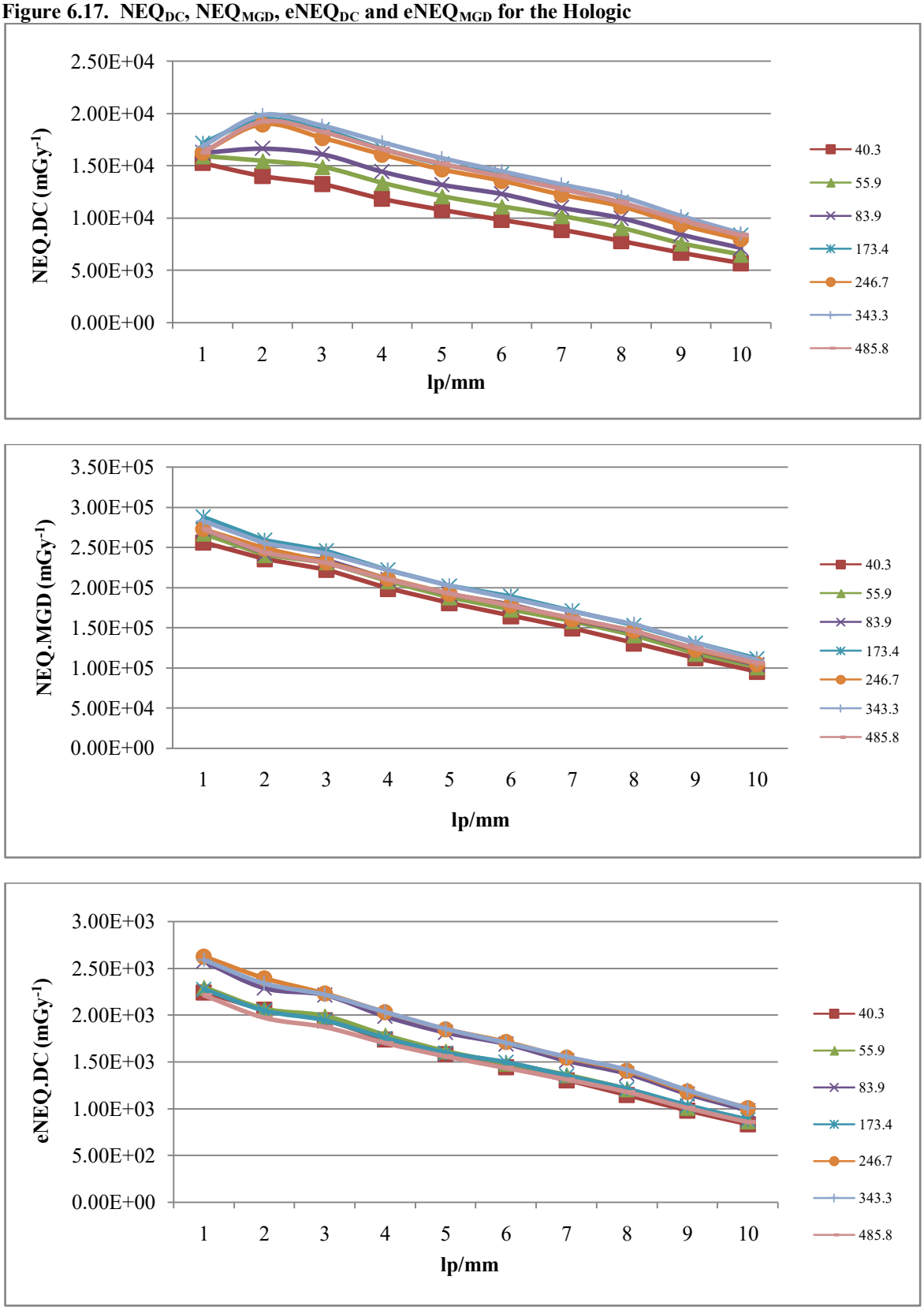






With reference to figure 6.17, the Hologic appeared to manifest a different behaviour in NEQ_{MGD}, NEQ_{DC}, $eNEQ_{MGD}$ and $eNEQ_{DC}$ with spatial frequency. The largest $eNEQ_{MGD}$ and $eNEQ_{DC}$ resulted in the AEC exposures with a DAK of 173.7 µGy. The $eNEQ_{MGD}$ and NEQ_{DC} showed a different response. Primarily the values were lower than the respective non-effective counterparts. With regards to $eNEQ_{DC}$ the higher values were achieved with larger DAK values. The NEQ_{MGD} values in figure 6.17 appear closer while the $eNEQ_{MGD}$ values appear more distinct and follow a pattern whereby the largest value resulted with the lowest DAK and vice versa.

Figure 6.17. NEQ_{DC}, NEQ_{MGD}, $eNEQ_{DC}$ and $eNEQ_{MGD}$ for the Hologic



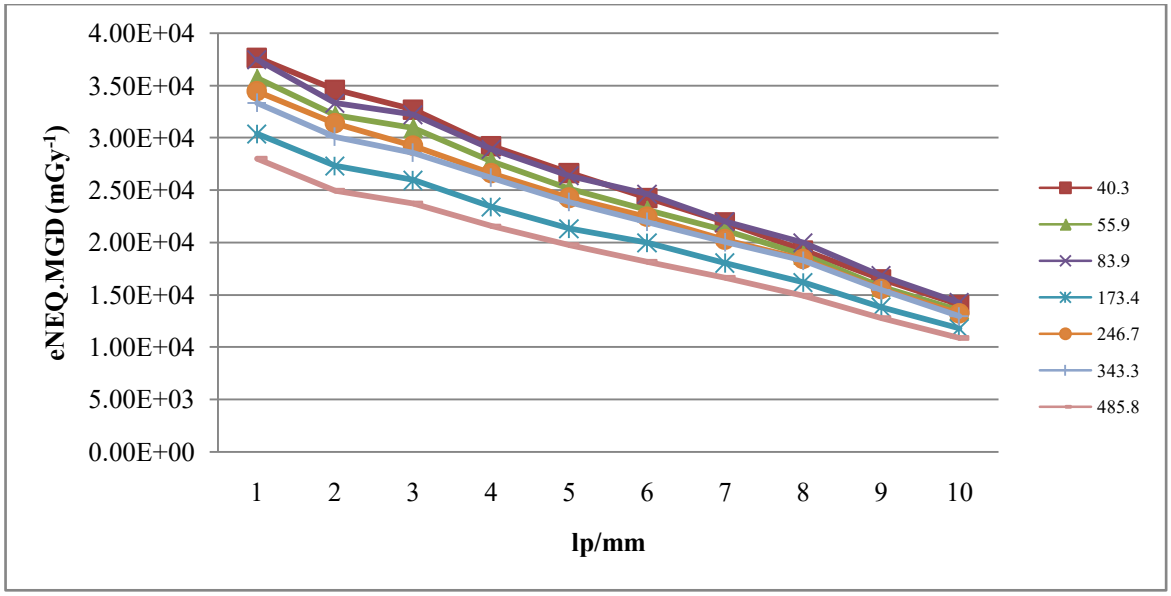
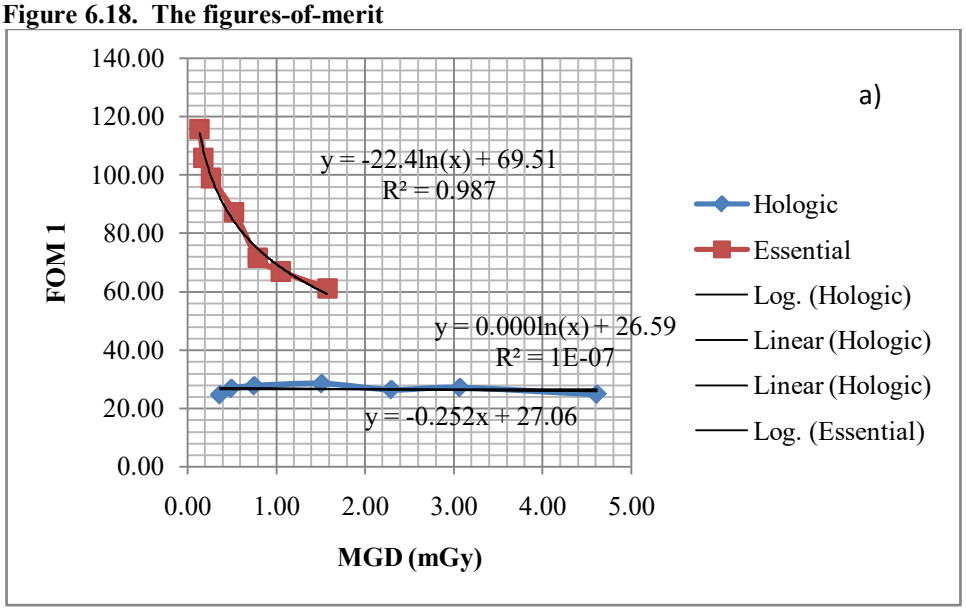
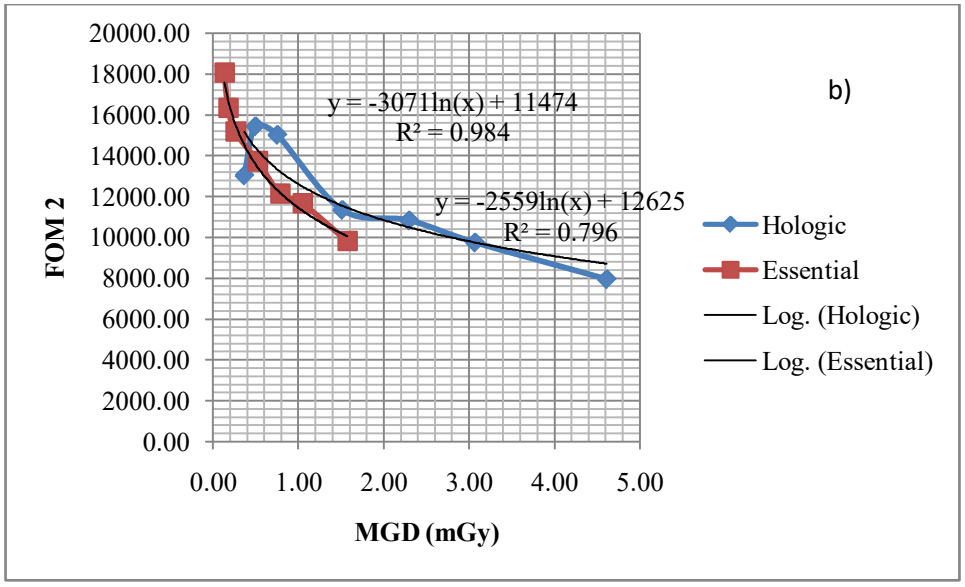
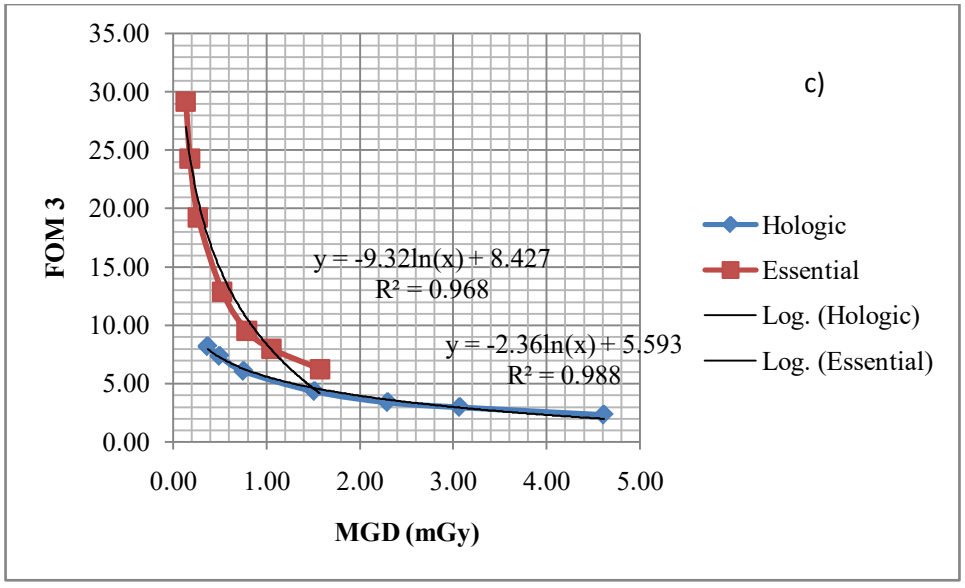
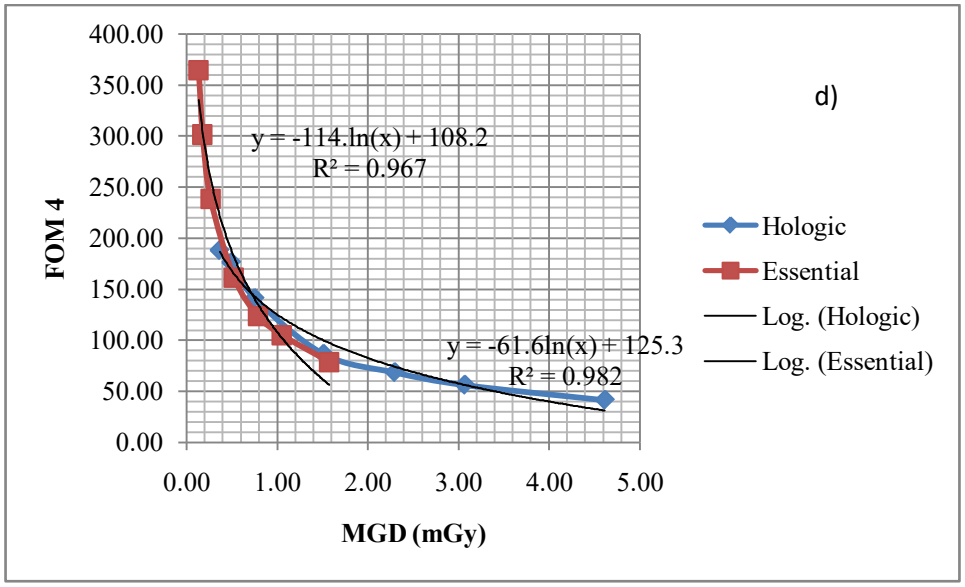


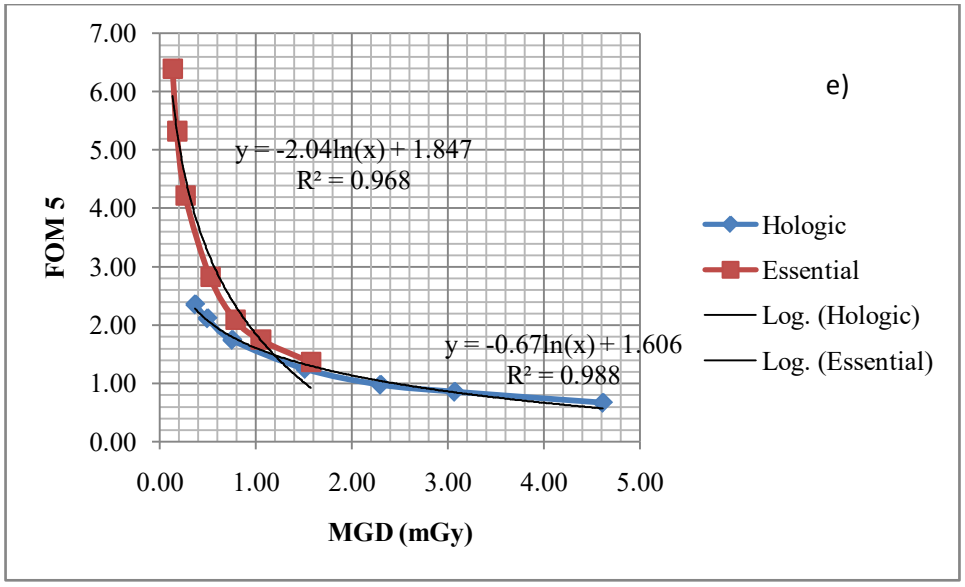
Figure 6.18. The figures-of-merit

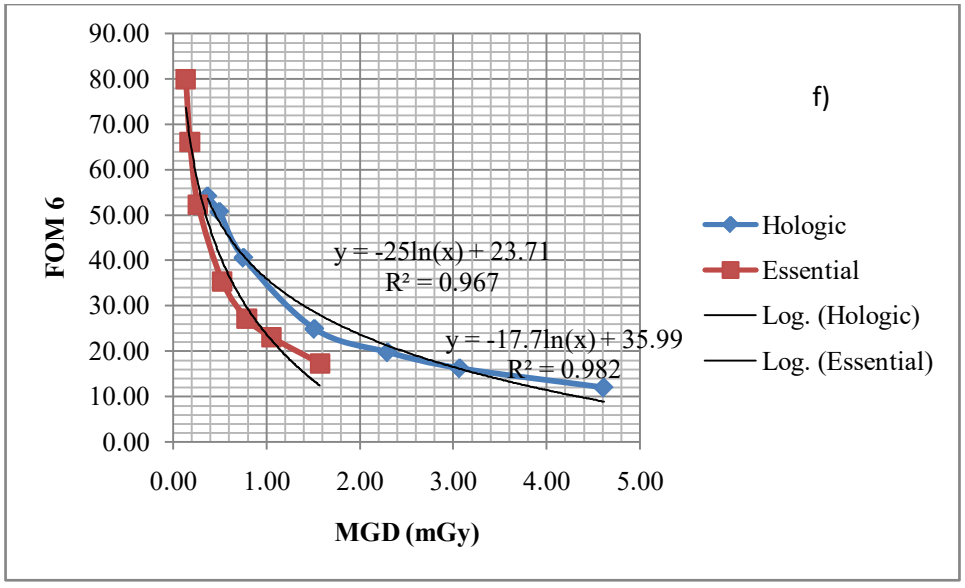


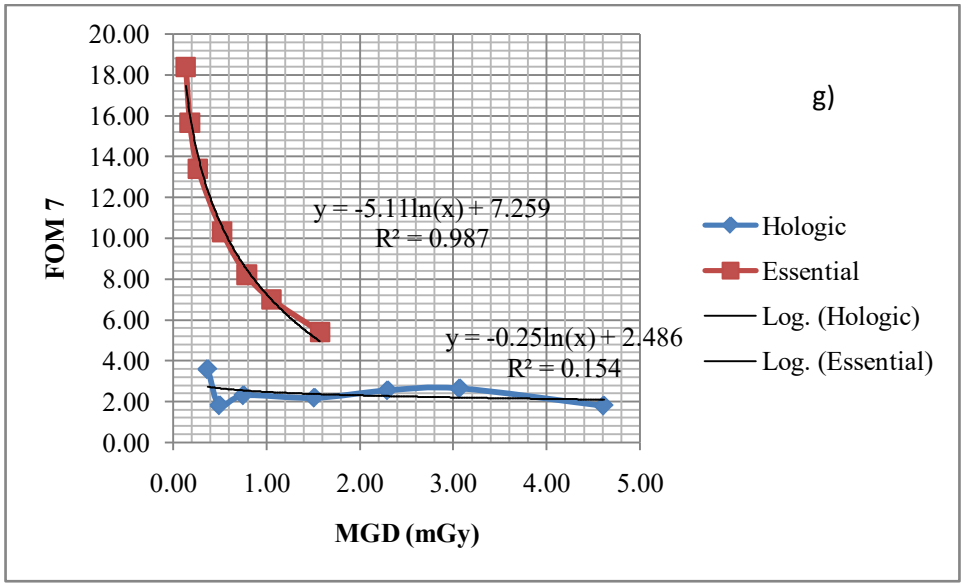

b)
FOM 2
MGD (mGy)
y = -3071ln(x) + 11474
R² = 0.984
y = -2559ln(x) + 12625
R² = 0.796
Hologic
Essential
Log. (Hologic)
Log. (Essential)

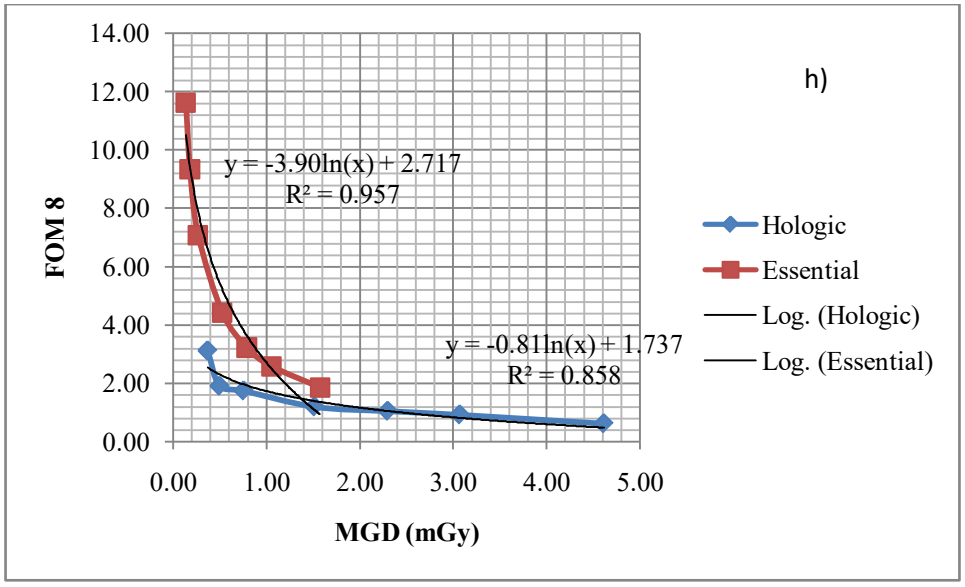

c)
FOM 3
MGD (mGy)
y = -9.32ln(x) + 8.427
R² = 0.968
y = -2.36ln(x) + 5.593
R² = 0.988
Hologic
Essential
Log. (Hologic)
Log. (Essential)

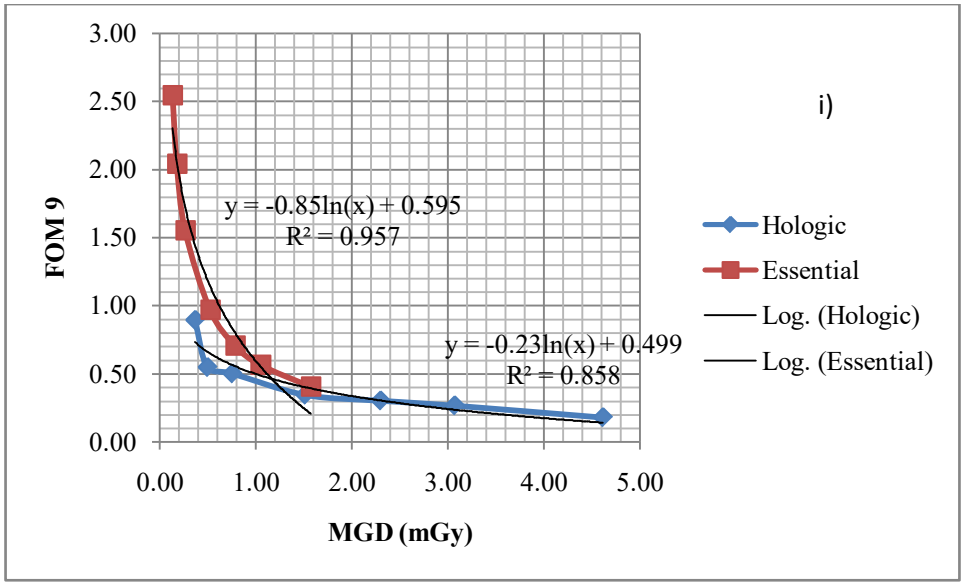

d)
FOM 4
MGD (mGy)
y = -114.ln(x) + 108.2
R² = 0.967
y = -61.6ln(x) + 125.3
R² = 0.982
Hologic
Essential
Log. (Hologic)
Log. (Essential)

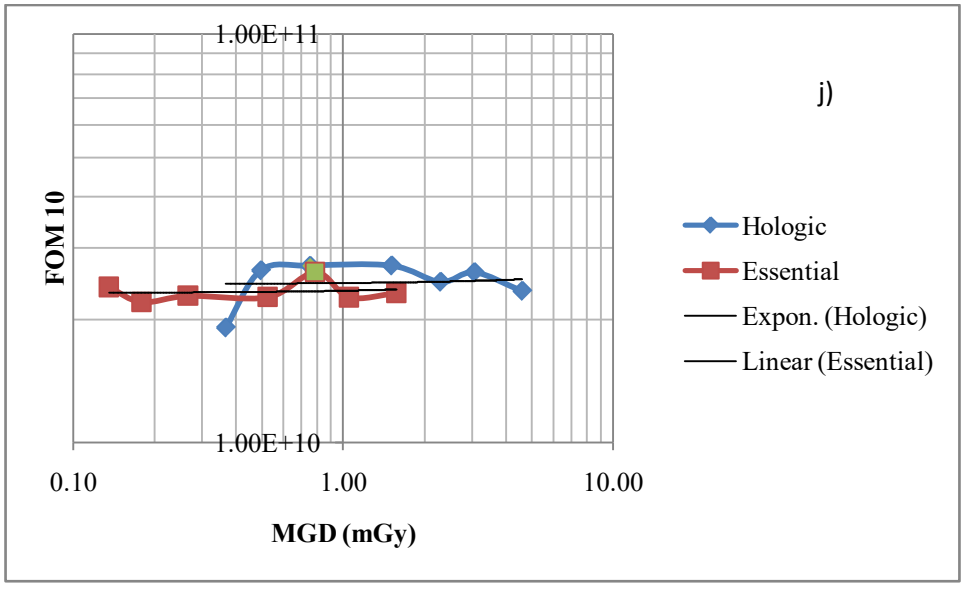

e)
FOM 5
MGD (mGy)
y = -2.04ln(x) + 1.847
R² = 0.968
y = -0.67ln(x) + 1.606
R² = 0.988
Hologic
Essential
Log. (Hologic)
Log. (Essential)

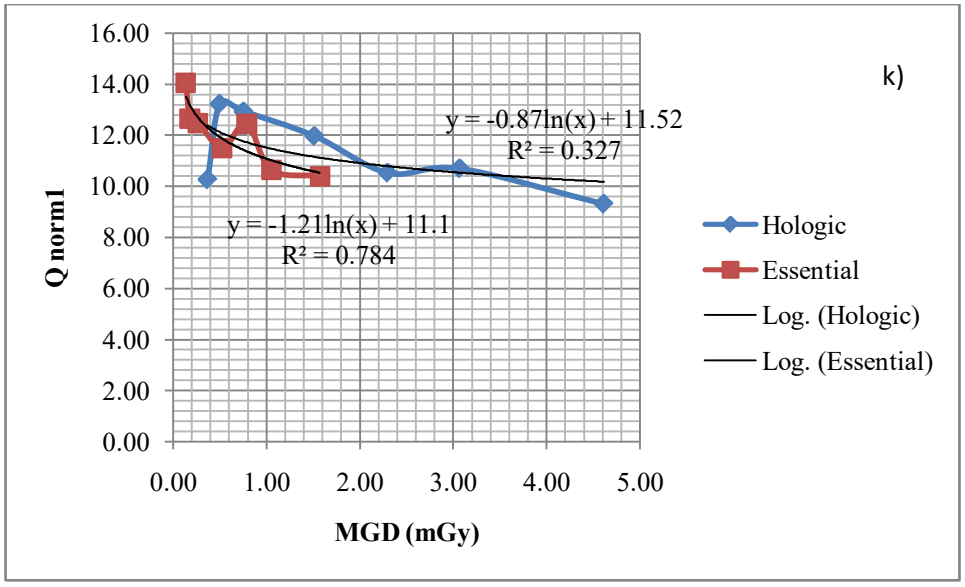

f)
FOM 6
MGD (mGy)
y = -25ln(x) + 23.71
R² = 0.967
y = -17.7ln(x) + 35.99
R² = 0.982
Hologic
Essential
Log. (Hologic)
Log. (Essential)

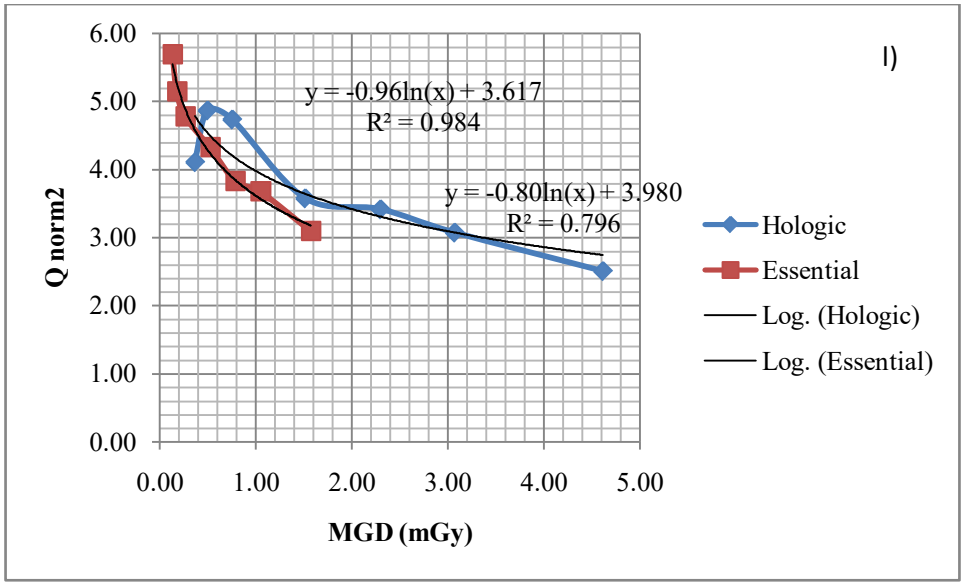

g)
FOM 7
MGD (mGy)
y = -5.11ln(x) + 7.259
R² = 0.987
y = -0.25ln(x) + 2.486
R² = 0.154
Hologic
Essential
Log. (Hologic)
Log. (Essential)

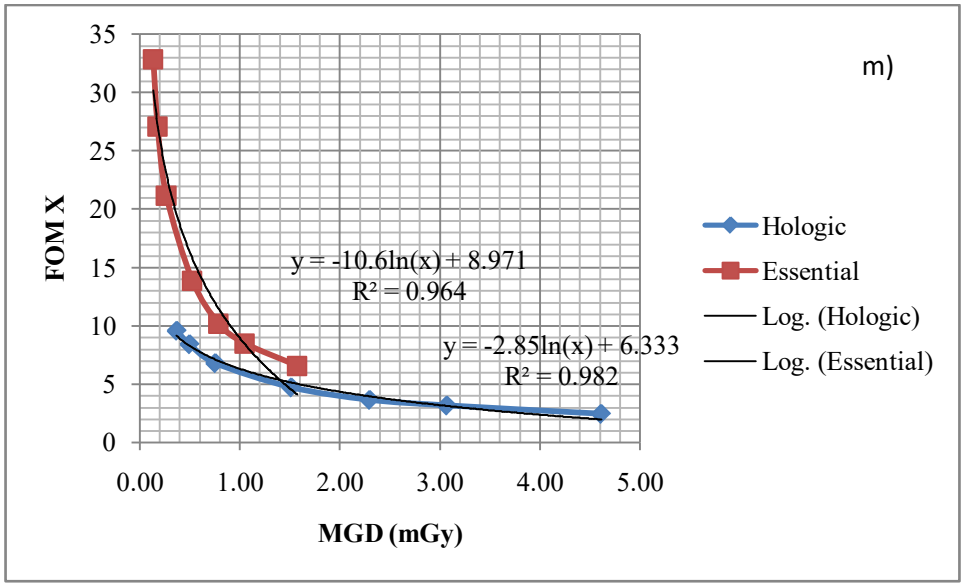

h)
FOM 8
MGD (mGy)
y = -3.90ln(x) + 2.717
R² = 0.957
y = -0.81ln(x) + 1.737
R² = 0.858
Hologic
Essential
Log. (Hologic)
Log. (Essential)


i)
FOM 9
MGD (mGy)
y = -0.85ln(x) + 0.595
R² = 0.957
y = -0.23ln(x) + 0.499
R² = 0.858
Hologic
Essential
Log. (Hologic)
Log. (Essential)


j)
FOM 10
MGD (mGy)
1.00E+11
1.00E+10
Hologic
Essential
Expon. (Hologic)
Linear (Essential)


k)
Q norm1
MGD (mGy)
y = -0.87ln(x) + 11.52
R² = 0.327
y = -1.21ln(x) + 11.1
R² = 0.784
Hologic
Essential
Log. (Hologic)
Log. (Essential)


l)
Q norm2
MGD (mGy)
y = -0.96ln(x) + 3.617
R² = 0.984
y = -0.80ln(x) + 3.980
R² = 0.796
Hologic
Essential
Log. (Hologic)
Log. (Essential)


m)
FOM X
MGD (mGy)
y = -10.6ln(x) + 8.971
R² = 0.964
y = -2.85ln(x) + 6.333
R² = 0.982
Hologic
Essential
Log. (Hologic)
Log. (Essential)

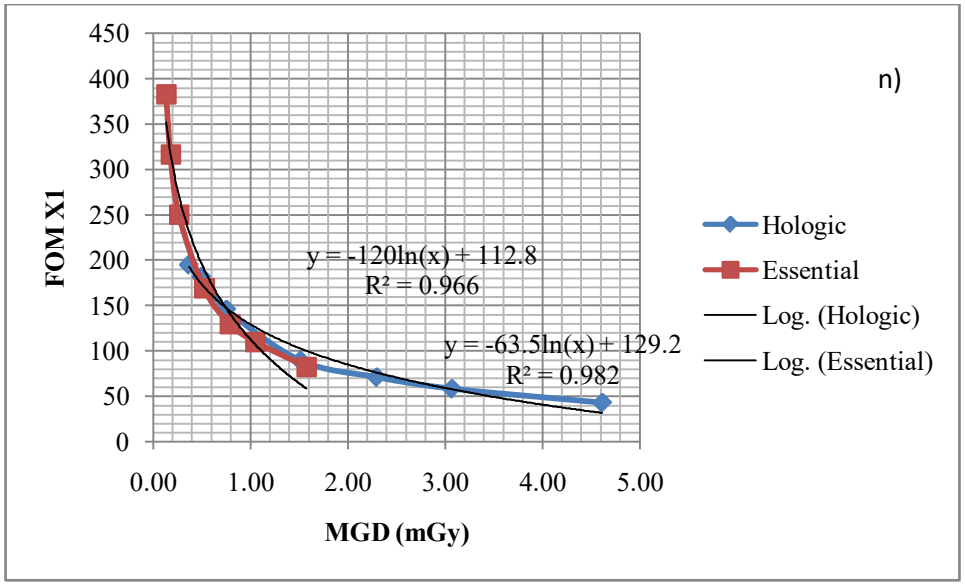
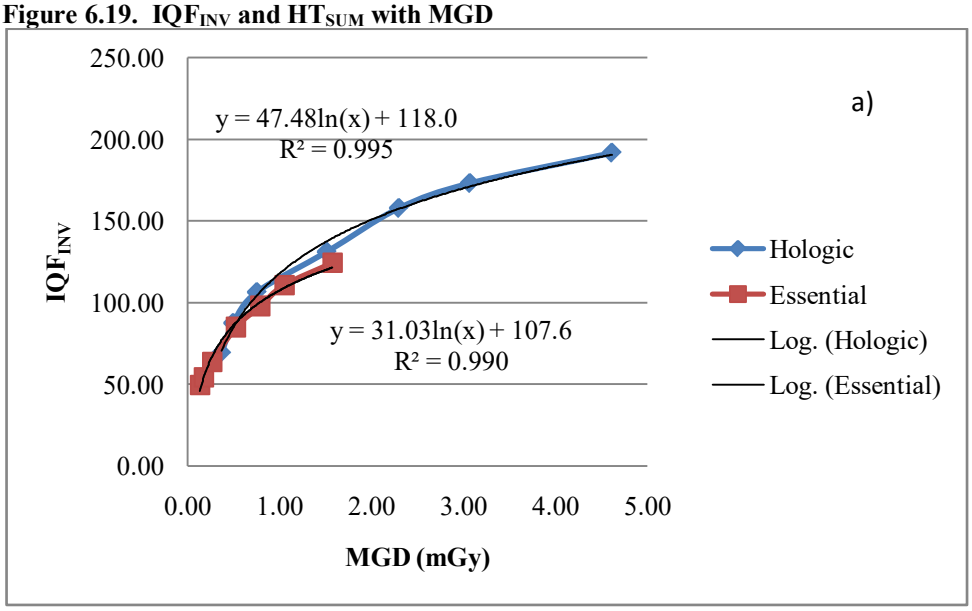


Figure 6.19. IQF_{INV} and HT_{SUM} with MGD



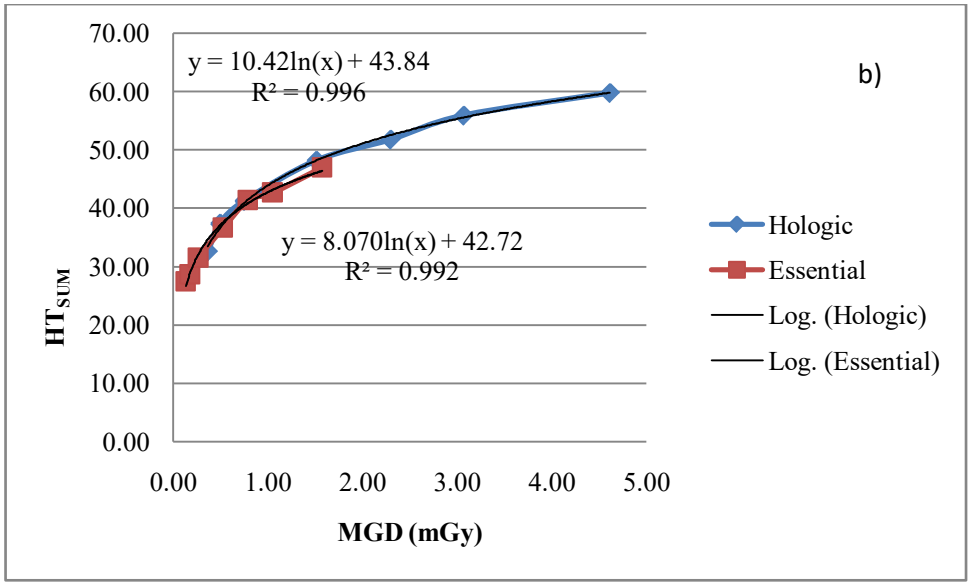
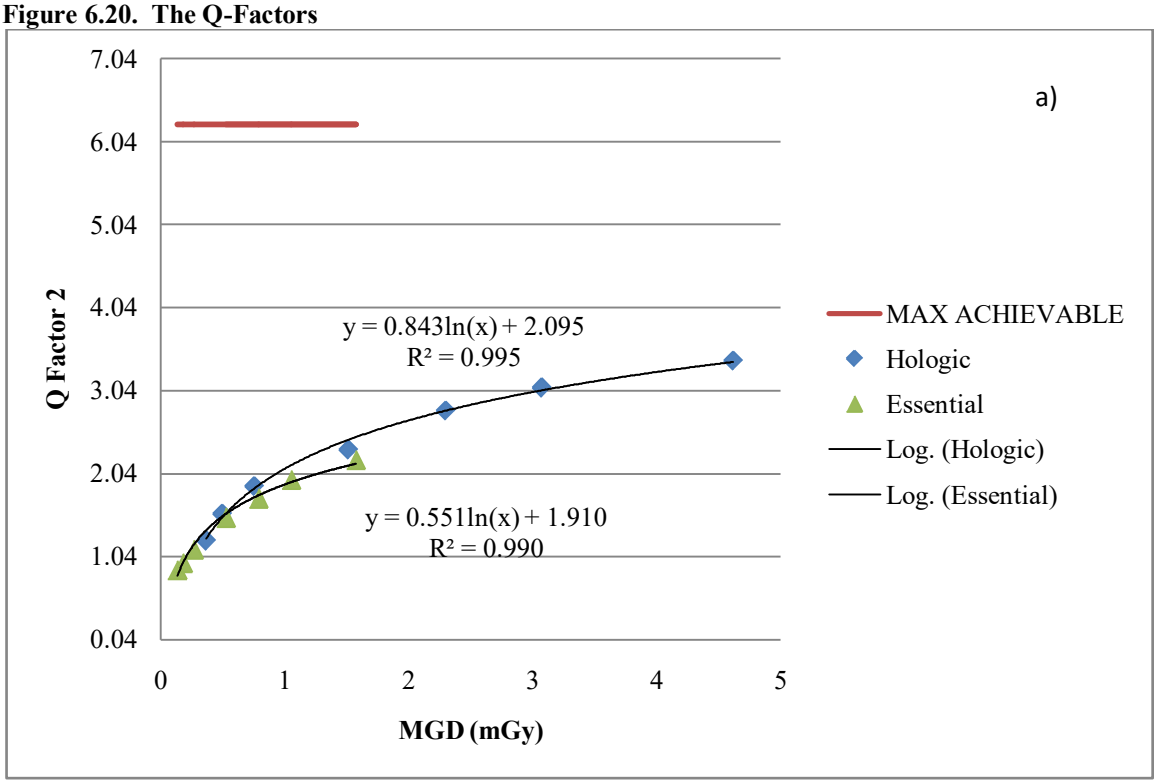
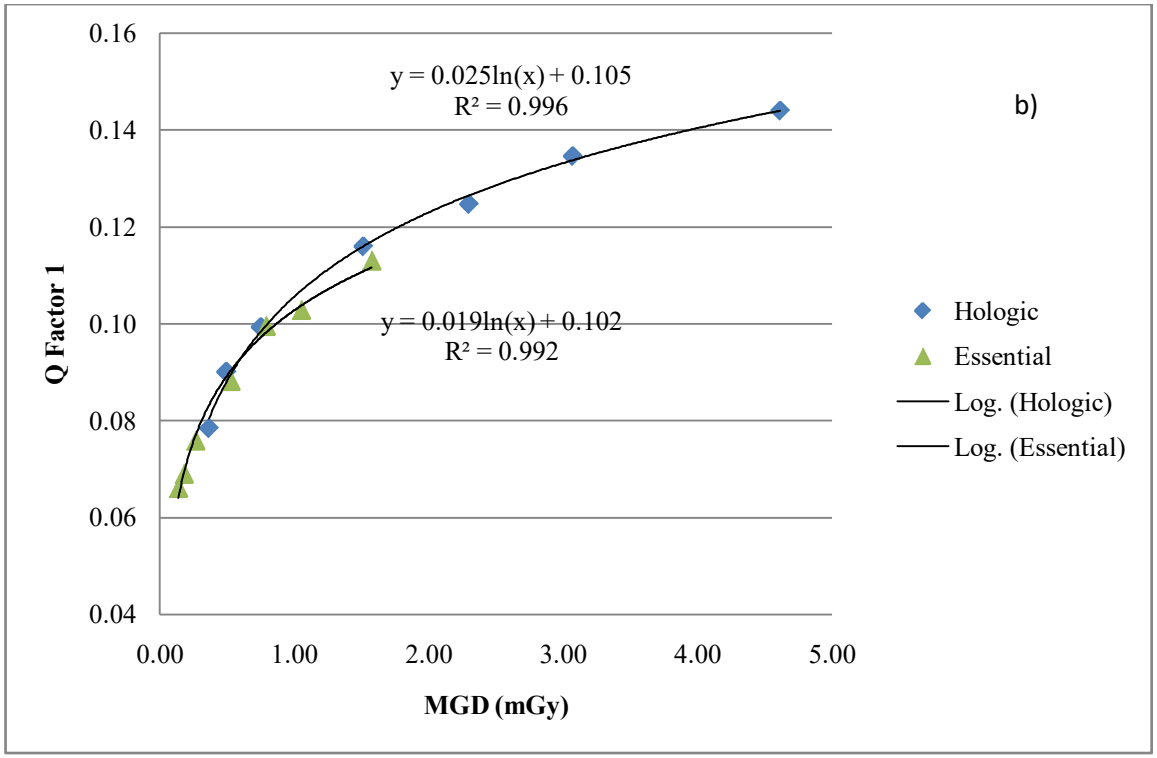
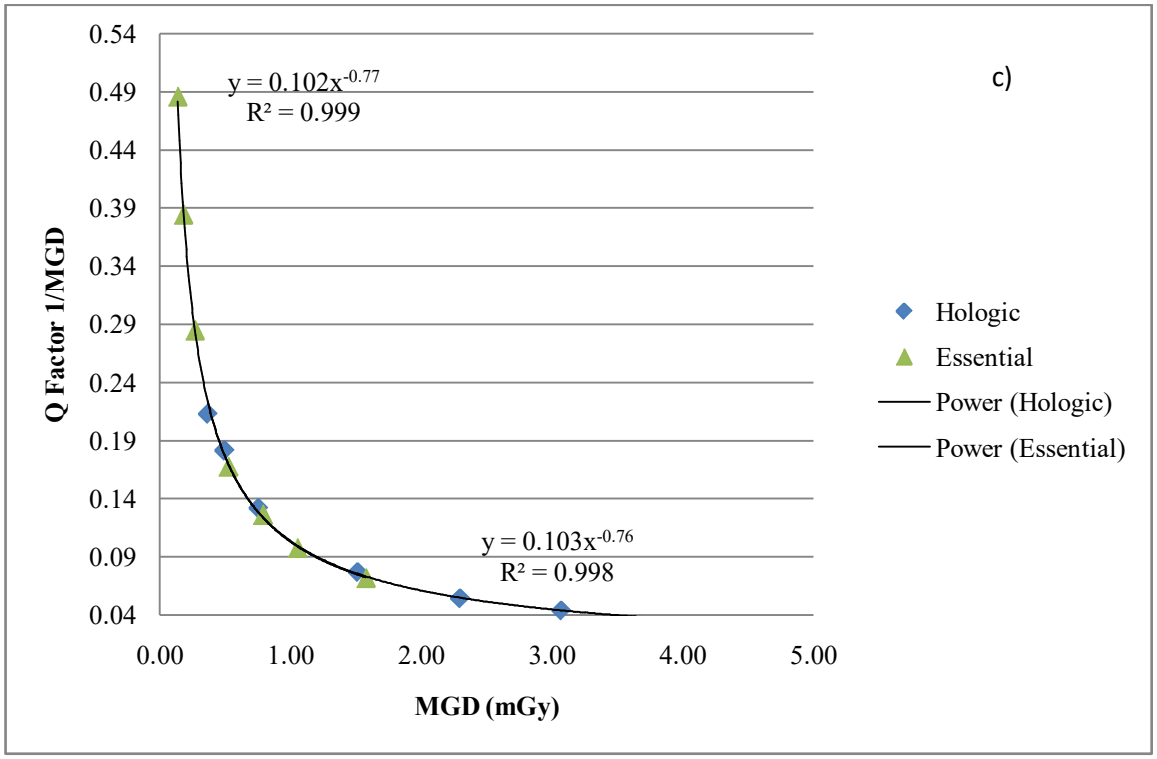


Figure 6.20. The Q-Factors




b)
y = 0.025ln(x) + 0.105
R² = 0.996
y = 0.019ln(x) + 0.102
R² = 0.992
Q Factor 1
MGD (mGy)
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.00
1.00
2.00
3.00
4.00
5.00
Hologic
Essential
Log. (Hologic)
Log. (Essential)


c)
y = 0.102x-0.77
R² = 0.999
y = 0.103x-0.76
R² = 0.998
Q Factor 1/MGD
MGD (mGy)
0.54
0.49
0.44
0.39
0.34
0.29
0.24
0.19
0.14
0.09
0.04
0.00
1.00
2.00
3.00
4.00
5.00
Hologic
Essential
Power (Hologic)
Power (Essential)

FIGURES-OF-MERIT

With reference to figure 6.18, the various FOM defined in table 6.1b show a variation in response with increased MGD. In other instances the systems showed similar patterns and the author feels that these can be compared or investigated in further depth.

It is immediately realised that the Essential resulted in FOMs higher in magnitude when compared to the Hologic despite the lower DAK values and hence the lower MGD values employed. It is easily noticeable that FOM 3, 4, 5, 6, 8, 9, X and X1 yielded a similar response in both systems. Regression lines were applied between the mentioned FOM and MGD showing a strong logarithmic relationship with R^2 values >0.86 for all plots. These FOMs decreased in magnitude with increased dose to the patient.

The CNR or SDNR in FOM 1 was measured as defined in the European protocol. FOM 7 represents the same concept however the CNR was measured using the CDMAM phantom as described previously. The graph shapes for both FOM 1 and 7 appear identical for the Essential and it is clearly noticeable that the relationship between the FOM value with increasing MGD is strongly logarithmic. This relationship appears somewhat different in the Hologic. The relationship between FOM 1 and 7 and MGD appears to be non-existent.

With reference to figure 6.18b, the relationship between FOM 2 and MGD appears strongly logarithmic in the Essential. In the Hologic, the FOM reached a peak at 0.50 mGy and then decreased logarithmically with increased MGD. The same pattern was seen in Qnorm1 and Qnorm2 in figures 6.18k and l respectively. It was noticed that the standard AEC dose was nowhere near the peak FOM value but was actually situated at a higher dose with reference to the graphs. This observation might require further investigation before any conclusions may be drawn. However, given this information one would be inclined in saying that the FOM 2 may be useful in system optimization of the Hologic, even with such early results.

In figure 6.19a, the IQF_{INV} was also plotted with MGD showing a strong logarithmic relationship in both systems. The same relationship was observed when the summed H_T (figure 6.19b) value and the Q factors (figure 6.20) were plotted against MGD. This relationship is probably attributed to the CDMAM configuration of gold objects whereby each gold detail increases in size and diameter logarithmically too.

The relationship observed between FOM 10 and MGD was different and less predictable. The image quality parameter selected was the summed H_T value from the automatically scored CDMAM phantom. With reference to figure 6.18j, FOM 10 revealed a poor R^2 value with MGD and there appears to be a slight or no change in FOM 10 with MGD. Interesting to note was that there appeared to be a match in FOM 10 at the same MGD, shown as the green data point in figure 6.18j.

The best overall R^2 value for both systems resulted in FOM 3, 4, 5, 6, X and X1 (>0.96). The Essential showed a strong logarithmic relationship between the FOMs and MGD except in FOM 10. Further still, the magnitude of each explored FOM for the Essential was larger in magnitude than those in the Hologic probably attributed to the lower DAK/MGD values employed. The latter unit not only showed a different magnitude in FOM, but it also seemed that its response with increasing MGD varied too. This was particularly observed in FOM 1, 2, 6, Qnorm1 and Qnorm2.

In FOM X and X1, with reference to figure 6.18m and n, it is observed that both systems showed a very strong logarithmic response. As in previous instances, the Essential appeared to show an overall higher FOM magnitude of X and X1 even at matched dose values (40-50% higher in the Essential).

Q factor and normalized Q factor – Approach 1

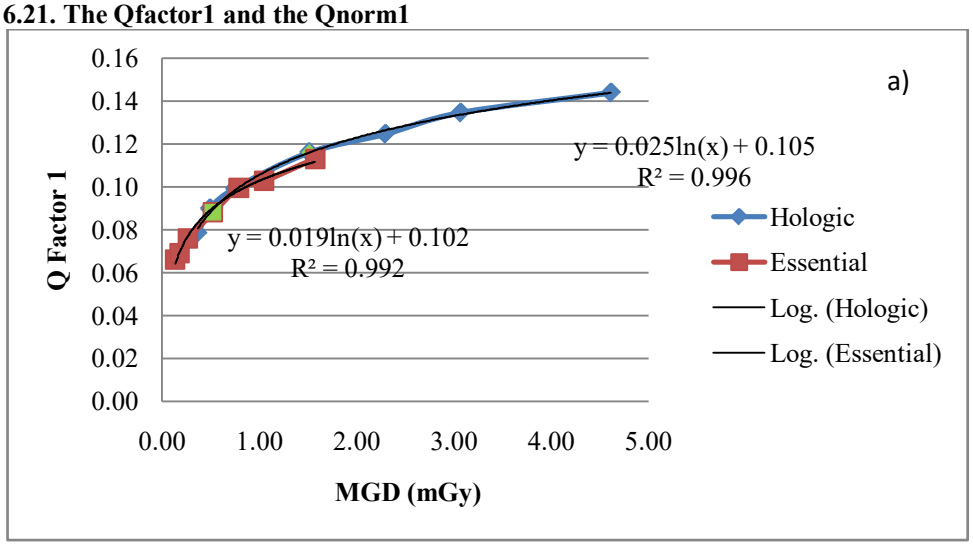
Unlike the well-known CDMAM-related image quality indicator such as the H_T and IQF_{INV} and also the CNR values, the Q factors defined in this study, as already stated, are taking a reference value into account. This implies that apart from measuring the image quality value, it can be compared to a theoretical value or a known value and in this way different systems can be potentially compared.

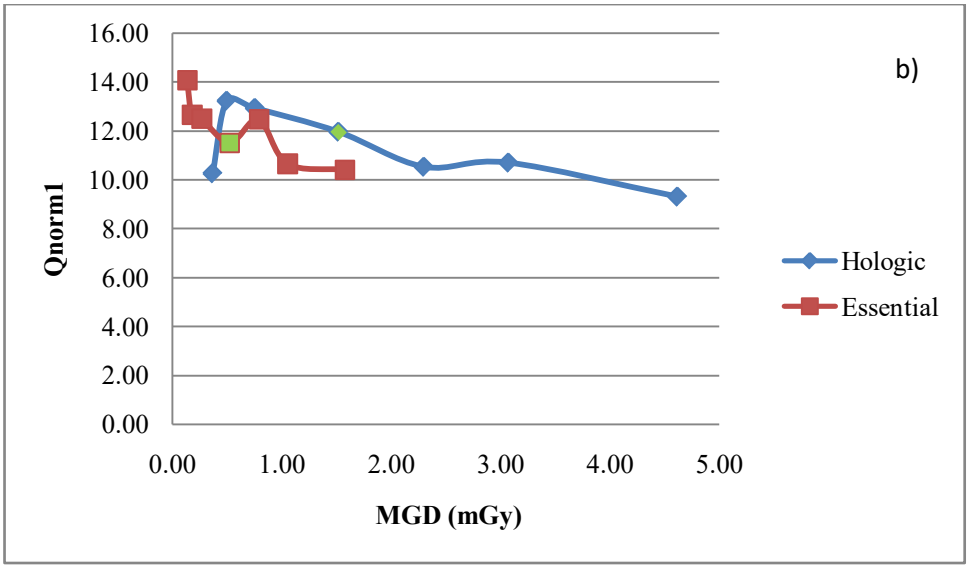
With reference to figure 6.21a, the Qfactor1 was plotted against MGD for the Essential and the Hologic. As it can be observed there is again a very strong logarithmic relationship between the two values. The green data point represents the AEC exposure and the respective Qfactor1. As can be seen in both systems, the higher the dose the better the image quality through this approach. Interestingly, both systems showed a very similar response in this regard. Consequently, the Qfactor1 can be confidently predicted from or for a given MGD.

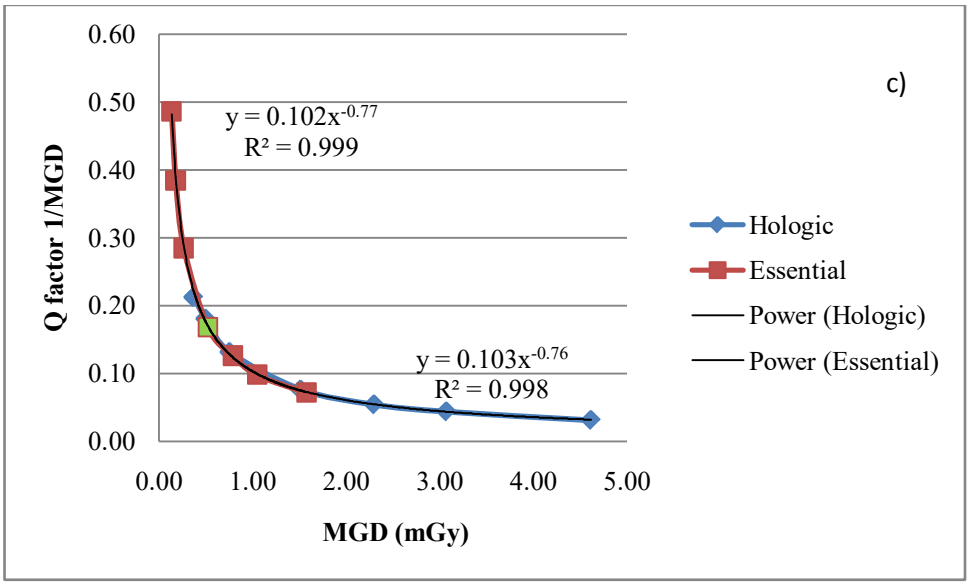
The Qnorm1 in equation 38 was plotted against the calculated MGD in figure 6.21b. The resulting graph showed that Qnorm1 decreased with MGD for both systems but its relationship with MGD was less remarkable. Interesting to note, that in the Hologic system the Qnorm1 revealed a shoulder at 0.5 mGy and then decreased constantly. In the Essential a peak in Qnorm1 was observed at 0.79 mGy. It can be observed that AEC exposures in both systems resulted in Qnorm1 values that are different from each other and also they did not result in the highest Qnorm1 values. This could be an indication that these exposures may require optimization in both systems.

In figure 6.21c the Qfactor1/MGD was plotted against MGD showing a very good power response. It is again seen from the graph that the Essential showed a higher normalised Qfactor1 at lower MGD, including the AEC exposure when compared to the Hologic. Once again, this is probably attributed to the lower doses employed. Nevertheless, at matched MGD levels (1.0 mGy) the Qfactor1/MGD levels difference between the two units was less than 1%.

Figure 6.21. The Qfactor1 and the Qnorm1







Q factor and normalized Q factor – Approach 2

Plots of the Qfactor2*100 in figure 6.22c and d show that there is a very good relationship between this image quality parameter and all the other previously defined CDMAM image quality parameters. The AEC exposure in figure 6.22 is again manifested by a green data point. It is worth noting that Qfactor2*100 relates better in terms of R^2 value with the CNR value achieved from the European protocol method then the method using the CDMAM phantom itself (figures 6.22a and b respectively). Figure 6.22e shows the Qfactor2*100 as a function of MGD and also shows the maximum achievable Qfactor2*100 value as specified earlier. The large R^2 in this graph shows that with a given MGD value in the Essential and the Hologic, the Qfactor2*100 is also very predictable. This means that an expected image quality indicator

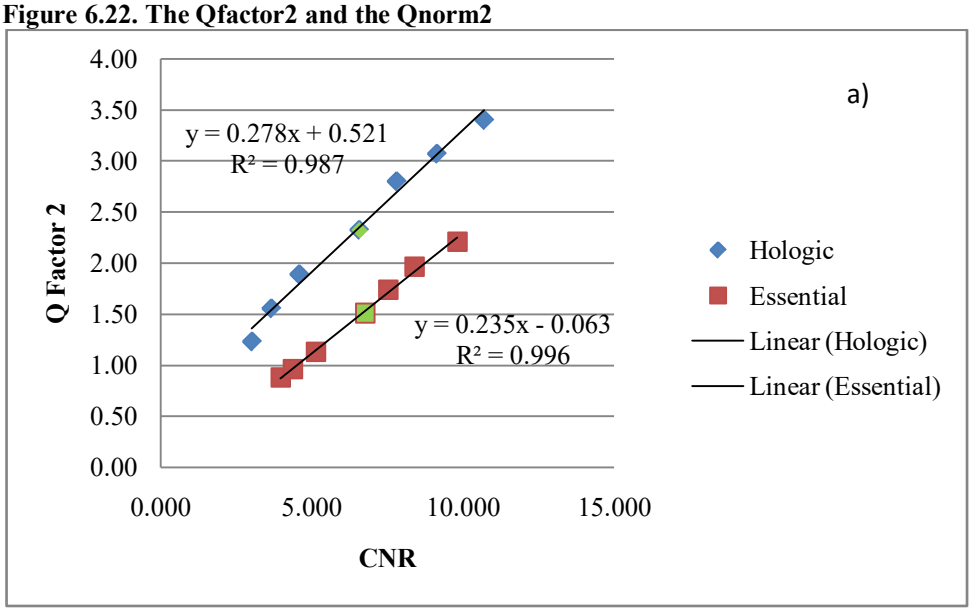
such as IQF_{INV}, CNR or H_T can be predicted too. This is has also proved to be a promising factor for system inter-comparison and also for system optimization.

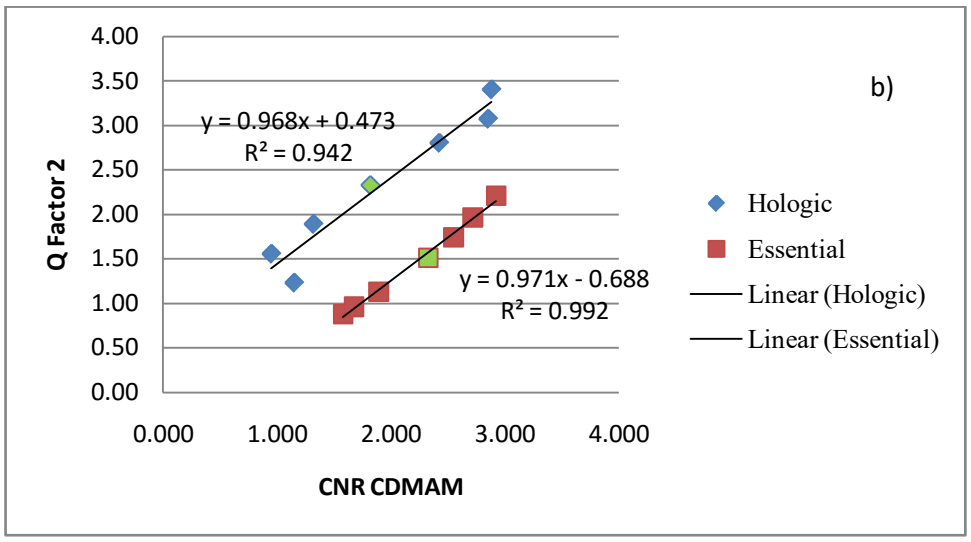
Table 6.7. Predicted threshold contrast and H_T values for the CDMAM phantom

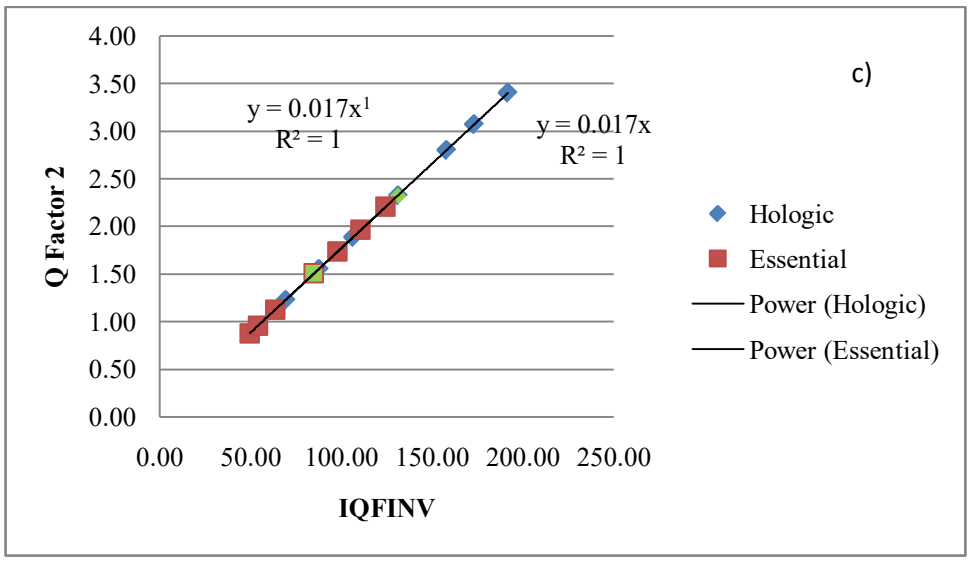
	Threshold Contrast (%)		H_T	
Diameter (mm)	**Acceptable**	**Achievable**	**Acceptable**	**Achievable**
0.06	56.78	38.62	2.49	3.02
0.08	34.15	23.23	2.41	2.92
0.1	23.17	15.84	2.34	2.83
0.13	14.87	10.25	2.25	2.71
0.16	10.60	7.37	2.16	2.60
0.2	7.49	5.25	2.06	2.46
0.25	5.41	3.81	1.94	2.31
0.31	4.05	2.85	1.81	2.16
0.4	2.98	2.07	1.63	1.96
0.5	2.36	1.60	1.47	1.78
0.63	1.91	1.26	1.29	1.59
0.8	1.60	1.01	1.11	1.40
1	1.39	0.85	0.96	1.23
1.25	1.25	0.72	0.81	1.06
1.6	1.13	0.62	0.66	0.89
2	1.05	0.55	0.55	0.76
		H_TSum	**25.95**	**31.69**
		H_TAv	**1.62**	**1.98**

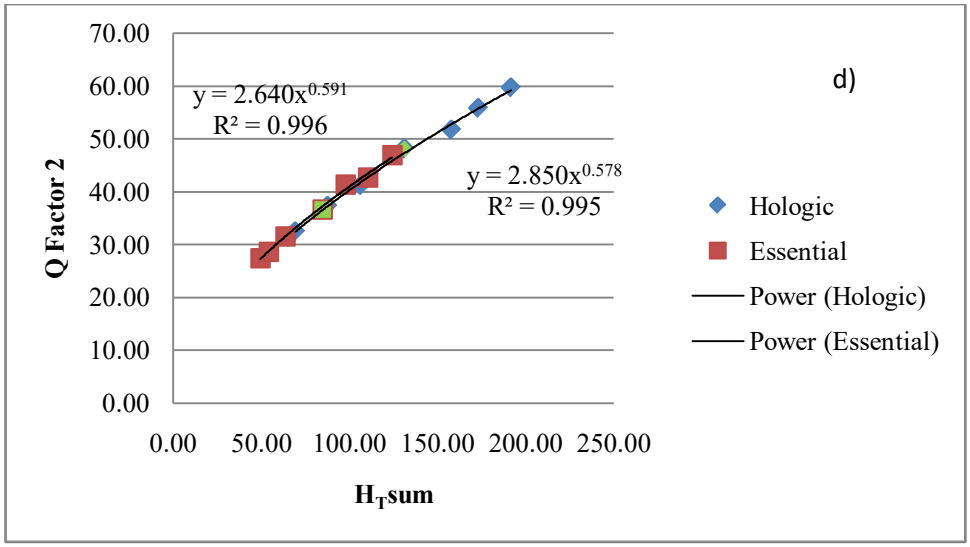
This signifies that if we are in a position to confidently predict an image quality indicator value for a given MGD, and we then compare this predicted value to an achieved or measured value we can determine whether optimization is required for that system (figure 6.23). Consequently, knowing this information, system image quality can be also inter-compared. With reference to figure 6.22f, the Qnorm2 showed a relationship with MGD relatively similar to Qnorm1 despite the different approach. Once again it can be observed that AEC exposures in both systems resulted in Qnorm2 values that are different from each other and also did not result in the highest Qnorm2 numbers. This could be another indication that these systems may require optimization.

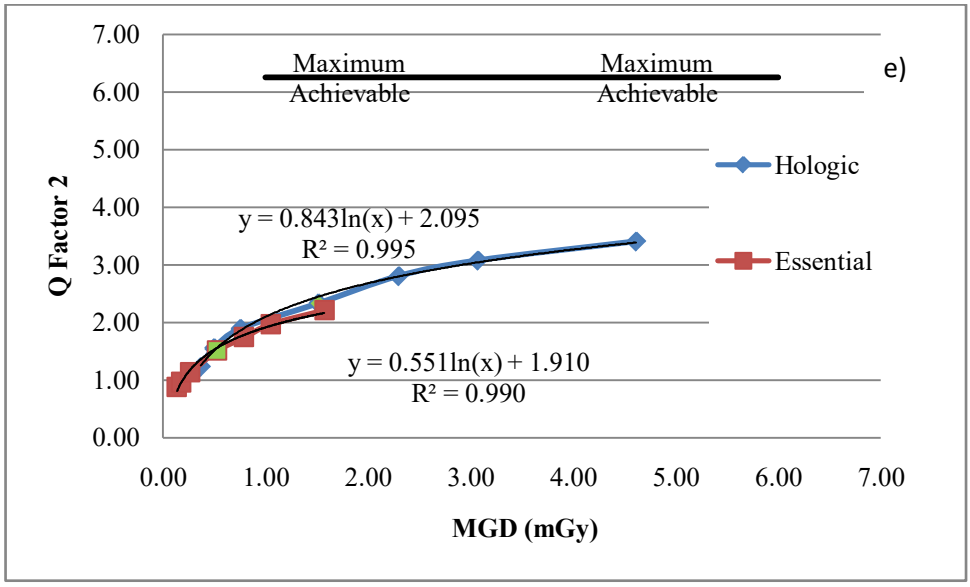
Figure 6.22. The Qfactor2 and the Qnorm2








d)
y = 2.640x^0.591
R² = 0.996
y = 2.850x^0.578
R² = 0.995
Hologic
Essential
Power (Hologic)
Power (Essential)
Q Factor 2
H_T sum
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
0.00
50.00
100.00
150.00
200.00
250.00


e)
Maximum Achievable
Maximum Achievable
Hologic
Essential
y = 0.843ln(x) + 2.095
R² = 0.995
y = 0.551ln(x) + 1.910
R² = 0.990
Q Factor 2
MGD (mGy)
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00

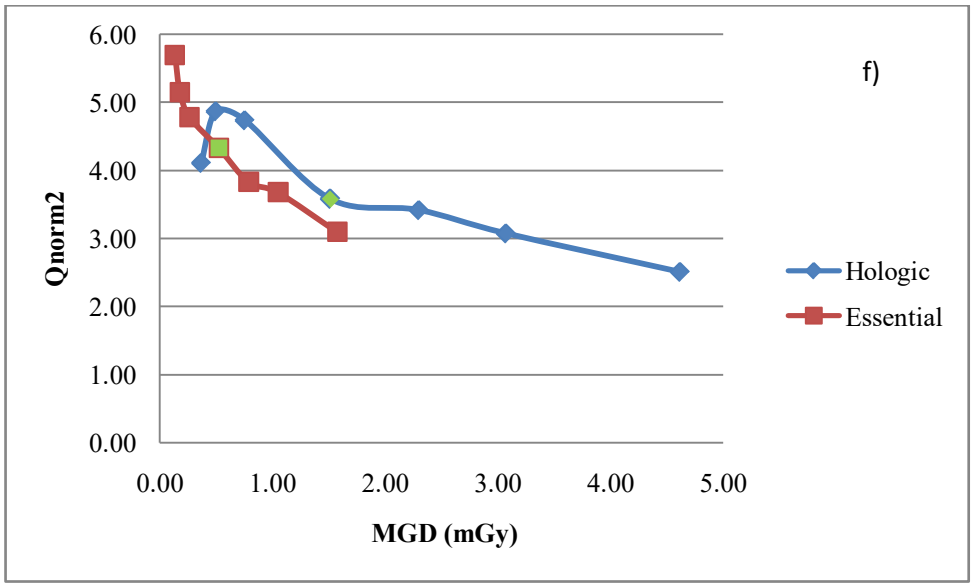
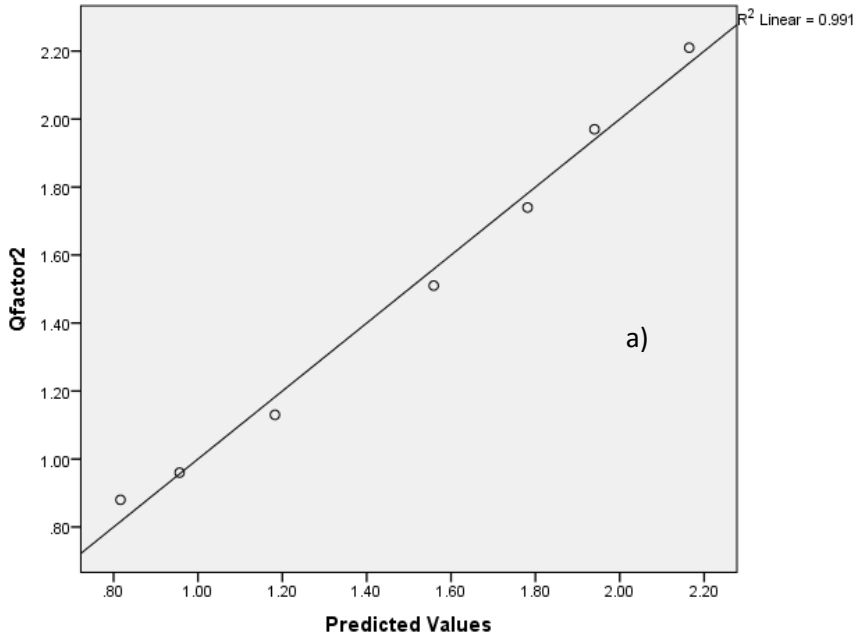
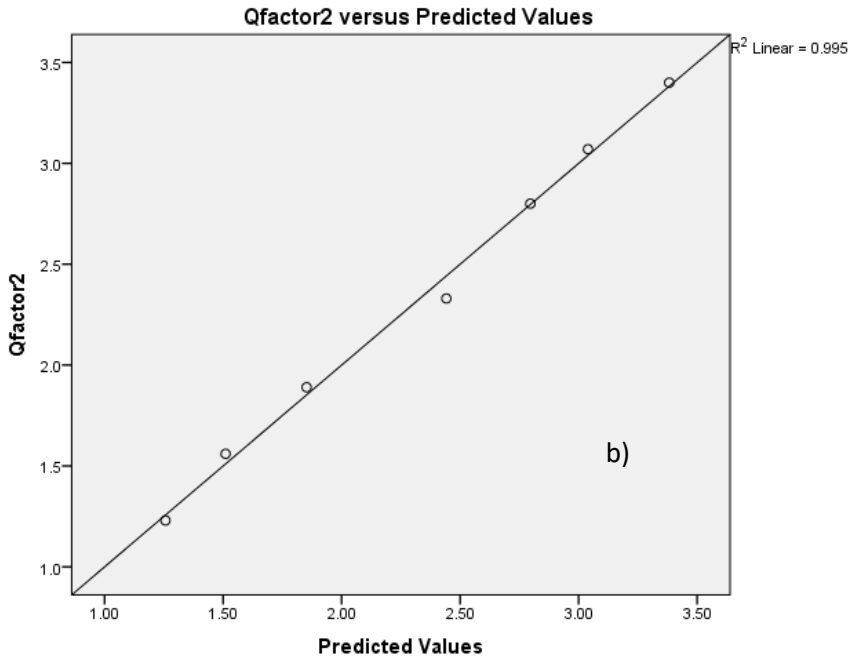

f)
Hologic
Essential
Qnorm2
MGD (mGy)
6.00
5.00
4.00
3.00
2.00
1.00
0.00
0.00
1.00
2.00
3.00
4.00
5.00

Figure 6.23. QFactor2 Versus Predicted Values a) Essential b) Hologic





Correlation Analysis

The FOM data presented within this study were all normally distributed. This was verified by the Kolmogorov Smirnov test using IBM SPSS version 20. The Kolmogorov Smirnov test tries to determine if two datasets differ significantly. The Kolmogorov Smirnov test has the advantage of making no assumption about the distribution of data. A Pearson r correlation analysis was eventually done between all the FOM presented, the Q factors, Qnorm1, Qnorm2, MGD, K and DAK and the results are presented and summarised in tables 6.8 and 6.9. Primarily, this analysis is important in establishing how well each FOM inter-correlates, corresponds and relates with one another. The analysis is equally useful in identifying how the FOMs, image quality indicators and/or MGD correlate, whether good or not so good, showing how well or to what extent all these factors move in relation to one another. A poor correlation indicates poor correspondence or equivalence and is also of interest within the scope of this exploration. As stated earlier the most reputable FOM is FOM 1. Consequently, FOMs that correlate poorly with FOM 1 also raise interest and are subject to further investigation.

With reference to table 6.8, there appears to be a good inter-correlation between most FOMs with Pearson r values greater than $\pm$0.8 in the Essential. Most of the results were statistically significant ($p<0.05$). However, equally interesting were those FOM that resulted in low Pearson r values. Particularly noticeable was FOM 10, which showed no correlation with all the other factors explored although p values were higher than the 0.05 level of significance. This was also the case in table 6.9 for the Hologic system.

The Hologic, in table 6.9 showed an overall different correlation pattern when compared to the Essential. There was a larger number of 'red boxes' showing a bigger number of poor inter-correlations. Interesting to note was that in the Hologic FOM 1 correlated significantly well with FOM 10 very much unlike the Essential.

There were also some differences in correlation between the quality parameters themselves. It resulted that CNR, IQF_{INV} and HT_{SUM} correlated poorly with FOM 10 in both systems, however all results were statistically insignificant. The Hologic showed more statistically insignificant poor inter-correlations between the aforementioned quality parameters and FOM 1, 7 and Qnorm1. All the image quality parameters correlated extremely well ($r>0.94$) with MDG, K and DAK. The latter results were all statistically significant.

There was again a mixed-up response in the correlation analysis between the image quality parameters and the detector quality factors in both systems. The Essential showed a relatively poor correlation (r=<0.5) between CNR, IQF_{INV} and HT_{SUM} and DQE and eDQE at 0.5 lp/mm; but all results had p values >0.05 hence statistically insignificant. In the Hologic there appeared to be two extremes; a very good correlation (r>0.88) with DQE and a very poor correlation (r<0.20) with eDQE. The former was significant (p<0.05) and the latter result was statistically insignificant. A good relationship was seen in both systems with NEQ and NNPS (r>0.75) which was remarkably better in the Essential (r>0.92). The same good correlation was seen between the effective counterparts (r>0.96) in both systems. All the results concerning NEQ and NNPS and their effective counterparts with the mentioned image quality parameters were statistically significant.

With reference to table 6.9, there appears to be no remarkable correlation between CNR, CNR with CDMAM, IQF_{INV} and HT_{SUM} and NEQ_{MGQ}, NEQ_{DC} and their effective counterparts for both systems. The only exception, with reference to table 6.9b, were the values shaded in green for the Hologic system. The CNR, IQF_{INV} and HT_{SUM} showed a relatively good statistically significant correlation with $eNEQ_{MGD}$ (r>0.77).

The NEQ_{MGQ}, NEQ_{DC} and their effective counterparts for both systems were also correlated with all the respective FOMs, Q factors and Qnorms. In the Essential, a significant correlation was recorded between $eNEQ_{DC}$ using the CDMAM phantom and FOMs 2, 3, 4, 5, 6, 8, 9, Qfactor/MGD, X and X1. In the Hologic system significant correlations were found to occur between $eNEQ_{DC}$ using the CDMAM phantom and FOM 7 and between $eNEQ_{MGD}$ and Qfactors 1 and 2, FOM 2, 3, 4, 5, 6, 8, 9, X, X1 and Qnorm1.

Of particular interest were Qfactors 1 and 2, Qfactor1/MGD, Qnorm1 and 2, FOM X and X1. These parameters were correlated with all the FOMs, all the image quality factors, the NEQ_{DC}, NEQ_{MGD} and their effective matches. This was done for both systems in this project. In the Essential all the investigated parameters showed a statistically significant correlation (r>0.85) with all the FOMs and all the image quality factors. A statistically insignificant correlation was observed between these parameters and FOM 10, NEQ_{DC}, NEQ_{MGD} and their effective counterparts.

In the Hologic the results appeared slightly different. Qfactors 1 and 2, Qfactor1/MGD, Qnorm 2, FOM X and X1 all correlated significantly well (r>0.80) with all the mentioned factors above. The only exception was Qnorm 1 which did not correlate with IQF_{INV}, HT_{SUM}, FOM1, FOM3,

FOM4, FOM5, FOM6, FOM7, FOM8, FOM9 and FOM10. As with the Essential a statistically insignificant correlation was observed between these parameters and FOM 10, the NEQ_{DC}, NEQ_{MGD} and their effective components.

Very interesting to note was that FOM 1, the most used and reputable FOM, correlated extremely well with all factors except with FOM 10 and the NEQ_{DC}, NEQ_{MGD} and their effective matches in the Essential. While in the Hologic this correlation was not present with any factor including the NEQ_{DC}, NEQ_{MGD} and their effective counterparts. This correlation was however statistically insignificant ($p>0.05$).

Calculation of CNR using the CDMAM phantom

The data from the CNR results, including the conventional method and the method using the CDMAM phantom, are also all normally distributed. A correlation analysis using the Pearson method was then done on the data.

Table 6.8. Correlation between CNR values a) Essential b) Hologic

a)		CNR	CNR_CDMAM
CNR	Pearson Correlation	1	.994**
	Sig. (2-tailed)		.000
CNR_CDMAM	Pearson Correlation	.994**	1
	Sig. (2-tailed)	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

b)		CNR	CNR_CDMAM
CNR	Pearson Correlation	1	.975**
	Sig. (2-tailed)		.000
CNR_CDMAM	Pearson Correlation	.975**	1
	Sig. (2-tailed)	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

From table 6.8 there is a strong correlation between the two values with a Pearson r value of >0.97 for both systems. The result is also statistically significant with a p value <0.05 level of significance. Given this information a stepwise linear regression analysis was performed to investigate to what extent one value can predict the other for 50 mm of PMMA. After the removal of one significant outlier for the Essential using the studentized residual method the results achieved are seen in table 6.9 and 6.10.

Table 6.9. Regression analysis results between the CNR values - Essential.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.024	.169		-11.997	.000
	CNR_CDMAM	3.790	.078	.999	48.710	.000

a. Dependent Variable: CNR

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999[a]	.998	.998	.08252

a. Predictors: (Constant), CNR_CDMAM

b. Dependent Variable: CNR

Table 6.10. Regression analysis results between the CNR values - Hologic.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.166	.722		-.230	.827
	CNR_CDMAM	3.477	.351	.975	9.899	.000

a. Dependent Variable: CNR

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.975[a]	.951	.942	.69833

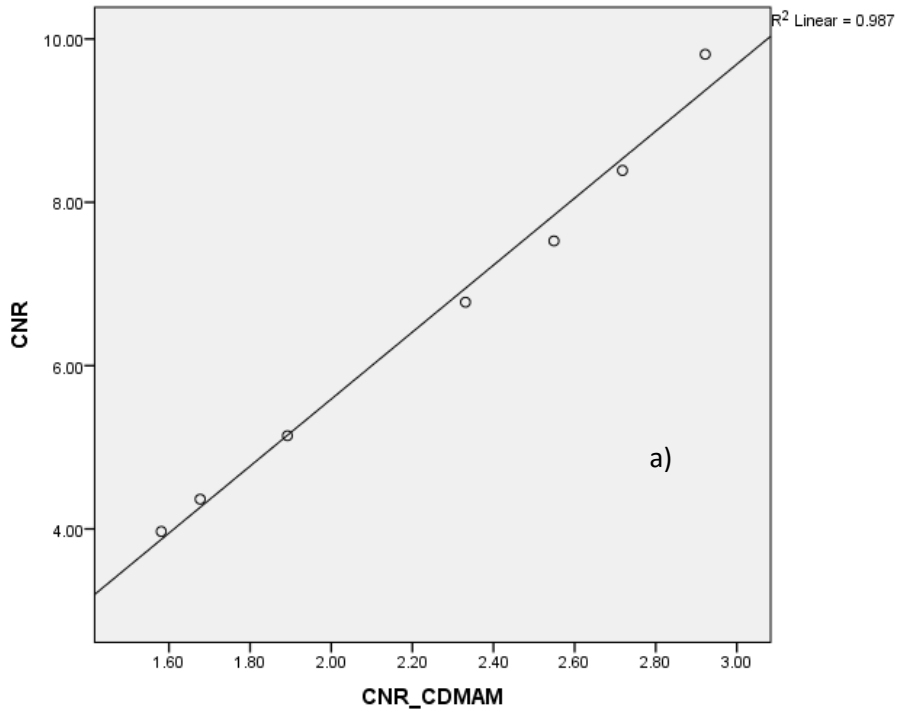
a. Predictors: (Constant), CNR_CDMAM

b. Dependent Variable: CNR

The CNR plots in figure 6.24 shows a resulting R^2 value >0.95 suggesting that the CDMAM phantom may be confidently used for the calculation of CNR at 50 mm PMMA thickness for a range of different doses. Further still, the CNR achieved from the CDMAM phantom may be converted to the conventional CNR method with the provided parameter estimates from this regression analysis in tables 6.9 and 6.10. The reliability of the predicted values is shown in the Q-Q plots in figure 6.25 whereby the R^2 value between the CNR values and the predicted values using the regression models in tables 6.9 and 6.10 is greater than 0.98.

For future reference it may be also worth exploring both CNR methods at other PMMA thicknesses. Consequently, the possibility of using the CDMAM phantom for tests like thicknesses compensation could be determined.

Figure 6.24. Plot of the CNR values using the conventional method and the values using the CDMAM phantom a) Essential b) Hologic



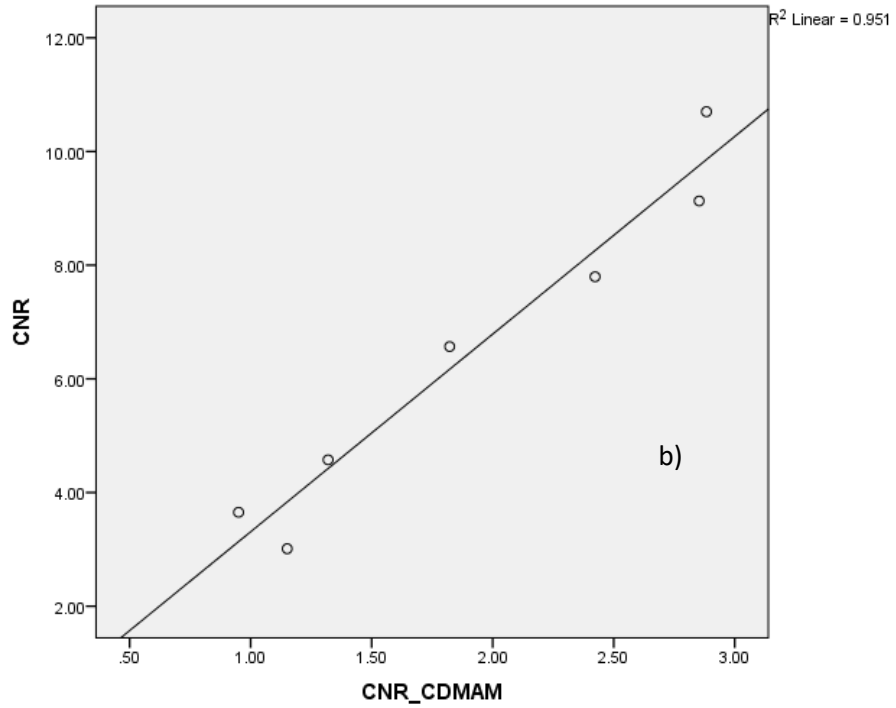
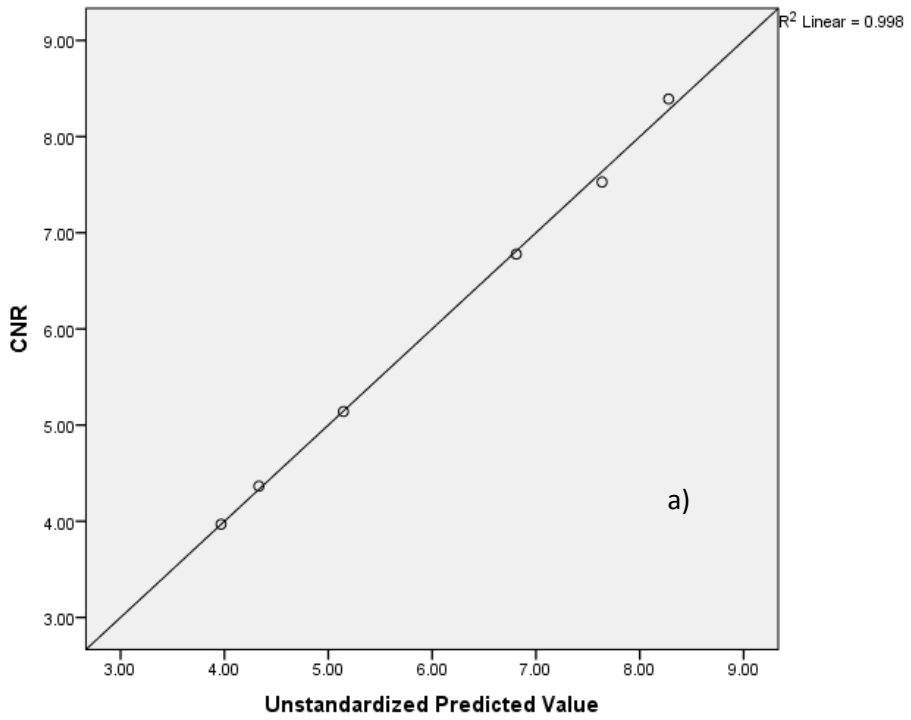
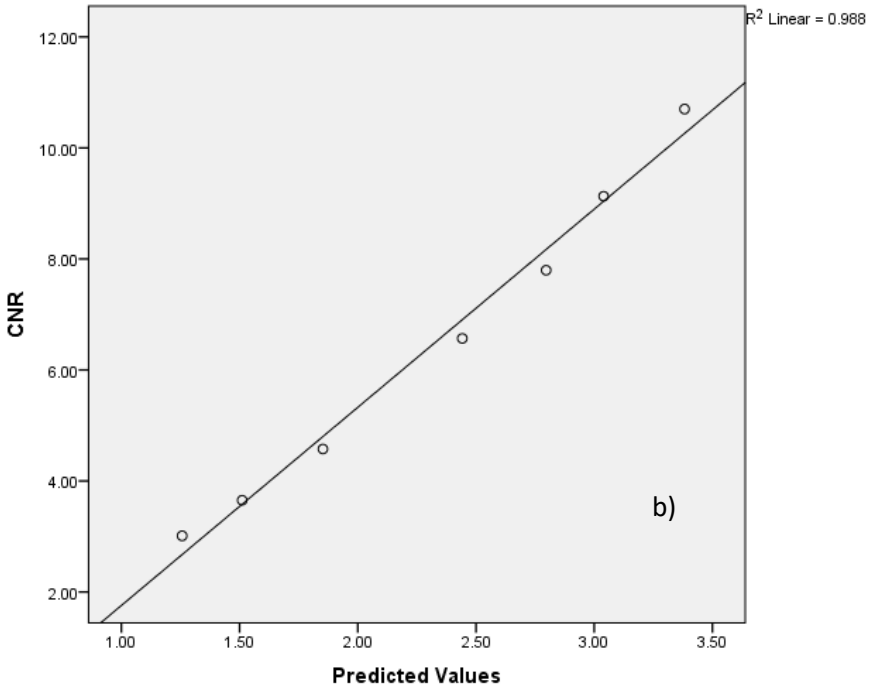


Figure 6.25. Q-Q Plot of the CNR values against the predicted values using the models in tables 6.9 and 6.10. a) Essential b) Hologic





Conclusion

The limitations of CNR, especially when one needs to compare anyone digital mammography system to another have been pointed out. If threshold contrast detail detectability using the CDMAM phantom is considered as the image parameter of choice, one also needs to consider the various interpretations and variations in which results can be presented. Parameters such as IQF_{INV} and H_T have been discussed and their various relationships with other parameters have been explored. Baring in mind that CDMAM scoring can be both manual and automated, it is believed that using such parameters based on threshold contrast detail detectability can make inter-system comparison in terms of image quality possible since both object detail and contrast are taken into account. With this principle in mind, the Q factors have been introduced which go a step further and make the measured image quality parameter relative to a theoretical or reachable value. The final step taken was to establish the relationship of the various image quality indicators, including the Q factors with MGD. This last step was essential in defining the various FOM included in this study including Qnorm1 and Qnorm2.

It is felt that this exploration has showed three essential components.

1. Other FOMs besides the reputable and well-known FOM 1 are valid factors that may be used within the scope of optimization of digital mammography;

2. At this stage, a solitary FOM alone may not be used confidently to compare different digital mammography systems;
3. The CDMAM may be used to calculate SDNR or CNR or predict SDNR levels within the objectives of this study. This can be done instead of the classical method described by the European protocol. Further investigation in this regard at different PMMA thickness and using different spectra would be recommended.

The author feels that the Q factors, the Qnorms and FOMs X and X1 were very effective in system characterisation in the units included in this study. All the parameters, including any of the Q factors comprised a theoretical reachable or target value which is otherwise not present in any other FOM. In FOM X the most desirable attributes were placed as numerator normalised to MGD. Amongst these desirable attributes is DQE, which also characterises the detector and is absent in any other approach. FOM X1 focuses on the use of the CDMAM phantom but follows the same principle of FOM X. The FOMs X and X1 and the Q factors were very predictable in both systems investigated. However, their magnitude was different and one cannot be sure that any of the results are or can be compared between systems unless more research is done in this regard. Nevertheless, having a baseline of such parameters and comparing the performance of these systems or similar systems based on these parameters will be useful in regular quality assurance. Each system could have a predetermined value of such parameters at known or using known conditions. Any new similar system can be compared to these established values.

Table 6.8. Correlation Results: Essential

		DAK	MGD	K	FOM1	FOM2	FOM3	FOM4	FOM5	FOM6	FOM7	FOM8	FOM9	FOM10	QFACTOR1_div_MGD	FOM_x	FOM_x1	Qfactor1	Qfactor2	Qnorm1	Qnorm2
DAK	Pearson Correlation	1	1.000**	1.000**	-.938**	-.949**	-.884**	-.885**	-.884**	-.885**	-.921**	-.873*	-.872*	0.138	-.860*	-.879**	-.885**	.948**	.972**	-.851*	-.948**
	Sig. (2-tailed)		0	0	0.002	0.001	0.008	0.008	0.008	0.008	0.003	0.01	0.01	0.767	0.013	0.009	0.008	0.001	0	0.015	0.001
MGD	Pearson Correlation	1.000**	1	1.000**	-.938**	-.949**	-.885**	-.885**	-.885**	-.885**	-.921**	-.874*	-.872*	0.139	-.860*	-.880**	-.885**	.948**	.972**	-.851*	-.948**
	Sig. (2-tailed)	0		0	0.002	0.001	0.008	0.008	0.008	0.008	0.003	0.01	0.01	0.766	0.013	0.009	0.008	0.001	0	0.015	0.001
K	Pearson Correlation	1.000**	1.000**	1	-.938**	-.948**	-.884**	-.885**	-.884**	-.885**	-.921**	-.873*	-.872*	0.139	-.860*	-.879**	-.885**	.948**	.972**	-.851*	-.948**
	Sig. (2-tailed)	0	0		0.002	0.001	0.008	0.008	0.008	0.008	0.003	0.01	0.011	0.767	0.013	0.009	0.008	0.001	0	0.015	0.001
FOM1	Pearson Correlation	-.938**	-.938**	-.938**	1	.991**	.983**	.981**	.983**	.981**	.993**	.977**	.976**	-0.232	.969**	.981**	.981**	-.984**	-.987**	.866*	.991**
	Sig. (2-tailed)	0.002	0.002	0.002		0	0	0	0	0	0	0	0	0.616	0	0	0	0	0	0.012	0
FOM2	Pearson Correlation	-.949**	-.949**	-.948**	.991**	1	.981**	.982**	.981**	.982**	.994**	.977**	.977**	-0.174	.973**	.979**	.982**	-.982**	-.983**	.893**	1.000**
	Sig. (2-tailed)	0.001	0.001	0.001	0		0	0	0	0	0	0	0	0.709	0	0	0	0	0	0.007	0
FOM3	Pearson Correlation	-.884**	-.885**	-.884**	.983**	.981**	1	1.000**	1.000**	1.000**	.996**	.999**	.999**	-0.192	.997**	1.000**	1.000**	-.971**	-.963**	.882**	.981**
	Sig. (2-tailed)	0.008	0.008	0.008	0	0		0	0	0	0	0	0	0.68	0	0	0	0	0	0.009	0
FOM4	Pearson Correlation	-.885**	-.885**	-.885**	.981**	.982**	1.000**	1	1.000**	1.000**	.996**	1.000**	1.000**	-0.181	.998**	1.000**	1.000**	-.970**	-.962**	.887**	.982**
	Sig. (2-tailed)	0.008	0.008	0.008	0	0	0		0	0	0	0	0	0.697	0	0	0	0	0.001	0.008	0

FOM5	Pearson Correlation	-.884**	-.885**	-.884**	.983**	.981**	1.000**	1.000**	1	1.000**	.996**	.999**	.999**	-0.192	.997**	1.000**	1.000**	-.971**	-.963**	.882**	.981**
	Sig. (2-tailed)	0.008	0.008	0.008	0	0	0	0		0	0	0	0	0.68	0	0	0	0	0	0.009	0
FOM6	Pearson Correlation	-.885**	-.885**	-.885**	.981**	.982**	1.000**	1.000**	1.000**	1	.996**	1.000**	1.000**	-0.181	.998**	1.000**	1.000**	-.970**	-.962**	.887**	.982**
	Sig. (2-tailed)	0.008	0.008	0.008	0	0	0	0	0		0	0	0	0.697	0	0	0	0	0.001	0.008	0
FOM7	Pearson Correlation	-.921**	-.921**	-.921**	.993**	.994**	.996**	.996**	.996**	.996**	1	.993**	.993**	-0.174	.990**	.995**	.996**	-.980**	-.980**	.897**	.994**
	Sig. (2-tailed)	0.003	0.003	0.003	0	0	0	0	0	0		0	0	0.708	0	0	0	0	0	0.006	0
FOM8	Pearson Correlation	-.873*	-.874*	-.873*	.977**	.977**	.999**	1.000**	.999**	1.000**	.993**	1	1.000**	-0.177	.999**	1.000**	1.000**	-.965**	-.955**	.885**	.978**
	Sig. (2-tailed)	0.01	0.01	0.01	0	0	0	0	0	0	0		0	0.704	0	0	0	0	0.001	0.008	0
FOM9	Pearson Correlation	-.872*	-.872*	-.872*	.976**	.977**	.999**	1.000**	.999**	1.000**	.993**	1.000**	1	-0.176	.999**	1.000**	1.000**	-.964**	-.954**	.885**	.977**
	Sig. (2-tailed)	0.01	0.01	0.011	0	0	0	0	0	0	0	0		0.706	0	0	0	0	0.001	0.008	0
FOM10	Pearson Correlation	0.138	0.139	0.139	-0.232	-0.174	-0.192	-0.181	-0.192	-0.181	-0.174	-0.177	-0.176	1	-0.152	-0.19	-0.181	0.297	0.186	0.277	-0.175
	Sig. (2-tailed)	0.767	0.766	0.767	0.616	0.709	0.68	0.697	0.68	0.697	0.708	0.704	0.706		0.745	0.684	0.697	0.518	0.69	0.547	0.708
QFACTOR1_div_MGD	Pearson Correlation	-.860*	-.860*	-.860*	.969**	.973**	.997**	.998**	.997**	.998**	.990**	.999**	.999**	-0.152	1	.998**	.998**	-.954**	-.945**	.891**	.973**
	Sig. (2-tailed)	0.013	0.013	0.013	0	0	0	0	0	0	0	0	0	0.745		0	0	0.001	0.001	0.007	0
FOM_x	Pearson Correlation	-.879**	-.880**	-.879**	.981**	.979**	1.000**	1.000**	1.000**	1.000**	.995**	1.000**	1.000**	-0.19	.998**	1	1.000**	-.969**	-.960**	.882**	.979**
	Sig. (2-tailed)	0.009	0.009	0.009	0	0	0	0	0	0	0	0	0	0.684	0		0	0	0.001	0.009	0
FOM_x1	Pearson Correlation	-.885**	-.885**	-.885**	.981**	.982**	1.000**	1.000**	1.000**	1.000**	.996**	1.000**	1.000**	-0.181	.998**	1.000**	1	-.970**	-.962**	.887**	.982**

	Sig. (2-tailed)	0.008	0.008	0.008	0	0	0	0	0	0	0	0	0	0.697	0	0		0	0.001	0.008	0
Qfactor1	Pearson Correlation	.948**	.948**	.948**	-.984**	-.982**	-.971**	-.970**	-.971**	-.970**	-.980**	-.965**	-.964**	0.297	-.954**	-.969**	-.970**	1	.987**	-.829*	-.982**
	Sig. (2-tailed)	0.001	0.001	0.001	0	0	0	0	0	0	0	0	0	0.518	0.001	0	0		0	0.021	0
Qfactor2	Pearson Correlation	.972**	.972**	.972**	-.987**	-.983**	-.963**	-.962**	-.963**	-.962**	-.980**	-.955**	-.954**	0.186	-.945**	-.960**	-.962**	.987**	1	-.881**	-.983**
	Sig. (2-tailed)	0	0	0	0	0	0	0.001	0	0.001	0	0.001	0.001	0.69	0.001	0.001	0.001	0		0.009	0
Qnorm1	Pearson Correlation	-.851*	-.851*	-.851*	.866*	.893**	.882**	.887**	.882**	.887**	.897**	.885**	.885**	0.277	.891**	.882**	.887**	-.829*	-.881**	1	.892**
	Sig. (2-tailed)	0.015	0.015	0.015	0.012	0.007	0.009	0.008	0.009	0.008	0.006	0.008	0.008	0.547	0.007	0.009	0.008	0.021	0.009		0.007
Qnorm2	Pearson Correlation	-.948**	-.948**	-.948**	.991**	1.000**	.981**	.982**	.981**	.982**	.994**	.978**	.977**	-0.175	.973**	.979**	.982**	-.982**	-.983**	.892**	1
	Sig. (2-tailed)	0.001	0.001	0.001	0	0	0	0	0	0	0	0	0	0.708	0	0	0	0	0	0.007	

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Good Correlation Poor/No Correlation Correlation <0.89

Table 6.9. Correlation Results: Hologic

		DAK	MGD	K	FOM1	FOM2	FOM3	FOM4	FOM5	FOM6	FOM7	FOM8	FOM9	FOM10	QFACTOR1_div_MGD	FOM_x	FOM_x1	Qfactor1	Qfactor2	Qnorm1	Qnorm2
DAK	Pearson Correlation	1	1.000**	1.000**	-0.272	-.922**	-.911**	-.904**	-.910**	-.904**	-0.366	-.818*	-.817*	0.036	-.875**	-.900**	-.904**	.928**	.955**	-0.704	-.921**
	Sig. (2-tailed)		0	0	0.555	0.003	0.004	0.005	0.004	0.005	0.42	0.024	0.025	0.94	0.01	0.006	0.005	0.003	0.001	0.078	0.003
MGD	Pearson Correlation	1.000**	1	1.000**	-0.274	-.921**	-.910**	-.903**	-.909**	-.903**	-0.365	-.817*	-.816*	0.034	-.874*	-.899**	-.903**	.927**	.954**	-0.705	-.921**
	Sig. (2-tailed)	0		0	0.552	0.003	0.004	0.005	0.005	0.005	0.421	0.025	0.025	0.942	0.01	0.006	0.005	0.003	0.001	0.077	0.003
K	Pearson Correlation	1.000**	1.000**	1	-0.274	-.921**	-.910**	-.904**	-.910**	-.904**	-0.365	-.817*	-.817*	0.035	-.874*	-.900**	-.903**	.927**	.954**	-0.704	-.921**
	Sig. (2-tailed)	0	0		0.553	0.003	0.004	0.005	0.005	0.005	0.421	0.025	0.025	0.941	0.01	0.006	0.005	0.003	0.001	0.077	0.003
FOM1	Pearson Correlation	-0.272	-0.274	-0.274	1	0.307	-0.089	-0.089	-0.09	-0.089	-0.341	-0.281	-0.279	.880**	-0.16	-0.117	-0.092	0.09	-0.018	0.717	0.309
	Sig. (2-tailed)	0.555	0.552	0.553		0.503	0.85	0.849	0.848	0.849	0.454	0.542	0.544	0.009	0.731	0.803	0.844	0.847	0.969	0.069	0.501
FOM2	Pearson Correlation	-.922**	-.921**	-.921**	0.307	1	.861*	.880**	.860*	.880**	0.079	0.678	0.675	0.16	.824*	.847*	.878**	-.869*	-.877**	.845*	1.000**
	Sig. (2-tailed)	0.003	0.003	0.003	0.503		0.013	0.009	0.013	0.009	0.866	0.094	0.096	0.731	0.023	0.016	0.009	0.011	0.009	0.017	0
FOM3	Pearson Correlation	-.911**	-.910**	-.910**	-0.089	.861*	1	.998**	1.000**	.998**	0.398	.941**	.941**	-0.318	.996**	1.000**	.998**	-.989**	-.989**	0.516	.861*
	Sig. (2-tailed)	0.004	0.004	0.004	0.85	0.013		0	0	0	0.377	0.002	0.002	0.488	0	0	0	0	0	0.236	0.013
FOM4	Pearson Correlation	-.904**	-.903**	-.904**	-0.089	.880**	.998**	1	.997**	1.000**	0.358	.925**	.924**	-0.292	.994**	.997**	1.000**	-.988**	-.982**	0.539	.879**
	Sig. (2-tailed)	0.005	0.005	0.005	0.849	0.009	0		0	0	0.431	0.003	0.003	0.526	0	0	0	0	0	0.212	0.009
FOM5	Pearson Correlation	-.910**	-.909**	-.910**	-0.09	.860*	1.000**	.997**	1	.997**	0.397	.942**	.941**	-0.319	.997**	1.000**	.998**	-.989**	-.989**	0.515	.859*
	Sig. (2-tailed)	0.004	0.005	0.005	0.848	0.013	0	0		0	0.378	0.002	0.002	0.486	0	0	0	0	0	0.237	0.013
FOM6	Pearson Correlation	-.904**	-.903**	-.904**	-0.089	.880**	.998**	1.000**	.997**	1	0.358	.925**	.924**	-0.291	.994**	.997**	1.000**	-.988*	-.982*	0.539	.880**

																		*	*		
	Sig. (2-tailed)	0.005	0.005	0.005	0.849	0.009	0	0	0		0.431	0.003	0.003	0.526	0	0	0	0	0	0.212	0.009
FOM7	Pearson Correlation	-0.366	-0.365	-0.365	-0.341	0.079	0.398	0.358	0.397	0.358	1	0.664	0.664	-0.699	0.4	0.411	0.362	-0.459	-0.407	-0.32	0.079
	Sig. (2-tailed)	0.42	0.421	0.421	0.454	0.866	0.377	0.431	0.378	0.431		0.104	0.104	0.081	0.374	0.359	0.424	0.3	0.364	0.484	0.867
FOM8	Pearson Correlation	-.818*	-.817*	-.817*	-0.281	0.678	.941**	.925**	.942**	.925**	0.664	1	1.000**	-0.574	.948**	.948**	.927**	-.945**	-.927**	0.235	0.677
	Sig. (2-tailed)	0.024	0.025	0.025	0.542	0.094	0.002	0.003	0.002	0.003	0.104		0	0.178	0.001	0.001	0.003	0.001	0.003	0.611	0.095
FOM9	Pearson Correlation	-.817*	-.816*	-.817*	-0.279	0.675	.941**	.924**	.941**	.924**	0.664	1.000**	1	-0.574	.948**	.948**	.926**	-.944**	-.927**	0.234	0.674
	Sig. (2-tailed)	0.025	0.025	0.025	0.544	0.096	0.002	0.003	0.002	0.003	0.104	0		0.178	0.001	0.001	0.003	0.001	0.003	0.613	0.097
FOM10	Pearson Correlation	0.036	0.034	0.035	.880**	0.16	-0.318	-0.292	-0.319	-0.291	-0.699	-0.574	-0.574	1	-0.374	-0.344	-0.296	0.331	0.257	0.647	0.162
	Sig. (2-tailed)	0.94	0.942	0.941	0.009	0.731	0.488	0.526	0.486	0.526	0.081	0.178	0.178		0.408	0.45	0.52	0.468	0.579	0.116	0.729
QFACTOR1_div_MGD	Pearson Correlation	-.875**	-.874*	-.874*	-0.16	.824*	.996**	.994**	.997**	.994**	0.4	.948**	.948**	-0.374	1	.998**	.994**	-.980**	-.976**	0.461	.823*
	Sig. (2-tailed)	0.01	0.01	0.01	0.731	0.023	0	0	0	0	0.374	0.001	0.001	0.408		0	0	0	0	0.298	0.023
FOM_x	Pearson Correlation	-.900**	-.899**	-.900**	-0.117	.847*	1.000**	.997**	1.000**	.997**	0.411	.948**	.948**	-0.344	.998**	1	.997**	-.988**	-.986**	0.491	.847*
	Sig. (2-tailed)	0.006	0.006	0.006	0.803	0.016	0	0	0	0	0.359	0.001	0.001	0.45	0		0	0	0	0.263	0.016
FOM_x1	Pearson Correlation	-.904**	-.903**	-.903**	-0.092	.878**	.998**	1.000**	.998**	1.000**	0.362	.927**	.926**	-0.296	.994**	.997**	1	-.988**	-.982**	0.535	.878**
	Sig. (2-tailed)	0.005	0.005	0.005	0.844	0.009	0	0	0	0	0.424	0.003	0.003	0.52	0	0		0	0	0.216	0.009
Qfactor1	Pearson Correlation	.928**	.927**	.927**	0.09	-.869*	-.989**	-.988**	-.989**	-.988**	-0.459	-.945**	-.944**	0.331	-.980**	-.988**	-.988**	1	.983**	-0.5	-.868*
	Sig. (2-tailed)	0.003	0.003	0.003	0.847	0.011	0	0	0	0	0.3	0.001	0.001	0.468	0	0	0		0	0.254	0.011
Qfactor2	Pearson Correlation	.955**	.954**	.954**	-0.018	-.877**	-.989**	-.982**	-.989**	-.982**	-0.407	-.927**	-.927**	0.257	-.976**	-.986**	-.982**	.983**	1	-0.564	-.876**
	Sig. (2-tailed)	0.001	0.001	0.001	0.969	0.009	0	0	0	0	0.364	0.003	0.003	0.579	0	0	0	0		0.187	0.01

Qnorm1	Pearson Correlation	-0.704	-0.705	-0.704	0.717	.845*	0.516	0.539	0.515	0.539	-0.32	0.235	0.234	0.647	0.461	0.491	0.535	-0.5	-0.564	1	.845*
	Sig. (2-tailed)	0.078	0.077	0.077	0.069	0.017	0.236	0.212	0.237	0.212	0.484	0.611	0.613	0.116	0.298	0.263	0.216	0.254	0.187		0.017
Qnorm2	Pearson Correlation	-.921**	-.921**	-.921**	0.309	1.000**	.861*	.879**	.859*	.880**	0.079	0.677	0.674	0.162	.823*	.847*	.878**	-.868*	-.876**	.845*	1
	Sig. (2-tailed)	0.003	0.003	0.003	0.501	0	0.013	0.009	0.013	0.009	0.867	0.095	0.097	0.729	0.023	0.016	0.009	0.011	0.01	0.017	

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Good Correlation Poor/No Correlation Correlation <0.89

Table 6.9. Correlation between CNR, CNR using CDMAM, IQF_{INV} and H_TSUM a) Essential b) Hologic

Essential		**CNR**	**CNR_CDMAM**	**IQF_INV**	**HT_SUM**	**SNR**
eNEQ.MGD	Pearson Correlation	.033	.026	.073	.052	.009
	Sig. (2-tailed)	.944	.956	.876	.912	.985
eNEQ.DC	Pearson Correlation	-.160	-.179	-.124	-.139	-.185
	Sig. (2-tailed)	.733	.701	.790	.766	.691
eNEQ.DC_CDMAM	Pearson Correlation	-.733	-.707	-.706	-.705	-.742
	Sig. (2-tailed)	.061	.075	.076	.077	.056
NEQ.MGD	Pearson Correlation	.327	.373	.338	.297	.340
	Sig. (2-tailed)	.475	.410	.459	.518	.455
NEQ.DC	Pearson Correlation	.214	.256	.232	.184	.224
	Sig. (2-tailed)	.644	.579	.617	.693	.629
NEQ.DC_CDMAM	Pearson Correlation	-.643	-.595	-.622	-.633	-.640
	Sig. (2-tailed)	.119	.159	.136	.127	.121

Hologic		**CNR**	**CNR_CDMAM**	**IQF_INV**	**HT_SUM**	**SNR**
eNEQ.MGD	Pearson Correlation	-.811*	-.701	-.773*	-.802*	-.803*
	Sig. (2-tailed)	.027	.079	.041	.030	.030
eNEQ.DC	Pearson Correlation	.151	.258	.244	.145	.222
	Sig. (2-tailed)	.746	.576	.597	.756	.633
eNEQ.DC_CDMAM	Pearson Correlation	-.172	.041	-.147	-.161	-.187
	Sig. (2-tailed)	.712	.931	.752	.730	.688
NEQ.MGD	Pearson Correlation	.053	.152	.127	.036	.102
	Sig. (2-tailed)	.910	.746	.787	.939	.828
NEQ.DC	Pearson Correlation	.242	.329	.316	.224	.295
	Sig. (2-tailed)	.600	.471	.491	.629	.521
NEQ.DC_CDMAM	Pearson Correlation	.179	.304	.245	.165	.216
	Sig. (2-tailed)	.700	.508	.597	.723	.641

CHAPTER 7

CONLUSIONS

This thesis highlights the fact that there is still a lot of work that needs exploration and analysis with respect to the FOM in digital mammography. Nevertheless, it provided a landmark for mammography practice in Malta and served as a foundation for any future work in digital mammography within the scope of this study.

In this first Maltese national mammography survey, the effectiveness of DR mammography in breast cancer screening has been confirmed. Patient data were made available from three clinics out of the participating nine. A dose survey of MGD calculated for 759 patients examined in the state-owned mammography facilities was performed. An MGD national DRL was set at 1.87 mGy for patients with a BCT between 5.0 and 7.0 cm. This range was selected since patient data were retrieved from three clinics only and the results showed that other international BCT reference levels may be unsuitable for the Maltese population. In fact, the overall average BCT was 5.75+1.4 cm. The survey results have shown that the technical standard of mammographic equipment in the Malta National Breast Screening Programme is on a par with other countries, including its Western European counterparts. It also highlighted the importance of regular quality assurance and optimization.

Modern FFDM units display the MGD and K to the breast following each exposure. However, information on how these values are calculated is absent and knowing how displayed MGD values compare and correlate to conventional Monte Carlo-based methods is useful especially if these values are used for DRL purposes. From measurements on PMMA phantoms, it has been shown that displayed and calculated MGD values are similar for thin to medium thicknesses and appear to differ with larger PMMA thicknesses. Consequently, a multiple linear regression analysis on the data was performed to generate models by which displayed MGD values on the two FFDM units included in chapter 4 may be converted to the conventional Monte Carlo values. These models should be a useful tool for medical physicists requiring MGD data from these FFDM units and should reduce the survey time spent on dose calculations.

The purpose of chapter 5 was to compare a number of measured image quality parameters using processed and raw images in two DR units and one CR mammography system. This is very important for the calculation of image quality parameters and the FOMs in digital mammography. This work showed that the difference between raw and processed image data is system-specific. The results have shown that there are no significant differences between raw and processed data in the mean threshold contrast values using the CDMAM phantom in all the systems investigated; however, these results cannot be generalised to all available systems. Notable differences were noted in CNR and in other tests including: response function, MTF, NEQ, NNPS and DQE as specified in IEC 62220-1-2. It is therefore emphasized that quantitative image quality analyses in digital mammography should be done using raw image data only, especially when performing detector characterization tests. The use of processed images may results in completely wrong or unpredictable values and may be inconsistent due to possible future software upgrades in any particular system.

The use of image quality parameters in digital mammography such as CNR, SNR and DQE have been widespread, with the intention of detector evaluation and/or quantitative evaluation of the system performance. These parameters are useful in ensuring adequate system performance when tests are done against international standards or guidelines. Nevertheless, parameters like CNR are relative quantities that lie within a range that is manufacturer and system-dependent.

Since, the CDMAM phantom can be scored automatically; threshold contrast detail detectability can make inter-system comparison possible since both object detail and contrast are taken into account objectively. Further still, the Q factors make the measured image quality parameter relative to a theoretical or reachable value. In addition, the relationship of the various image quality indicators, including the Q factors with MGD was established. This last step was essential in defining the various FOM included in this study including Qnorm1 and Qnorm2. Therefore, the use of a FOM permits quantitative assessment in terms of image quality and patient dose.

With reference to the conclusions drawn in chapter 6, it is felt that other FOMs besides the reputable and well-known FOM 1 are valid. However, the use a solitary FOM cannot be used to compare different digital mammography systems. Nevertheless, it is felt that the Q factors, the Qnorms and FOMs X and X1 are very promising and are effective in system characterisation and are worth further future exploration. Further still, the FOMs X and X1 and

the Q factors were very predictable from a given MGD in both systems investigated. Although their use for inter-system comparison is still uncertain they are still very valid for constancy and possibly for system optimization, especially if their use is explored with different PMMA thicknesses and different spectra.

Future work related to this project discussed in chapter 8 will focus on further evaluation of the defined FOMs, potential use of the FOM including the Q factors and Qnorm1 and 2 in optimization and the application of the discriminability index (d') in the evaluation and scoring of the CDMAM phantom.

CHAPTER 8

FUTURE WORK

Signal Detection Analysis

The starting point for the signal detection theory (SDT) is that nearly all reasoning and decision making takes place in the presence of some uncertainty. Signal detection theory provides a precise language and graphic notation for analyzing decision making in the presence of uncertainty. The general approach of SDT has direct application for us, within the scope of this study, in terms of sensory experiments but it also offers a way to analyze many different kinds of decision problems. One of the concepts behind the SDT is the discriminability index (d'). The discriminability of a signal depends both on the separation and the spread of the noise alone and the signal plus the noise curves. The most widely used measure in this regard is defined as follows:

$$d' = \frac{separation}{spread} \qquad 54$$

The number d' is an estimate of the strength of the signal. Increasing the stimulus strength separates the two probabilities defined above.

It is envisaged that the d' concept can be adapted to the CDMAM scoring methodology. The reasoning can be summarized in the table 8.1 below.

Table 8.1. The signal detection theory principle.

(Gold Disc) / **Signal**	**Response (Disc present ?)**	
	Yes	**No**
Present	Hit	Miss
Absent	False Alarm	Correct Negative

The CDMAM phantom version 3.4 manual provides users with a 'map' of the locations of the each gold detail. Consequently, with reference to table 8.1 this map determines whether a signal (gold disc in its respective location) is present or not. The score of the image for a given exposure can be a **Hit** (correct identification of the disc location), **Correct Negative** (correct identification of absence of disc), **Miss** (unidentified disc) or a **False Alarm** (incorrect location identification).

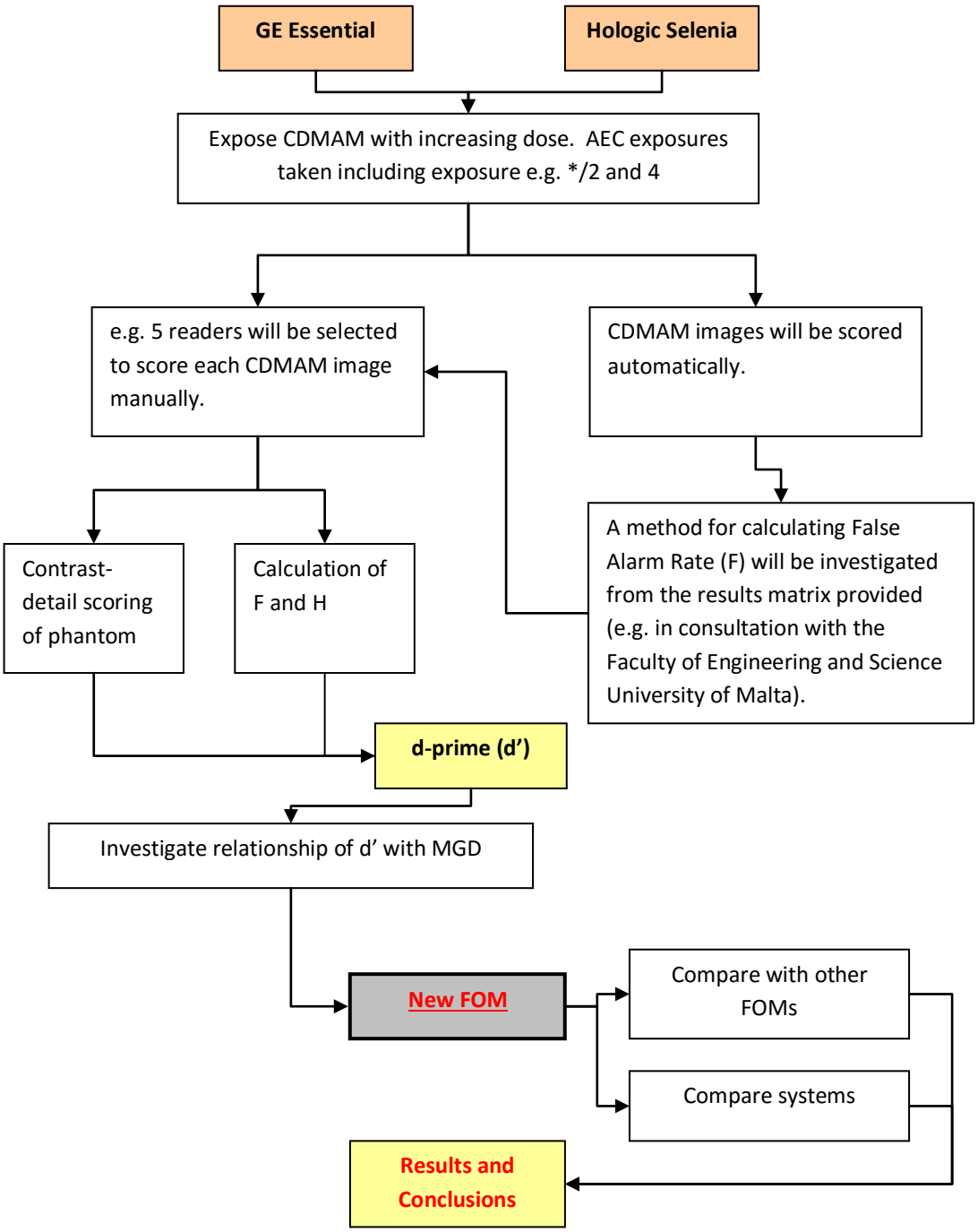
This means that using the above reasoning, the **hit rate (H)** and the **false alarm rate (F)** can be determined from any given CDMAM image. The H in this case would be the proportion of correctly identified discs in relation to the total number. The F would be the identification of discs in which no disc is present in the CDMAM map provided. With this in mind, a perfect subject's performance (human or automated) would be 100% H and 0% F. Intuitively and practically, the best subject aims to maximize H and minimize F and thus the larger the difference between H and F, the better the subject's sensitivity. The d' is a measure of the difference between H and F. Consequently, hit rate H could be considered being the signal and the false alarm rate could be seen as the signal including the noise component. However, d' is actually the difference between the z-transforms of these two rates and expressed in equation **55.**

$$d' = z(H) - z(F) \qquad 55$$

The objective of any future work will be to select a group of observers and score a selection of CDMAM images with increasing dose for any system that may be included in this study. The d' for each exposure/image will be measured and calculated and the relationship of this index with increasing dose will be verified. It is hoped that this exercise could form a sound foundation for a similar but automated process. The major limitation in using available CDMAM automated software in calculating the d' is that whilst they conveniently score the images in terms of hits and misses, they fail to specify false alarms. For instance, by using the CDMAM Analyzer Software H can be calculated however the F value cannot be established.

With further progression in this work the possibility or need of introducing other quality parameters or modify the existing equations in this document can be investigated and explored. Flowchart 8.1 addresses the proposed structure for any future work involving d'.

Flowchart 8.1. Flow chart of future work related d' analyses



FUTURE OBJECTIVES WITH THE Q-FACTORS

In this document it has been seen that all Q factors (Qfactor1 and Qfactor2*100) in this study showed a very strong logarithmic relationship with the MGD for the Essential and the Hologic at the standard breast. It will be interesting to note how the same Q factors relate to dose using other detector technologies and possibly with varying PMMA thickness. Until now, the Q factor and the FOMs that include the Q factors, are the preferred image quality indicator/s (within the scope of this study) since they are relative terms i.e. they compare the measured image quality to a reference value or a theoretical value. Further still, for the digital mammography systems included in this study, the Q factors can be predicted by knowing the MGD for a given exposure e.g. an AEC exposure to the standard breast. Until now, this can be only be stated for the Essential the Hologic and cannot be generalized for the other mammography systems. We can actually go a step further when knowing or when we are able to predict the Q factors (refer to flowchart 8.2). Since the latter was also found to correlate very well with IQF_{INV}, CNR and H_T, these image quality parameters can also be predicted from the MGD value employed. This means that systems whose IQF_{INV}, CNR and H_T measured/calculated values or whose Q factors themselves fall short or do not match the expected values, (within predetermined established limits) might need optimization or need some kind of adjustment/correction.

A nonlinear regression analysis was performed to create a model that could predict the Qfactor1 and Qfactor2*100 for the Essential and the Hologic for the MGD of the various exposures used in this project. As shown from the Q-Q plot in figures 8.1 and 8.2 it is clear that the predicted values are very close to the calculated values. This makes the assumptions and objectives related to the Q factor very promising and reachable.

Conclusion

The work in this project in not only useful for the Maltese population but also for the community of medical physicists who specialise in digital mammography. The ideas and results of this exploration should give another perspective of assessing or viewing the performance of digital mammography systems. Inter-system comparison based on the FOMs may not be possible at the moment but the FOMs explored, especially the ones including the Q

factors are valid contributors in the quality assurance and constancy of digital mammography systems.

Flowchart 8.2. Optimization through Q factor prediction.

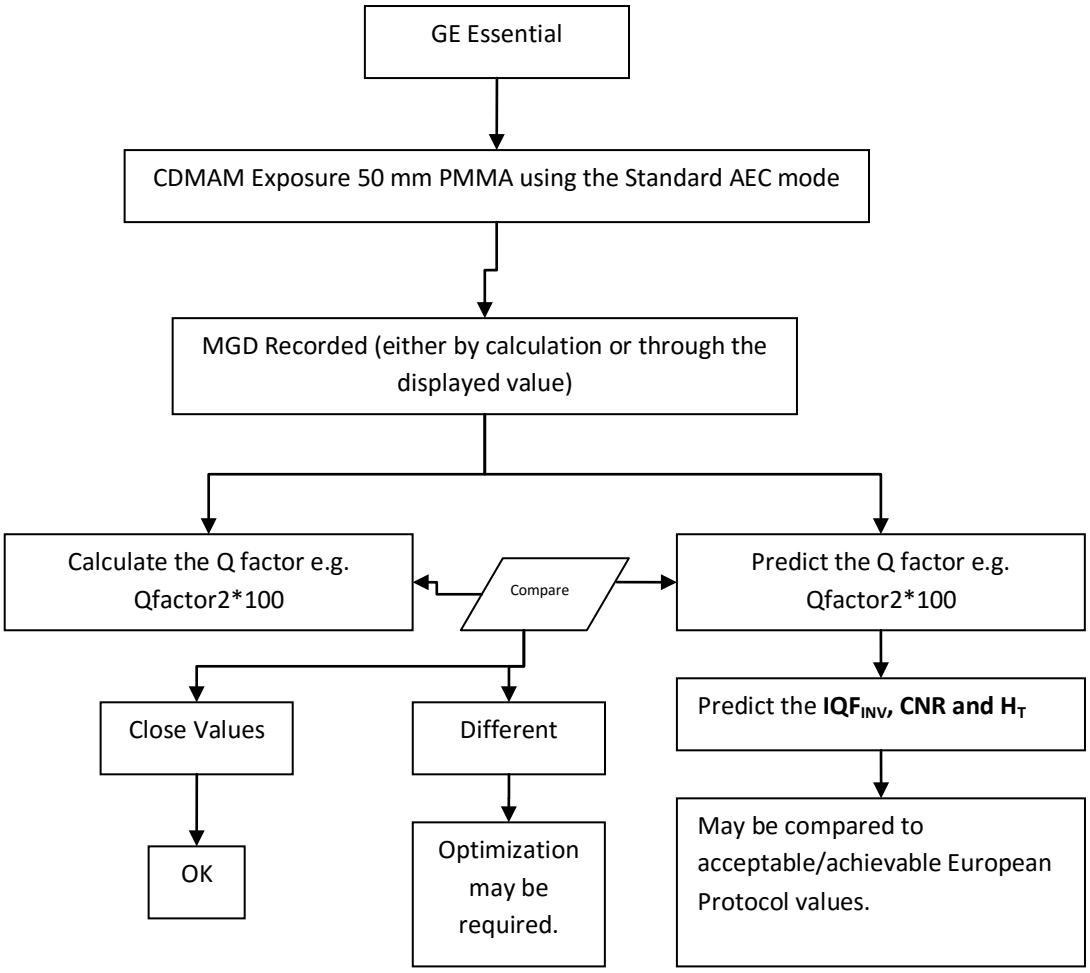
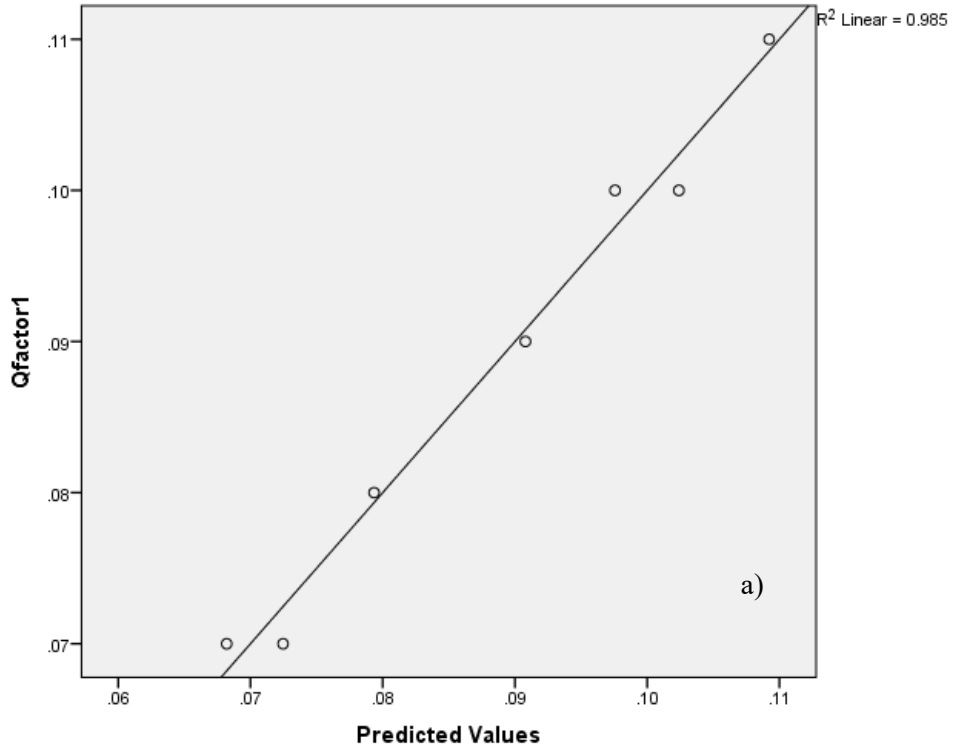


Figure 8.1. Q-Q plot of calculated versus predicted a) Qfactor1 and b) Qfactor2*100 values - Essential



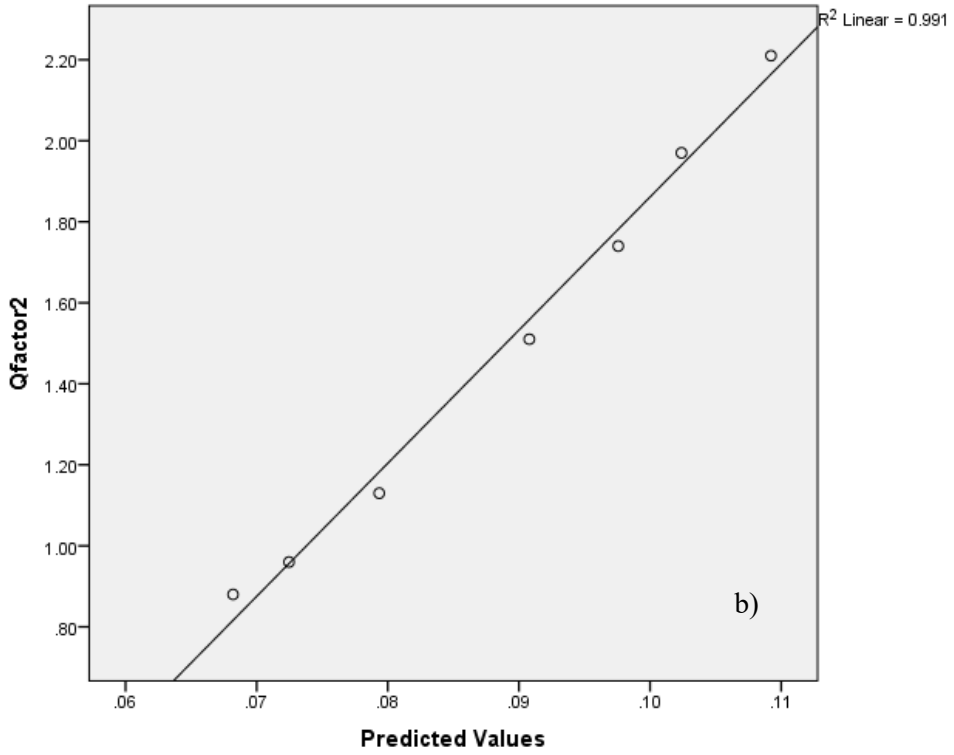
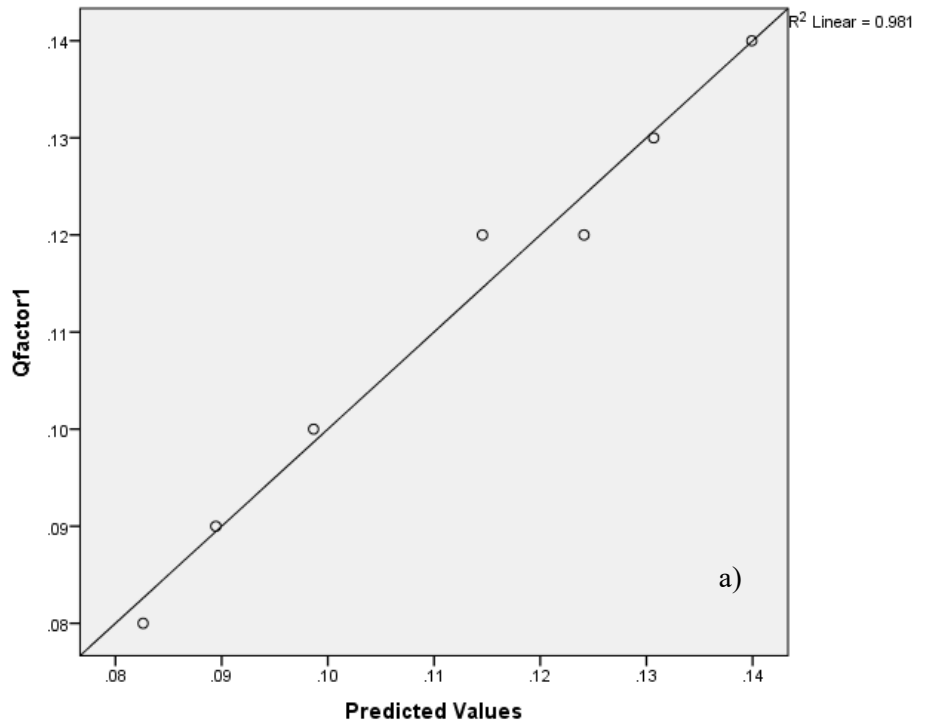
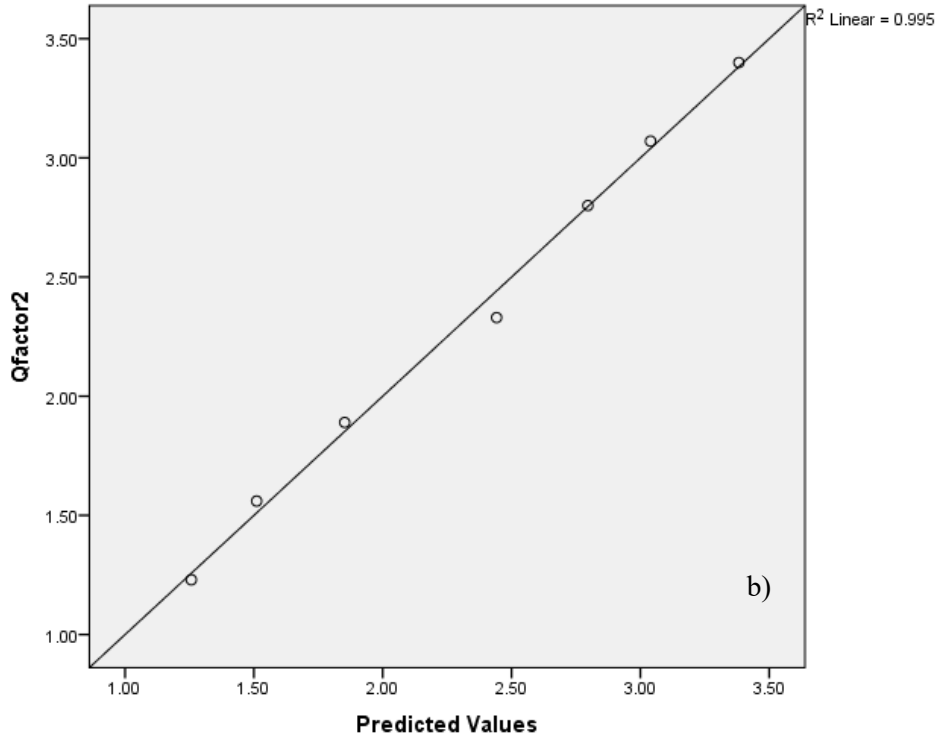


Figure 8.2. Q-Q plot of calculated versus predicted a) Qfactor1 and b) Qfactor2*100 values - Hologic





REFERENCES

1. van Engen, R., et al., *European protocol for the quality control of the physical and technical aspects of mammography screening 2a Screen-Film Mammography in: European guidelines for quality assurance in breast cancer screening and diagnosis, 4th Edition.* 2006.
2. van Engen, R., et al., *European protocol for the quality control of the physical and technical aspects of mammography screening 2b Digital Mammography in: European guidelines for quality assurance in breast cancer screening and diagnosis, 4th Edition.* 2006.
3. Perry, N., et al., *European guidelines for quality assurance in breast cancer screening and diagnosis Fourth Edition Supplements*. 2013, Luxembourg: Office for Official Publications of the European Union, 2013.
4. Borg, M., I. Badr, and G.J. Royle, *Mammography equipment performance, image quality and mean glandular dose in Malta.* Radiat Prot Dosimetry, 2013. **156**(2): p. 168-83.
5. Borg, M., I. Badr, and G. Royle, *Should Processed or Raw Image Data Be Used in Mammographic Image Quality Analyses? A Comparative Study of Three Full-Field Digital Mammography Systems.* Radiat Prot Dosimetry, 2014.
6. Borg, M., I. Badr, and G.J. Royle, *A Study to Determine the Differences between the Displayed Dose Values for Two Full-Field Digital Mammography Units and Values Calculated Using a Range of Monte-Carlo-Based Techniques: A Phantom Study.* Radiat Prot Dosimetry, 2012.
7. Borg, M., I. Badr, and G.J. Royle, *The Use of a Figure-of-Merit (Fom) for Optimisation in Digital Mammography: A Literature Review.* Radiat Prot Dosimetry, 2012.
8. Ministry of Health, t.E.a.C.C.M. *Breast Screening Programme*. 2012; Available from: https://ehealth.gov.mt/HealthPortal/health_institutions/Units/breastscreening/indexpage_brstscr.aspx.
9. IEC, *IEC 62220-1-2 ed1.0. Medical electrical equipment - Characteristics of digital X-ray imaging devices - Part 1-2: Determination of the detective quantum efficiency - Detectors used in mammography.* 2007.
10. *ICRU Report 82. Mammography - Assessment of Image Quality.* Journal of the ICRU, 2009. **9**(2): p. 65-75.
11. Ranger, N.T., J.Y. Lo, and E. Samei, *A technique optimization protocol and the potential for dose reduction in digital mammography.* Medical Physics, 2010. **37**(3): p. 962-969.
12. Baldelli, P., N. Phelan, and G. Egan, *Investigation of the effect of anode/filter materials on the dose and image quality of a digital mammography system based on an amorphous selenium flat panel detector.* Br J Radiol, 2010. **83**(988): p. 290-295.
13. Physics, N.M.S.U.D.o. *Types of Experimental Errors*. 2010; Available from: http://www.physics.nmsu.edu/research/lab110g/html/ERRORS.html.
14. Gregory, K.J., J.E. Pattison, and G. Bibbo, *Uncertainties of exposure-related quantities in mammographic x-ray unit quality control.* Medical Physics, 2006. **33**(3): p. 687-698.
15. (IPEM), I.o.P.a.E.i.M., *Report 89 The Commissioing and Routine Testing of Mammographic X-Ray Systems.* 2005.
16. 62220–1-2, I., *International Electrotechnical Commission, Medical electrical equipment—Characteristics of digital x-ray imaging devices—Part 1-2: Determination of the detective quantum efficiency, detector used in mammography.* 2007.
17. Nations, U., *Scientific Committee on the Effects of Atomic Radiation (UNSECAR) Sources and Effects of Ionising Radiation.* 2000, United Nations (UN): New York.

18. Hendrick, E., et al., *Technical quality control in mammography screening programmes in 22 countries.* Int J Qual Health Care, 2002. **14**(3): p. 219-226.
19. Moore, A.C., et al., *IPEM Report 89. The Commissioing and Routine Testing of Mammographic X-ray Systems.* 2005, York Publishing Services: UK.
20. Vassileva, J., et al., *Implementation of the European protocol for quality control of the technical aspects of mammography screening in Bulgaria.* Radiation Protection Dosimetry, 2005. **114**(1-3): p. 403-405.
21. Young, K.C. and J.M. Oduko, *NHSBSP Equipment Report 0801: TECHNICAL EVALUATION OF THE HOLOGIC SELENIA WITH A TUNGSTEN TUBE.* **NHSBSP**, 2008.
22. (IEC), I.S., *Evaluation and routine testing in medical imaging departments – Part 3-2: Acceptance tests – Imaging performance of mammographic X-ray equipment.* IEC Publications, 2007.
23. Moore, A.C., Dance, D.R., Evans, D.S., Lawinski, C.P., Pitcher, E.M., Rust, A. and Young, K.C *IPEM Report 89. The Commissioning and Routine Testing of Mamographic X-Ray Systems. A Technical Quality Control Protocol.* . 2005, Institute of Physics and Engineering in Medicine.: York, UK.
24. Workman, A., L. Castellano, and E. Kulama, *Commissioning and Routine Testing of Full Field Digital Mammography Systems.*, in *Equipment Report*, NHSBSP, Editor. 2009, NHSBSP.
25. Faj, D., et al., *Survey of mammography practice in Croatia: equipment performance, image quality and dose.* Radiation Protection Dosimetry, 2008. **131**(4): p. 535-540.
26. Ciraj-Bjelac, O., et al., *Good Reaso to implement quality assurance in nationwide breast cancer screening programs in Croatia and Serbia: Results from a pilot study.* European Journal of Radiology, 2009.
27. Kosutic, D., O. Ciraj-Bjelac, and D. Arandjic, *Mammography practice in Serbia: evaluation and optimisation of image quality and the technical aspects of the mammographic imaging chain.* Radiation Protection Dosimetry, 2010. **139**(1-3): p. 293-297.
28. Thierens, H., et al., *Typetesting of physical characteristics of digital mammography systems for screening within the Flemish breast screening programme.* European Journal of Radiology, 2009. **70**: p. 539-548.
29. Institute, N.C. *Digital vs. Film Mammography in the Digital Mammographic Screeing Trial (DMIST): Questions and Answers.* 2005; Available from: www.cancer.gov/newscenter/qa/dmistqandA.
30. Bick, L. and F. Diekmann, *Digital Mammography.*, ed. A.L. Baert. 2010, New York: Springer.
31. Vedantham, S., et al., *Full breast digital mammography with an amorphous silicon-based flat panel detector: physical characteristics of a clinical prototype.* Medical Physics, 2000. **27**(3): p. 558-567.
32. Shaw, J., et al. *Enhanced a-Si/CsI-based flat-panel x-ray detector for mammography*. 2004.
33. Ghetti, C., et al., *Physical characteristics of GE Senographe Essential and DS digital mammography detectors.* Medical Physics, 2008. **35**(2): p. 456-463.
34. Yorker, J.G., et al. *Characterization of a full-field digital mammography detector based on direct x-ray conversion in selenium*. 2002.
35. Bissonnette, M., et al. *Digital breast tomosynthesis using an amorphous selenium flat panel detector*. 2005.
36. Åslund, M., et al., *Physical characterization of a scanning photon counting digital mammography system based on Si-strip detectors.* Medical Physics, 2007. **34**(6): p. 1918-1925.

37. Thunberg, S.J., et al. *Evaluation of a photon-counting mammography system*. 2002.
38. Krug, K.B., et al., *Image Quality of Digital Direct Flat-Panel Mammography Versus an Analog Screen-Film Technique Using a Phantom Model.* Am. J. Roentgenol., 2007. **188**(2): p. 399-407.
39. Center, S.W.I. *True Digital Mammography vs. Computed Mammography*. 2010; Available from: www.coastalxray.com.au/SWIC_TrueDigitalMammography.html.
40. Programme, N.B.C.S. *Digital Mammography and Tomosynthesis*. 2014; Available from: http://www.cancerscreening.nhs.uk/breastscreen/digital-mammography.html.
41. (NEMA), N.E.M.A., *Digital Imaging and Communications in Medicine (DICOM).* in *Greyscale standard display function*. 2001.
42. Cush, S., *Evaluation of Digital Mammography: Update of the UK position.* Breast Cancer Resaerch, 2006. **8**: p. 2.
43. Muhogora, W.E., et al., *Application of European protocol in the evaluation of contrast-to-noise ratio and mean glandular dose for two digital mammography systems.* Radiation Protection Dosimetry, 2008. **129**(1-3): p. 231-236.
44. Morán, P., et al., *A survey of patient dose and clinical factors in a full-field digital mammography system.* Radiation Protection Dosimetry, 2005. **114**(1-3): p. 375-379.
45. Williams, M.B., P. Raghunathan, and M.J. More, *Optimization of Exposure Parameters in Full Field Digital Mammography.* Medical Physics, 2008. **35**(6): p. 2414-2423.
46. Schueller, G., et al., *Image Quality, lesion detection, and diagnostic efficacy in digital mammography: Full-field digital mammography versus computed radiography-based mammography using digital storage phosphor plates.* European Journal of Radiology, 2008. **67**(3): p. 487-496.
47. *Breast Imaging Reporting and Data System (BI-RADS). American College of Radiology.* 2003.
48. K. Schubauer-Berigan, M., et al., *Mammography Dose in Relation to Body Mass Index, Race, and Menopausal Status.* Radiation Protection Dosimetry, 2002. **98**(4): p. 425-432.
49. Kalathaki, M., et al., *Comparison of Full Field Digital (FFD) and Computed Radiography (CR) mammography system in Greece.*, in *International Conference on Radiation Protection in Medicine*. 2010: Varna, Bulgaria.
50. Berns, E. and E. Hendrick, *Performance comparison of full-field digital mammography to screen-film mammography in clinical parctice.* Medical Physics, 2002. **29**(5): p. 830-834.
51. Gennaro, G., et al., *Patient dose in full-field digital mammography: an Italian survey.* European Radiology, 2004. **14**: p. 645-652.
52. Gennaro, G. and C. di Maggio, *Dose comparison between screen/film and full-field digital mammography.* European Radiology, 2006. **16**(11): p. 2559-2566.
53. Fischer, et al., *Comparative study in patients with microcalcifications: full-field digital mammography vs screen-film mammography.* European Radiology, 2002. **12**(11): p. 2679-2683.
54. Barr, N., et al., *Comparison of Microcalcification Detection Rates and Recall Rates in Digital and Analogue Mammography*, in *Digital Mammography*, J. Martí, et al., Editors. 2010, Springer Berlin / Heidelberg. p. 513-517.
55. Young, K.C. and D. Kitou, *Review of literature on digital mammography screening.*, NHSBSP, Editor. 2005, NHBSP Digital Imaging Technologies Steering Group on behalf of the National Coordinating Centre for Physics in Mammography.
56. Schulz-Wendtland, R., et al., *Phantom study for the detection of simulated lesions in five different digital and one conventional mammography system. .* Rofo, 2004. **176**: p. 1127-1132.

57. Obenauer, et al., *Screen-film vs. full-field digital mammography: image quality, detectability and characterization of lesions.* European Radiology, 2002. **12**(7): p. 1697-1702.
58. Skaane, P., et al., *Breast Lesion Detection and Classification: Comparison of Screen-Film Mammography and Full-Field Digital Mammography with Soft-copy Reading—Observer Performance Study1.* Radiology, 2005. **237**(1): p. 37-44.
59. Fischmann, A., et al., *Comparison of full-field digital mammography and film-screen mammography: image quality and lesion detection.* Br J Radiol, 2005. **78**(928): p. 312-315.
60. Pisano, E.D., et al., *Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening.* The New England Journal of Medicine, 2005. **353**(17): p. 1773-1783.
61. Buist, D.S.M., et al., *Factors contributing to mammographic failure in women aged 40-49 years.* J Natl Cancer Inst, 2004. **96**: p. 1432-1440.
62. Charney, P.A., et al., *Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography.* Annals of Internal Medicine, 2003. **138**: p. 168-175.
63. Kerlikowske, K., et al., *Effect of age, breast density, and family history on the sensitivities of first screening mammography.* JAMA, 1996. **276**: p. 33-38.
64. Wolfe, J.N., *Risk for breast cancer development detrmined by mammographic parenchymal pattern.* Cancer, 1976. **37**: p. 2486-2492.
65. Byrne, C., et al., *Mammographic features and breast cancer risk: effects with time, age and menopausal status.* J Natl Cancer Inst, 1995. **87**: p. 1622-1929.
66. Boyd, N.F., G.S. Dite, and J. Stone, *Heritability of mammographic density, a risk for breast cancer.* New England Journal of Medicine, 2002. **347**: p. 886-894.
67. Lewin, J.M., et al., *Comparison of Full-Field Digital Mammography with Screen-Film Mammography for Cancer Detection: Results of 4,945 Paired Examinations1.* Radiology, 2001. **218**(3): p. 873-880.
68. Lewin, J.M., et al., *Clinical Comparison of Full-Field Digital Mammography and Screen-Film Mammography for Detection of Breast Cancer.* Am. J. Roentgenol., 2002. **179**(3): p. 671-677.
69. Skaane, P., K. Young, and A. Skjennald, *Population-based Mammography Screening: Comparison of Screen-Film and Full-Field Digital Mammography with Soft-Copy Reading—Oslo I Study1.* Radiology, 2003. **229**(3): p. 877-884.
70. Skaane, P. and A. Skjennald, *Screen-Film Mammography versus Full-Field Digital Mammography with Soft-Copy Reading: Randomized Trial in a Population-based Screening Program—The Oslo II Study1.* Radiology, 2004. **232**(1): p. 197-204.
71. Skaane, P., S. Hofvind, and A. Skjennald, *Randomized Trial of Screen-Film versus Full-Field Digital Mammography with Soft-Copy Reading in Population-based Screening Program: Follow-up and Final Results of Oslo II Study1.* Radiology, 2007. **244**(3): p. 708-717.
72. Pisano, E.D., et al., *Diagnostic Accuracy of Digital versus Film Mammography: Exploratory Analysis of Selected Population Subgroups in DMIST1.* Radiology, 2008. **246**(2): p. 376-383.
73. Hambly, N.M., et al., *Comparison of Digital Mammography and Screen-Film Mammography in Breast Cancer Screening: A Review in the Irish Breast Screening Program.* Am. J. Roentgenol., 2009. **193**(4): p. 1010-1018.
74. Heddson, B., et al., *Digital versus screen-film mammography: A retrospective comparison in a population-based screening program.* European Journal of Radiology, 2007. **64**(3): p. 419-425.

75. Del Turco, M.R., et al., *Full-Field Digital Versus Screen-Film Mammography: Comparative Accuracy in Concurrent Screening Cohorts.* Am. J. Roentgenol., 2007. **189**(4): p. 860-866.
76. Vigeland, E., et al., *Full-field digital mammography compared to screen film mammography in the prevalent round of a population-based screening programme: the Vestfold County Study.* European Radiology, 2008. **18**(1): p. 183-191.
77. Zdešar, U., *Reference levels for image quality in mammography.* Radiation Protection Dosimetry, 2008. **129**(1-3): p. 170-172.
78. Dance, D.R., et al., *Additional factors for the estimation of mean glandular doses using the UK mammography dosimetry protocol.* Physics in Medicine and Biology, 2000. **45**: p. 3225-3240.
79. Protection., I.C.o.R., *Radiological Protection and Safety in Medicine.* ICRP Publication 73, 1996. **26**: p. 1-47.
80. Smans, K., et al., *Towards a proposition of a diagnostic (dose) reference level for mammographic acquisitions in breast screening measurements in Belgium.* Radiation Protection Dosimetry, 2005. **117**(1-3): p. 321-326.
81. Baldelli, P., et al., *Comprehensive Dose Survey of Breast Screening in Ireland.* Radiation Protection Dosimetry.
82. Countries, C.o.t.E., *Council Directive 97/43 Euratom on health protection of individuals against dangers of ionizing radiation in relation to medical exposure.* Off. J.Eur. Commun., 1997. **L180 (40)**.
83. Union, E., *Council Directive 2013/59/EURATOM*, O.J.o.t.E. Union, Editor. 2014.
84. (IPEM), I.o.P.a.E.i.M., *Report 91 Recommended Standards for the Routine Performance Testing of Diagnostic X-Ray Imaging Systems.* 2005.
85. (IPEM), I.o.P.a.E.i.M., *Report 88 Guidance on the Establishment and Use of Diagnostic Reference Levels for Medical X-Ray Examinations. .* 2004.
86. Young, K.C., *Radiation Doses in the UK trial of breast screening in women aged 40-48 years. .* Br J Radiol, 2002. **75**: p. 362-370.
87. Alizadeh Riabi, H., P. Mehnati, and A. Mesbahi, *Evaluation of mean glandular dose in a full-field digital mammography unit in Tabriz, Iran.* Radiation Protection Dosimetry, 2010. **142**(2-4): p. 222-227.
88. Wolfe, J.N., et al., *Breast Parenchymal Patterns: analysis of 332 incident breast carcinomas.* American Journal of Radiology, 1982. **138**: p. 113-118.
89. Sutton, D., et al., *Diagnostic Reference Levels - a UK percective In:.* Proceedings of Symposium - Radiation Protection of the North West European RP Societies, Utrecht., 2003: p. 225-228.
90. Young, K.C., A. Burch, and J.M. Oduko, *Radiation doses received in the UK Breast Screening Programme in 2001 and 2002.* Br J Radiol, 2005. **78**(927): p. 207-218.
91. Michielsen, K., et al., *Results of a European dose survey for mammography.* Radiation Protection Dosimetry, 2008. **129**(1-3): p. 199-203.
92. Samei, E. and M.J. Flynn, *An experimental comparison of detector performance for computed radiography systems.* Med Phys, 2002. **29**(4): p. 447-59.
93. Samei, E. and M.J. Flynn, *An experimental comparison of detector performance for direct and indirect digital radiography systems.* Med Phys, 2003. **30**(4): p. 608-22.
94. Samei, E., et al., *A framework for optimising the radiographic technique in digital X-ray imaging.* Radiation Protection Dosimetry, 2005. **114**(1-3): p. 220-229.
95. Kanga, K.C., et al., *A critical comparison of three full field digital mammography systems using figure of merit.* Med J Malaysia, 2010. **65**(2): p. 119-122.
96. Williams, M.B., et al., *Beam Optimization for Digital Mammography -II.* LNCS, 2006. **4046**: p. 273-280.

97. Boone, J.M., *Normalized glandular dose (DgN) coefficients for arbitrary X-ray spectra in mammograhy: computer fit values of Monte Carlo derived data.* Medical Physics, 2002. **29**: p. 869-875.
98. Delis, H., et al., *Evaluating the Figure of Merit in mammography utilizing Monte Carlo simulation.* Nuclear Instruments and Methods in Physics Research., 2007. **A 580**: p. 493-496.
99. Morgan, M.P., M.M. Cooke, and G.M. McCarthy, *Microcalcifications associated with breast cancer: an epiphenomenon or biologically significant feature of selected tumors?* J Mammary Gland Biol Neoplasia, 2005. **10**(2): p. 181-7.
100. Marshall, N.W., et al., *Image quality assessment in digital mammography: part I. Technical charceterization of the systems.* Phys Med Biol, 2011. **56**: p. 4201-4220.
101. Garcia-Molla, R., R. Linares, and R. Ayala, *Study of DQE dependence with beam quality on GE essential mammography flat panel.* J Appl Clin Med Phys, 2011. **12**(1): p. 3176.
102. Levison, D.A., et al., *Oesophageal neoplasia in male Wistar rats due to parenteral di(2-hydroxypropyl)-Nitrosamine (DHPN): a combined histopathological, histochemical and electron microscopic study.* J Pathol, 1979. **129**(1): p. 31-6.
103. Berns, E.A., R.E. Hendrick, and G.R. Cutter, *Optimization of technique factors for a silicon diode array full-field digital mammography system and comparison to screen-film mammography with matched average glandular dose.* Medical Physics, 2003. **30**(3): p. 334-340.
104. Lo, J.Y., et al., *Radiographic techniques optimization for an amorphous selenium FFDM system: Phantom measurements and initial patient results.* Proceedings of the IWDM. 2004, North Carolina: Chapel Hill, NC, University of North Carolina at Chapel Hill.
105. Klausz, R. and N. Shramchenko, *Dose to population as a metric in the design of optimised exposure control in digital mammography.* Radiation Protection Dosimetry, 2005. **114**(1-3): p. 369-374.
106. Young, K.C., et al., *Optimal beam quality selection in digital mammography.* Br J Radiol, 2006. **79**(948): p. 981-990.
107. Dance, D.R., et al., *Influence of anode/filter material and tube potential on contrast, signal-to-noise ratio and average absorbed dose in mammography: a Monte Carlo study.* British Journal of Radiology, 2000. **73**: p. 1056-1067.
108. Poyry, P., *Sentinel teaching course on digital mammography*, in *OPtimisation of exposure settings in digital mammography.*, F. STUK - Radiation and Nuclear Safety Authority, Editor. 2006.
109. Vedantham, S., et al., *Breast imaging using an amorphous silicon-based full-field digital mammographic system: stability of a clinical prototype.* J Digit Imaging, 2000. **13**(4): p. 191-9.
110. Vedantham, S., et al., *Full breast digital mammography with an amorphous silicon-based flat panel detector: physical characteristics of a clinical prototype.* Med Phys, 2000. **27**(3): p. 558-67.
111. Yaffe, M.J., *Detectors for Digital Mammography*, in *Digital Mammography*, L. Bick and F. Diekmann, Editors. 2010, Springer: Berlin. p. 13-31.
112. Rivetti, S., et al., *Comparison of different commercial FFDM units by means of physical characterization and contrast-detail analysis.* Med Phys, 2006. **33**(11): p. 4198-209.
113. Thomas, J.A., et al., *Contrast-detail phantom scoring methodology.* Med Phys, 2005. **32**(3): p. 807-14.
114. System, A.M. *CDMAM 3.4*. 2015; Available from: http://www.artinis.com/product/cdmam_34.

115. Young, K., J. Cook, and J. Oduko, *Use of the European Protocol to Optimise a Digital Mammography System Digital Mammography*, S. Astley, et al., Editors. 2006, Springer Berlin / Heidelberg. p. 362-369.

116. Suryanarayanan, S., et al., *Flat-panel digital mammography system: contrast-detail comparison between screen-film radiographs and hard-copy images.* Radiology, 2002. **225**(3): p. 801-7.

117. Karssemeijer, N. and H.O. Thijssen, *Determination of contrast-detail curves of mammography systems by automated image analysis. .* Digital Mammography '96. Proceedings of the 3rd International Workshop on Digital Mammography, 1996: p. 155-160.

118. Zoetelief, J., J.T.M. Jansen, and N.J.P. de Wit, *Determination of Image Quality in Relation to Absorbed Dose in Mammography.* Radiation Protection Dosimetry, 1993. **49**(1-3): p. 157-161.

119. Verbrugge, B., *Validation of Analysis Methods for Automated CDMAM Reading.*, in *Katholieke Universiteit Leuven*. 2007, FACULTY OF SCIENCE DEPARTMENT OF PHYSICS AND ASTRONOMY SECTION OF NUCLEAR AND RADIATION PHYSICS, KATHOLIEKE UNIVERSITEIT LEUVEN: Leuven.

120. Young, K.C., et al., *Evaluation of software for reading images of the CDMAM test object to assess digital mammography systems.* Proc. of SPIE, 2008. **6913**(6913C): p. 1-11.

121. B.V., A.M.S., *Manual CDMAM Analyser Version 1.1*. 2006.

122. Wichmann, F.A. and N.J. Hill, *The psychometric function: II. Bootstrap-based confidence intervals and sampling.* Percept Psychophys, 2001. **63**(8): p. 1314-29.

123. Wichmann, F.A. and N.J. Hill, *The psychometric function: I. Fitting, sampling, and goodness of fit.* Percept Psychophys, 2001. **63**(8): p. 1293-313.

124. Klein, S., *Measuring, estimating, and understanding the psychometric function: A commentary.* Perception & Psychophysics, 2001. **63**(8): p. 1421-1455.

125. Veldkamp, W.J., M.A. Thijssen, and N. Karssemeijer, *The value of scatter removal by a grid in full field digital mammography.* Med Phys, 2003. **30**(7): p. 1712-8.

126. Wu, X., et al., *Normalised average glandular dose in molybedenum target-rhodium filter and rhodium target-rhodium filter mammography.* Radiology, 1994. **193**: p. 83-89.

127. Hendrick, R.E., et al., *Comparison of Acquisition Parameters and Breast Dose in Digital Mammography and Screen-Film Mammography in the American College of Radiology Imaging Network Digital Mammographic Imaging Screening Trial.* Am. J. Roentgenol., 2010. **194**(2): p. 362-369.

128. IAEA. *https://rpop.iaea.org/RPoP/RPoP/Content/index.htm*. 2011.

129. Radcal. *ACCU-Family Instruments -ACCU-PRO*. 2009; Available from: www.radcal.com/accupro3.html.

130. Unfors. *Unfors Xi for mammography*. Available from: www.unfors.com/products.

131. Robson, K.J., *A parametric method for determining mammographic X-ray tube output and half value layer.* Br J Radiol, 2001. **74**(880): p. 335-40.

132. Hologic, *Selenia Dimensions Quality Control Manual MAN-01116 Revision 002*. 2008: Bedford, USA.

133. Young, K.C. and J.M. Oduko, *Technical evaluation of Hologic Selenia Dimensions 2-D Digital Breast Imaging System with softwarw version 1.4.2 NHSBSP equipment report 1201* N.C.C.f.t.P.o. Mammography, Editor. 2012.

134. Young, K.C., et al., *Technical Evaluation of the GE Essential Full Field Digital Mammography System NHSBSP Equipment Report 0803*, N.C.C.f.t.P.o. Mammography, Editor. 2008.

135. Baldelli, P., et al., *COMPREHENSIVE DOSE SURVEY OF BREAST SCREENING IN IRELAND.* Radiation Protection Dosimetry, 2010.
136. Thiele, D.L., et al., *Diagnostic reference levels for mammography in BreastScreen Queensland.* Australas Phys Eng Sci Med, 2011. **34**(3): p. 415-8.
137. Eklund, S.A., *The impact of insurance on oral health.* J Am Coll Dent, 2001. **68**(2): p. 8-11.
138. Young, K.C. and M.L. Ramsdale, *Image Quality and Dose Measurement Phantoms in the UK Breast Screening Programme.* Radiation Protection Dosimetry, 1993. **49**(1-3): p. 175-177.
139. Young, K.C., M.L. Ramsdale, and F. Bignell, *Review of Dosimetric Methods for Mammography in the UK Breast Screening Programme.* Radiation Protection Dosimetry, 1998. **80**(1-3): p. 183-186.
140. Jamal, N., K.H. Ng, and D. McLean, *A study of mean glandular dose during diagnostic mammography in Malaysia and some of the factors affecting it.* Br J Radiol, 2003. **76**(904): p. 238-45.
141. Ciraj-Bjelac, O., et al., *Mammography radiation dose: initial results from Serbia based on mean glandular dose assessment for phantoms and patients.* Radiation Protection Dosimetry, 2010. **140**(1): p. 75-80.
142. Heggie, J., *Survey of dose in screening mammography.* Australas Phys Eng Sci Med, 1996. **19**: p. 207-216.
143. Eklund, S., et al., *The Impact of Anatomic Variations on Absorbed Radiation Doses in Mammography.* Radiation Protection Dosimetry, 1993. **49**(1-3): p. 167-170.
144. Moran, P., M. Chevalier, and E. Vano, *Comparative study of dose values and image quality in mammography in the area of Madrid.* Br J Radiol, 1994. **67**(798): p. 556-63.
145. Klein, R., et al., *Determination of average glandular dose with modern mammography units for two large groups of patients.* Phys Med Biol, 1997. **42**(4): p. 651-71.
146. Gentry, J.R. and L.A. DeWerd, *TLD measurements of in vivo mammographic exposures and the calculated mean glandular dose across the United States.* Med Phys, 1996. **23**(6): p. 899-903.
147. Tsapaki, V., et al., *Investigation of breast dose in five screening mammography centres in Greece.* J Radiol Prot, 2008. **28**(3): p. 337-46.
148. Wu, X., G.T. Barnes, and D.M. Tucker, *Spectral dependence of glandular tissue dose in screen-film mammography.* Radiology, 1991. **179**(1): p. 143-8.
149. Boone, J.M., *Glandular Breast Dose for Monoenergetic and High-Energy X-ray Beams: Monte Carlo Assessment.* Radiology, 1999. **213**: p. 23-37.
150. Sobol, W.T. and X. Wu, *Parametrization of mammography normalized average glandular dose tables.* Med Phys, 1997. **24**(4): p. 547-54.
151. D.A., F., *Statistical Models: Theory and Practice*. 2nd Edition ed. 2009.
152. Wikipedia. *Pearson product-moment correlation coefficient*. 2015; Available from: https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient.
153. Kramer, R., et al., *Backscatter factors for mammography calculated with Monte Carlo methods.* Phys Med Biol, 2001. **46**(3): p. 771-81.
154. Hemdal, B., *Forward-scattered radiation from the compression paddle should be considered in glandular dose estimations.* Radiat Prot Dosimetry, 2011. **147**(1-2): p. 196-201.
155. Petrie, A. and C. Sabin, *Medical Statistics at a Glance*. 2000, Oxford: Blackwell Science Ltd.
156. Handbook, N.E.S. *Quantile-Quantile Plot*. 2013; Available from: http://www.itl.nist.gov/div898/handbook/eda/section3/qqplot.htm.

157. Zanca, F., et al., *Evaluation of clinical image processing algorithms used in digital mammography.* Med Phys, 2009. **36**(3): p. 765-75.
158. Bosmans, H., et al., *Image quality measurements and metrics in full field digital mammography: an overview.* Radiation Protection Dosimetry, 2005. **117**(1-3): p. 120-130.
159. Smans, K., et al., *Simulation of image detectors in radiology for determination of scatter-to-primary ratios using Monte Carlo radiation transport code MCNP/MCNPX.* Med Phys, 2010. **37**(5): p. 2082-91.
160. Fletcher-Heath, L. and R.V. Metter. *Quantifying the performance of human and software CDMAM phantom image observers for the qualification of digital mammography systems*. 2005. SPIE.
161. Yang, C.-Y.J. and R.V. Metter. *The variability of software scoring of the CDMAM phantom associated with a limited number of images*. 2007. SPIE.
162. BV, A.M.S., *CDMAM Analyser Version 1.1*. 2006.
163. Troiano, A.R., et al., *Low disease risk in relatives of north african lrrk2 Parkinson disease patients.* Neurology, 2010. **75**(12): p. 1118-9.
164. Dainty, J.C. and R. Shaw, *Image Science: principles, analysis and evaluation of photographic-type image processing.* 1974, London: Academic Press Inc.
165. Rivetti, S., et al., *Physical and psychophysical characterization of a novel clinical system for digital mammography.* Med Phys, 2009. **36**(11): p. 5139-48.
166. Rivetti, S., et al., *Comparison of different computed radiography systems: physical characterization and contrast detail analysis.* Med Phys, 2010. **37**(2): p. 440-8.
167. Marshall, N.W., *Detective quantum efficiency measured as a function of energy for two full-field digital mammography systems.* Phys Med Biol, 2009. **54**(9): p. 2845-61.
168. Salvagnini, E., et al., *Effective detective quantum efficiency for two mammography systems: Measurement and comparison against established metrics.* Medical Physics, 2013. **40**(10): p. 101916.
169. Protection., I.C.o.R., *1990 Recommendations of the International Commission on Radiological Protection.* ICRP Publication 60, 1991. **21**: p. 1-201.
170. Hendrick, R.E., *Quality assurance in mammography. Accreditation, legislation, and compliance with quality assurance standards.* Radiol Clin North Am., 1992. **30**(1): p. 243-55.
171. Muramoto, H., et al., *Influence of monitor luminance and room illumination on soft-copy reading evaluation with electronically generated contrast-detail phantom: comparison of cathode-ray tube monitor with liquid crystal display.* Nagoya J Med Sci, 2006. **68**(3-4): p. 115-20.
172. Shimamoto, K., et al., *Interobserver agreement and performance score comparison in quality control using a breast phantom: screen-film mammography vs computed radiography.* Eur Radiol., 2002. **12**(9): p. 2192-7. Epub 2002 Apr 18.
173. Samei, E., et al., *Effective DQE (eDQE) and speed of digital radiographic systems: an experimental methodology.* Med Phys, 2009. **36**(8): p. 3806-17.
174. Carton, A.-K., et al., *The effect of scatter and glare on image quality in contrast-enhanced breast imaging using an a-Si/CsI(Tl) full-field flat panel detector.* Medical Physics, 2009. **36**(3): p. 920-928.
175. Kimme-Smith, C., *New digital mammography systems may require different X-ray spectra and, therefore, more general normalized glandular dose values.* Radiology., 1999. **213**(1): p. 7-10.
176. Huda, W., et al., *Experimental investigation of the dose and image quality characteristics of a digital mammography imaging system.* Medical Physics, 2003. **30**: p. 442-448.

177. Bernhardt, P., T. Mertelmeier, and M. Hoheisel, *X-ray spectrum optimization of full-field digital mammography: simulation and phantom study.* Med Phys., 2006. **33**(11): p. 4337-49.
178. Samei, E., et al., *Effective DQE (eDQE) and speed of digital radiographic systems: an experimental methodology.* Med Phys., 2009. **36**(8): p. 3806-17.
179. Samei, E., et al., *Effective DQE (eDQE) and speed of digital radiographic systems: An experimental methodology.* Medical Physics, 2009. **36**(8): p. 3806-3817.
180. Boone, J.M., T.R. Fewell, and R.J. Jennings, *Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography.* Medical Physics, 1997. **24**(12): p. 1863-1874.
181. Morrison, G.D., Radiography. **4**(3): p. 228-229.
182. David, M.G., et al., *Experimental and Monte Carlo-simulated spectra of standard mammography-quality beams.* The British Journal of Radiology, 2012. **85**(1013): p. 629-635.
183. Salvagnini, E., et al., *Effective detective quantum efficiency for two mammography systems: measurement and comparison against established metrics.* Med Phys, 2013. **40**(10): p. 101916.
184. Marshall, N., E. Salvagnini, and H. Bosmans, *Effective Detective Quantum Efficiency (eDQE) Measured for a Digital Breast Tomosynthesis System*, in *Breast Imaging*, H. Fujita, T. Hara, and C. Muramatsu, Editors. 2014, Springer International Publishing. p. 127-133.
185. Salvagnini, E., et al. *Effective detective quantum efficiency (eDQE) and effective noise equivalent quanta (eNEQ) for system optimization purposes in digital mammography*. 2012.
186. Yaffe, M., *Basic Physics of Digital Mammography*, in *Digital Mammography*, U. Bick and F. Diekmann, Editors. 2010, Springer Berlin Heidelberg. p. 1-11.
187. KONSTANTINIDIS, A.C., *EVALUATION OF DIGITAL X-RAY DETECTORS FOR MEDICAL IMAGING APPLICATIONS* in *DEPARTMENT OF MEDICAL PHYSICS AND BIOENGINEERING* 2011, UNIVERSITY COLLEGE LONDON.
188. Michail, C.M., et al., *Figure of Image Quality and Information Capacity in Digital Mammography.* BioMed Research International, 2014. **2014**: p. 11.