A comparative study evaluating the performance of diagnostic radiography units and protocols for paediatric and adult chest radiography examinations.

Sadeq Haleem Mohsin Al-Murshedi

PhD. Thesis **2019**

A comparative study evaluating the performance of diagnostic radiography units and protocols for paediatric and adult chest radiography examinations.

Sadeq Haleem Mohsin Al-Murshedi

School of Health Sciences,

University of Salford, Salford, UK

Submitted in Partial Fulfilment of the Requirements for the

Degree of Doctor of Philosophy (PhD), July 2019

(Medical Physics)

Table of Contents

List of Figures

List of Tables

List of publications

NO.	Title	Status
1	An investigation into the validity of utilising the CDRAD 2.0 phantom for optimisation studies in digital radiography. Sadeq Al-Murshedi, Peter Hogg, and Andrew England. DOI: 10. 1259/ bjr. 20180317	Journal paper published in the British journal of radiology (BJR)
2	A novel method for comparing radiation dose and image quality, between and within different X-ray units in a series of hospitals. Sadeq Al-Murshedi, Luis Lanca, Peter Hogg, and Andrew England. DOI: 10.1088/1361-6498/aae3fa	Journal paper published in the journal of radiological protection (JRP)
3	Relationship between body habitus and image quality and radiation dose in chest x-ray examinations: a phantom study. Sadeq Al-Murshedi, Peter Hogg, and Andrew England. DOI:10.1016/j.ejmp.2018.12.009	Journal paper published in Physica Medica
4	Comparative analysis of radiation dose and low contrast detail detectability using routine paediatric chest radiography protocols. Sadeq Al-Murshedi, Peter Hogg, Annemieke Meijer, Hendrik Erenstein ,and Andrew England. DOI:10.1016/j.ejrad.2019.02.017	Journal paper published in the European Journal of Radiology (EJR)
5	A comparative study to evaluate radiation dose and image quality for neonate and adult chest radiography among several diagnostic radiography X-ray units. Sadeq Al-Murshedi, Peter Hogg, and Andrew England.	Conference paper presented in UKRC congress 2018
6	An investigation of the impact of image viewing parameter settings on the performance of 2.4 MP colour monitor in visualising low contrast detail using the CDRAD phantom. Sadeq Al-Murshedi,	Conference paper presented in ECR congress 2018

	Peter Hogg, Mohamed Benhalim, Maily J Alrowily and Andrew England.	
7	Comparison of Image Quality and Radiation Exposure in Hospitals. Sadeq Al-Murshedi, Peter Hogg, and Andrew England.	Conference paper presented in SPARC congress /University of Salford 2018
8	Impact of the anode heel effect on image quality and effective dose for AP pelvis: a pilot study. Carst Buissink, Matthew Bowdler, Sadeq Al-Murshedi, S. Custódio, Andrea Huhn, J. Pires-Jorge, Ali Mohammed Ali, A. Peters, Y. Rey, T. Urdahl and Peter Hogg	Conference paper presented in ECR congress 2017
9	A Comparison of the Performance of Digital Radiography Systems and Chest Imaging Protocols for Paediatrics in Hospitals across the Northwest of England. Sadeq Al-Murshedi, Peter Hogg, and Andrew England.	Conference paper presented in SPARC congress /University of Salford 2017
10	Evaluation of image quality, lesion visibility and entrance surface dose for routine adult chest X-ray examinations in 8 hospitals. Sadeq Al-Murshedi, Peter Hogg, and Andrew England.	Poster presented in UKRC congress 2018
11	A comparative assessment of pathology visibility and radiation dose for routine neonatal chest radiography examinations in eight hospitals. Sadeq Al-Murshedi, Peter Hogg, and Andrew England.	Poster presented in ECR congress 2018
12	A phantom study comparing image quality and radiation exposure for adult chest radiography examinations in hospitals using 17 diagnostic radiography x-ray units. Sadeq Al-Murshedi, Peter Hogg, and Andrew England.	Poster presented in ECR congress 2018
13	A comparison of the performance of a 2.4 MP colour monitor and a 5.0 MP monochrome	Poster presented in ECR congress 2018

	monitor in visualising low contrast detail using the CDRAD phantom. Sadeq Al-Murshedi, Peter Hogg, Ahmed khalid Abdullah and Andrew England.	
14	A novel method to investigate the correlation between the improvement of low contrast detail detectability using CDRAD phantom and visual lesion visibility and visual image quality in Lungman chest phantom. Sadeq Al-Murshedi, Peter Hogg, and Andrew England.	Poster presented in ECR congress 2018
15	A comparative study to evaluate dose and image quality for paediatrics phantom chest radiography using 17 diagnostic radiography x-ray units. Sadeq Al-Murshedi, Peter Hogg, and Andrew England.	Poster presented in ECR congress 2018
16	Evaluation of image quality and entrance surface dose for routine adult chest radiography examinations in 8 hospitals. Sadeq Al-Murshedi, Peter Hogg, and Andrew England.	Poster presented in ECR congress 2018
17	An investigation into the impact of aging on the performance of LCD 2.4 MP colour display monitor when visualising low contrast detail using a CDRAD phantom. Sadeq Al-Murshedi, Peter Hogg, and Andrew England.	Poster presented in ECR congress 2018
18	Impact of the anode heel effect on image quality and effective dose for AP pelvis: a pilot study. Carst Buissink, Matthew Bowdler, Ahmed khalid Abdullah, Sadeq Al-Murshedi, S. Custódio, Andrea Huhn, J. Pires-Jorge, Ali Mohammed Ali, A. Peters, Y. Rey, T. Urdahl and Peter Hogg	Book chapter OPTIMAX 2016

Acknowledgements

Firstly, I would like to express my deepest appreciation to my supervisors, **Dr Andrew England** and **Professor Peter Hogg,** for their help, guidance and encouragement during my journey to completing my PhD, and for evoking so much excitement and interest in their teaching. Words would not be enough to express my feelings for their tremendous help and support.

I would like to thank professor **David Manning** for providing some formative inspiration without which this thesis would not have occurred.

I am also grateful to all the staff and the PhD students in the Radiography Directorate / University of Salford for their help and support.

I also offer great thanks to my family for their persistent support and encouragement.

Finally, I would like to take the opportunity to thank the Higher Committee for Education Development in Iraq (HCED-Iraq) for their financial support.

Abbreviations

American College of Radiology	ACR
Amorphous selenium	a-Se
Alternative Free-response Receiver Operating Characteristics Analysis	AFROC
Area under the curve	AUC
Alpha level of significance	Alpha
Automatic exposure control	AEC
Air kerma area product	KAP
Backscatter factor	BSF
Chest radiography	CXR
Contrast to noise ratio	CNR
Conspicuity index	CI
Commission of the European Communities	CEC
Contrast detail	CD
Computed radiography	CR
Caesium iodide	CsI
Detective quantum efficiency	DQE
Direct digital radiography	DDR
Diagnostic reference level	DRL
Dose area product	DAP
Digital Imaging and Communications in Medicine	DICOM
Entrance surface air kerma	ESAK
Effective dose	ED
Entrance surface dose	ESD
Figure of merit	FOM
Four-alternative forced-choice	4AFC
Focus to skin distance	FSD
Free-response Receiver Operating Characteristics Analysis	FROC
Grayscale standard display function	GSDF
Hounsfield unit	HU
Image quality	IQ
Incident air karma	IAK
Image quality figure inverse	IQF_{inv}
Image quality figure	IQF
Indirect digital radiography	IDR
International Commission on Radiation Units and Measurements	ICRU
Institute of Physics and Engineering in Medicine	IPEM
Interclass correlation coefficient	ICC
The jack-knife Free-response Receiver Operating Characteristics Analysis	JAFROC

Kilovoltage peak	kVp
Low contrast detail	LCD
Modulation transfer function	MTF
Milliampere second	mAs
Metal oxide semiconductor field effect transistors	MOSFETs
National Health Service	NHS
Noise power spectrum	NPS
National patient dose database	NPDD
Polymethyl methacrylate	PMMA
Priori difference of the mean	APD
Posterior anterior	PA
Quality control	QC
Quality index	Q
Receiver Operating Characteristics Analysis	ROC
Regions of interest	ROIs
Signal to noise ratios	SNR
Simulated lesion visibility	LV
Signal-difference-to-noise ratio	SdNR
Source to image detector distance	SID
Source to object distance	SOD
Standard deviation	SD
Threshold detection index	H_T
Thermoluminescent dosimeters	TLDs
Thin film transistor	TFT
Visual grading analysis	VGA
World Health Organization	WHO

Abstract

Purpose: Little is known about the variations in image quality (IQ) and radiation dose for paediatric and adult chest radiography (CXR), between and within hospitals. Large variations in IQ could influence the diagnostic accuracy, and variations in radiation dose could affect the risk to patients. This thesis aims to develop, validate and then use a novel method for comparing IQ and radiation dose for paediatric and adult CXR imaging examinations and report variation between a series of public hospitals.

Method: A Figure of Merit (FOM) concept was used for the purposes of comparing IQ and radiation dose, between and within hospitals. Low contrast detail (LCD) detectability, using the CDRAD 2.0 phantom, was utilised as the main method for IQ evaluation. The validity of utilising LCD detectability, using CDRAD 2.0 phantom, for evaluating visual IQ, simulated lesion visibility (LV) and CXR optimisation studies, was investigated. This was done by determining the correlation between the LCD detectability and visual measures of IQ and LV for two lesions with different locations and visibility in the Lungman chest phantom.

The CDRAD 2.0 phantom and two anthropomorphic phantoms (adult Lungman and the neonatal Gammex phantom) were used to simulate the chest region. Radiographic acquisitions were conducted on 17 X-ray units located in eight United Kingdom (UK) public hospitals within the North-west of England using their existing CXR protocols. The CDRAD 2.0 phantom was combined with different thicknesses of Polymethyl methacrylate (PMMA) slabs to simulate the chest regions of 5 different age groups: neonate, 1, 5, 10 years and adults. A Lungman phantom, with and without the fat jacket, was used to simulate average and larger sized patients.

IQ was evaluated using a number of methods, including: 1) physically, by calculating LCD detectability as represented by an image quality figure inverse (IQF_{inv}) using the CDRAD analyser software; 2) using images acquired from the anthropomorphic phantoms – for this, a relative visual grading analysis (VGA) method was used. Additionally, signal to noise ratios (SNR), contrast to noise ratios (CNR) and conspicuity indices (CI) were calculated for all phantom image data in this study. Incident air karma (IAK) was measured using a solid-state dosimeter.

Results: Regarding the validation of utilising LCD detectability for evaluating visual IQ and LV, and CXR optimisation studies, a strong positive correlation ($r = 0.91$; $p < 0.001$)

was observed between IQF_{inv} and the visual IQ scores from the Lungman phantom. A good correlation was observed between IQF_{inv} and visual LV from the Lungman phantom for both lesions (lesion 1 (with low visibility) ($r = 0.79$; $p < 0.001$); lesion 2 (with high visibility) ($r = 0.68$; $p < 0.001$), respectively).

Considerable variation in standard imaging protocols/techniques, radiation dose, IQ and FOM were observed between the hospitals, while within hospital variation was lower. A weak correlation between IQ and radiation dose was observed across most of the age groups studied.

Conclusion: A novel method has been established to evaluate and compare IQ and radiation dose between and within hospitals based on an FOM concept. This combines IQ and radiation dose into a single factor and is the first of its kind to reported within the field of medical imaging. It can be confirmed that LCD detectability using the CDRAD 2.0 phantom is valid for evaluating visual IQ and LV and can be of use within routine quality assurance and optimisation studies in digital radiography. Further radiation dose optimisation for the paediatric age groups and adult group, especially in hospitals /X-ray machines with low IQ and high IAK, are required.

Chpter 1: Introduction and Thesis Outline

1.1 Introduction

An essential technical priority of any X-ray examination is the production of an image with sufficient quality to answer the clinical question. Balanced against this is the requirement to avoid unnecessary radiation risk to the patient (Busch & Faulkner, 2005). The importance of this is even more apparent for paediatric radiology since children are approximately 10 times more radiosensitive than adults and the danger of developing cancer in children could be up to 3 times more than that for adults (Mathews et al., 2013; Raissaki, 2004). Within national legislature, a key aim is to identify acquisition parameters which deliver acceptable image quality (IQ) with the lowest radiation dose practicable (Britain, 2000). The balance between radiation dose and IQ is often referred to as optimisation and is one of the fundamental principles in radiographic practice (Willis, 2009). Optimisation is not easy to achieve because of the differences in the commercially available digital detectors and imaging equipment characteristics together with patient size variations (Almén, Lööf, & Mattsson, 1996). Differences are also likely to exist in the acquisition parameters that are utilised during imaging examinations and the underlying reasons for this are likely to relate to radiographer/radiologist preferences as well as the inherent differences between the technologies.

Paediatric dose optimisation is even more challenging than that of adults, with the wide variation in paediatric sizes making the determination of optimum exposure parameters more difficult (Willis, 2009). The above factors could lead to IQ and radiation dose differences, between and within hospitals, for the same clinical investigations. IQ differences may also affect the diagnostic outcome. Furthermore, differences in radiation dose may affect the risk to the patient. Among the different X-ray examinations, chest radiography (CXR) is the most common examination in both children and adults as it is an invaluable first line imaging test for investigating a broad range of clinical problems (Bacher et al., 2003; Raoof et al., 2012; Veldkamp, Kroft, & Geleijns, 2009). However, the optimisation of radiation dose for CXR is considered to be difficult because of the wide range of tissue types and pathologies potentially present within the chest region. These tissue types vary between high (e.g. mediastinum) and low (e.g. lung) X-ray attenuation characteristics (Doyle, Martin, & Gentle, 2006; G. Singh, Martin, McCurrach, & Phanco, 2013). Often there are a number of clinical protocols that are available and suitable for undertaking CXRs (G. Singh et al., 2013). The Commission of the European Communities

(CEC) (Commission of the European Communities (CEC), 1996a) and the American College of Radiology (ACR) (American College of Radiology, 2014) have independently published 2 guidelines on paediatric and adult radiographic imaging which refers to the chest region. Both guidelines have limitations. The CEC guideline is the most detailed with regard to paediatric and adult radiography and it provides information on the image criteria necessary for accurate diagnosis as well as for the selection of optimal acquisition parameters. A major limitation of this report is that it was designed in an era of analogue film/screen systems and provides no specific recommendations for digital radiography. The ACR guideline is adapted for digital systems but it does not provide instructions regarding the optimal method for selecting acquisition parameters. No guidelines exist to identify an optimal protocol for paediatric and adult CXR. Consequently, questions have arisen around the extent to which standard clinical protocols for undertaking paediatric and adult CXR examinations vary between and within hospitals, and the resultant impact on IQ and radiation dose. Specific to the United Kingdom (UK) is the lack of data about the likely differences in radiation dose and IQ for both paediatric and adult CXR, between and within hospitals.

This doctoral thesis will develop and validate a novel method for comparing radiation dose and IQ between and within hospitals, based on a Figure of Merit (FOM) concept. With this in mind, using a combination of commercially available phantoms with a focus on paediatric and adult CXR, this thesis seeks to produce novel data about potential differences in radiation dose, IQ and standard clinical imaging protocols for CXR in both paediatric and adult patients when imaged in several hospitals within the North-west of England. Low contrast detail (LCD) detectability, using the CDRAD 2.0 phantom, will be utilised as the main method for IQ evaluation. Prior to its use, an experiment was undertaken to determine the validity of utilising LCD detectability and the CDRAD 2.0 phantom for visual IQ and simulated lesion visibility (LV) evaluations. Its further validity for use in optimisation studies of CXR was also tested in order to help achieve the aims of this thesis.

1.2 Research Question

To what extent do standard protocols for undertaking paediatric and adult CXR examinations vary, between and within hospitals, and what is the resultant impact on IQ and radiation dose?

1.3 Aims and Objectives

The aim of thesis is to determine the variability of imaging protocols and X-ray imaging systems, between and within hospitals, when undertaking paediatric and adult CXR. In doing so, this study will develop and validate a method to assess the impact of any variability on IQ and radiation dose. Also, this thesis will determine the impact of patient thickness on both IQ and radiation dose.

To achieve the above aims, the following objectives were formulated:

Objectives:

1. To develop and validate a method for comparing IQ and radiation dose in phantoms across a range of paediatric and adult CXR protocols at different hospitals.
2. To evaluate the variability in standard imaging protocols, IQ and radiation dose in phantoms for paediatric and adult CXR between/within hospitals.
3. To investigate the influence of patient thickness on IQ and radiation dose between and within hospitals.
4. To determine if a correlation exists between the LCD detectability using CDRAD 2.0 phantom and visual measures of IQ and LV.

1.4 Study novelty

The main novel contributions of this PhD thesis are listed below:

1. A novel method for comparing radiation dose and IQ between X-ray units based on the use of an FOM concept has been established. This new method is also likely to be beneficial for the optimisation of X-ray units, between and within hospitals. Establishing this novel method was necessary since it was observed that there is no standardised method which considers both IQ and dose, between and within hospitals. The existing methods either depends on Diagnostic Reference Levels (DRLs), which could be considered a self-limiting approach because they only take radiation dose and not IQ into account, or depended on separate IQ and dose

evaluations. All these have no common link, which makes the comparison between hospitals difficult and complex.

2. Use of a novel FOM formula {FOM= $(IQF_{inv})^2$ /IAK} which can be used with contrast detail (CD) phantoms and utilises the LCD detectability as a metric for IQ evaluation. This novel formula is useful because existing FOM formulas depend on SNR and CNR as metrics for IQ evaluation. These are considered a limitation since they do not take into account the object size during IQ evaluation. Also, they are not a standardised method for IQ evaluations because of the variations in selecting sizes and locations for regions of interest (ROIs).
3. The provision of new information about how routine standard size paediatric and adult CXR protocols result in variation in IQ and radiation dose between different X-ray machines, between and within hospitals.
4. The provision of new information about how routine large size adult CXR protocols result in variation in IQ and radiation dose between and within hospitals.
5. Information on the relationship between body habitus, IQ and radiation dose when undertaking adult CXR examinations using routine clinical protocols.
6. Information on the validity of utilising LCD detectability, using a CDRAD 2.0 phantom, for evaluating visual IQ and a visual LV, and CXR optimisation studies.

1.5 Location of the thesis

The theme of this PhD thesis sits within the discipline of imaging in medical physics. It is especially focused on the performance of imaging technologies and protocols and their influence on image quality. This work has value to professionals involved in medical imaging using X-ray technologies including radiologists, radiographers, medical physicists and engineers.

1.6 Thesis Outline

This PhD thesis includes six chapters, and they are summarised diagrammatically in **Figure (1-1)**.

Chapter 1 includes a brief introduction on the thesis topic, research problem, aims and objectives for the study, proposed contributions of the study and finally the thesis outline.

Chapter 2 introduces background information about digital X-ray imaging systems, including their characteristics, physical principles, advantages and disadvantages. Following this, IQ in digital radiography is presented in detail and this includes: 1) definitions of IQ, an explanation of the key components of IQ and the factors which influence them. 2) IQ assessment methods and their respective advantages and disadvantages. 3) background information on the concept of signal detection theory and the Rose model with their applications in IQ evaluation. 4) explanation of the LCD detectability method for IQ evaluation and test objects which can be used, and their respective advantages and disadvantages.

Chapter 3 provides an introduction to chest radiography and chest characteristics. A critical literature review of the previous studies comparing IQ and radiation dose for CXRs among hospitals is also presented. Following this, a definition for the FOM concept, a description of existing formulas and a discussion of using FOM for evaluating and comparing radiation dose and IQ are all presented. Previous studies that attempted to investigate the correlation between the LCD detectability using a CDRAD 2.0 phantom with visual IQ methods in an anthropomorphic phantoms or clinical images are introduced and analysed critically. Finally, the last section of this chapter focuses on the gap in the literature which this thesis seeks to address.

Chapter 4 explains the methods used for achieving the aims this PhD thesis. It focuses on the methods used for assessing the variation in IQ and radiation dose between X-ray units. This includes: 1) the participating X-ray units; 2) quality control (QC) tests; 3) phantoms used for IQ evaluations; 4) radiation dose measurements; 5) procedure for image acquisition; 6) IQ evaluation; 7) FOM calculation. Next, the method used for the validation of LCD detectability using a CDRAD 2.0 phantom in optimisation studies and its validity in visual IQ and LV is presented. Finally, the statistical analysis used for analysing the obtained data is reported.

Chapter 5 commences with the results of the validation of LCD detectability using a CDRAD 2.0 phantom in optimisation studies. Next, data on the assessment IQ, radiation dose, FOM and standard clinical protocols between X-ray units is presented under different subsections based on the age group. Finally, the last section reports the results of the influence patient thickness has on IQ and radiation dose.

Chapter 6 is organised in the same manner as the results section and as such is divided into different sections and subsections. The data presented in chapter 5 (the Results chapter), which includes the results of the validation LCD detectability experiment, are discussed and compared with the literature. Next, the suggested 'new' novel method for comparing IQ and radiation dose between X-ray units based on FOM is discussed. Following this, the obtained results of the variation in IQ and radiation dose between and within hospitals is discussed. Following that, the results obtained from the influence phantom thickness on IQ and radiation dose is discussed. Finally, this chapter summaries the key conclusions of this thesis in a concise manner and presents the limitations of the research along with planned future works.

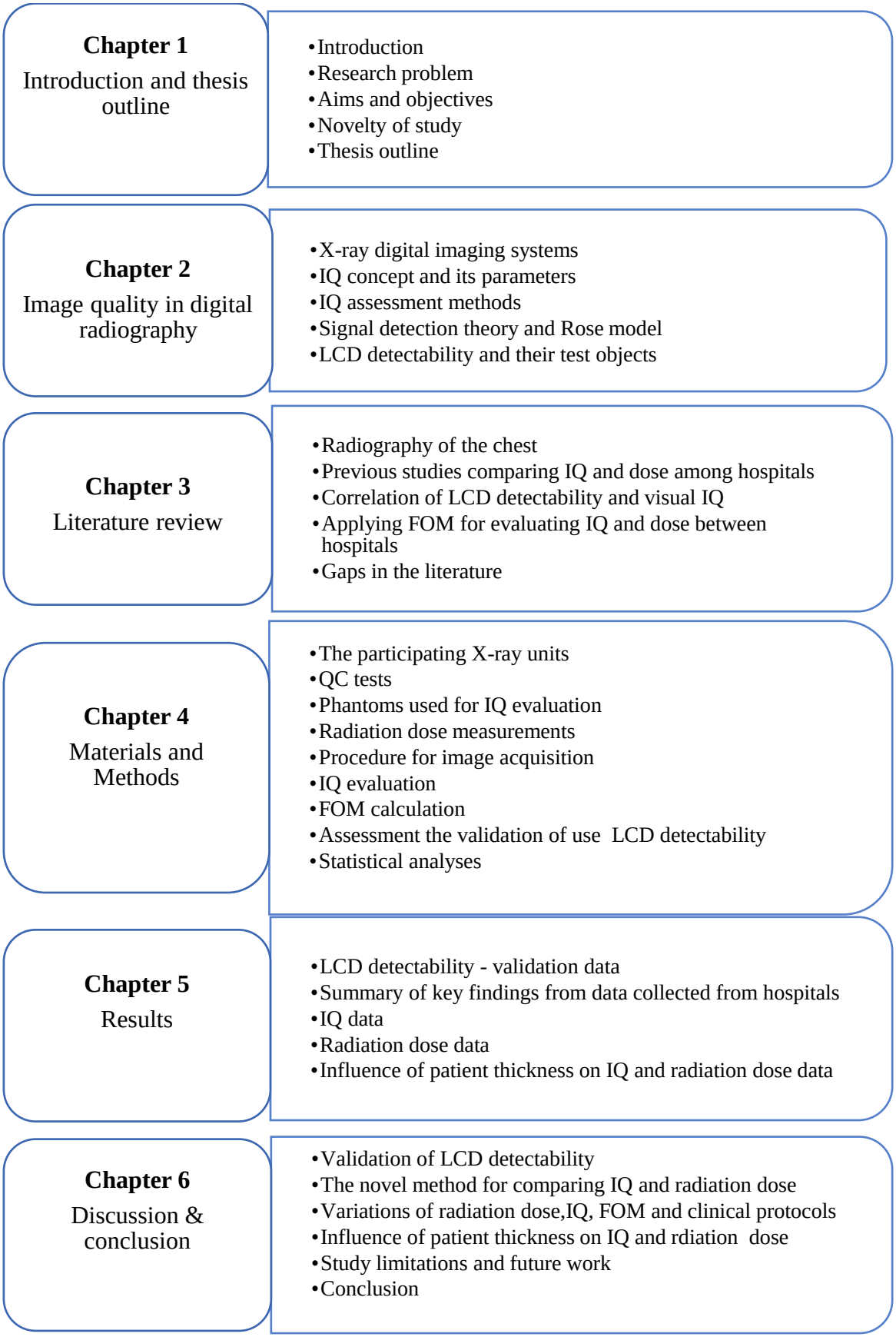

Figure 1-1 Schematic diagram illustrating the outline of this PhD thesis.

Chapter 2: Image Quality in Digital Radiography

2.1 Overview of Chapter

IQ in digital radiography is the major focus of this chapter. The first main section of this chapter describes the various digital imaging systems used in radiography and their impact on IQ and patient dose. This includes a basic conception of the physics principles, technical characteristics, advantages and disadvantages of the imaging systems. The next section focuses on the concept of IQ and its parameters. It will start with a detailed discussion on the concept and definition of medical IQ in diagnostic radiology and an explanation of why it is difficult to define an optimal or perfect image. This will then be linked to the idea of the dose optimisation of protocols and techniques based on the clinical indications for the acquired image. This section will also discuss the different parameters of IQ (e.g. noise, spatial resolution, contrast and artefacts). This includes the definition of these parameters, the factors that influence them and their impact on the final image. The next section in this chapter presents the different types of options utilised for IQ evaluation, e.g. observer performance and physical methods. This includes a definition of each method, their application and their advantages and disadvantages. In terms of physical methods, a particular emphasis will be placed on detective quantum efficiency (DQE), SNR, CNR and CI. Regarding the observers' performance, the general concept of visual IQ assessment will be described, and both the visual evaluation of patients/anthropomorphic phantom images and the visual detectability of physical measures using test objects (psychophysical method) will be explained in detail. The receiver operating characteristics analysis (ROC) and VGA methods will be examined in detail as examples of visual IQ evaluations of patients/anthropomorphic phantom images. Following this, signal detection theory and the Rose model will be discussed. This will also include an explanation of the concept, application and their limitations. Finally, this chapter will focus on the LCD detectability method and the required test objects. This includes an explanation of the theoretical framework for this theory, usefulness, applications and limitations. Also, this section will provide an overview of the most common CD phantoms available (i.e. Leeds TO.10 phantom, CDMAM phantom and CDRAD 2.0 phantom) and their principle work, applications and advantages and disadvantages. Furthermore, the physical and visual methods utilised with these CD phantoms will be discussed and their benefits and drawbacks will be highlighted.

2.2 X-ray Digital Imaging Systems

This section will present an abbreviated characterisation of the principle work, physics, features, benefits and drawbacks of the different types of X-ray digital imaging systems: computed radiography (CR), indirect digital radiography (IDR) and direct digital radiography (DDR). This helps to understand the influence of these different types of imaging systems on performance and the resultant IQ, particularly on the detectability of LCD objects as well as their impact on the patient dose.

2.2.1 Computed Radiography (CR)

The CR system is constructed from storage phosphor plates surrounded by a cassette. The storage phosphor, such as barium fluorohalide activated with divalent europium ions or powder-based materials (BaFBr:Eu), is used for detecting and storing X-ray photons. The CR system generates images during an indirect conversion process that uses storage phosphor plates linked with a separate image readout process, and this is applied in two stages (**Figure (2-1)**). First, the X-ray photons are absorbed by storage phosphor plates. Second, a laser digitiser reads the exposed phosphor plates, the light emitted is acquired by photo detectors and the resultant luminescence is converted into a digital image (Lança & Silva, 2009a; Samei et al., 2004). The introductions of needle-crystalline CR detectors and the dual reading of CR plates are considered to be a new development in CR technology. Introducing the new detectors (needle-crystalline detectors) leads to an increase in quantum efficiency and improvements in low contrast resolution with the further possibility of reducing radiation dose (Cowen, Davies, & Kengyelics, 2007; Schaefer-Prokop, Neitzel, Venema, Uffmann, & Prokop, 2008). On the other hand, the dual reading CR system has also contributed to improvements in quantum efficiency which helps to reduce patient radiation dose. These systems rely on diaphanous detector materials, double-sided storage phosphors and light collection optics on the sides, front and back of detectors (Cowen et al., 2007; Schaefer-Prokop et al., 2008).

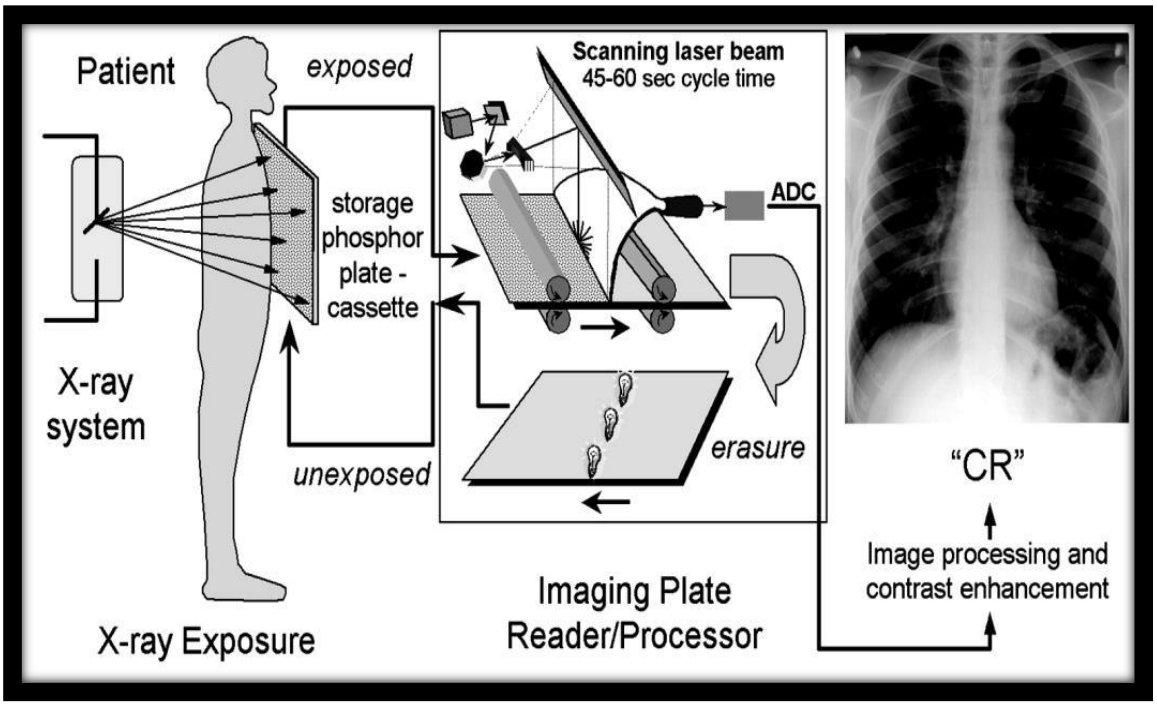

Figure 2-1 The process of a CR imaging system: exposure of a storage phosphor plate, readout of the plate and erasure of the plate (Samei et al., 2004).

2.2.2 Digital Radiography (DR)

In contrast to the CR system which utilises a phosphor storage plate, the digital radiography (DR) system uses thin film transistor (TFT) arrays for detecting and processing X-ray photons attenuated from the object to apply a direct readout process (**Figure (2-2)**). DR systems are divided into 2 types based on the process used to convert X-ray photons into electric signals.1) Indirect conversion, which uses a light-sensitive TFT photodiode and a scintillator layer. 2) Direct conversion, which uses a TFT charge collector and an X-ray-sensitive photoconductor layer (**Figure (2-3)**) (Kotter & Langer, 2002; Lanca & Silva, 2012).

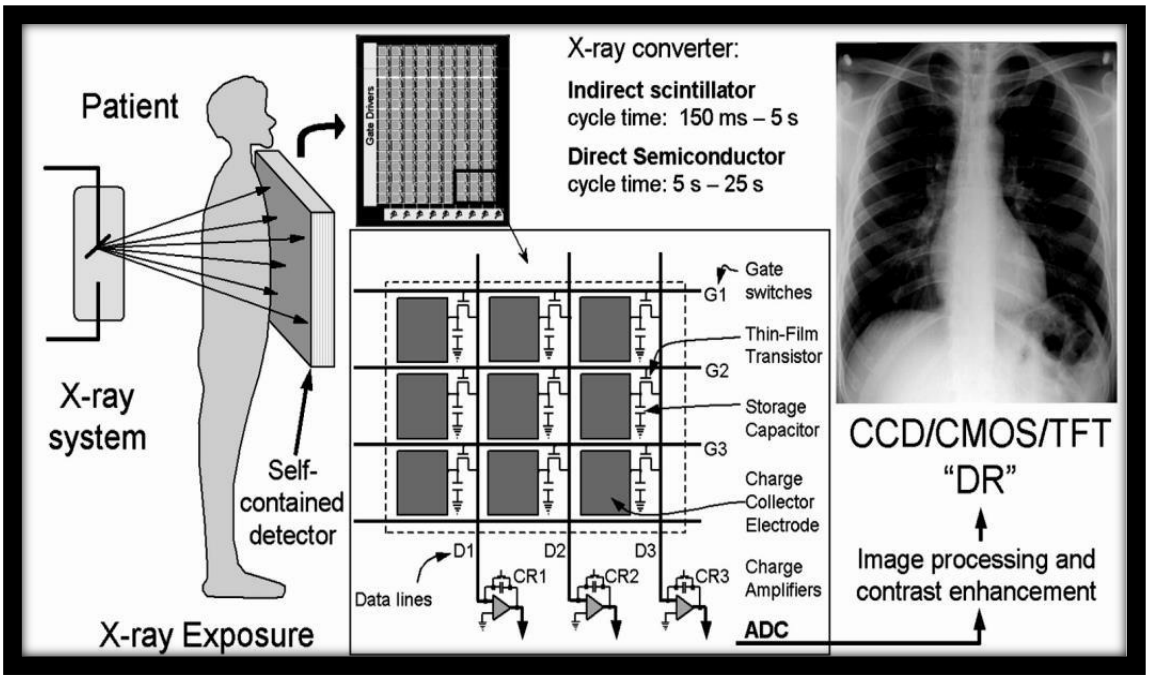

Figure 2-2 The process for direct and indirect DR imaging systems (Ehsan Samei et al., 2004).

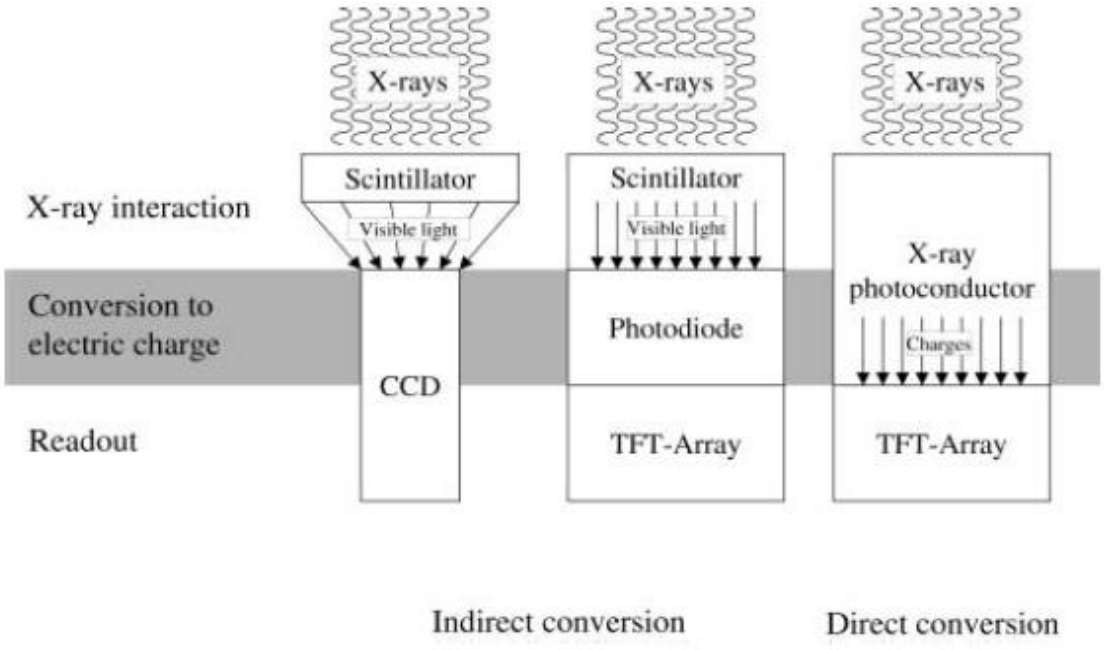

Figure 2-3 Comparison of work principles between direct and indirect DR imaging system (Kotter & Langer, 2002).

2.2.2.1 Indirect Digital Radiography (IDR)

IDR utilises gadolinium oxysulphide, or cesium iodide (CsI), as an X-ray detector. The phosphor and scintillation are utilised in the indirect transformation system and is divided into 2 types, structured and unstructured (**Figure (2-4)**). The phosphor material in the structured scintillator is a needle-like structure that is vertical to the screen surface. This leads to an increase in the number of X-ray photon interactions and a reduction in the lateral scattering of light photons (Lanca & Silva, 2012). In contrast, high levels of light scatter result from the unstructured scintillators and this leads to a decrease in spatial resolution. In IDR, the X-ray photons are absorbed and captured in a scintillation layer which are then converted into light photons by fluorescent material. Following this, the amorphous silicon (a-Si) photodiode array converts the fluorescent light to an electric signal (**Figure (2-3)**) (Lanca & Silva, 2012).

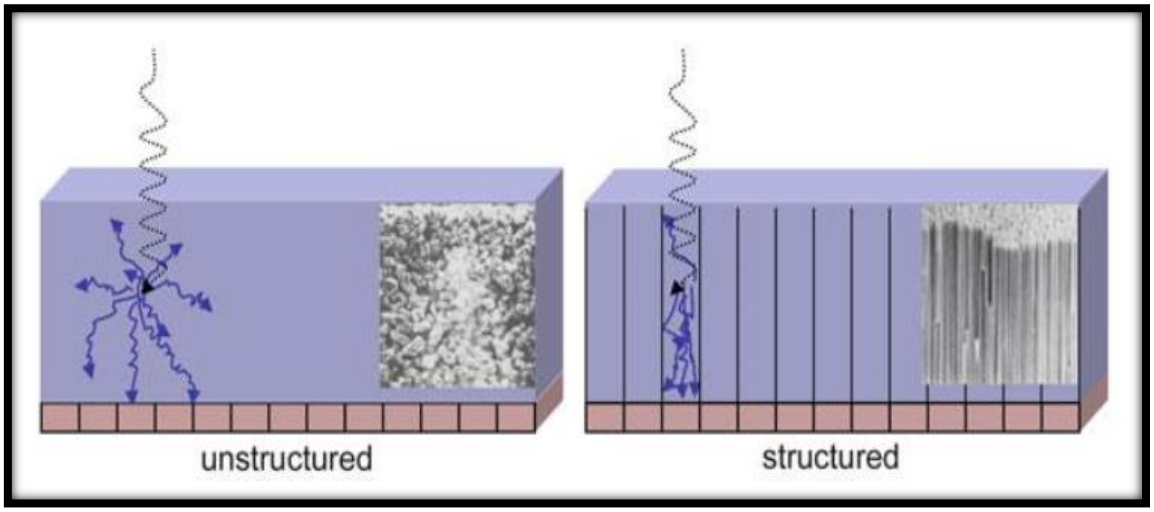

Figure 2-4 Unstructured scintillation and structured scintillation, respectively (Lanca & Silva, 2012).

2.2.2.2 Direct Digital Radiography (DDR)

DDR utilises X-ray photoconductor materials, such as amorphous selenium (a-Se) alloyed with re-crystallised arsenic, since a-Se displays high spatial resolution and characteristic X-ray absorption (Kotter & Langer, 2002 ;Seibert, 2009). In the DDR detector, there is no middle stage of image acquisition in which X-ray photons are captured directly and transformed into an electrical signal. An electric field is applied across the a-Se layer prior to exposing X-ray photons. As a result of the electric field, the absorbed X-ray photons in the a-Se layer are converted to electric signals and then drawn directly to the signal-collecting electrode (Lanca & Silva, 2012). The signals are then stored at the TFT signal collector and then they are read out (**Figure (2-5)**. Finally, the signals are quantified and amplified into a digital code value representing the corresponding pixels (Lanca & Silva, 2012). As the DDR detectors do not require any light-to-signal transformation, they are possibly less susceptible to transformation noise in comparison with IDR detectors (Samei, 2003).

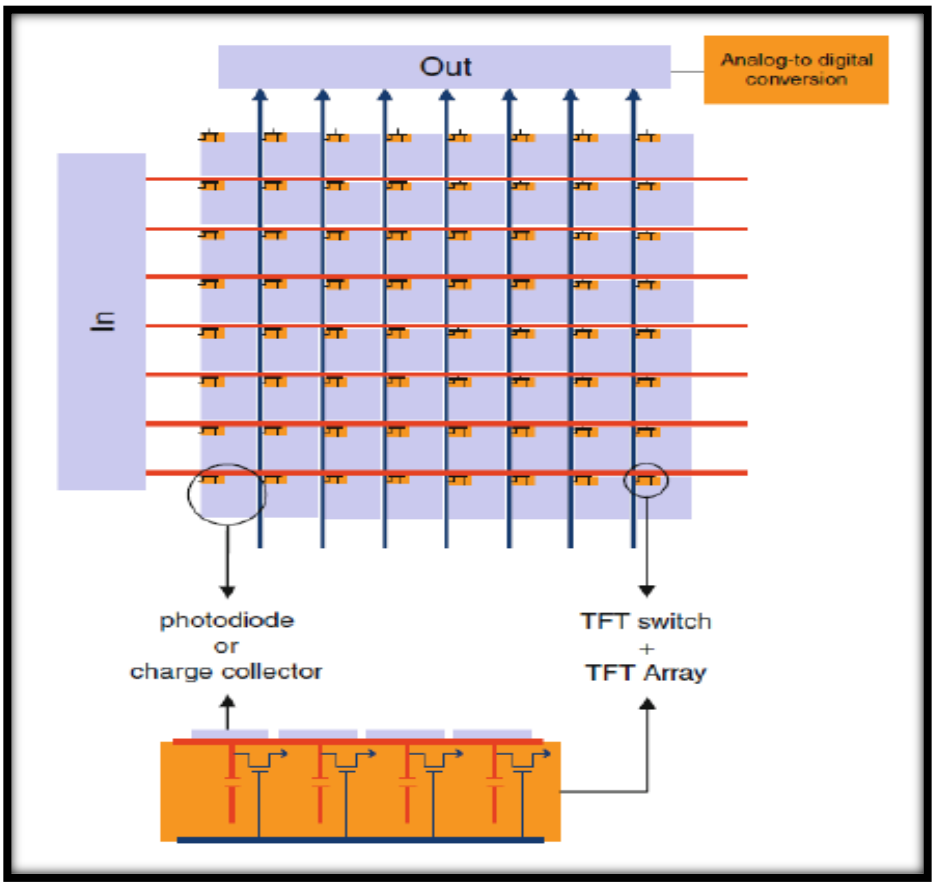

Figure 2-5 TFT array (Lanca & Silva, 2012).

2.2.3 Comparison between systems

Comparing between DR systems, CR systems have several benefits such as portability and flexibility. This is especially useful in the operating theatre, the neonatal imaging zone and the trauma zone (Alghamdi, 2016). The CR system also has a lower cost compared with DR, providing access to the lower end of the market. In contrast, DR systems offer better performance than CR (Bacher et al., 2003; Hamer, Völk, Zorger, Feuerbach, & Strotzer, 2003) and this is related to the high DQE for DR systems compared with that of CR (Bertolini et al., 2012; Schaefer-Prokop et al., 2008). Also, the delay in image reporting because of the process of cassette handling and scanning, which can take a lot of time, is considered to be another disadvantage of CR systems (Cowen et al., 2007). The DR system can provide images with IQ comparable to that of CR systems but with a lower radiation dose (Hamer et al., 2003; Jansson, Geijer, Persliden, & Andersson, 2006; McEntee, Frawley, & Brennan, 2007).

2.3 Image Quality Concept and its parameters

IQ is a broad term in which it is difficult to define in ways that can be readily integrated into specified goals for improving medical imaging or that set metrics through which quality can be measured, compared and evaluated. IQ's meaning is likely to differ from person-to-person and there is no specific or widely accepted definition for the expression of IQ (Shet, Chen, & Siegel, 2011; Singh & Pradhan, 2015). In medical imaging, there is no viable subjective or objective definition that can identify a typical or a perfect image. The reason behind this is that the medical images are acquired for different clinical indications and this tends to focus concentration on various features within a given image. Consequently, an image that is perfect or acceptable for one specific purpose may be not acceptable for another purpose. This causes enormous variation in the evaluation of acceptable IQ and results in difficulties in determining optimum imaging protocols and the optimum radiation exposure. For instance, the optimum imaging protocol and its resultant radiation dose to determine the position of a nasogastric tube could be extremely different from that needed for the detection of a subtle lung lesion, pneumothorax or a rib fracture in the chest (Shet et al., 2011). The CEC has referred to this concept within their reports on IQ: "the aim of the Quality Criteria is to characterize a level of acceptability of normal basic radiographs which could address any clinical indication," (Commission of the European Communities (CEC), 1996b), these "Quality Criteria cannot be applied in all cases" (Commission of the European Communities (CEC), 1996a). "For certain clinical

indications a lower level of IQ may be acceptable, but this should ideally always be associated with a lower radiation dose to the patient" (Shet et al., 2011). To summarise, the IQ is determined by the observer's ability to utilise the image for a specific diagnostic problem (Burgess, 1995). With this in mind, the general definition of IQ could be 'a measured attempt to demonstrate how well an image should demonstrate physiology and/or anatomy and any alteration that could happen to the anatomical structure as result of an abnormality' (Bourne & Kagadis, 2010). On the other hand, the utility of radiologic images and the precision of diagnosis rely on two prime factors: the quality of radiologic images and the performance of the observers. Images with a good quality can improve task-related performance. However, this is not sufficient in itself to get a precise and correct diagnosis from the observer (Barrett, Myers, & Rathee, 2004; M nsson, 2000; M J Tapiovaara, 2008). For instance, a missed lesion might be related to the observer's incorrect decision instead of the limited detectability. An image with moderate or low IQ could be acceptable by the observer as sufficient for the clinical task, while an image with good IQ could require technical modifications (Kundel, 1979). From reading the literature, it appears that there are three factors that can lead to errors in the detection task: search or scanning errors, recognition errors, and decision making errors. The first one occurs as a result of a failure of the observer to fixate on the lesion, the second one occurs because the fixated lesion is not detected as a lesion, and the third one occurs due to incorrect interpretation of the detected lesion (Kundel, Nodine, & Carmody, 1978). Decision making was found to be the most significant factor that contributes to lesion detection errors (Kundel et al., 1978). A study was conducted by Manning, Ethell, & Donovan (2004) who investigated the errors concerning missed lung nodules on CXRs and concluded that these were a result of decision making instead of a failure in the detection the nodules. The reason behind this is that the anatomical structures surrounding the abnormalities are extremely complex. This causes difficulties in determining the abnormality from the normal anatomical structures. This might suggest that the essential problem is related to perception rather than the quality of the diagnostic image (Szczepura, Tomkinson, & Manning, 2017). In addition, another perceptual error can result from a 'satisfaction of search error', in which the attention of the observers is distracted from the lesion by a more prominent finding (Berbaum, Dorfman, Franken, & Caldwell, 2000; Berbaum, Franken, Dorfman, Caldwell, & Krupinski, 2000). By contrast, it was stated by Pinto (2010) that although there are numerous reasons for errors in the detection of lung lesions, the diagnostic quality of an image is one that must not be overlooked.

IQ is not specified by a single factor, but it consists of a combination of factors such as noise, contrast, resolution and existence or absence of any distortion or artefacts (Jones, 2006). **Figure (2-6)** demonstrates the influence of different IQ parameters on the appearance of the resultant image.

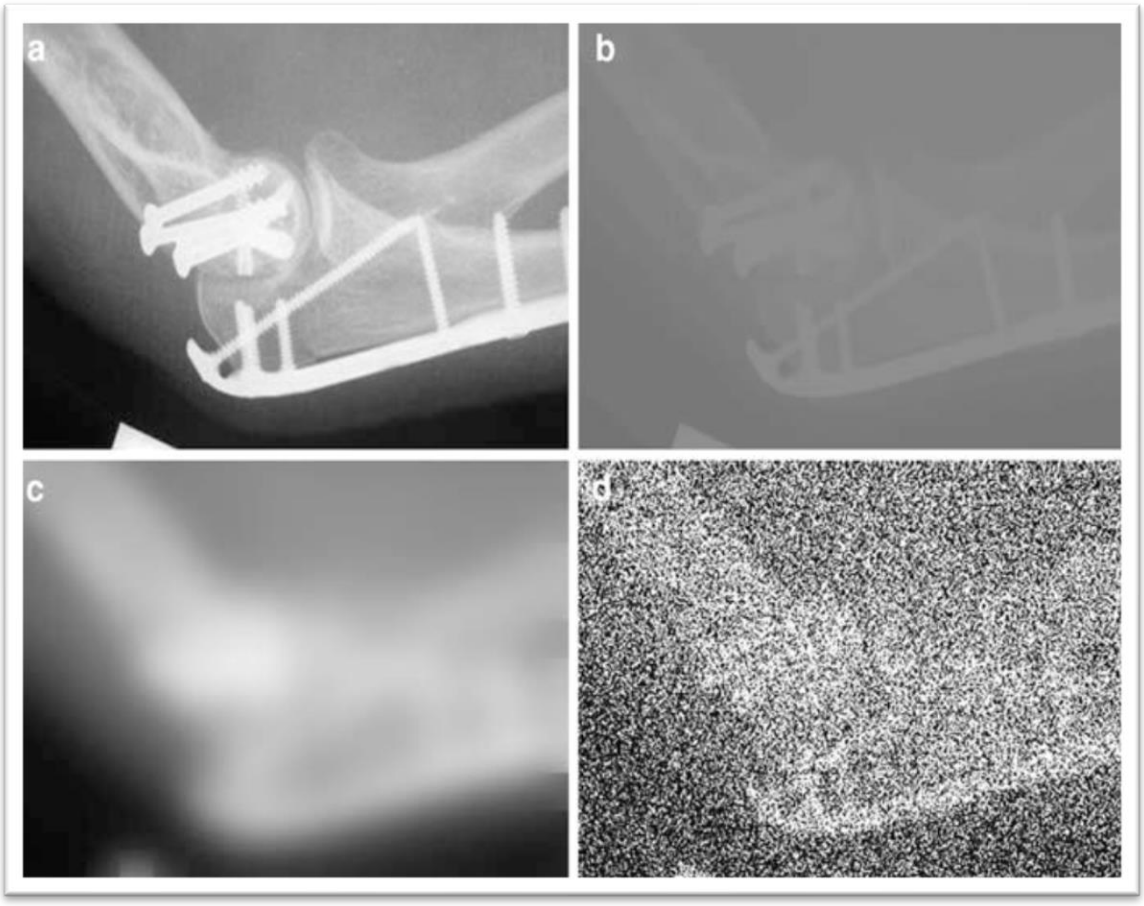

Figure 2-6 Demonstrates the influence of each of the different parameters of IQ on the resultant IQ. a) Image with suitable contrast, resolution, and with low noise. b) Image with high spatial resolution and low noise, but is useless since it has low contrast. c) I Image with high contrast, low levels of noise, but with very poor spatial resolution. d) Image with high spatial resolution, but with extremely high levels of noise which limits the contrast information (Bourne, 2010).

2.3.1 Noise

In radiography, noise is defined as uncertainty or imprecision in the recording of an image. In other words, it represents the unwanted stochastic fluctuations in the image. It is considered the most influential factor on IQ since it can decrease the visibility of some structures- particularly for low contrast objects (Berbaum, Franken, et al., 2000; Samei, Eyler, & Baron, 2000). Its negative influence on IQ is a result of reducing the visibility of the anatomical structures and abnormal areas (e.g. pathology), particularly for the low contrast detail. This has the potential to impact negatively on diagnosis and interpretation. There are 2 main types of noise: anatomical noise and radiographic noise (Samei, Flynn, Peterson, & Eyler, 2003). The latter, generally, is divided into 2 types: quantum noise and system noise (e.g., electronic/detector noise and computational/quantisation noise) (Aichinger, Dierker, Joite-Barfuß, & Säbel, 2012; Samei, & Flynn, 2003). Anatomical noise can be described as the effect of overlaying (superimposed) anatomical features on an image, for example ribs, lung vessels, heart, mediastinum, and diaphragm on a CXR image (McAdams, Samei, Dobbins, Tourassi, & Ravin, 2006). It was shown that anatomical noise can have a negative influence on observer performance for pathology detection; this is especially the case in CXR because the chest region is located in a quantum-saturated zone of the body in which anatomical structures are considered the main source of noise (Båth, Håkansson, Börjesson, Kheddache, Grahn, Bochud, et al., 2005; Håkansson, Båth, Börjesson, Kheddache, Allansdotter Johnsson, et al., 2005; Samei, Flynn, & Eyler, 1999). Radiographic noise is defined as undesirable differences within an image that do not originate within the imaged subject (Samei, & Flynn, 2003). It is characterised by the variance in pixel values of a uniform object in an image. The existence of radiographic noise gives the image a mottled and grainy appearance. Increasing the level of radiographic noise leads to a reduction in contrast resolution. In turn, this increases the difficulty of differentiating fine details and low-contrast objects (Alsleem, 2014).

Quantum noise appears as result of limited number of photons being received by the image detector. Quantum noise increases when the number of photons reaching the detector reduces and vice versa (Samei, & Flynn, 2003). Such a reduction could be due to the X-ray tube output being set too low and/or the object overly attenuating the X-ray beam. Furthermore, the quantum noise increases when increasing the distance of object from the image detector. To minimise quantum noise, the selection of suitable acquisition parameters is paramount, as is correct image post-processing (Williams et al., 2007).

System noise arises as a result of the statistical fluctuations inside the imaging system during image formation (e.g., detector noise and quantisation noise), and it has an influence on the final image noise (Aichinger et al., 2012).

Detector noise is generated due to the non-uniform response of the detector to the incident X-ray beam. This type of noise is occasionally called electronic noise and is related to inherent defects that are imposed during the formation of the internal structure of the image detector. Such defects sometimes create undesirable signal in the image. Prime reasons for detector noise, especially with DR systems, are the variance in linearity and sensitivity between pixels, dead pixels, and detector-response non-uniformities (Samei, & Flynn, 2003).

With respect to quantisation noise, this occurs during digitisation and includes the system incorrectly interpreting the analogue output voltage of the image detector and converting it to discrete pixel values (grayscale values) (Williams et al., 2007).

2.3.2 Spatial Resolution

Spatial resolution is the ability of an imaging system to distinguish between 2 adjacent objects within an image, and it is also known as image sharpness which is related to distinguishing the image's edges (Williams et al., 2007). It is usually evaluated by means of a test pattern composed of tight strips and reported in terms of the number of line pairs per millimetre (lp/mm) that can be distinguished on the image (Williams et al., 2007).

Spatial resolution can be influenced by image blurring which causes spatial resolution loss. This is further influenced by numerous factors, including the movement of a patient during image acquisition, detector element effective aperture size and geometric factors such as the focal spot size of the X-ray tube (Williams et al., 2007).

In CR systems, the scattered laser light emitted from the image receptor reading process leads to a reduced spatial resolution (Seibert, Boone, Cooper, & Lindfors, 2004; Williams et al., 2007). In DR systems there are 2 factors that can affect spatial resolution and detector blurring. First, there is the IDR system, which influences the spreading of light photons through the process of converting X-rays into light. This problem can be minimised by utilising structured converters (e.g. caesium iodide), which have narrow, parallel columnar structures. The incident X-ray photons travel along the long dimensions of these columns, leading to an improvement in spatial resolution through increased detection efficiency and reduced scattered light (Williams et al., 2007). However, DDR systems do not suffer from

this problem because the electrons inside the photoconductor material is directed towards the TFT array. This minimises the spread of electrons (Williams et al., 2007).

Second, within DR the size of the detector element has a negative influence on the spatial resolution of the system (better resolution with smaller detector element dimensions). The signal quantity or the charge read from the detector elements contribute to all of the absorbed X-ray energy inside an individual detector element during radiation exposure. As a result, image contrast is reduced as a result of smearing an object's structures within the image when the size of the detector element is larger than the structures of the object being imaged (Williams et al., 2007).

2.3.3 Contrast

Contrast (radiographic contrast) is a term used to indicate the magnitude of the signal difference between an object in the image and its surroundings (Williams et al., 2007). Contrast resolution can be influenced by four factors: 1) the object contrast, 2) the imaging procedure and techniques used, 3) the image detector contrast and 4) the display contrast (Hendee & Ritenour, 2002). Object contrast indicates the characteristics of the imaged object and this includes the anatomical and physiological characteristics. The second main factor that affects contrast resolution is the imaging procedure/technique used. Variables that can impact on contrast resolution include X-ray beam filtration, the inclusion of an anti-scatter radiation grid, tube voltage and tube current. For instance, using more filtration can decrease contrast resolution, while improvements can be sought through the use of an anti-scatter radiation grid (Hendee & Ritenour, 2002). Commonly, contrast can be improved by introducing a higher or lower density substance into the object. Such an example would include the use of contrast media in certain imaging procedures (Hendee & Ritenour, 2002). Detector contrast, which is specified by the physical characteristics of the image detector, can be considered another factor that can influence the contrast resolution of the medical image. The detector contrast can be demonstrated by how the image detector detects and converts the X-ray photon energy into a signal output. The dynamic range of the image detector, which represents the ratio of the maximum to minimum input X-ray intensities on the detector surface, is the factor that impacts on the contrast resolution of the image (Hendee & Ritenour, 2002). The dynamic range of CR and DR varies from 1,000:1 to 10,000:1 (Williams et al., 2007). DR systems have a higher potential for improving contrast detail when compared with CR systems.

2.3.4 Artefacts

Image artefacts can be defined as features that mimic pathology, hide pathology or generally degrade IQ (Drost, Reese, & Hornof, 2008). Several factors can result in image artefacts. One of these factors relates to the appropriate selection of acquisition factors (Shetty, Barthur, Kambadakone, Narayanan, & Kv, 2011). For instance, even though DR systems have detectors with wide dynamic ranges capable of displaying good IQ, over- or under-exposure during the imaging process can still lead to artefactual appearances on the resultant image (Drost et al., 2008). Image artefacts can also arise because of the object itself, for instance motion or incorrect patient position (Drost et al., 2008).

For DR systems, ghost artefacts can be a common occurrence (Bushong, 2013). For DR images, the image is generated after the photodiodes digitise emitted light from the scintillation layer. However, the released charge in the photodiode can persist beyond the readout phase, this causes a ghost artefact on the image (Drost et al., 2008).

With respect to CR systems, artefacts can result from inadequate deletion of the prior image. A further cause of ghost artefacts can arise from background radiation. Furthermore, faulty rollers in the digitiser of CR image detector can lead to defective scanning which can result in artefacts (Bushong, 2013). Artefacts from afunctional or faulty detector pixels are also possible (Shetty et al., 2011). Artefacts can also arise from image post-processing software. For example, the unsuitable use of filters, which are actually intended to improve IQ and/or enhance particular features within an image, can lead to artefacts being introduced into the image (Honey & Mackenzie, 2009).

2.3.5 Influence of Digital Image Processing on Image quality

The raw imaging data acquired after exposure needs to be processed for initial display using different techniques and operations which have a high impact on the final appearance of the displayed image. Considerable differences can be detected in the appearance of the images when using computer programmes produced by different manufacturers, although they use comparable algorithms such as edge enhancement, noise reduction, and contrast enhancement to alter the appearance of the image.(Körner et al., 2007).

The image processing is utilised for improving the image quality by decreasing the noise, artefacts and optimising the contrast for the optimal viewing. However, the image processing cannot be impact on the spatial resolution by utilising processing software since it depends on the technical variables such as the size of the pixel but the lack of spatial resolution can be partially overcome by optimising the other processing variables. In

contrast, the image processing should be optimised for each digital system and anatomical examination area independently. For instance, the type of the processing algorithms required for lateral chest radiography is different from that for the posteroanterior projection (Körner et al., 2007). An example of the influence of the image processing tool on the final appearance of the viewed image illustrated in **Figure (2-7)**.

On the other hand, the choice and characteristics of the image display monitor (monitor resolution, luminance, type [monochrome or colour; liquid crystal display or cathode ray]) and the image viewing conditions (with and without using post-processing such as windowing and magnification) have an influence on final displayed image. For instance, the contrast of the images can be adjusted during display and this can improve the contrast resolution of the image (Hendee & Ritenour, 2002). Image post-processing is often utilised by the radiologists during image quality evaluation for diagnostic interpretation because these techniques permit for manipulation the grayscale of the image to enhance diagnostic interpretation (Seeram, 2019). Several studies demonstrate that utilising windowing and magnification during image quality evaluation can have a significant influence on improve the observer performance for the LCD detectability (Bacher et al., 2006; Nishimura et al., 2010).

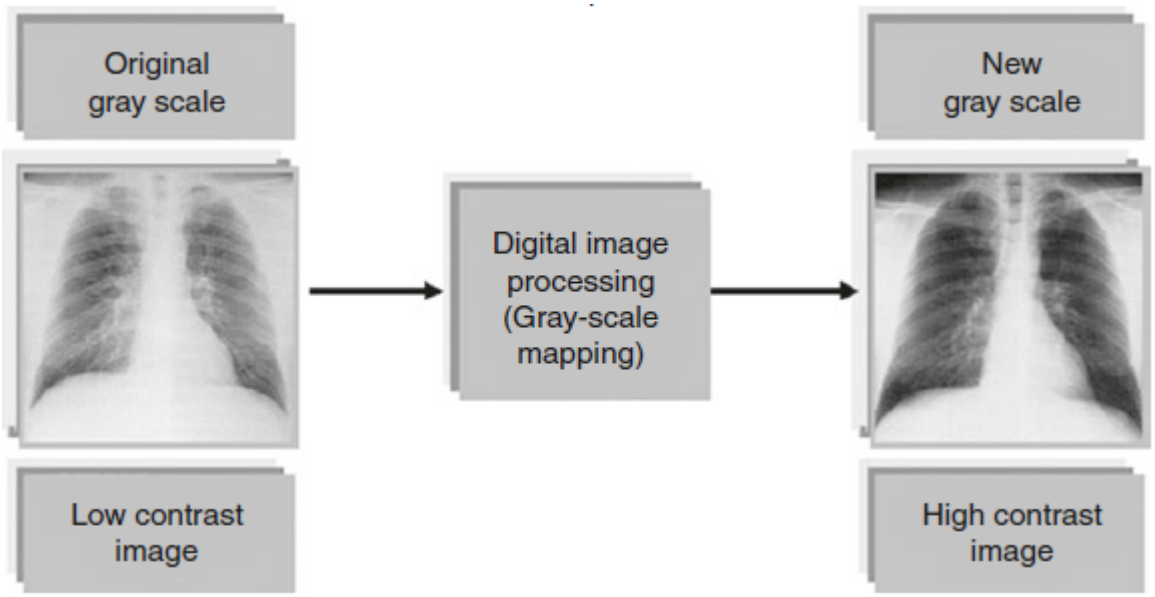

Figure 2-7 This figure demonstrates the impact of using image processing tool 'greyscale mapping' on the final appearance of the image (Seeram, 2019).

2.4 Image Quality Assessment Methods

To achieve optimisation of an imaging technique it is necessary to determine and measure the quality of the resultant images, and to evaluate whether they are fit for diagnostic purposes.

There are several methods in medical imaging through which IQ and the performance of imaging systems can be evaluated. Some of these approaches examine one aspect of IQ whilst some of them examine several aspects of IQ. In general, there are two types of IQ evaluation methods: physical and observer performance. Physical methods such as SNR, CNR and DQE are commonly used for the quality control of imaging systems since they have high reliability (Krupinski, 2010). Nevertheless, these methods are limited in only considering only one or two aspects of IQ. The second type of method involves the use of human observers in visually analysing images. These methods have high validity and are closer to the actual clinical situation, however have a lower reliability due to inter- and intra-observer variability (Vennart, 1997). **Figure (2-8)** illustrates the types of IQ evaluation methods available.

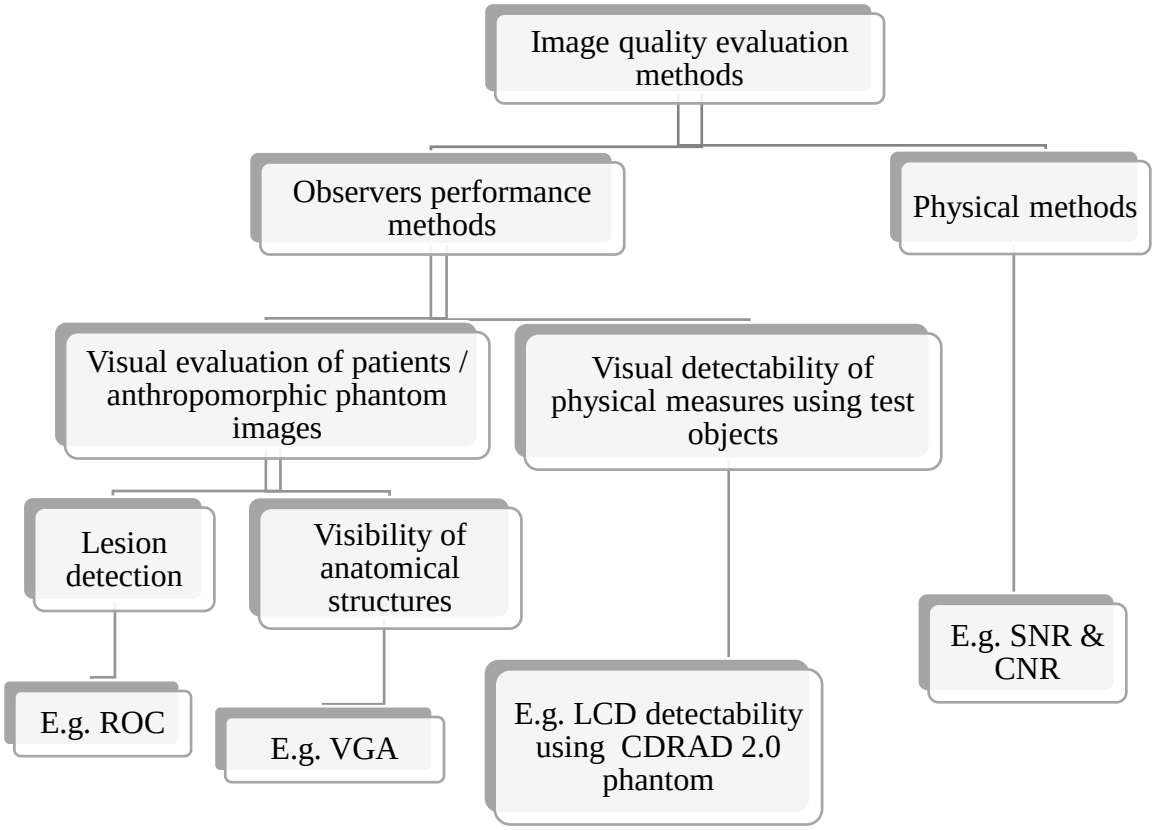

Figure 2-8 Types of IQ evaluation methods.

2.4.1 Physical Methods

Physical methods of IQ assessment are designed for assessing the 'total' X-ray imaging system performance and also for evaluating the performance of individual components. These methods form the basis of acceptance testing prior to commissioning a new piece of equipment in clinical practice. They also form the basis on which decisions are made for assessing equipment performance over time (Vennart, 1997). Such methods have the benefit of being repeatable means of evaluating IQ and, if done in a consistent fashion, will yield results that are not subject to human variations as is the case in visual IQ measures (Morrell, 2006). Physical methods permit the performance of an imaging system to be characterised by measuring specific physical parameters and compiling them according to the demand of a specific imaging task. Parameters such as SNR, CNR, modulation transfer function (MTF), noise power spectrum (NPS) and DQE are also physical measures of IQ and are routinely used as quality assurance measurements to ensure that the performance

of an imaging system is both accurate and consistent (Vennart, 1997). The physical IQ measures have also been widely utilised in the literature for evaluating IQ and in an optimisation studies (Ekpo, Hoban, & McEntee, 2014; Samei, Dobbins, Lo, & Tornai, 2005; Smans, Struelens, Smet, Bosmans, & Vanhavere, 2010). The remainder of this section will focus on the most common physical parameters used within medical imaging.

2.4.1.1 Detective Quantum Efficiency (DQE)

DQE assesses the overall ability of the imaging system to utilise the information of the incident X-ray photons to form the radiographic image (Månsson, 2000; Davidson, Bushong, Seeram, & Swan, 2014). DQE is valuable when comparing the performance of several imaging systems- particularly in the evaluation of new image detectors. It considers both MTF and NPS. MTF indicates an imaging system's capability to reproduce and maintain the information of frequency included in the incident X-ray signal. NPS indicates the frequency content of the noise of an imaging system (Bath, 2010; Tsai, Lee, & Matsuyama, 2008).

DQE values range from 0 to 1, where '1' suggests the detector has perfect performance which means that the information content is exactly identical to that of the X-ray beam (Miracle & Mukherji, 2009). DQE is considered to be an extremely useful physical measure for evaluating detectors, however, it does have several drawbacks. It provides only a small amount of information about the final appearance of the resultant image due to it being a descriptor of a single component in the imaging chain (i.e. the detector). It does not therefore consider all the factors that can influence IQ such as the level of the radiation dose, the image processing and the image display characteristics (Bath, 2010).

2.4.1.2 Signal to Noise Ratio (SNR)

SNR demonstrates the relationship between the contrast of an object in an image and the general level of noise. It is considered to be a simplistic approach and is often used for characterising the visibility of an object in the image (Lança & Silva, 2009b). SNR is determined by the ratio of the mean signal of the object to the standard deviation (SD) of the background pixels (Smans et al., 2010). SNR is considered to be one of the most frequent methods used in the literature to evaluate IQ for digital imaging systems and has been described in numerous optimisation studies as a major factor in determining IQ (Ekpo et al., 2014; Mraity, Hogg, & England, 2015; Smans et al., 2010). SNR has many advantages including its high reliability and its ease to implement. However, utilising SNR

calculations for evaluating IQ has been criticised in the literature since it does not assess the observation stage, therefore potentially decreasing the validity of the method.
Another limitation of SNR is that it does not consider the size and shape of the object (Bath, 2010) Such factors have been found to have an influence on the detectability of objects by observers (McAdams et al., 2006). Furthermore, the characterisation of SNR is extremely simple when considering that human observers are sensitive to textual noise. To clarify this, SNR is dependent on quantum noise, which is influenced by the amount of the X-ray photons that reach the detector. Observers are familiar with the background texture of the medical images which can be influenced by other sources of noise (Bath, 2010); Several studies reported that anatomical noise is more clinically important compared with image detector or system noise (Samei et al., 1999). In addition, to obtain similar SNR values, a higher exposure (larger number of X-ray photons) is required for an image detector with smaller pixels when compared with a detector with larger size pixels. However, observers are typically not interested with the values of the single pixel; instead they integrate information over a region in the radiographic image and are not affected by the pixel-to-pixel alterations (Bath, 2010).

2.4.1.3 Contrast to Noise Ratio (CNR)

CNR can be considered a simple measure of IQ when seeking to detect objects (Moore, Wood, Beavis, & Saunderson, 2013). In comparison with SNR, a CNR approach would provide, for certain conditions, valuable information regarding the influence of noise on contrast detection. This is related to the fact that a radiographic image that has a high SNR value does not necessarily have a good IQ unless it has a high value of CNR. However, this requires sufficient variation between the area of interest and the surrounding tissue (Lyra, Kordolaimi, & Salvara, 2010). Practically, CNR is calculated from the pixel value variation between the test object (the signal inside the region of interest) and its neighbour (background signal). This is then divided by the pixel value SD of the background which represents the noise (Bosmans et al., 2006; Desai, Singh, & Valentino, 2010; Tang et al., 2012). CNR is similar to the concept of the signal-difference-to-noise ratio (SdNR) (Samei et al., 2005); this kind of measure is particularly beneficial for consistency testing and might also be useful for the optimisation of IQ and radiation dose for X-ray imaging systems (Moore et al., 2013). Similar to SNR, CNR suffers from several limitations, including: 1) it is considered that quantum noise is the limiting noise source and CNR does not take into account the influence of anatomical noise; and 2) it does not consider the size

and shape of the object during IQ evaluations (Bath, 2010). This decreases the validity of utilising CNR as an indicator for IQ performance (Vaishnav, Jung, Popescu, Zeng, & Myers, 2014).

2.4.1.4 Conspicuity Index (CI) Measurements

Lesion conspicuity can be defined as the visibility of the lesion against its surroundings (Revesz, G., Kundel, H.L. and Graber, 1974). Several factors can affect the lesion conspicuity, including: 1) the physical properties of the lesion structure such as the contrast, size, and edge sharpness; 2) the nature of the anatomical tissue that surrounds the lesion and its associated complexity, such as the ribs and blood vessels; 3) the image detector characteristics; 4) the acquisition factors (Kundel, H. L., & Revesz, 1976; Neitzel, Pralow, Schaefer-prokopb, & Prokopb, 1998; Szczepura & Manning, 2016). The concept of lesion conspicuity can be useful to determine the sources of the observation error for radiologists (Kundel, H. L., & Revesz, 1976). In addition, it provides a potentially useful way to investigate the effects of acquisition factors on IQ. The measurement of lesion conspicuity can be divided into two types: physical and visual measurements. In terms of the physical measurements, the first attempt was undertaken by Kundel, H. L., & Revesz (1976) and it was measured using the following equation:-

$$Lesion\ conspicuity = \frac{lesion\ contrast}{surrounding\ complexity}$$

The above equation can be considered crude because it does not consider other factors that affect lesion conspicuity such as the size of the lesion and lesion edge sharpness (Manning et al., 2004; Szczepura & Manning, 2016). The first equation for lesion conspicuity measurement, which successfully considers all of the relevant factors, was reported by Manning et al. (2004):

$$CI = \frac{dtan(\theta - 1)\Delta GL}{\sqrt{\sigma_s^2 + \sigma_n^2}}$$

Where:

CI = conspicuity index

d = maximum lesion dimension

θ = maximum edge angle (the maximum slope angle to the edge of the visible hole profile) in degrees

ΔGL = mean contrast (difference in grey level)

σ_s = mean noise within the lesion

σ_n = mean background noise

However, this study has some limitations in their use of CI. These were discussed in the conference paper by Szczepura & Manning (2016). In 2016, the first software programme which allowed the measurement of lesion conspicuity became available (Szczepura & Manning, 2016). The mathematical calculations for this software are based on the equation reported by Manning et al. (2004). The authors reported that the perceptual occurrence of focal image features is likely to be more accurate and repeatable when using their software (Szczepura & Manning, 2016).

2.4.2 Observer Performance Methods

These methods of IQ evaluation utilise the decisions from human observers. In other words, IQ evaluation is related to the visual interpretation of image data by observers (Martin, 2007). This section will discuss in detail two methods of observer performance studies. The first is conducted with patient/anthropomorphic phantom images and can be subdivided into 2 categories: lesion detection and the visibility of anatomical structures. The second looks at the visual detection of physical measures utilising test objects.

2.4.2.1 Visual Evaluation of Patients / Anthropomorphic Phantom Images

2.4.2.1.1 Receiver Operating Characteristics Analysis (ROC)

The essential task for an observer in medical imaging is to determine whether the radiographic image displays evidence of a pathology or not and identifies the location and type of the pathology. Therefore, it is important to have a method that can identify the performance of the observers in diagnosis and the performance of the whole imaging chain in improving pathology visibility (Bath, 2010). The ROC approach is often utilised in radiology to measure the observer performance in pathology detection and diagnosis. According to the International Commission on Radiation Units and Measurements' (ICRU) report 54, this method is the only approach that permits the assessment of imaging system performance without observational bias (Vennart, 1997). The principle work of this method depends on signal detection theory, in which the observer tries to reveal an object (signal)

with low contrast from a noisy background. The clinical relevance of this theory is invoked by the observer's seeking to differentiate abnormal cases from a group of normal cases (Mnsson, 2000).

To achieve this, observers are provided with set of images in which some of them are normal and some of them are abnormal and they are asked to point out the presence or absence of a pathology. The observers' performance can be specified by the number of true and false answers. For each level of image evaluation, the sensitivity, which represents the true positive rate, is plotted versus the 1-specificity, which represents the false positive rate since the specificity represents the true negative rate. **Figure (2-9)** illustrates an example of an ROC plot. The resultant curve is called a ROC curve, and it represents the relationship between the true positive rate and the false positive rate (Obuchowski, 2003). By improving the test accuracy, the ROC curve shifts towards the top left-hand corner of the graph. On the other hand, the blue line in **Figure (2-9)** represents the line of pure chance, also known as the total insufficiency of the test to differentiate between normal and abnormal cases (Park, Goo, & Jo, 2004). The area under the curve (AUC) can be used to evaluate the comprehensive performance of the diagnostic test, observer or imaging system because it illustrates the average value of sensitivity for all the possible specificity values. The values of the AUC can range from 0 to 1 and the higher values (closer to 1) represent superior performance of the diagnostic test. A value of 1 represents the perfect diagnostic test (Park et al., 2004).

In the basic ROC approach, the observers are required to provide a binary yes/no decision for presence or absence of a pathology. This is a limitation because it does not reflect the clinical reality (Morrell, 2006). Furthermore, ROC does not take into account the position of a pathology and this could lead to a loss of statistical power (Crop, 2015). For example, if the observer misses the true lesion in a radiographic image and instead selects an erroneous position as a lesion, this means the observer has performed two errors (a false negative and a false positive). However, these two errors will cancel out each other, and this would mean that at the case level the observer would be rated with a true positive (Crop, 2015).

To overcome the limitations of a general ROC approach, several developments have been proposed to improve the performance of ROC, increase its statistical power and minimise the number of images required. The free response ROC (FROC) was suggested by Bunch,

Hamilton, Sanderson, & Simmons (1977), wherein the location of a pathology also needs to be taken into consideration. In an FROC method, many lesions could be utilised in each image, and the position of each lesion should be identified (Crop, 2015). This approach provides excellent statistical power with a smaller number of images and fewer observers (Zarb, Rainford, & McEntee, 2010). The FROC curve indicates the fraction of true positive lesions as a function of the mean number of false positive answers per image (Crop, 2015). An alternative FROC (AFROC) approach is considered to be an alternate way of analysing the data generated from the FROC method. This method has been described and used in several studies (Chakraborty & Winter, 1990; Chakraborty & Berbaum, 2004a; Thompson, Manning, & Hogg, 2013). The jack-knife free-response ROC (JAFROC) considers the recent advances in ROC methods and overcomes a flaw in ROC by increasing its statistical power (Chakraborty, 2005; Chakraborty & Berbaum, 2004b). This method has been used widely in the literature (Fletcher et al., 2015; Jessop et al., 2015). In contrast, utilising ROC methods for IQ evaluations has some limitations; such as that it is time consuming and requires a large number of images. This makes it difficult to implement practically, especially in quality control purposes.

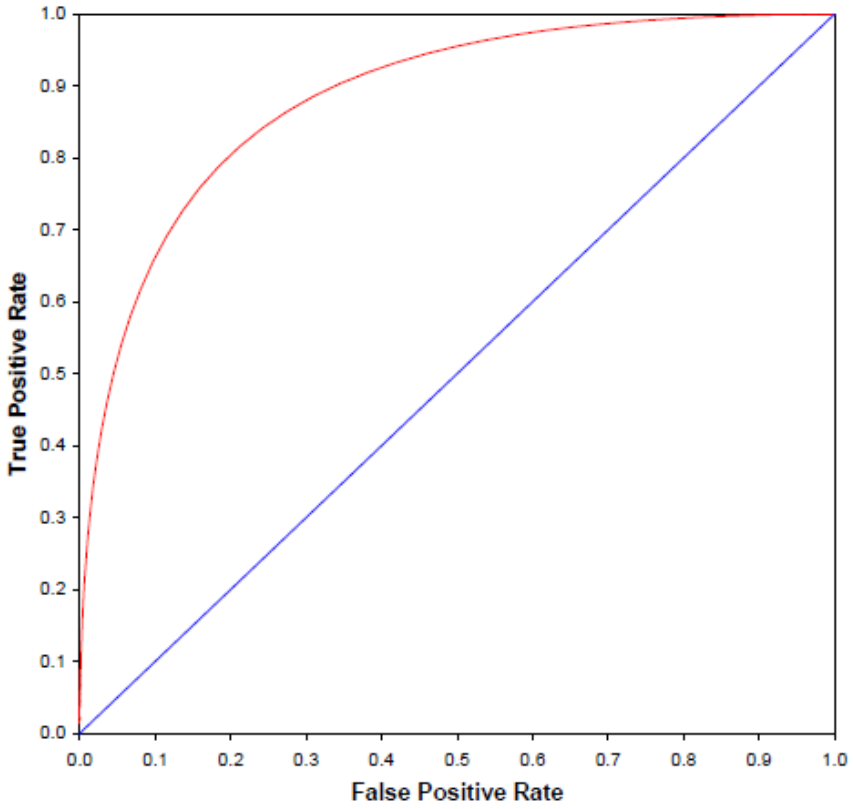

Figure 2-9 Example of a ROC plot (Morrell, 2006).

2.4.2.1.2 Visual Grading Analysis (VGA)

The visualisation of anatomical structures or pathologies in an image, using human observers, is considered to be an acceptable, firmly-established, valid approach for assessing IQ in medical imaging (Seeram et al., 2014). Its application is dependent on to what extent the anatomical structures and/or pathologies can be seen by the observers, such as by asking the observers to rate the visibility of the required details in the medical images (Smedby & Fredrikson, 2010). A human-based method, such as visual grading, makes the evaluation clinically relevant and therefore a preferred approach for evaluating IQ in medical imaging for many reasons. Firstly, the visual grading method is considered to have high validity, providing that the anatomical structures are selected based on their clinical relevance (Bath, 2010). Secondly, several studies have demonstrated a correlation between pathology detection and the clarity of the anatomical structures (Sund et al., 2000; Tingberg, 2000; Tingberg, Båth, et al., 2000; Tingberg, Herrmann, et al., 2000) as well as with physical measures of IQ evaluation (Sandborg et al., 2001; Sandborg, McVey, Dance, & Carlsson, 2000; Sandborg, Tingberg, Ullman, Dance, & Alm Carlsson, 2006). This can potentially be considered as a validation of the presumed potential of pathology detection

from the visibility of anatomical structures based on the idea of a visual grading approach. Thirdly, in contrast with ROC studies, visual grading studies are relatively simple to implement and this is related to the need to include fewer images and observers when compared with ROC studies (Bath, 2010). Furthermore, comparatively moderate amounts of time are required for the implementation of visual grading studies and this might allow the possibility of implementing this approach into hospitals (Bath, 2010). Finally, in contrast with ROC studies, this method does not require any specific preparation, for example ensuring the images contain pathology (lesions) and that the correct software is available (Bath, 2010). VGA methods have been utilised widely to compare different imaging techniques within an X-ray department (Almen et al., 2000; Tingberg et al., 2004; Anders Tingberg & Sjöström, 2005). There are two types of VGA methods that can be utilised to evaluate medical images, and these are as follows.

Relative VGA:

The observer is requested to score the quality of the image depending on the visibility of the anatomical structures in the image against the same anatomical structures within a reference image. A Likert scale is sometimes used for scoring the anatomical structures (whether the structure visibility is better, worse or the same as in the reference image (Ludewig, Richter, & Frame, 2010) , as illustrated in **Table (2-1)**.

Table 2-1 An example of the five point Likert scale for image scoring using a relative VGA method.

Scores	Meaning
5	Much better than the reference image
4	Better than the reference image
3	Equal to the reference image
2	Worse than the reference image
1	Much worse than the reference image

During image evaluation, the image under evaluation and the reference image need to be displayed on side by side on monitors that have been calibrated to the same standard; an alternative is to display the images on one monitor side by side. Image scoring can be calculated using the following equation (Seeram et al., 2014):

$$VGAS_{arel} = \frac{\sum_{i=1}^{I}\sum_{s=1}^{S}\sum_{o=1}^{O} G_{(rel)i,s,o}}{I \times S \times O}$$

Where:

$G_{(rel)i,s,o}$: represents the relative rating for a given image (i), structure (s) and observer (o).

O, I and S: represent the number of images, structures and observers, respectively.

Absolute VGA:

Here, the observer is requested to score the quality of the image depending on the visibility of anatomical structures in the image and there is no reference image to aid comparison and scoring (Båth & Månsson, 2007) , as illustrated in **Table (2-2)**. The data from this approach is used to calculate the ($VGAS_{abs}$) using the following equation (Seeram et al., 2014):

$$VGAS_{abs} = \frac{\sum_{i=1}^{I}\sum_{s=1}^{S}\sum_{o=1}^{O} G_{(abs)i,s,o}}{I \times S \times O}$$

Where:

$G_{(abs),o,i,s}$: represents the absolute rating for a given image (i), structure (s) and observer (o).

I, S and O: represent the number of images, structures and observers, respectively.

Table 2-2 An example of the five-point Likert scale for image scoring using an absolute VGA method.

Scores	Meaning
5	Excellent
4	Good
3	Moderate
2	Poor
1	Very poor

For minimising the risk of bias when using visual grading methods, an international group of radiologists and physicists established the CEC (Commission of the European Communities (CEC), 1996a) which proposed a set of IQ criteria for each specific type of imaging examination (general radiography, paediatric radiography and CT) for addressing the variations in IQ evaluations.

This method (VGA) has several limitations. Firstly, it does not take into account the clinical indication for the examination, from which the observer will need to distinguish between the normal and abnormal cases (Crop, 2015). Secondly, although this method and other visual IQ evaluation approaches have high validity, they suffer from low reliability due to the inter and intra observer variability. This limitation can be improved by involving several observers and taking the average value of their respective scores.

2.4.2.2 Visual Detectability of Physical Measures Using Test Objects (Psychophysical)

Psychophysical measures are tools that are utilised for measuring perception and performance. They are utilised to disclose basic perceptual processes, evaluate the performance of observers and for specifying the required characteristics of the imaging equipment (Pelli, D. G., & Farell, 2010). In this method, observers are required to give a subjective response regarding the impact of a physical stimulus on the imaged test object. Normally, the visible stimuli (e.g. test object) utilised in such measurements is very simplistic, such as a set of bar patterns, discs or square-shaped objects made with areas of various contrasts and diameters (Zarb et al., 2010). For obtaining greater reliability utilising this method, the variance among observers must be taken into account and the average results across the different observers are recommended.

2.5 Signal Detection Theory and Rose Model

The general signal detection theory was first developed in 1940 as a psychophysical technique in the study of audition (Manning, 1998). For several years it has been utilised as a theoretical framework to understand decision making when presented with uncertainty (Jang, Wixted, & Huber, 2009). It was originally utilised for analysing the detectability and recognition procedures of human observers viewing a signal (object) and attempting to discover the decision criterion (threshold) that was adopted by these observers in the detectability procedure (Jäkel & Wichmann, 2006). The presumption of this theory was that noise from various sources intervenes with the sensory stimuli of the human observer (Green, D. M., and Swets, 1988). According to this theory, detection performance is related to the internal SNR of the observer (Swets, Green, Getty, & Swets, 1978; Uffmann et al., 2004). The basic principles of the signal detection theory is known as the "Rose model" (Guerra, 2004). The Rose model characterises the relation between the noise (quantum noise), size of the object and contrast of the object. According to this model, the contrast of the object should be increased when the size (diameter) of the object is decreased in order for a detailed object to be visualised. This is normally found practically when phantoms with homogeneous backgrounds are used (Hemdal, 2009). Rose, in his work, tried to find the minimum level of noise that is required for an image to be visualised by the human eye, he observed that the values of SNR ranged from 5 to 7 and these represent the minimal values that are required to differentiate between a low-contrast object (disk) and a noisy background (threshold for detectability) (Rzeszotarski, 1999). This model is used commonly in radiology for IQ evaluation and comparing the performance of medical imaging systems. However, this model has been criticised (Bath, 2010) for the SNR being too simple for human observers who are sensitive to noise features; observers are generally not concerned with single pixel values and are not influenced by the pixel-to-pixel variation. Therefore the validity of using this method for IQ evaluation and the comparison of medical imaging systems could be considered low as argued by Bath (2010).

2.6 Low Contrast Detail Detectability (LCD) Method and their Test Objects

The LCD detectability method provides an assessment of the detection ability of imaging systems for low contrast and small details. LCD detectability depends on several factors such as noise, spatial resolution, contrast resolution, radiation exposure level and the visual response of the observer (Lu, Nickoloff, So, & Dutta, 2003). The LCD detectability

approach can be evaluated using a CD phantom and can be determined visually using observers or physically using automated computer software. These phantoms are comprised of a series of discs with different diameters to represent objects with different levels of contrast. The task of the observer or the computer software is to detect the faintest disc that can be distinguished on the image (Tapiovaara & Sandborg, 2004).

LCD detectability methods which utilise CD phantoms are well-established in the literature as an accepted method for comparing the IQ from different radiography systems (Veldkamp, Kroft, Boot, Mertens, & Geleijns, 2006 ; Bacher et al., 2003; Rong, Shaw, Liu, Lemacks, & Thompson, 2001; Geijer, Beckman, Andersson, & Persliden, 2001). Several studies have investigated and reported that the LCD detectability method is considered a reliable method for identifying large variations in the performance of the imaging systems and is able to rank them based on their contrast and noise characteristics (Marshall, Faulkner, Kotre, & Robson, 1992 ; Cohen et al., 1984; Tapiovaara & Sandborg, 2004). Moreover, several studies have been utilised this method in optimisation studies (Alzimami, Sassi, Alkhorayef, Britten, & Spyrou, 2009; Hamer et al., 2004; Tung et al., 2007). An advantage of this approach is its ability to assess the whole imaging chain, including the image display and the visual evaluation from using human observers (Peer et al., 2001; Veldkamp et al., 2009).

In contrast, the LCD detectability approach suffers from some limitations largely because it uses a homogeneous phantom which does not take into account the influence of the anatomical structures in the background (Månsson, Båth, & Mattsson, 2005; Veldkamp et al., 2009).

2.6.1 Leeds TO.10 Phantom

The Leeds TO.10 phantom is one of the most common phantoms reported in the literature and was developed by Hay, Clarke, & Eng (1985). The phantom was designed originally for routine quality assurance testing of fluoroscopic imaging equipment and it is constructed from a Perspex slab which contains 108 circular objects with different diameters and levels of contrast (Dragusin, Smans, Jacobs, Inal, & Bosmans, 2008). **Figure (2-10)** illustrates the Leeds test object TO.10 and its radiographic appearances. In this detection task, the observers search for and count the number of the visible circular objects in each row, then the scoring of the observers is compared with the correct scoring of standard data. This type of detection task is quick and easy to achieve. The output of the

phantom, a threshold of contrast-detail detectability, can be presented in two different ways: either graphically on a logarithmic scale or as the threshold detection index (H_T) versus the square root of the object area (Dragusin et al., 2008):

$$H_T(A) = \frac{1}{(C_T \times \sqrt{A})}$$

Where:

C_T: the threshold contrast

A (cm^2): the area of the details.

The Leeds test object provides nominal contrast values for object diameters and calibrated contrast values over a range of various X–ray beam conditions. The scores for each observer need to be averaged to generate the CD curves. The performance of the imaging system is determined by the position and shape of the curve, whereas the visibility of a particular object's size is determined by the value of (H_T). There is a very strong relationship between them; when the value of H_T increases the visibility of the object will also increase and vice versa (Dragusin et al., 2008). A study was undertaken by Gallacher, MacKenzie, Batchelor, Lynch, & Saunders (2003), who proposed a new method for presenting LCD detectability results by converting the CD curve to a single value: quality index (Q).

Q is calculated by taking the average of the ratio of the measured detection index $H_T(A)$ (from the imaging system under evaluation) to that of the reference value for all the object diameters as shown in the following equation:

$$Q = \frac{1}{n} \times \sum_{i=1}^{n} \frac{H_T(A_i)}{H_{Tref}(A_i)}$$

Where:

Q: quality index

n: number of object diameter groups

$H_T(Ai)$: the threshold detection indexes for the imaging system under evaluation

$H_{Tref}(Ai)$: the threshold detection indexes for the reference curve.

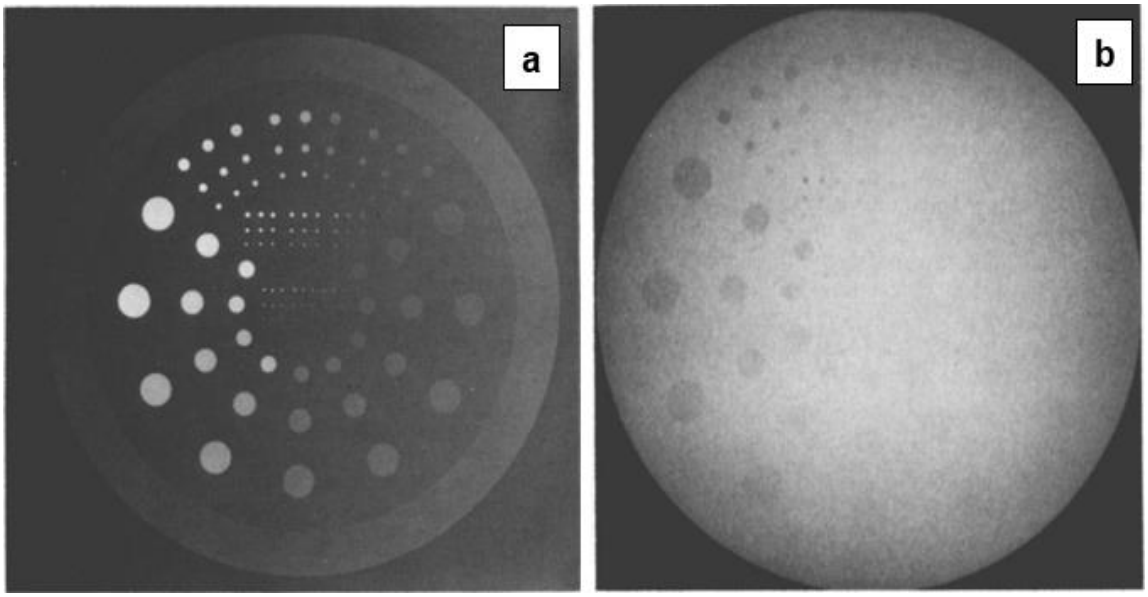

Figure 2-10 (a) Leeds TO.10 test object phantom. (b) Leeds test object - resultant X-ray image. (Hay et al., 1985).

2.6.1.1 CDMAM 3.4 Phantom

The CDMAM 3.4 phantom has been designed specifically for evaluating IQ in mammography examinations. The phantom is constructed from an aluminium slab which contains gold objects of various thicknesses (contrasts) and sizes. These gold objects are distributed in a square matrix of 16 rows and 16 columns. Within the row the object size is constant, while the object thickness increases logarithmically. Within the columns the object thickness remains constant yet the object size increases logarithmically. The discs decrease in diameter (0.06–2.00 mm) and thickness (0.03–2.00 μm) in logarithmic steps (Bacher, Smeets, De Hauwere, et al., 2006). The range in the object sizes and thicknesses has been designed to simulate microcalcifications and small soft tissue masses (Lanconelli et al., 2007; Paulana, Wigati, & Soejoko, 2015). Within each square in the phantom there are two identical objects (with the same size and thicknesses). One is placed in the centre of the square and the second is placed in a randomly chosen corner. The detection task within this phantom, whether visually by observers or automatically by the computer software, requires the observer to indicate the position of the corner that has the eccentric object for each square cell in the phantom. This phantom is very useful for evaluating the performance of mammography devices according to their resolution and threshold contrast-detail detectability. **Figure (2-11**) illustrates the CDMAM 3.4 phantom and its resultant X-ray image.

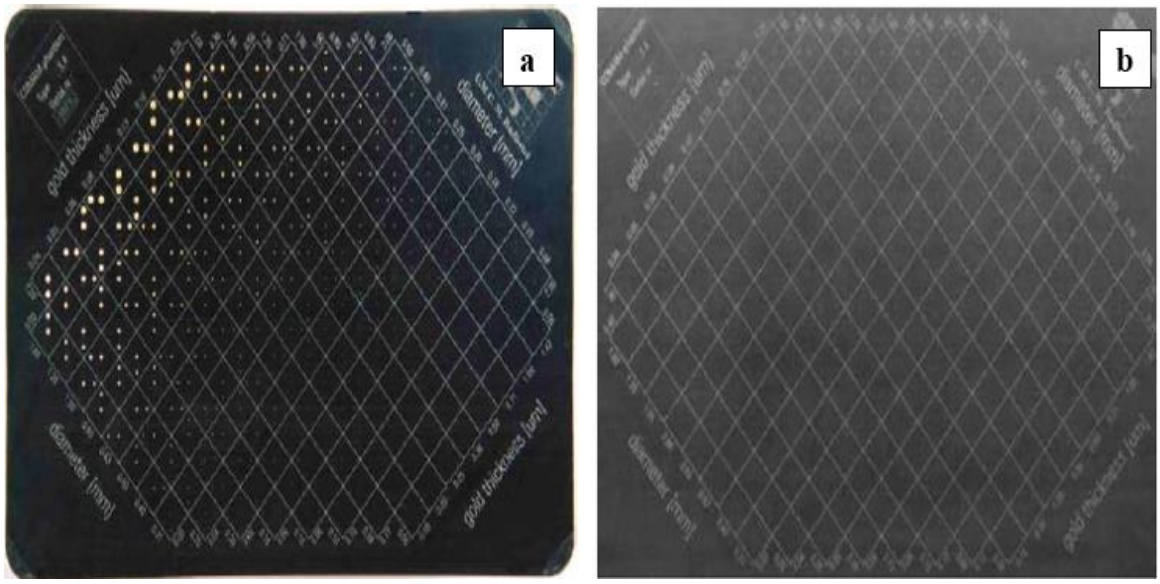

Figure 2-11 (a) The CDMAM 3.4 phantom. (b) The resultant CDMAM phantom X-ray image. (Riverti di Val Cervo, 2011).

2.6.1.2 CDRAD 2.0 Phantom

The CDRAD 2.0 phantom (Artinis Medical Systems, Zetten, the Netherlands) (Burght, Floor, Thijssen, & Bijkerk, 2014) is a special physical test object that has been designed to evaluate LCD threshold detectability. LCD threshold detectability is the lowest contrast needed to visualise objects of various sizes above the noise threshold for general radiography systems. The output of the CDRAD 2.0 phantom, which represents the LCD detectability, is considered to be one of the most important features in clinical disease detection and is a large challenge for any radiographic system and radiological display (Mc Laughlin et al., 2012; Peer et al., 2003). The CDRAD 2.0 phantom consists of a square acrylic plastic plate (265× 265 mm), with a thickness of 10 mm, and a series of flat drilled holes of various depths and diameters. A lead grid divides the phantom into 15 rows and 15 columns resulting in a total of 255 squares; the diameter of the drilled holes in each of the columns and their depth in each row varies logarithmically from 0.3 to 8.00 mm. In each square, there are either one or two visible holes present: the first three rows contain only one visible hole in the centre of each square, whilst the rest rows (from the fourth row to the fifteenth row) have two identical visible holes (one in the centre of the square and the other positioned randomly in one of the four corners) (Burght et al., 2014). **Figure (2-12)** shows the CDRAD 2.0 phantom and its resultant X-ray image. Commercially, there

are five versions of the CDRAD 2.0 phantom available, each with different locations for the visible holes in the four corners to minimize the influence of locational familiarity on the observers (Burght et al., 2014).

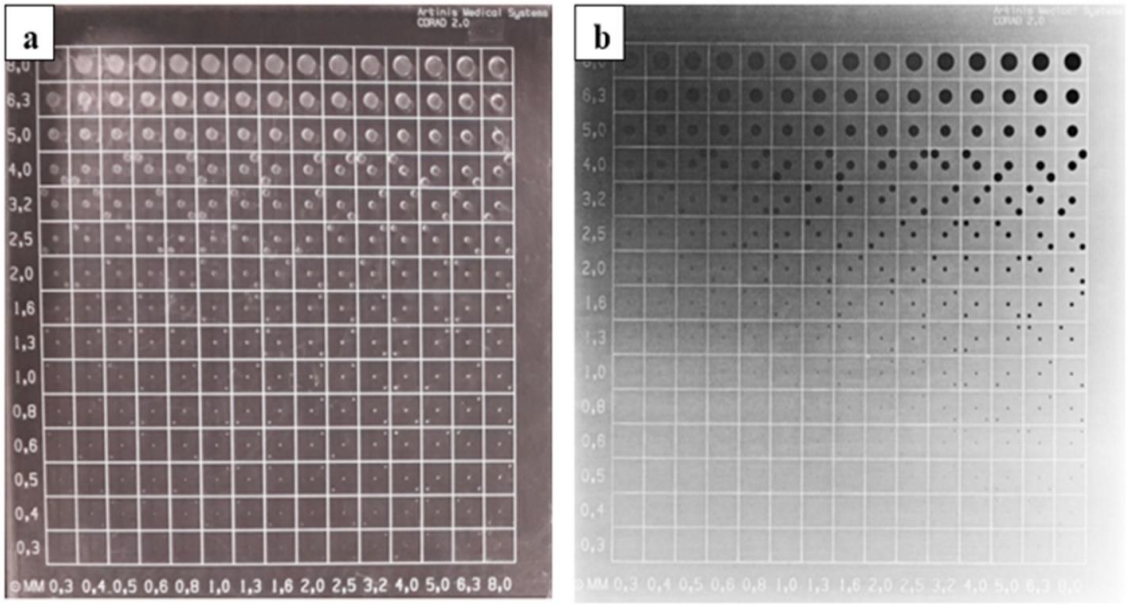

Figure 2-12 (a) CDRAD 2.0 phantom. (b) CDARD 2.0 phantom resultant X-ray image.

Two factors determine the threshold of LCD detectability: 1) the size of the object (hole); 2) the noise of the imaging system under investigation. Rose has shown that smaller sized objects require high contrast, while lower contrast is needed for visualising larger sized objects (Rose, 1974); this relationship can be expressed using the following equation:

C. D = k

Where:

C: contrast of objects.

D: size of the object.

K: constant value that represents the threshold of object visibility.

In the CDRAD 2.0 phantom the LCD detectability can be measured by calculating the CD curve or image quality figure (IQF) based on the method that was proposed by Burght et al. (2014).

The CD curve is a graphical representation of the LCD detectability. It is constructed by combining the minimum depth (contrast) of each visible hole in each column (for each diameter) in the image (Burght et al., 2014). When the CD curve is closer to the origin point in the graph a superior LCD detectability performance will be recorded as shown in **Figure (2-13)**. In general, the CD curve is hard to construct and complicated to compare between cases. As a result, IQF which is a numerical summary of the CD curve, is commonly used as indicator for LCD detectability performance (Alzimami et al., 2009). The IQF value is easy to utilise for comparisons between the different images that are acquired with different acquisition parameters, or with different diagnostic imaging systems. IQF can be calculated from the summation of the product of the lowest diameter (threshold diameter) in each of the fifteen columns that has had an object correctly detected along with their intrinsic depths (contrast). The IQF value can be calculated using the following equation (Burght et al., 2014):

$$IQF = \sum_{i=1}^{15} C_i D_{(i,th)}$$

Where:

$D_{(i, th)}$ represents the lowest (threshold) diameter in the column (i) that has a visible hole correctly detected, C_i represents the depth (contrast) value of the hole in the column (i).

IQF is inversely proportional to the IQ (lower IQF denotes a high IQ and the ability to visualise minimal size and lower contrast lesions). The IQF_{inv} can be calculated by taking the inverse value of the IQF, as show in the equation below. IQF_{inv} has an extreme relationship with the IQ (IQ increases with increasing IQF_{inv}) (Burght et al., 2014).

$$IQF_{inv} = \sum_{i=1}^{15} \frac{1}{C_i D_{(i,th)}}$$

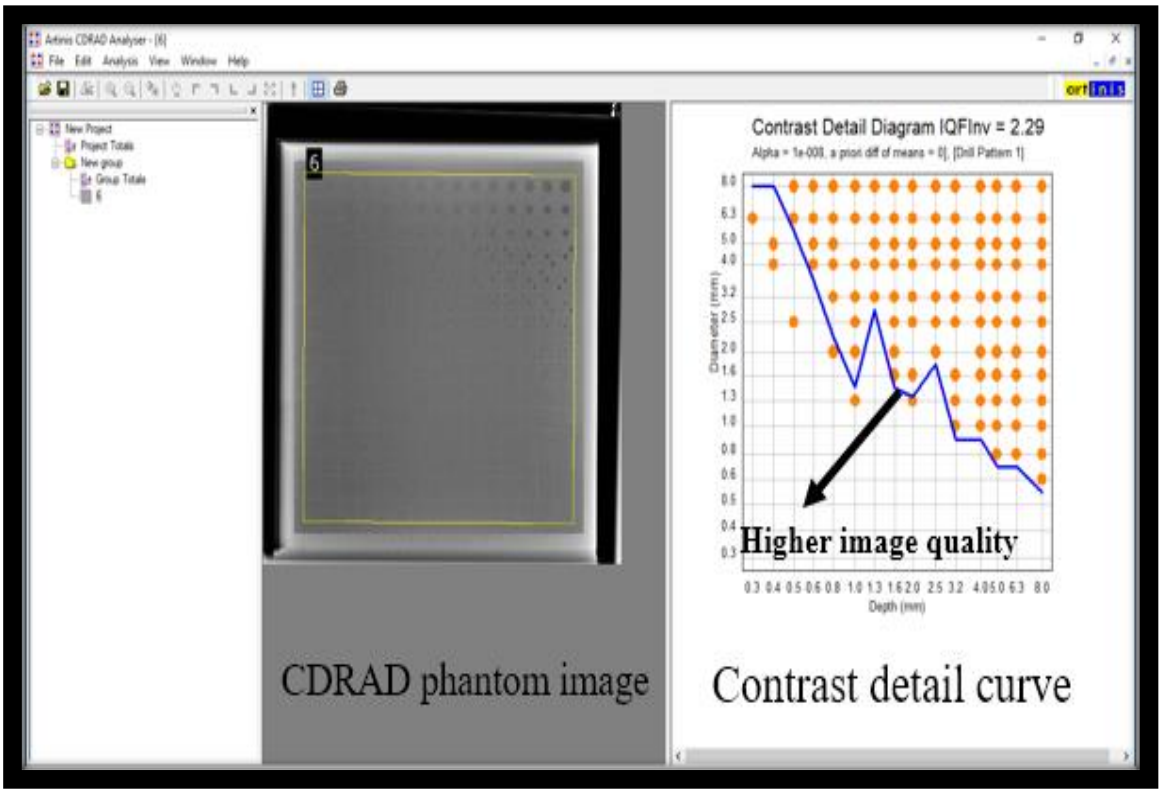

Figure 2-13 CDRAD 2.0 phantom X-ray image and resultant contrast details curve calculations generated using CDRAD 2.0 phantom analyser software.

2.6.1.2.1 Visual LCD Detectability Assessment Method

Visually evaluating the resultant CDRAD 2.0 phantom images was suggested by the manufacturer as a viable option for assessing IQ. Such evaluations are then based on the human perception and decision criteria (Pascoal, Lawinski, Honey, & Blake, 2005). This method involves a number of observers and each one individually scores the CDRAD 2.0 phantom images. The average score for all the observers represents the quality of the image which combines all of the information about LCD detectability.

With a visual (perceptual) method the CDRAD 2.0 phantom images are assessed in the regions where the holes are visible. This method involves inviting each observer to identify the position of the corner that has the visible hole for each square cell (hole diameter) and for each column on the resultant image. The holes are visually detectable in both the centre of the cell (for the first three rows) or the corner (for the remaining rows). The count is only correct if the visible hole is correctly identified. The final individual score for each image and from each observer needs to be corrected according to the nearest neighbour correction scheme, which is recommended by the CDRAD 2.0 phantom manufacturer.

According to the nearest neighbouring correction scheme, a true detection for a region requires two or more correctly detected neighbouring regions to be valid. A false or undetected object in a specific region is converted to true detection if three or more neighbouring regions are truly detected. These rules are utilised for all the regions on the image, except those in the four corners because they have only two neighbouring regions. The true detection for any of these four cells requires one neighbour cell to be detected correctly and the false detection for any of them converted to true detection if the two neighbouring cells are detected correctly (Burght et al., 2014). Following the correction scheme, the results (CD curve or IQF_{inv}) represent the threshold LCD detectability.

2.6.1.2.2 Physical LCD Detectability Assessment Method

CDRAD 2.0 phantom images can be evaluated physically using the automated CDRAD analyser software and the results are displayed as a CD curve and an IQF_{inv}. In each square cell of the CDRAD 2.0 phantom image, the software identifies the central visible hole and examines the four corners to discover the centre of the peripheral visible holes. Subsequently, the software utilises a statistical method to determine whether a hole has been detected or not. This statistical method utilises the mean pixel signal value and its SD, and is done for both the visible hole under evaluation and its background. The Welch Satterthwhaite statistical model (Student t test with Welch correction) is used in this software to decide whether the average signal level in the square cell is larger than that of the background or not. If there is a statistically significant difference between the signal and its background, the visible hole is considered detected (Burght et al., 2014).

To achieve the above statistical calculation correctly, the locations of the central and peripheral visible holes in all of the 255 square cells of the CDRAD 2.0 phantom need to be identified. The algorithm that is used in the software analyser for determining hole positions uses a 3 stage computer algorithm: -

1. Determination of borders of the CDRAD 2.0 phantom image.

The software determines the four corners of the phantom image (lead grids surrounding the outer square cells) by assuming that only the phantom is illuminated and the background is black. Based on this assumption, the lead incised outline of the phantom will be detected by a search algorithm and the phantom's four corners and overall location will be recognised as illustrated in **Figure (2-14 a)** (Burght et al., 2014).

2. Determination of the centre of the square cells and visible holes of the CDRAD 2.0 image.

After determining the phantom outline (the four corners of the phantom) in the first step, the four sides of each of the 255 square cells are located automatically. Next, the centre of each square cell is located at the middle of the four sides which also represents the centre of the visible hole in that square cell. In terms of the peripheral visible hole, the software searches the four corners to identify the peripheral visible hole using a statistical calculation based on the ideal observer model. These statistical calculations include measuring the average pixel value for each of the four corners in the cell, and choosing the corner that has the highest average value to represent the visible hole (Burght et al., 2014).

3-Determination of the background and visible hole signals.

The background and visible hole signals are measured from two different locations in each square and are represented by the red and white regions respectively, as illustrated in the **Figure (2-14 b)**. The locations of the background regions depend on the position of the second visible hole (white region) in the square cell and the four regions of the background and signal (red and white regions). The average pixel value ($\mu_{background}$) and SD for the background ($\sigma_{background}$) are calculated from the red regions, whereas their signal values ($\mu_{background}$ and $\sigma_{background}$) are calculated from the white regions. (Burght et al., 2014).

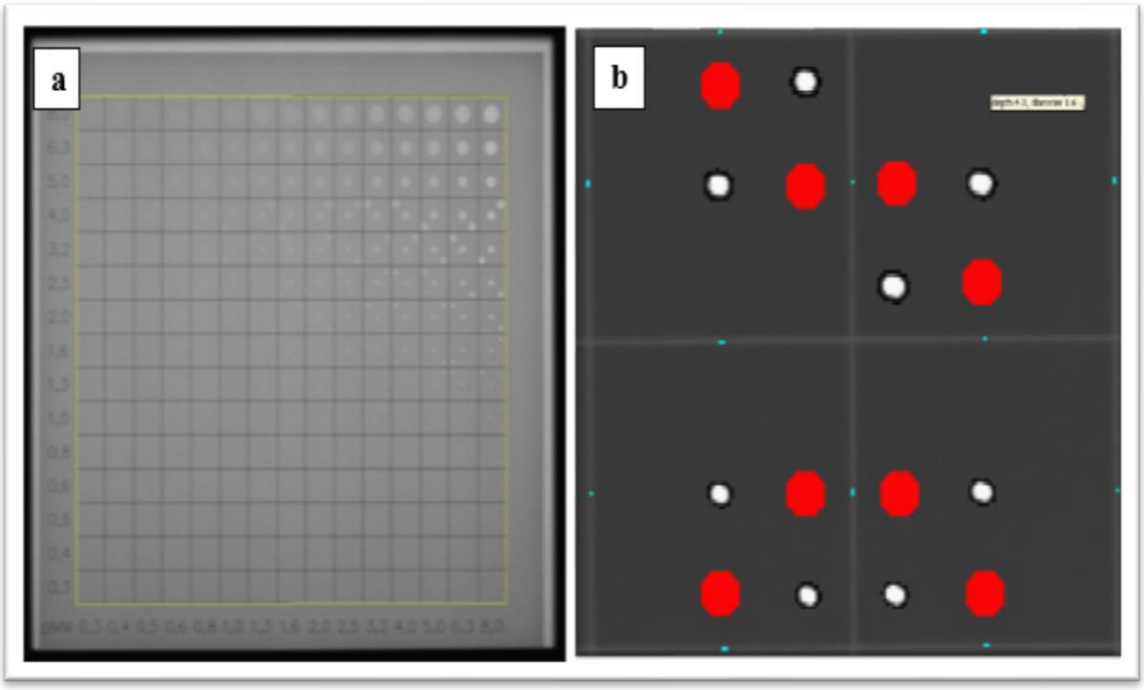

Figure 2-14 (a) Determination of the borders of the CDRAD 2.0 phantom images. (b) The background and visible spot signals measured from two different locations in each square cell and are represented by the red and white regions, respectively (Burght et al., 2014).

Prior to the analysis, three parameters in the software need to be set by the user: the Alpha level of significance (Alpha), the a priori difference of the mean (APD) and the source to image detector distance (SID). In terms of the Alpha level, the statistical test used in the CDRAD phantom analyser software uses a 95% confidence level. The significance level can be any values ranging from 0 to 0.5 (Pascoal et al., 2005). Any changes in the values of Alpha can lead to an increase or decrease in the IQF_{inv} values and thus leads to an upward or downward shift in the CD curve. Setting low values of Alpha value leads to an increase in the confidence level for the detection of the objects. Numerous objects are detected at lower confidence levels because the threshold level decreases and a higher quality is assigned to the resultant image (Norrman, Gårdestig, Persliden, & Geijer, 2005; Pascoal et al., 2005). APD should be considered (and included in the calculation) during the automated scoring method and set relative to the image bit depth. The APD needs to be set at 0 if the images under comparison have different bit depths. This ensures for a valid comparison between the images (Brosi, Stuessi, Verdun, Vock, & Wolf, 2011; Burght et al., 2014; Lança, 2011). Finally, the SID needs to be set based on the values employed during the data acquisition.

2.6.1.2.3 Physical Versus Visual LCD Detectability Evaluation Method

The LCD detectability assessment method based visual interpretation is essential for the IQ evaluation and has high validity and is considered to be a critical factor in the evaluation of the whole imaging chain. However, the variability of this method within and between observers is considered to be the main limitation that could contribute to variation in the results (Aufrichtig, 1999). This variation could influence the threshold detectability test and thus limits its ability of accurately evaluating IQ and system performance (Crop, 2015). The source of the variation in the results is grouped into two categories: inter- and intra-observer variability, based on terminology of Swets and Pickett (Swets & Pickett, 1982).

Pascoal et al., (2005) reported that the visual evaluation of IQ is time consuming, tiring and could possibly introduce bias in some cases wherein the images are viewed on more than one occasion because of learning effects (Pascoal et al., 2005). This is, especially the case for CDRAD 2.0 phantom images.

The above limitations could be partly addressed by applying a physical method, such as using the analyser software instead of relying on observers. The main advantages of the automated method are that it has high reliability and consistency in the scoring criteria it uses to evaluate the threshold contrast detail detection and that it does not suffer from the subjectivity of the human visual and cognition systems (Pascoal et al., 2005). The automatic method described has been reported in the literature. It is believed to be more sensitive than that of the visual/perceptual method and has the ability to report small low contrast variability (Pascoal et al., 2005). Alongside the above, several studies (De Crop et al., 2012; Norrman et al., 2005) have argued that there is good agreement and correlation between the two different scoring methods (visual and physical methods) when evaluation CD images.

In a study by De Crop et al., (2012), sixteen CDRAD 2.0 phantom images were acquired using a chest protocol. They used automatic exposure control (AEC) with a range of different tube voltage values (70,81, 90, 100, 113, and 125 kilovoltage peak (kVp)). The SID was set at 150 cm and different beam filtrations (0.1, 0.2, and 0.3 mm Cu) were used. The CDRAD 2.0 images were evaluated visually by four observers (all medical physicists) and physically using a CDRAD 2.0 phantom analyser software. A strong correlation was found between the visual evaluation method by observers and the physical evaluation method by the CDRAD analyser software (Spearman r= 0.895; P =0.001). In a study by

Norrman et al. (2005) , several images of the CDRAD 2.0 phantom were acquired using 77 kV. This is similar to acquisitions for the lumbar spine and urography and with wide range of milliampere second (mAs) values (2, 5, 10, 16,20, 25, 32, 40, 63, 80, and 100 mA s). 3 observers (2 medical physicists and 1 radiologist) assessed the CDRAD 2.0 phantom images visually, and then the images were assessed physically using the computer software (Computer Calculated Image Quality [CoCIQ]). It was found that there was a very strong linear relationship between the visual assessment using observers and the physical assessment using the software (R^2 = 0.98). Moreover, the physical LCD detectability method displays time saving, which is an important factor in optimisation and survey studies; ease of use and a decrease of human labour; and repeatablability and precision. (Fernandez, Ordiales, Guibelalde, Prieto, & Vano, 2008).

2.6.1.3 CDRAD 2.0 Phantom Versus the Other Low Contrast Detail Test Objects

In comparison with other CD phantoms, the CDRAD 2.0 phantom has been designed to allow the use of a 4-alternative forced-choice (4AFC) method for the evaluation of visual IQ. During the image evaluation, the observers are required to determine not only whether the hole is visible or not, but also indicate the position of the corner (between four possibilities) in which the visible hole is present (Burght et al., 2014). Using this method, 4AFC, can decrease the probability of a false positive error by 25%, minimise the subjectivity of the reader's threshold and measure the correct target detection based on a statistical base (De Hauwere, Bacher, Smeets, Verstraete, & Thierens, 2005). This improves reliability and accuracy (because it requires many repetitions) when compared with conventional CD phantoms which do not use this 4AFC approach, such as Leeds TOR phantom (Aufrichtig & Xue, 2000). Furthermore, with 4AFC the detectability mistakes made by the observers can be controlled using the nearest neighbouring correction scheme (Burght et al., 2014). The large number of different diameter holes and the random placement of the details leads to the CDRAD 2.0 phantom being more complicated than the others. The main limitation with the other CD phantoms, such as the Leeds TOR, relates to the detectability process. As a result of the crude design of these phantoms, observers have prior knowledge about the placement of the holes which could have an impact on the image evaluation by increasing the potential bias (Rong et al., 2001). Moreover, the CDRAD 2.0 phantom has the advantage of its using automatic physical scoring methods (analyser software) to evaluate the IQ. This is not available with some of the other types of CD phantoms.

2.7 Chapter Summary

In this chapter, an overview of the digital imaging system and IQ concepts have been presented together with the different types of evaluation methods (**Table 2-3**).

Table 2-3 This table summarises the key elements of advantages and limitations of the different IQ evaluation methods.

IQ evaluation method type	Advantages	Limitations
Physical measures of IQ evaluation (SNR,CNR, DQE and LCD detectability)	1)They have a high reliability (repeatability) if undertaken in a consistent fashion. 2)They are easy to implement. 3) They are considered time saving. 4) The LCD detectability method can give an indication of visibility of low contrast detail objects such as subtle pathology. 5)They do not require detail human observers.	1) They have a low validity because they often measure only one or two aspects of image quality such as noise or contrast. 2)They do not take into account the influence of anatomical noise arising when imaging human anatomy. 3) They do not evaluate the entire imaging system (eg, the observation stage is not included).
Visual measure of IQ evaluation (ROC and VGA)	1)They have a high validity because they measure all the aspects of image quality such as noise, contrast,..etc. 2) They take into account the influence of anatomical noise arising when imaging human anatomy. 3)They are evaluating the entire imaging system including the observation stage	1)They have a low reliability as a result of inter- and intra-observer variability. 2)They are not easy to implement in clinical practice because they require trained human observers. 2)They are time consuming. 3)ROC methods can be used for evaluating the pathology detection performance only and not for the general IQ evaluation. 4) VGA method does not consider the evaluation of pathology detection performance and focusing only on the general IQ.

The information in this chapter can be summarised as follows:

IQ is a broad expression and is hard to specify through one particular definition. This is because it is determined by an observers' ability to use the information in the image for a specific problem or task. Each method of IQ evaluation has advantages and disadvantages. The CDRAD 2.0 phantom was found to be a good choice for IQ evaluation using the LCD detectability method when compared with other CD phantoms. The physical method of LCD detectability evaluation for CDRAD 2.0 phantom images (software analyser) has high reliability and has been found to have very good correlation with the visual evaluation method undertaken by observers.

Chapter 3: Literature Review

3.1 Chapter Overview

This chapter will start with an introduction to radiography of the chest and its main characteristics, its importance and its challenges for both paediatric and adult patients. For the purposes of this thesis, the literature review is presented in a critical review format, arranged into 3 sections. The first section (presented in 3.4.) concentrates on the previous approaches that have been utilised for comparing IQ and radiation dose for CXRs among hospitals. This section presents a critical review of the available literature that has been used to evaluate the level of variation in IQ and radiation dose among hospitals for both paediatric and adult CXR. The reviewed literature is divided into 3 topics and listed under 3 subsections (3.4.1., 3.4.2. and 3.4.3.) based on the type of the methods used for IQ evaluation and radiation dose measurements. The first topic is patient studies; the second is anthropomorphic phantoms studies; and the third is physics phantom studies. For each topic a definition, a description of the method used as well as their advantages and disadvantages will be presented. Also, a critical review of the available literature will be presented.

The second section (presented in 3.5.) focuses on introducing a novel method based on FOM to compare radiation dose and IQ for evaluative purposes, within and between hospitals. In this section, the literature will focus on FOM and its meaning, its importance in optimisation and the idea of utilising FOM for comparing the performance of imaging systems and protocols amongst X-ray units. Also, this section will highlight existing FOM formulas along with their disadvantages and suggestions for new FOMs.

The third section (presented in 3.6.) focuses on investigating the validity of using LCD detectability with an CDRAD 2.0 phantom in visual IQ evaluation and LV evaluation along with its validity in CXR optimisation studies. In this section, the literature that considers the correlation between LCD detectability and IQ will be reviewed and critiqued. The critique will consider the IQ evaluation methods that were used and their image viewing conditions. Also in this section will be three specific questions that are crucial for achieving the aim of this thesis and have not been answered in the literature will be considered. Finally, a summary will be presented of the knowledge gap that will be addressed by this thesis**.**

3.2 Search Strategy

In order to identify literature that is relevant to this thesis, a wide literature search was executed on several databases: ScienceDirect, PubMed, Medline, online journals and Google Scholar. To acquire scientific literature relevant to assessing the variability in IQ and radiation dose between hospitals (section 3.4.,section 3.5.), the following key words were utilised: *paediatric chest radiography, adult chest radiography, optimisation, image quality, radiation dose, chest, CDRAD phantom, low contrast-detail, radiologic technologist/technician, plain x-ray, radiography, entrance dose, monitoring, dose measurement, imaging techniques, reference dose levels, technical factor, DRL, diagnostic reference levels, UK national patient dose, quality assurance, image quality and dose in UK, CXR examinations in UK, radiation, FOM, figure of merit, radiographer.*

For the purpose of this thesis, research studies were excluded if they did not compare image quality and radiation dose between or within hospitals. Regarding the research studies that were carried out using the CDRAD phantom as a vehicle for image acquisition, there was no restriction on the date of publishing and the examined anatomical area. This is because the CDRAD phantom is the main method for IQ evaluation used in this thesis. Research studies that were not carried out using the CDRAD phantom were excluded if they were not related to the CXR imaging because they are not related to the objectives of the thesis.

To obtain literature relevant to evaluating the validity of utilising LCD detectability using a CDRAD 2.0 phantom in IQ and LV evaluation and its validity for use in optimisation studies (section 3.6.), the following key words were utilised: *image quality, low contrast-detail (LCD) detectability, visual grading analysis, CDRAD phantom, digital radiography, image quality figure inverse, IQF_{inv}, correlation of visual and physical image quality, CDRAD phantom validation, correlation of CDRAD phantom with visual image quality, CDRAD phantom in chest radiography, CDRAD phantom in optimisation studies, contrast-detail phantoms, contrast-detail evaluation*. There were no restrictions on the date of publishing in order to ensure that the substantial seminal studies were identified and included.

During the search process, only the articles that were in English were included in the literature review. The Boolean search operators (AND, OR, NOT) were utilised where needed to revise the search.

3.3 Radiography of the Chest

CXR is considered the most frequently performed diagnostic X-ray examination and it is valuable for helping to resolve a broad range of clinical problems (Bacher et al., 2003; Raoof et al., 2012). X-ray images of the chest can generate a substantial quantity of information and can assist in establishing a diagnosis, guiding treatment and have a role in follow-up. Even with the recent developments in cross-sectional imaging, for example CT, CXR still remains important as a first pathway for diagnosing pulmonary, cardiovascular and skeletal (thoracic) disorders. Furthermore, there are additional advantages for CXR, including its relatively low cost and low radiation dose (Veldkamp, Kroft, & Geleijns, 2009; Speets et al., 2006).

The CXR examination is considered to be one of the most difficult examinations to obtain an adequate diagnostic image from. This is generally a result of the wide range of tissue types within the chest region that have both high and low X-ray attenuation characteristics. By way of an example, the variation in the attenuation characteristics between the mediastinum and the lungs is large and at the same time the visualisation of the fine lung texture is especially demanding (Doyle, Martin, & Gentle, 2006; Singh, Martin, McCurrach, & Phanco, 2013). As a result there are often a number of acquisition parameters that are available and suitable (Singh et al., 2013). In general, paediatric dose optimisation can be more challenging than that for adults. The wide variation in paediatric sizes makes the determination of optimum exposure parameters more difficult (Almen, Lööf, & Mattsson, 1996; Willis, 2009). The chest region adds further complexity when performing paediatric radiography because, as stated, the attenuation within chest tissue varies widely. In neonates, CXR is important for the evaluation of potentially life-threatening conditions and is often performed as a mobile examination with the patient in an incubator. Neonatal patients often have daily CXR examinations and this generates further issues around dose optimisation (Singh & Pradhan, 2015). Several studies have been conducted for both paediatrics and adult patients which demonstrated that there are different acquisition protocols which are commonly utilised for CXR examinations. As a result there can be large variations in radiation dose and IQ for the same examination (Johnston & Brennan, 2000; Carroll & Brennan, 2003; Hart, Hillier, & Shrimpton, 2012). Legislation advocates that the radiation doses resulting from X-ray examinations must be as low as practicable (ALARP) and without any degradation of IQ (Britain, 2000).

3.4 Previous Studies Comparing Image Quality and Radiation Dose for CXRs among Hospitals

This section presents a critical literature review of the studies that have compared IQ and radiation dose between and within hospitals. This includes a description of the methods that have been used along with their advantages and disadvantages.

Several studies have been conducted to compare the radiation dose and IQ for different examinations and different modalities, both between and within hospitals. These studies were conducted using different methods for evaluating IQ and radiation dose. For IQ, several of these studies have acquired images using patients, whilst others have used anthropomorphic phantoms. Images have been evaluated using different visual methodologies, such as ROC, absolute VGA and relative VGA; some have used physical methods, such as SNR and CNR. A number of studies have used different types of physical phantoms and evaluated the resultant images, both physically and visually. Much like IQ assessments, dosimetry measurements have been undertaken using several methods- for example, using direct measurements with solid-state dosimeters, thermoluminescent dosimeters (TLDs), metal oxide semiconductor field effect transistors (MOSFETs) or through the use of mathematical modelling techniques such as Monte Carlo. There are are variations in the dose quantities reported within these studies- for example, incident air kerma (IAK), entrance surface air kerma (ESAK), air kerma area product (KAP) and effective dose (ED).

Published studies undertaken to compare differences in IQ and radiation dose among imaging centres can be classified based on their methods and fall into three groups: 1) patient study methods; 2) anthropomorphic phantom study methods; 3) physics phantom study methods.

3.4.1 Group One: Patient Studies

This subsection starts with a definition of the patient studies method and how IQ evaluation and radiation dose measurements can be obtained. Advantages and disadvantages will then be highlighted. Finally, the survey studies that were used for evaluating IQ and radiation dose between and within hospitals will be presented.

The patient study method uses patients as a tool to obtain the required information regarding IQ and radiation dose. For dosimetry measurements, the main dose quantities that are reported within the literature include IAK, ESAK and KAP; these can be measured using TLDs that are placed on the skin of the patient or from the output of the X-ray

machine (Billinger, Nowotny, & Homolka, 2010a). Regarding IQ, images acquired from patients tend to have been assessed using the absolute or relative VGA method based on the criteria published in the CEC (Commission of the European Communities (CEC), 1996a) or using an ROC method.

Although the patient-based method tends to have higher validity because they deal with data generated from real patients, it does have some limitations. Firstly, this method has low reliability when comparing the IQ and radiation dose among hospitals because of the expected variation in patient size within the same age group, and the variation in the tissue characteristics of the patients, such as presence or absence of disease. In other words, the data is not cross-matched, i.e. the same patient is not imaged in the different comparator hospitals. These two elements will have an influence on both IQ and radiation dose and potentially lead to an inaccurate comparison. Secondly, it is very difficult to implement optimisation studies using patient studies because several images need to be acquired from each patient and this means exposing the patient to numerous exposures which is not ethically acceptable.

The patient-based method is currently one of the main methods used to generate DRL data for a specific country or area, and also for a specific examination.

DRL concept is used for protecting patients from an excessive radiation dose and it was proposed initially in the ICRP publication No. 73 (ICRP, 1996). The idea of DRL is based on identifying the dose values for the radiographic examinations in order to utilise it as a benchmark for judging the level of radiation protection (Suliman & Mohammedzein, 2014). DRL values for each examination are determined from a percentile point for the recorded radiation dose from patients or equivalent (such as a phantom) and normally represented by a simple dosimetry measure such as entrance surface dose (ESD) or dose area product (DAP) (ICRP, 1996).

Several large DRL studies have been presented using this method (patient study method) (Akpochafor, Omojola, Adeneye, Aweda, & Ajayi, 2016; Hart, Hillier, & Shrimpton, 2012; Hart, D. W. B. F., Wall, B., & Shrimpton, 2000). Also, many surveys using the patient study method have been undertaken to investigate radiation dose levels between hospitals (Agarwal & Newbery, 2016; Elshiekh, Suliman, & Habbani, 2015; Ujah et al., 2012; Egbe, Inyang, Eduwem, & Ama, 2009).

Several survey studies have been also conducted in the UK to investigate the variation in IQ and radiation dose for CXRs between hospitals (Brennan & Johnston, 2002; Hart, 2003;

Hart et al., 2012; Johnston & Brennan, 2000; Shrimpton et al., 1986; Simpson, Martin, Darragh, & Abel, 1998; Wall & Hart, 1997; Warren-Forward, Haddaway, McCall, & Temperton, 1996; Warren-Forward & Millar, 1995). The majority of these studies have been carried out using film screen systems. Such systems are now considered to be outdated. The most recent survey study in the UK, the national patient dose database (NPDD), was reported by Hart et al. (2012). Data was collected over 5 years (from January 2006 to December 2010) from 320 hospitals across the UK. This represents approximately 25% of all the hospitals that have diagnostic X-ray facilities. Data collected included 165,000 ESD and 185,000 DAP measurements for a range of X-ray examinations (abdomen AP, chest AP, chest LAT, chest PA, knee PA, knee LAT, lumber spine AP, lumber spine LAT, pelvis AP, shoulder AP, skull AP/PA, skull LAT, thoracic spine AP, thoracic spine AP). Furthermore, 221,000 DAP measurements and 146,000 fluoroscopy times for complete examinations.

In terms of CXR, the PA adult examination results saw a large variation in ESD, ranging from 0.02 mGy to 0.56 mGy (mean 0.11 mGy, 3rd quartile 0.14 mGy). This study is the latest national patient dose survey in the UK and it is an extremely useful resource for providing an indication about the difference in radiation doses between hospitals, and the recommended DRLs (the third quartile values). However, this survey study was conducted on adults only and paediatric age groups were not included. There was also no IQ evaluation.

3.4.2 Group Two: Anthropomorphic Phantom Studies

This subsection starts with a definition of the anthropomorphic phantom study method. Then, the IQ assessment and radiation dose measurements undertaken using this method will be presented. Next, an explanation for the difference between the anthropomorphic phantoms used for dosimetry and that used for IQ assessment will be given, along with their advantages and disadvantages.

Finally, studies that used this method for evaluating IQ and radiation dose between and within hospitals will be outlined.

The anthropomorphic phantom study method is based on using anthropomorphic phantoms as a simulation for patients. In terms of dosimetry measurements, the dosimetry phantoms-such as ATOM dosimetry phantoms (CIRS Inc, Norfolk, Virginia, USA)- can be used to investigate the radiation dose using TLD or MOSFET dosimeters placed inside the phantom. They can also provide a measure of the organ dose, or a measurement of other

dose quantities such as IAK, ESAK and KAP. These can be determined using TLDs or solid-state dosimeters placed on the surface of the phantom or from the output of the X-ray machine. Regarding IQ evaluation, images from imaging anthropomorphic phantoms- such as a Lungman chest phantom (Kyoto Kagaku Co., Japan)- can be used to evaluate physical IQ (such as SNR and CNR) or visual attributes (e.g. VGA); the latter are often based on the criteria of CEC (Commission of the European Communities (CEC), 1996a) or through ROC techniques for the overall diagnostic performance. Furthermore, the anthropomorphic imaging phantoms can also be used for measuring IAK, ESAK, and KAP using several methods such as TLD, solid state dosimeters (placed on the surface of the phantom), or the output of the X-ray machine.

It is important to mention that there are some differences between dosimetry and the imaging anthropomorphic phantoms. Dosimetry phantoms were originally designed for dosimetry only and are considered not suitable for IQ evaluation since they do not adequate represent human geometry or anatomy, particularly in the chest region. Furthermore, these phantoms are constructed as a series of axial sections with many small holes distributed within each slice to enable the placement of TLD or MOSFET dosimeters for the purpose of organ dose measurements. These axial sections unfortunately create artefactual shadows that impact the resultant images. In contrast, the imaging phantoms are originally designed for IQ evaluation since they approximate patient anatomy and geometry well. They not comprised of a series of axial sections.

The anthropomorphic phantom study method is considered as a good alternative to studies involving patients and a good approximation of clinical situations with regard to the anatomical structures. These phantoms are considered suitable for optimisation studies and for investigating the radiation dose and IQ among hospitals. This is because they provide a good approximation of patient anatomy and consequently provide excellent radiological image fidelity. They have a 3 dimensional structure which permits the emulation of body size and can demonstrate the three typical chest X-ray positions (posterior-anterior, antero-posterior and lateral projections) (Veldkamp et al., 2009). In addition, these phantoms are controlled in terms of the body size and tissue characteristics, such as presence or absence of disease, and this provides consistency and reproducibility in simulated organ structures. This allows an extremely accurate comparison of IQ (using imaging phantoms) and also radiation dose (using dosimetry phantoms) among hospitals since the phantom has a standard size. Any variations in IQ (using imaging phantoms) and radiation dose (using dosimetry phantoms) should relate to the protocol used and the characteristics of the

imaging systems utilised. A final reason for using anthropomorphic phantoms is that there are no ethical issues for image acquisition and radiation dose measurements; i.e. direct irradiation of human volunteers. Ethical approval would only be required for the visual IQ evaluation.

Imaging and dosimetry anthropomorphic phantoms have been used in the literature for survey and research work for evaluating IQ (imaging phantoms) and radiation dose (dosimetry phantoms) between and within hospitals.

Several studies have been conducted using anthropomorphic phantoms to investigate the variation in radiation dose levels and IQ among different imaging centres (Seo et al., 2014; Hintenlang, Williams, & Hintenlang, 2002; M. D. Cohen et al., 2012).

A survey dosimetry study was conducted by Seo et al. (2014) to investigate the variation in radiation dose for 14 radiographic examinations in the central region of Korea. It included 47 imaging centres using film-screen systems and 74 imaging centres using digital radiography systems. The radiation dose measurements were carried out using an adult anthropomorphic dosimetry phantom (RANDO). Study findings showed a wide variation in radiation dose for both film-screen and DR systems. For the film-screens, the ESD ranged from 0.11 to 0.68 mGy for PA chest projections and from 0.46–6.93 mGy for lateral views. For DR systems, the dose ranged from 0.03 to 0.69 mGy for PA projections and from 0.13 to 11.48 mGy for the lateral projections. The other examinations also had a wide ranging radiation dose. The choice of the phantom used for measuring the variation in radiation dose is optimal because this phantom has a standard size and tissue similar to that of a human's. This increases the study's validity and reliability. However, this could be difficult practically if the study involved paediatric age groups since this would require many dosimetry phantoms to cover all of the age groups. A limitation of this study was that IQ evaluation was not included.

Another study has been conducted by Hintenlang et al. (2002) to investigate the variation from the local standard examination used in each hospital and the effective dose that was delivered to one year old patients for the lateral, PA and AP chest projections in ten hospitals. An anthropomorphic dosimetry phantom replicating a one-year old child was used and the effective dose was calculated from MOSFET dosimeter data. Findings from this study show that the effective dose for the same examination varied between different institutions: AP chest ranged from 0.0027 to 0.059 mSv, PA chest ranged from 0.002 to 0.041 mSv and lateral chest ranged from 0.004 to 0.062 mSv. IQ was assessed both visually (visual grading) and physically (optical density throughout the lung); visual grading was

found to be ‘clinically acceptable’, while the optical density ranged from 0.62 to 2.44 for AP chest, ranged from 0.42 to 2.65 for PA chest and ranged from 0.47 to 1.99 for lateral chest. Although this study has many merits in terms of phantom selection, it was conducted with only a single phantom that replicated a 1 year old child and thus did not cover a range of paediatric age groups. Furthermore, it might be difficult to judge the resultant IQ for the acquired images since the phantom was designed for dosimetry purposes and not for IQ evaluation. The conclusion of the visual IQ is therefore questionable.

In a report by Cohen et al. (2012a), the neonatal chest phantom (Gammex phantom) was used to investigate the variation in radiation exposure for neonatal portable chest images across 4 children’s hospitals in the USA. The variations in radiation dose were evaluated by determining the mean exposure index across the hospitals. The findings of this study show that there was no great variation in the exposure index across sites and there was no site with a mean exposure index twice that or less than half of the other sites. The authors stated that these variations in the exposure index across the sites were related to operational differences, exposure settings and inherent filtration. However, this study was carried out with a limited number of hospitals and any differences in IQ between hospitals were not investigated.

3.4.3 Group Three: Physics Phantom Studies

This subsection starts with a definition of the physical/IQ evaluation and radiation dose measurements. The studies that utilised this approach for evaluating IQ and radiation dose between and within hospitals will be presented. Finally, benefits and limitations of this method will be highlighted.

The physical phantom method is based on using physics phantoms as a method for simulating humans when investigating IQ and radiation dose, such as the CDRAD 2.0 phantom or Leeds TO.10 phantom in combination with PMMA slabs which simulates the thickness of the irradiated area. In terms of dosimetry measurements, the main dose quantities recorded by the physics phantom method are IAK, ESAK and KAP and can be measured using TLDs or solid-state dosimeters that are placed on the surface of the PMMA slabs or acquired from the output of the X-ray machine.

Several survey studies have been carried out using physical phantoms to investigate the variation in IQ and radiation dose among hospitals. Some of these studies were carried out for different examinations/anatomical areas, but not necessarily for the chest; others were carried out using film screen systems or digital imaging systems. As an initial step in this

literary analysis it is valuable to consider all the studies for several reasons. Firstly, to examine protocol and machine types and how they can influence IQ and radiation dose; secondly, to examine IQ evaluation methods when using physical phantoms; thirdly, to determine whether IQ and radiation dose variations exist between hospitals; fourthly, to see whether any correlation exists between IQ and radiation dose between hospitals.

An early attempt to investigate the variation in LCD detectability and radiation dose for PA CXR was conducted by Geleijns, Schultze Kool, Zoetelief, Zweers, & Broerse (1993) in fourteen Dutch hospitals. A LucAl chest phantom (Conway et al., 1984) was used for the dosimetry measurements and two different physical phantoms (Leeds test object TOR CDR test object and CDRAD phantom) were used to investigate the LCD detectability. The images were evaluated visually by observers. This study was carried out using film-screen and only for adults. A large variation in LCD detectability was observed, and the highest value of radiation dose was observed to be 9 times higher than the lowest value. The range of the IAK ranged from 0.028 to 0.260 mGy. Nevertheless, other physical parameters of IQ such as SNR and CNR were not investigated.

A survey was conducted by Van Soldt et al. (2003) in the Netherlands to investigate the variation in LCD detectability and radiation dose for PA CXR using different X-ray units, with both digital and film-screen systems, between different hospitals. A LucAl chest phantom was used to simulate the patients for dosimetry measurements and was accompanied by the CDRAD phantom for evaluating LCD detectability. The CDRAD phantom images were presented as a hardcopy and were evaluated visually by three observers. The findings of this study show that the LCD detectability with digital systems was superior to than that of the film, but its radiation dose was higher (3^{rd} quartile ESD 0.13mGy). They suggested that it is possible for greater dose reduction without losing LCD detectability. However, this study did not include paediatric age groups, nor were physical IQ parameters such as SNR and CNR measured. The correlation between the IQ and dose was also not calculated.

A comparison study by L pez, Morant, Geleijns, & Calzado (2000) was conducted at 5 hospitals in Tarragona, Spain, to investigate the LCD detectability and radiation dose in a number of paediatric age groups for different X-ray examinations (chest, pelvis and abdomen). The PMMA slabs were used to simulate the patients for dose measurements and the Leeds TOR (CDR) phantom was used for measuring LCD detectability which was evaluated visually by three observers. The ESD range was 75–729 µGy, 813–1600 µGy, 980–2300 µGy and 94–250 µGy for pelvis radiographs for the 5 months old and the 5-

year-old, abdomen radiographs for the 5-year-old and chest radiographs for the 5-year-old, respectively. The variation in LCD detectability between sites was also observed. This study was focused only on one age group (5 years) for chest radiographs and the study was conducted using film-screen systems. However, this study was limited by the type of physical phantom used. The Leeds TOR has a crude design and does not have a 4AFC option for LCD detectability evaluation such as the CDRAD 2.0 phantom does. In a 4AFC paradigm the observers are required to determine not only whether the hole is visible or not, but also indicate the position of the corner (between four possibilities) in which the visible hole is present. This reduces the false positive error to 25% (De Hauwere et al., 2005). With a Leeds TOR phantom, observers have prior knowledge about the place of the holes and this prior knowledge will likely influence the image LCD detectability evaluation. Again, the correlation between the IQ and dose was not calculated. Another study was conducted in the Netherlands by Veldkamp et al. (2006), who compared the LCD detectability and radiation dose of 8 different X-ray machines located in 6 hospitals (using CDRAD 2.0 phantom and a LucAl chest phantom). Standard protocols of adult CXR similar to clinical practice for each hospital were employed across all of the machines. The LCD detectability in this study was evaluated visually by 6 observers. The findings show there is a significant difference in LCD detectability and radiation dose between the different X-ray machines. The radiation dose represented by the ED was calculated from ESD using the PCXMC software program. The ED between the X-ray machines ranged from 0.010 mSv to 0.032 mSv. However, the correlation between the LCD detectability and dose was not calculated.

Anja Almén et al. (1996) investigated the variation in radiation dose and LCD detectability in 19 Swedish hospitals for pelvic examinations of a 1 year old child; other age groups and anatomical areas, such as the chest, were not included. The study was carried out using a CD phantom (CDRAD type 1.1) as a physical test object and TLDs were used for measuring the radiation dose to the surface of the phantom. The LCD detectability was evaluated visually by five observers. Study findings showed that there was considerable variation in the protocols used between hospitals. Similarly, the radiation dose ranged from 0.09 mGy to 1.7 mGy (mean 0.65 mGy) and there was no significant difference in LCD detectability for the high doses (0.4-1.7 mGy). Additionally, they observed a weak correlation between LCD detectability and radiation dose. The authors state that there is a large potential for reducing radiation dose by optimising clinical protocols. The study was carried out with a crude test object (CDRAD type 1.1) which does not have a 4AFC option.

Another group of researchers attempted to investigate the radiation dose and IQ for adult chest and abdomen examinations across the UK using a physical phantom (Cole et al., 2009). Fifteen CR X-ray machines for chest examinations and twenty X-ray machines for abdomen examinations were involved in this study and the IQ was represented by calculating the CNR. The phantoms that were used for IQ evaluations consisted of PMMA slabs with a contrast disc located in front of the PMMA phantom and the thickness of the phantoms were 9 and 18 cm for the chest and abdomen, respectively. The findings of this study show that there is a wide variation in ED ranging from 170 to 640 μSv for the abdomen examination. For the chest examination, the ED ranged from 9 to 35 μSv; detector dose varied from 2.4 μGy to 7.8 μGy for the abdomen examination and from 4.6 to 19.0 μGy for the chest examination. Furthermore, there was a wide variation in the clinical protocols used between sites, and the researchers indicated that many of the X-ray machines had not been optimised. This study was carried out for adult examinations only and paediatric groups were not included. Furthermore, only CR systems were included. Other limitations are that the phantom used for IQ evaluations was extremely crude, only one parameter of physical IQ (CNR) was evaluated, and other parameters such as SNR and the LCD detectability were not investigated. In addition, the number of X-ray machines would be not sufficient to be representative of the whole UK.

It should be noted that in the literature reviewed so far, no study has linked IQ to the radiation dose between hospitals and the evaluations of IQ and radiation dose have been presented separately. To facilitate the comparison of the reviewed literature mentioned above, **table (3-1)** summarises the prime elements of the studies that used physical phantom methods.

The physical phantom method has several benefits compared with patient studies and anthropomorphic phantom methods. When utilising this method, it can be possible to simulate all the desired age groups by increasing or decreasing PMMA slab thicknesses below and above the physics phantom. Furthermore, it is possible to evaluate the IQ represented by the LCD detectability physically using a software analyser and this method has high reliability, time saving capabilities and extremely useful for purpose of comparing IQ within and between hospitals. Nevertheless, physical phantoms have a uniform background that does not take into account anatomical structure noise.

Table 3-1 This table summarises the key elements of the previous studies that used physical phantoms.

Authors (years)	Study details	Findings / Conclusion
Gleans, J et al. (1993).	Study area: The Netherlands Survey type: radiation dose & IQ survey Phantom type: CDRAD phantom, Leeds TOR and LucAl chest phantom Examination type: Adults PA CXR Imaging system type: film-screen IQ evaluation type: visual by 2 observers	1-High variation in IQ 2-Dose was observed to be 9 times higher than the lowest value, and the IAK ranged from 0.028 to 0.260 mGy. 3- No correlation observed between IQ and dose.
Van Sold, R. T. M. et al (2003).	Study area: The Netherlands Survey type: radiation dose & IQ survey Phantom type: CDRAD phantom & LucAl chest phantom Examination type: Adults PA CXR Imaging system type: film-screen and digital systems IQ evaluation type: visual by 3 observers	High variation in IQ and dose, and the third quartile of ESD = 0.13 mGy.
Lopez, M. et al. (2000).	Study area: Tarragona/ Spain Survey type: radiation dose & IQ survey Phantom type: Leeds TOR Examination type: Paediatric for Chest, pelvic and Abdomen Imaging system type: film-screen IQ evaluation type: visual by 3 observers	High variation in IQ and radiation dose, and the ESD for CXR ranged 94–250 µGy.
Veldkamp et al. (2006).	Study area: Tarragona Survey type: radiation dose & IQ survey Phantom type: CDRAD phantom & LucAl chest phantom Examination type: Adult CXR Imaging system type: digital systems: CR and DR IQ evaluation type: visual by 6 observers	A significant difference in IQ and radiation dose were observed, and the effective dose ranged from 0.010 mSv to 0.032 mSv.

Almen et al. (1996).	Study area: Sweden Survey type: radiation dose & IQ survey Phantom type: crude CD phantom Examination type: pelvic of one-year age Imaging system type: film-screen IQ evaluation type: visual by 5 observers	1-There is considerable variation in protocols used among hospitals. 2-Radiation dose was ranged from 0.09 mGy to 1.7 mGy. 3-There is no significant difference in IQ for the high doses (0.4-1.7 mGy). 4-A weak correlation between IQ and dose.
Cole, Mackenzie et al. (2009)	Study area: UK Survey type: radiation dose & IQ survey Phantom type: PMMA phantom with a contrast disc Examination type: adult's chest and abdomen Imaging system type: Digital system: CR only IQ evaluation type: CNR was measured physically	1-Effective dose ranged 170 to 640 μSv for the abdomen examination and ranged from 9 to 35 μSv for the chest examination. 2-A wide variation in protocols used among sites was observed.

3.5 Applying FOM Concept for Evaluating and Comparing Radiation Dose and IQ between and Within Hospitals:

DRLs are normally used as a method for monitoring the radiation dose from X-ray imaging procedures (D'Ercole, Thyrion, Bocchiola, Mantovani, & Klersy, 2012;Aroua et al., 2007). However, DRLs were not primarily intended as a guide for optimisation and they only consider the radiation dose and not IQ. Practically, DRLs are helpful for narrowing the likely variation in radiation dose within and between hospitals, but they cannot identify and control variations in IQ. The reason behind this is that reports in the literature demonstrate no straightforward correlation between IQ and radiation dose (Almén et al., 1996; Geleijns et al., 1993; Struelens et al., 2008; Veldkamp et al., 2006). On the other hand, several studies (see section 3.4) have attempted to investigate variations in IQ and the patient radiation dose between hospitals. However, these studies presented variations in IQ and radiation dose separately, thereby failing to combine them as a single component. This makes any comparison of an X-ray unit's performance and determining the optimal protocol for imaging more difficult. Ultimately, there is always a trade-off between the radiation dose and IQ by which no one factor should be analysed separately from the other. A FOM is an extremely useful indicator for helping to identify the acquisition parameters for which the IQ is highest. The lowest possible radiation exposure is normally known as the endpoint of optimisation (Samei, Dobbins, Lo, & Tornai, 2005). FOMs attempt to combine IQ and radiation dose and can be calculated from the square of the IQ score divided by the dose. A higher value of FOM for a particular imaging system indicates a better performance and vice versa. FOM can be used to identify optimal imaging techniques that produce an appropriate IQ with a low dose (Mraity et al., 2015).

Utilising an FOM offers an extremely valuable and attractive way of comparing IQ and radiation dose for digital X-ray systems since it considers both IQ and radiation dose together. The FOM is commonly offered as a single number that reflects the efficiency of an imaging system operating under specific conditions (Borg, Badr, & Royle, 2012). According to the literature reviewed in this PhD thesis (see section 3.4), an FOM has never before been used for purpose of evaluation and comparing IQ and patient dose, between and within hospitals.

A FOM could potentially be useful in optimising the X-ray protocols and equipment, both between and within hospitals.

From the literature reviewed concerning optimisation, it has been identified that the common formulas for FOM when using LCD detectability with a CD phantom, such as a CDRAD 2.0 phantom, are: -

$$FOM = \frac{SNR^2}{dose} \text{ and } FOM = \frac{CNR^2}{dose}$$

(Bosmans et al., 2012a; Dragusin et al., 2008; Kepler & Vladimirov, 2008).

However, these two formulas use SNR and CNR as a metric. These are self-limiting in that they do not take object size into account (pathology or anatomical structure). This was found to be very important in detection tasks undertaken in clinical practice (McAdams et al., 2006).

3.6 Correlation between LCD Detectability and the Visual Image Quality in CXR

This section starts with an introduction to the characteristics of the chest region for conventional radiography examinations when compared with other body regions. These characteristics provide a link to the type of phantom that should be used to simulate this region for image evaluation purposes. The critical review of the studies looks at the correlation between LCD detectability as represented by the IQF_{inv} using a CDRAD 2.0 phantom. The visual IQ from anthropomorphic phantoms and clinical images (acquired using patients) will also be presented.

3.6.1 Introduction

The zones of the human body can be classified into 2 categories: the quantum-limited region (where the detectability of the lesions in this region is mainly affected by the quantum noise) and the quantum-saturated region (where the anatomical structures are the main factor that influence lesion detection when compared with that of the quantum noise). Medical images tend to sit in one of these two regions, and chest radiography is located in the latter region (Sund, Båth, Kheddache, & Månsson, 2004).

Clinically, the noise of the anatomical structures in the chest zone is considered the major limiting factor that influences the relevant detection tasks and is more important than quantum noise. Several studies have investigated this (Samei et al., 1999). By way of an example, the overlap of the ribs can influence the detection of lung nodules. This is because the ribs overlay about 75% of the area of the lungs (McAdams et al., 2006). Also, a fundamental part of the lungs is projected above the heart and diaphragm and this has

influence on the contrast of lung lesions. The appearance of lung lesions on CXR images is of great concern and the appearance of these lesions against the anatomical background leads to a decrease lesion conspicuity. This can cause approximately 30% of pulmonary nodules being missed on the initial CXR, although they may be detected retrospectively (McAdams et al., 2006). For instance, the nodules that have a small diameter of around 3 mm have to be visible on CXR, even with existence of scattered radiation. Nonetheless, detectability of the nodules that have diameters smaller than around 8 mm is rare if they are influenced by the anatomical noise (McAdams et al., 2006). A study was conducted by Håkansson, Båth, Börjesson, Kheddache, Grahn, et al. (2005) demonstrated that the quantum noise and noise from the detector in CXR does not limit the detectability of a nodule with a diameter equal to 10 mm in any section of the chest area, including the heart and mediastinum.

Several studies (Månsson, Båth, & Mattsson, 2005; Veldkamp et al., 2009) reported that anthropomorphic phantoms are considered to be the closest approximation to clinical reality and that optimisation studies should be achieved either using real patients or anthropomorphic phantoms with high fidelity. In contrast with CD phantoms which have a uniform background, such as the CDRAD 2.0 phantom and Leeds TO.10 test phantom, the anthropomorphic Lungman chest phantom represents an excellent approximation of the anatomical background.

The validity of using physical phantoms with a uniform background, such as CD phantoms, to represent the chest region for the purpose of IQ evaluation or for optimisation studies could be questionable.

To summarise, the selection of the type of phantom for IQ evaluation and optimisation studies for X-ray chest examinations has a large influence on findings and study validity.

3.6.2 Validity of using the CDRAD 2.0 Phantom for Quantifying the Visual Image Quality of Chest Radiographs

The LCD detectability method, using a CD phantom such as the CDRAD 2.0 phantom, has been frequently reported in the literature as a method for comparing the performance of imaging systems and for optimisation studies (Alzimami et al., 2009; Bacher et al., 2003; Geijer et al., 2001; Rong et al., 2001; Veldkamp et al., 2006). A benefit of this method is that it evaluates the entire imaging chain, including the visual assessment using observers (Veldkamp et al., 2009). However, the LCD detectability method using physical phantoms, such as the CDRAD 2.0 phantom has been criticised in the literature. It depends on homogeneous phantoms and does not consider the influence of anatomical noise which

arises from human anatomy (Månsson et al., 2005). The principle work of these phantoms is based on the Rose model which assumes that quantum noise is the limiting factor that can influence IQ (LCD detectability) (Rose, 1948). This assumption is true for the CDRAD 2.0 phantom, but it could be limited for several common tasks in diagnostic radiology (Månsson et al., 2005). Clinically, the limiting factor for observer performance in LCD detection results from the anatomical background rather than the quantum noise (Samei et al., 1999). Consequently, the validity of only using LCD detectability with a CDRAD 2.0 phantom for visual IQ evaluation, visual LV evaluation and optimisation studies is questionable and needs further investigation (Månsson et al., 2005).

To achieve the aim of this thesis, it was necessary to investigate the validity of utilising LCD detectability using the CDRAD 2.0 phantom for visual IQ evaluation, visual LV evaluation and its validity in CXR optimisation studies. To understand this, the main questions that need to be answered are:

1. **Is there a correlation between the physical measure of LCD detectability (using a CDRAD 20 phantom) and visual measures of IQ resulting from the anthropomorphic chest phantom?**
2. **Is there a correlation between the physical measure of LCD detectability (using a CDRAD 2.0 phantom) and the visual measure of LV in an anthropomorphic chest phantom?**
3. **Is the physical measure of LCD detectability (using CDRAD 2.0 phantom) valid for use in CXR optimisation studies?**

Without answering the questions above, it is difficult to achieve a number of thesis objectives.

To answer these questions, an analysis of the studies that have tried to investigate the consistency between the LCD detectability, using a CDRAD 2.0 phantom, and visual IQ was performed.

Only 3 studies have tried to describe the consistency of IQ behaviour between the CD phantom, which is represented by the CDRAD 2.0 phantom, and those from anthropomorphic phantoms or clinical images (Geijer & Persliden, 2005; Bacher et al., 2006; De Crop et al., 2012). These 3 studies suffer from similar limitations. They have been conducted using ‘normal’ anthropomorphic phantoms or ‘normal’ patients where no pathology was present. The inclusion of a pathology provides the opportunity for a more accurate evaluation of IQ in the detection task. In addition, including a pathology is

important in optimisation studies; for instance, a study by Anders et al., (2005) showed that the detectability of a pathology under specific levels of quantum noise was not changed, while the score of IQ using VGA method had been influenced significantly. Further reductions in radiation exposure may be achieved based on the score of pathology detectability rather than that from the VGA method. Secondly, in 2 studies (Bacher, Smeets, Vereecken, et al., 2006; De Crop et al., 2012) , the free adjustment on image viewing parameters was applied and the observers were allowed to adjust display parameters such as brightness, contrast and magnification. This choice of image viewing display might not be acceptable for this type of study (one which aims to investigate the correlation between the LCD detectability and visual IQ evaluation using clinical images or anthropomorphic phantoms), and therefore this method has high validity but lower reliability. The valid choice of an image viewing display here should have been consistent and not subject to free adjustment; in other words, the observers should not have been allowed to adjust the brightness, contrast and magnification of the image, as this increases the reliability. The reliability is more important than the validity in achieving the aim of this part of the study. A consistent display of images is necessary to ensure that any variation (in IQ assessment) between the images is only caused by the imaged object and not by the adjustment of the image viewing display such as by zooming or windowing. This is necessary to have a valid correlation result between the LCD detectability (using CDRAD 2.0 phantom) and visual IQ measured from patients/anthropomorphic phantoms.
A study by Geijer & Persliden (2005) investigated the influence of different tube potentials (ranging from 48 to 125 kV) on IQ for anteroposterior lumbar spine radiography. The mAs was adjusted to keep a constant value for the effective dose (0.11 mSv). 2 different phantoms were used: a CD phantom (CDRAD 2.0 phantom) and an adult anthropomorphic phantom (Alderson phantom). The LCD detectability, represented by the IQF_{inv} and using a CDRAD 2.0 phantom, was evaluated physically using a software analyser. The Alderson phantom images were assessed visually by 8 observers (radiologists) using a relative VGA method. The findings of this study demonstrate higher IQ scores for high kVp values (≥96 kVp) for both phantoms. In a study conducted by Bacher, Smeets, Vereecken, et al., (2006), they compared the variation in IQ and radiation dose for CXR between 2 different detector types: an amorphous silicon and amorphous selenium flat-panel detector. The comparison was performed using two different methods: one was done clinically, with a group of patients; the other was done with a CD phantom (CDRAD 2.0 phantom). The clinical part of the study showed that there was a significant decrease in effective dose when using the

amorphous silicon flat-panel system, compared with that of the amorphous selenium flat-panel system, without any reduction of IQ as evaluated using the absolute VGA method by five observers (experienced radiologists). The later finding was confirmed by the second part of the study using the CDRAD 2.0 phantom. Images were acquired using the CDRAD 2.0 phantom on both systems, at various dose levels. The images were evaluated visually by 5 observers. However, in these 2 studies the correlation between the IQ from the Alderson phantom, the images from patients and the IQF_{inv} from the CDRAD 2.0 phantom was not calculated.

In 2012, the correlation between clinical visual IQ evaluation for CXR using 3 cadavers and LCD detectability represented by the IQF_{inv} using the CDRAD 2.0 phantom were investigated by De Crop et al. (2012). The images were acquired using AEC. SID was equal to 150 cm, and various values of tube voltage (70, 81, 90, 100, 113, and 125 kVp) and additional beam filtration (0.1, 0.2, and 0.3 mm Cu) were utilised. The acquired images from the 3 cadavers were evaluated visually using an absolute VGA method by four experienced radiologists. The LCD detectability from the CDRAD 2.0 phantom was evaluated physically using the CDRAD analyser software and visually using 4 observers (medical physicists). Good correlation ($r=0.916$; $p<0.001$) was identified between the physical evaluation of IQF_{inv} and the visual IQ using the absolute VGA method on the cadaveric images. This study can be considered as a pivotal study in investigating the correlation between LCD detectability using CDRAD 2.0 phantoms and visual IQ evaluation from clinical images when compared with the previous two studies (H. Geijer, 2005; Bacher et al., 2006). De Crop et al. (2012) used a reliable, valid and controlled method and successfully calculated the correlation between the two different types of IQ. However, in addition to the 2 common limitations previously listed, further limitations can be suggested. Firstly, this study used the absolute VGA score rather than the relative VGA score. The latter could be less biased and more sensitive in its detection of subtle differences in IQ and this is related to present of the reference image. Secondly, the selected values of kVp and beam filtration when using an AEC are unlikely to reflect the full range of IQ that could be expected clinically. Therefore, the resultant correlation between the physical values of IQF_{inv} and the visual IQ (De Crop et al., 2012) does not cover the whole range of IQ that could be expected clinically.

The 3 studies evaluated above (Geijer, 2005; Bacher et al., 2006; De Crop et al., 2012) have not examined the consistency and the correlation between the LCD detectability using a CDRAD 2.0 phantom and the visual LV from either the phantom or clinical images.

3.7 Gap in the Literature which this Thesis seeks to address

This thesis will consider the limitations in the literature regarding the previous studies comparing IQ and radiation dose for CXRs among hospitals (see section 3.4.3.) which used physical phantoms. The proposed actions this thesis will take to address these limitations are as follows: using a novel method to evaluate and compare radiation dose and IQ between X-ray units based on an FOM concept instead of measuring only the radiation dose (DRL concept) or only presenting the results of radiation dose and IQ separately without linking them together; using a novel FOM formula that utilises CD as metric for IQ evaluation to avoid the highlighted limitations of the existing formulas of FOM that utilise SNR and CNR as a metric; using the physical evaluation method for LCD detectability (CDRAD analyser software) for a CDRAD 2.0 phantom image evaluation, because it is more reliable for comparing IQ among hospitals due to its repeatability and precision; covering a range of paediatric age groups (neonate,1 year,5 years and 10 years of age) and adults for CXR; investigating the variation of other physical IQ parameters such as SNR, CNR and CI and not just the LCD detectability.

Anthropomorphic phantoms will be included in this thesis to simulate neonate and adult patients. These phantoms will be necessary to achieving some objectives of the current study, including: to investigate the influence of phantom thickness on IQ and radiation dose for adult CXR examination across the different hospitals; to investigate the variability in IQ, LV and radiation dose between and within hospitals; to investigate the variability in visual IQ for neonates, between and within hospitals. This was why the combination of using anthropomorphic phantoms and physical phantoms are considered to be the optimum choice for this thesis.

Regarding the literature that relates to the correlation between LCD detectability using a CDRAD 2.0 phantom and visual IQ/visual LV in CXR (presented in section 3.6.2.), this thesis will consider the reported limitations. The following actions will be taken within to address these limitations: 1) investigating the validity of using LCD detectability using CDRAD 2.0 phantom for visual IQ evaluation by determining the correlation between the physical evaluation of IQF_{inv}, which results from a CDRAD 2.0 phantom, and the scores of the visual IQ evaluation using a relative VGA method which results from

anthropomorphic Lungman chest phantom images. 2) investigating the validity of using LCD detectability, using a CDRAD 2.0 phantom, for visual LV evaluation by determining the correlation between the physical evaluation of IQF_{inv}, resulting from the CDRAD 2.0 phantom, and the scores of the visual LV evaluations, using a relative VGA method for two selected lesions generated from a Lungman chest phantom.3) investigating the validity of using LCD detectability, using the CDRAD 2.0 phantom, in CXR optimisation studies.

3.8 Chapter Summary

The LCD detectability method, using the CDRAD 2.0 method, was found to be extremely reliable and applicable for IQ evaluation between hospitals when compared with other methods that used patients or anthropomorphic phantoms. Practically, the main limitation for the latter method (using anthropomorphic phantoms) was lack of phantoms that cover the different age groups. The main limitations for the former method (using patients) were the ethical issues and the variation of patient size and characteristics. From the reviewed literature, it was found that the IQ and radiation dose have not been linked together during comparisons between hospitals and were always presented separately. Also, the concept of FOM has not been utilised for the purpose of evaluating IQ and radiation dose between hospitals, and this is the focus of this thesis.

This chapter has indicated that there is no current study which has attempted to investigate the validity of using LCD detectability, using a CDRAD 2.0 phantom, in visual LV evaluation. In addition, the validity of using LCD detectability in visual IQ evaluation is not well investigated. Therefore, one of the decided objectives of this thesis was to focus on the necessity of proposing and implementing a new method for achieving the validation of using LCD detectability in optimisation studies and in visual IQ and LV evaluation.

Chapter 4: Material and Methods

4.1 Chapter Overview

Focusing on paediatric and adult CXR, this chapter outlines a novel method for evaluating and comparing radiation dose, IQ and clinical protocols for a range of X-ray machines located in several different hospitals. To achieve this, the FOM was used for a combination of commercially available phantoms. These phantoms include CDRAD 2.0, Lungman and Gammex. The FOM was utilised for assessing the differences in radiation dose and IQ for CXR in adults and paediatrics. Prior to the start of data collection within the hospitals, an experiment was undertaken to determine the validity of using LCD detectability, with the CDRAD 2.0 phantom, for evaluating visual IQ and LV, as well as to determine its validity for use in optimisation studies. This was necessary to ensure that the main method (LCD detectability) and phantom (CDRAD 2.0 phantom) used in this thesis were valid for achieving the aims.

This chapter contains a detailed explanation of the methods used in this thesis. It is divided into three sections in order to aid understanding: 1) an assessment of IQ, radiation dose and clinical protocols; 2) an assessment of the validity of using LCD detectability and the CDRAD 2.0 phantom for visual IQ and LV evaluation, and for the CXR optimisation studies; 3) statistical analyses. The first section (section 4.2.) focuses on the method used to evaluate the variability of IQ, radiation dose and clinical protocols among X-ray machines. The second section (section 4.3.) focuses on the method used to evaluate the validity of utilising LCD detectability using a CDRAD 2.0 phantom. The third section (section 4.4.) describes the statistical analyses that were performed on the data.

To achieve the aims of this thesis, precise measurements of the radiation dose received by the chest for the different age groups, and assessment of their resultant image qualities, were required. For this, a combination of physical and anthropomorphic phantoms were used together with a solid-state dosimeter (RaySafe X2). A flowchart (**Figure (4-1))** illustrates the main stages of the methods used for the evaluation of IQ, radiation dose and the clinical protocols for the X-ray machines (section 4.2.). **Figure (4-2)** illustrates the main stages of the methods (section 4.3.) utilised for achieving the validation of utilising LCD detectability with the CDRAD 2.0 phantom.

Section 4.2. is divided into seven subsections. First is the participating X-ray units subsection. This illustrates the number of hospitals and the technical characteristics of the

X-ray machines that were used. Second is the QC tests subsection which describes the QC tests that were conducted on the X-ray machines prior to the experimental work. Third is a discussion of the phantoms used for IQ evaluation, including a justification for using anthropomorphism phantoms and the CDRAD 2.0 phantom; this also contains details of PMMA thickness, and the characteristics of the anthropomorphic phantoms which were used. Fourth is the radiation dose measurement subsection describing the methods that were utilised for dose measurements. Fifth is the procedure and image acquisition subsection which demonstrates the stages that were used for image acquisition, including the phantoms that were used for image acquisitions, the coding of the acquired images and the mechanism for recording information. Sixth is the IQ evaluation subsection describing the different methods that were used for IQ evaluation. This includes descriptions of the visual IQ (observer) method used for evaluating the anthropomorphic phantom images, the criteria for image evaluation and the how the observers were selected. Also within this subsection are explanations of the physical measurements of IQ for the anthropomorphic phantoms and CDRAD 2.0 phantom images, represented by SNR, CNR and CI. In addition, the physical calculation of LCD detectability for CDRAD 2.0 phantom images, as represented by the IQF_{inv}, is justified and explained along with the selection of the input parameters for the CDRAD analyser software. Finally, the seventh subsection explains FOM and presents different formulas of FOM that will be used with the anthropomorphic phantoms, and the novel formula for FOM that will be used with CDRAD 2.0 phantom images. A justification for using this formula will also be given.

Section 4.3. contains information about the design of the experiment used for validating the LCD detectability using a CDRAD 2.0 phantom, the characteristics of the X-ray machines, information about the phantoms, acquisition parameters used to acquire images, and the rationale for selecting the given acquisition parameters. Furthermore, the IQ and LV evaluation methods, which were conducted for Lungman chest phantom images, are explained.

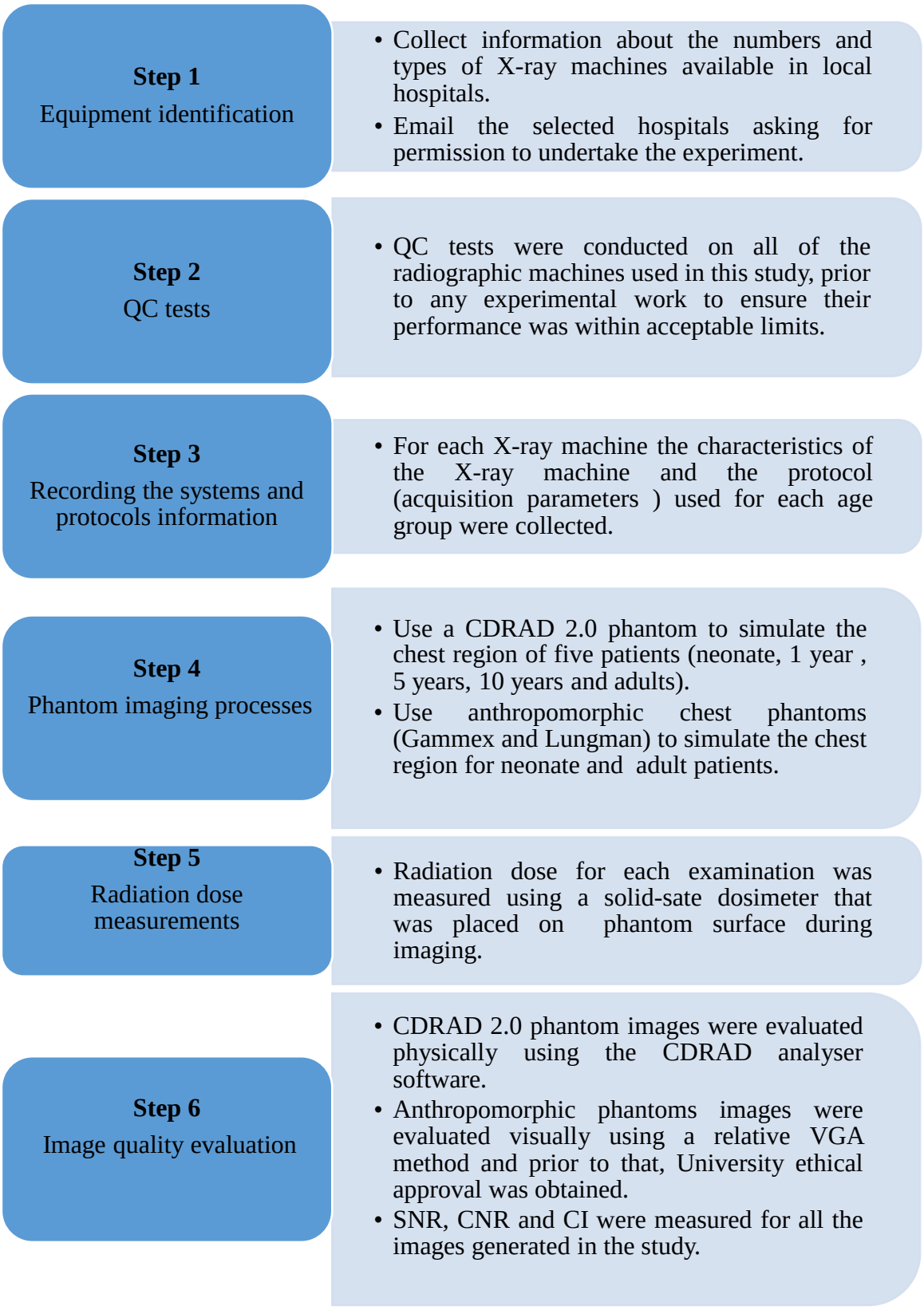

Figure 4-1 Flowchart demonstrating the stages of the first section of the method chapter (section 4.2.), focusing on assessing the likely difference in dose, IQ and clinical protocols between and within hospitals.

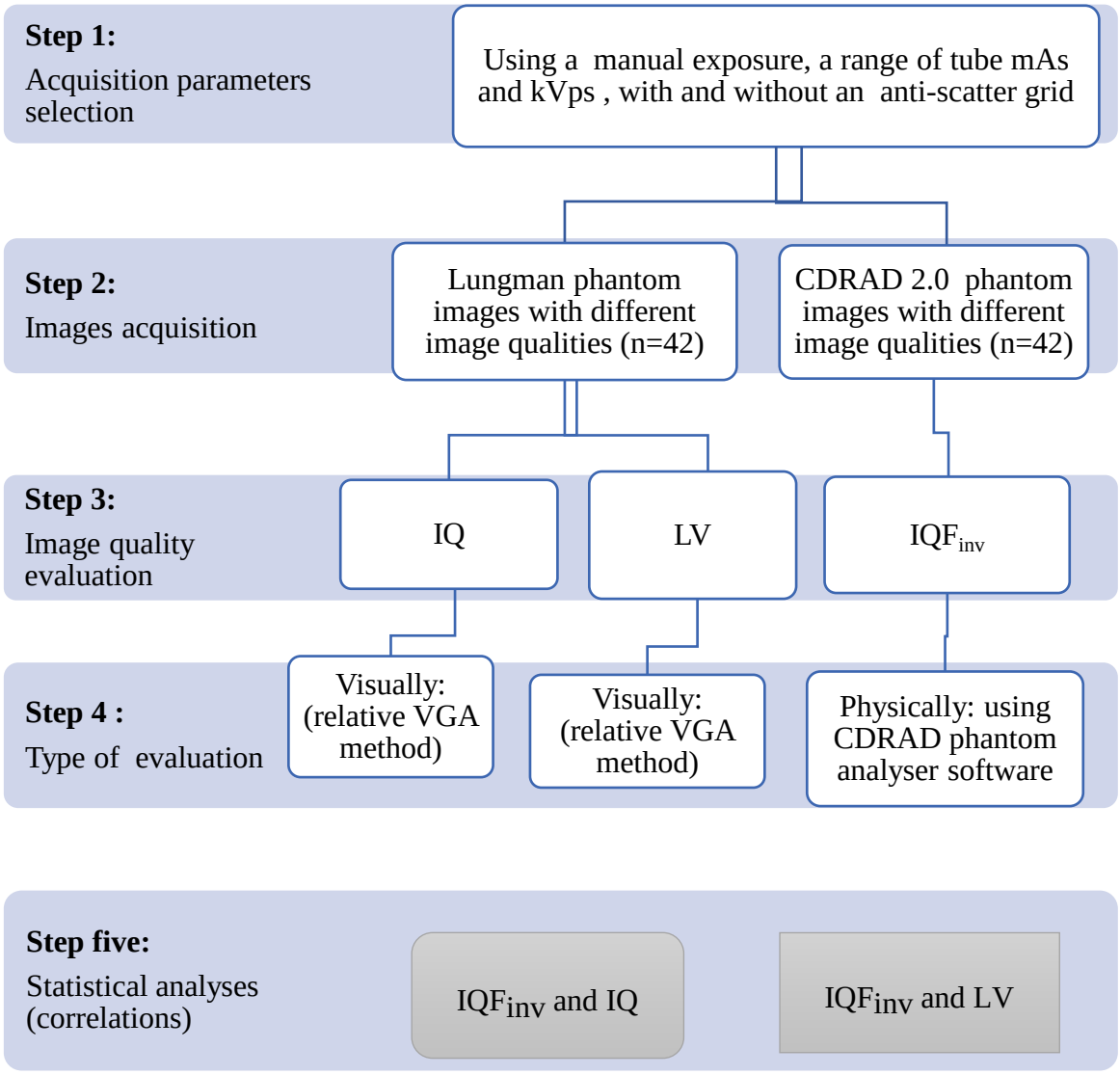

Figure 4-2 This diagram demonstrates the stages of the second section of the method chapter (section 4.3.) which focuses on the empirical work designed for validating LCD detectability using a CDRAD 2.0 phantom in IQ, LV evaluation and optimisation studies.

4.2 Assessment of Image Quality, Radiation Dose, and Clinical Protocols for the range of X-ray machines

4.2.1 Participating X-ray Units

Data collection started in December 2016 and continued until December 2017. The aim was to collect data from 8 hospitals and a total of seventeen X-ray machines. The X-ray machines were located throughout the Northwest of England. 8 NHS hospitals were selected to include machines from as many different manufacturers as possible and to reflect a good geographical spread along with a range of protocols that are currently utilised in paediatric and adult CXR. Seventeen X-ray machines were selected and included CR systems, DDR systems, IDR systems, mobile machines and stationary machines. **Table (4-1)** illustrates the participating hospitals, information about the X-ray machines, including machine type and their technical characteristics.

Table 4-1 This table demonstrates details of the equipment used from the participating hospitals and their X-ray machines alongside information regarding the manufacturer and the technical characteristics of the X-ray machines and their detectors.

Hospital number	X-ray machine number	Equipment type	Tube manufacturer	Detector information	
				Detector type	Manufacturer
1	1	Static	Philips	IDR	Philips
2	1	Mobile	Carestream	IDR	Carestream
	2	Static	Siemens	IDR	Siemens
	3	Static	Carestream	DDR	Carestream
	4	Static	Samsung	IDR	Samsung
3	1	Static	Siemens	IDR	Siemens
	2	Static	Siemens	IDR	Siemens
	3	Static	Siemens	IDR	Siemens
	4	Static	Siemens	IDR	Siemens
	5	Static	Siemens	IDR	Siemens
4	1	Static	Philips	DDR	Philips
	2	Static	Philips	CR	Carestream
5	1	Static	Siemens	CR	Carestream
6	1	Static	Philips	DDR	Philips
	2	Static	Philips	DDR	Philips
7	1	Static	Carestream	DDR	Carestream
8	1	Static	Siemens	CR	Carestream

4.2.2 Quality Control (QC) Tests

The optimum performance of an X-ray system is assured by planned testing and maintenance programs; QC is an integral part of these processes (Institute of Physics and Engineering in Medicine, 2005). QC tests were conducted on all X-ray machines used in this study prior to the experimental work to determine any errors, and to ensure that their technical performance fell within expected limits. The QC tests are outlined in the recommendations of the Institute of Physics and Engineering in Medicine (IPEM) (Institute of Physics and Engineering in Medicine, 2005). Tests typically include an assessment of radiation dose output variation with kV, radiation dose output variation with mA, radiation

dose output variation with time. and the overall reproducibility. A QC data collection form for the X-ray machines is shown in **Appendix A-1**. The results indicated that the machines were fit for routine clinical use; consequently, all the machines were included in the study.

4.2.3 Phantoms Used for Image Quality Evaluation

4.2.3.1 CDRAD 2.0 Phantom

The choice of phantom type for both IQ evaluations and optimisation studies can have significant effects on the results of any study. Anthropomorphic phantoms are the preferred choice for many optimisation studies because they provide a good approximation of patient anatomy and have excellent radiological fidelity. However, these phantoms are not commercially available for all age groups and there is a particular lack of these phantoms for paediatrics. Commercially, there are a group of available paediatric dosimetry phantoms (ATOM dosimetry phantoms) which simulate four paediatric ages: newborn, one, five and ten years old (Varchena, 2002). However, these phantoms have many limitations. They have been designed for dosimetry only and may not be suitable for IQ assessments as they do not provide a level of realism sufficient for representing human anatomy, especially for the chest region. In addition, these dosimetry phantoms are produced as a series of axial sections and the holes specified for TLD placement generate shadows on the resultant X-ray images from the gaps between the phantom slices. Furthermore, the geometry of these phantoms, particularly for chest region, is limited compared with that of the imaging phantoms. The final apparent limitation is that these phantoms have been constructed in a way which does not allow the researcher to include any simulated lesions or pathology inside the lung fields.

There is therefore an apparent lack of commercially available anthropomorphic paediatric phantoms. There are only 2 phantoms available to simulate the paediatric chest. These include the Lungman, by Kyoto Kagaku (LUNGBOY PH-1C Multipurpose Paediatric Chest Phantom; Kyoto Kagaku Company Ltd, Japan), which is designed to represent a 5-year-old child, and the Gammex 610 neonatal phantom. However, the cost of the first one is high and not available locally. The second has been designed to simulate only the neonatal chest region and as such it was used in this thesis.

Since this thesis focuses on all paediatric and adult groups, it was necessary to take into account the limitations of using only the currently available anthropomorphic phantoms and provide a method for covering all of the paediatric and adult age groups. The choice of using physical phantoms, such as the CDRAD 2.0 phantom, has been supported by the

previous survey studies that used LCD detectability using a CDRAD 2.0 phantom to compare IQ and radiation dose between hospitals (Almén et al., 1996; Geleijns et al., 1993; Van Soldt et al., 2003; Veldkamp et al., 2006). Also, this phantom has been utilised extensively in the literature for optimisation studies and comparing the performance of imaging systems (Alzimami et al., 2009; Bacher et al., 2003; Geijer et al., 2001; Rong et al., 2001; Veldkamp et al., 2006).

Several benefits were obtained in this thesis from implementing the LCD detectability using a CDRAD 2.0 phantom as the main method for IQ evaluation. Firstly, by using this method it was easy to simulate all the required age groups by increasing or decreasing the PMMA slab thicknesses. This was one of the main reasons for using LCD detectability and the CDRAD 2.0 phantom, since the aim of this thesis focused on covering all of the paediatric age groups and also including adults. It was therefore impossible to achieve this aim using current anthropomorphic phantoms because there were none available commercially to cover all the paediatric age groups. The second advantage of choosing to use LCD detectability with a CDRAD 2.0 phantom is the possibility of using the automatic analyser computer software for evaluating IQ physically. This provides a mechanism for extremely reliable comparisons for IQ between and within hospitals. Furthermore, this method is both easy and quick to use which is necessary for such survey studies as this, due to the large amount of data that is expected to be collected. Finally, an experiment was undertaken in section 4.3. prior to data collection which established that LCD detectability with the CDRAD 2.0 phantom is valid for IQ and LV evaluation; it was later concluded, in this thesis, that it was valid to be used in optimisation studies. A description of the CDRAD 2.0 phantom, its characteristics and its principle work is presented in detail in chapter 2, subsection 2.6.3.

The CDRAD 2.0 phantom was used as a chest phantom simulation for five age groups: neonate, 1-, 5-, 10-years olds and adult. These ages have been selected for three reasons: 1) medically, they cover the majority of the paediatric groups (neonate, 1, 5, 10 years) (Varchena, 2002) and adult groups; 2) within each age group, the size variation of the internal organs are no more than 15% (Varchena, 2002) and it is therefore not worthwhile selecting ages with relatively similar organ dimensions as this would not allow for any meaningful comparisons; 3) these age groups are identical to the paediatric mathematical phantoms which are frequently utilised in Monte Carlo programs for dosimetry calculations (Tapiovaara & , Siiskonen, 2008). The CDRAD 2.0 phantom was placed between different

thicknesses of medical grade plastic PMMA slabs. The number of slabs varied according to the age of the patient under simulation in order to simulate the thickness of the irradiated area of the patient.

According to the literature (Brosi et al., 2011; Ween, Olstad, Jakobsen, & Olsen, 2009) and the recommendations from the CDRAD 2.0 manufacturer (Burght et al., 2014), there is widespread agreement about the use of PMMA slabs alongside the CDRAD 2.0 phantom to simulate the thickness of the irradiated area of the patient. However, most of these reports have not provided any justification for selecting the thickness of the PMMA needed for each age group. As a result, there is wide variation in the PMMA thicknesses used for specific age groups between studies. In this thesis, the PMMA thicknesses for each standard age group were estimated by applying a conversion factor. This method allows the estimation of the PMMA thicknesses required for any region of human body and is based on its overall thicknesses. According to this principle, each 1 cm of PMMA equates to a 1.5 cm thickness of the chest region (Rassow, Schmaltz, Hentrich, & Streffer, 2000; Vano et al., 2008). The chest thickness of the patient for each age group can be determined using the following equation (Hart, D. W. B. F., Wall, B., & Shrimpton, 2000):

$$(ChestAPthickness) = 8.47 * (w|h) + 17.51 * (w|h)^{0.5} + 4.21$$

Where (w) and (h) represent the weight and height of the patient, respectively.

Several studies have reported clinical data about the weight and height differences between age groups of patients, but there is no consistency between them (Billinger, Nowotny, & Homolka, 2010b; International Atomic Energy Agency, 2013; Kiljunen, Tietäväinen, Parviainen, Viitala, & Kortesniemi, 2009). Within this work, the weight and height for each age group was obtained based on the dimensions of ATOM dosimetry phantoms and the mathematical phantom used in the PCXMC dosimetry software (Tapiovaara & , Siiskonen, 2008). Within these phantoms the values represent the average for each age group. This increases the reliability of the dose estimates as the thickness is more realistic for any corresponding age group and avoids the variations seen within the literature. However, it was necessary to select one of these two reference phantoms (ATOM dosimetry phantom) and the computer-based phantom (PCXMC) as a baseline for estimating the thicknesses of PMMA as a result of the relatively small variations in their thicknesses. Consequently, the thicknesses within the PCXMC phantom were selected because they are the same thicknesses that are used to calculate the effective dose. The chest thickness of each age

group (according to the ATOM dosimetry phantom and the PCXMC phantom) and their equivalent PMMA thickness are described in **Table (4-2)**.

Table 4-2 The thickness of the chest area for paediatric age groups and the adults (according to the ATOM dosimetry phantom and the PCXMC phantom) and their equivalent PMMA thickness.

Age groups	Chest AP thickness (cm)		Equivalent PMMA thickness (cm)	
	ATOM phantom	PCXMC phantom	ATOM phantom	PCXMC phantom
Neonate	9.00	9.80	6.00	6.53
One year	12.00	13.00	8.00	8.67
Five years	14.00	15.00	9.33	10.00
Ten years	17.00	16.80	11.33	11.20
Fifteen years	-	19.60	-	13.07
Adults	23.00	20.00	15.33	13.33

A decision to exclude the fifteen year old age group was made because all of the participating hospitals considered those within the fifteen year age group to be adults. This is because they have a relatively similar body size. In addition, using the above calculation (**Table (4-2)**, the PMMA thickness representative of a 15-year-old is very similar to that of the adults.

4.2.3.2 Lungman Phantom

The Lungman adult chest phantom is a multipurpose anthropomorphic chest phantom produced by Kyoto Kagaku, Japan **Figure (4-3)** (Dewerd L.A, 2014). It has a dimensional structure that permits the simulation of chest pathology and demonstration of the postero-anterior and lateral chest features. It provides an anatomical model of a torso with soft tissue substitute materials and artificial bones which have absorption ratios extremely close to human tissue. This phantom includes a removable mediastinum and pulmonary vessel structure and simulated lesions can be inserted into the lung field. The phantom comes with three groups of simulated lesions with differing sizes and densities. Its weight is approximately 18 Kg which is to simulate a standard size patient. The anterior and posterior chest fat jackets, which as 6 cm thick, represent human adipose tissue and can be added to the Lungman phantom to simulate a larger size body type (**Figure (4-3c)** (Murata & Nitta,

2016). The Lungman phantom was used in this thesis to simulate an adult chest region for average and larger sized patients. This phantom is suitable for achieving a number of objectives of this thesis. Firstly, for evaluating and comparing the IQ, LV and clinical protocols for average and larger sized adult patients (the larger size patient was simulated by attaching a fat jacket to the phantom). Secondly, for investigating the influence of phantom thickness on IQ and radiation dose.

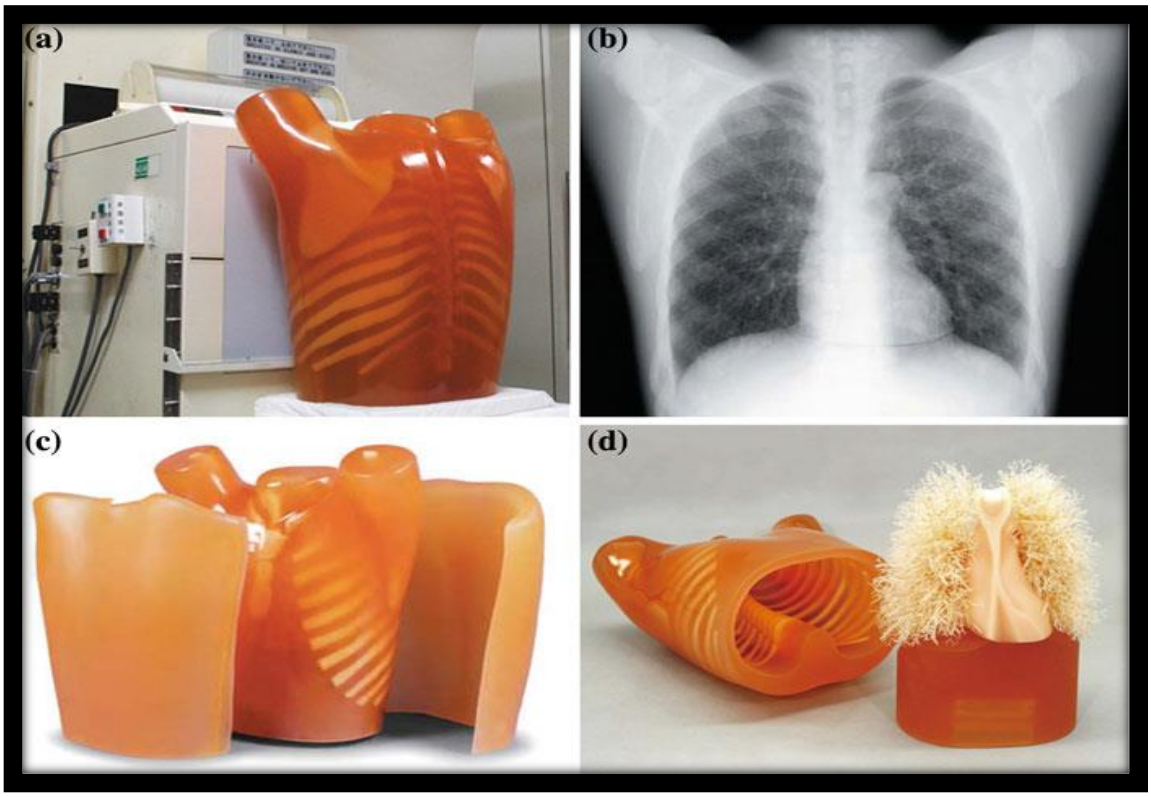

Figure 4-3 The multipurpose anthropomorphic chest phantom "Lungman" (Kyoto Kagaku Co., Japan). (a) Positioned for PA CXR. (b) The resultant X-ray image. (c) The phantom with the fat jacket to simulate a larger patient. (d) The inner components of the phantom (Dewerd L.A, 2014).

4.2.3.3 Gammex Phantom

The Gammex 610 phantom (Gammex Inc, USA) (**Figure (4-4)**) is an anthropomorphic neonatal chest phantom that simulates a 1-2 kg neonate. The structure of this phantom contains a torso, lungs, bronchial tree, spine, ribs and clavicle. Also, the Gammex phantom includes clinically relevant IQ features for resolution and noise in the form of a lung with a simulated pneumothorax with pleural thickening, and a lung with simulated hyaline

membrane disease (Dewerd L.A, 2014). It can be used for evaluating the whole imaging chain for routine QA using CR and DR systems (Dewerd, 2014). In addition, this phantom has been used widely in the literature for optimisation studies (Cohen et al., 2012b; Precht, Tingberg, Waaler, & Outzen, 2014; Singh & Pradhan, 2015; Smans et al., 2010; Smet et al., 2018). In this thesis, the Gammex phantom was used to simulate the chest of a neonate in order to achieve one of the objectives of this thesis: evaluating and comparing the visual IQ for neonatal patients between and within hospitals.

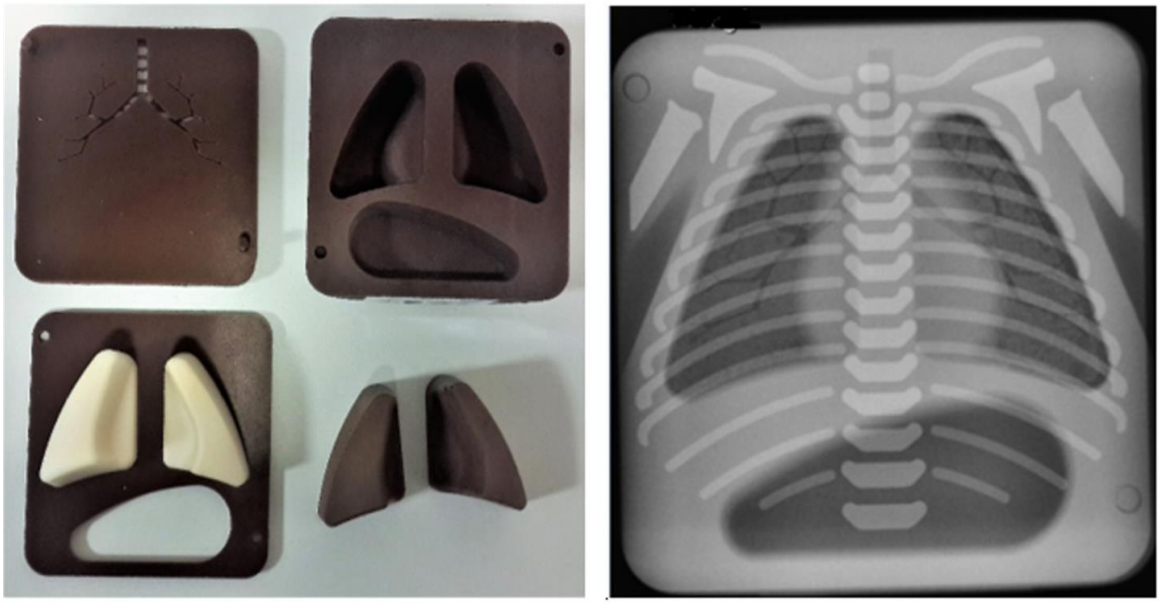

Figure 4-4 Illustrates the Gammex neonatal chest phantom (Model 610, Gammex Inc, USA). (a) The inner components of the phantom. (b) The resultant X-ray image.

4.2.4 Radiation Dose Measurements

A commercially available solid-state dosimeter (RaySafe X2, Unfors RaySafe AB, Billdal, Sweden) (**Figure (4-5)**) was used to measure the IAK (µGy) at the surface of the phantoms- the point of entry of the X-ray beam central ray. The RaySafe X2 dosimeter was used to ensure a precise measurement of the radiation dose was achieved. RaySafe X2 has a working range from 40 to 150 kVp and can detect radiation dose within a wide range (1nGy and 9999 Gy). The manufacturer suggests it has an accuracy of within ±5% of the calibrated values. In comparison with TLD, it gives a direct dose measurement which minimises the errors that can result from the TLD calibration process. In addition, TLD is time consuming, while RaySafe X2 gives an instant measure. Despite the fact that TLDs have high sensitivity for the low levels of radiation (e.g. scatter), this was not an issue in this thesis since the radiation dose was measured within the primary radiation field.

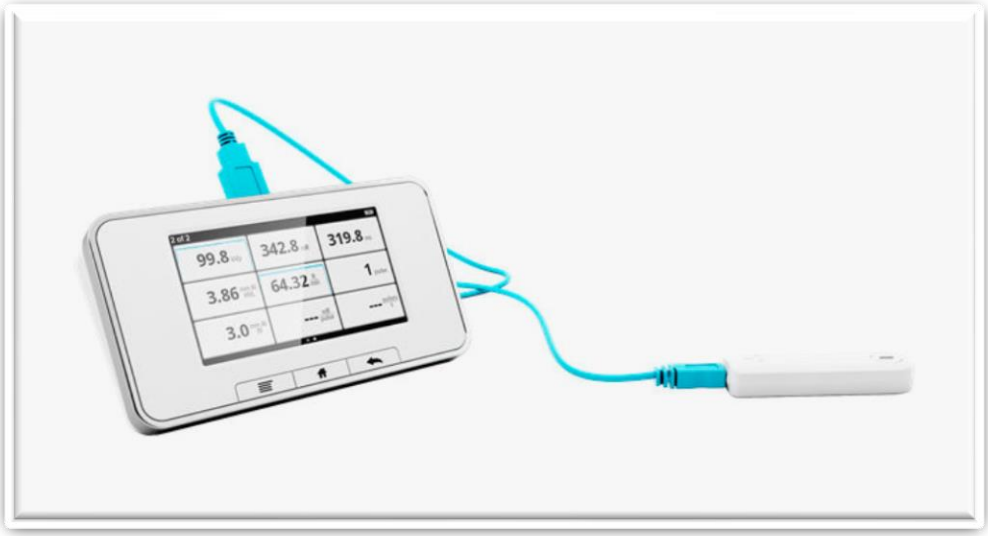

Figure 4-5 RaySafe X2 dosimeter (RaySafe X2, 2016).

The above method for measuring IAK was used only for radiographic examinations that were carried out using manual exposure control (non-AEC). However, it would not be appropriate to use the same technique for measuring IAK for examinations conducted using AEC. This is because the inclusion of the dosimeter is likely to increase the attenuation and, as a result, the exposures are likely to be higher. A simple extrapolation technique was therefore used to estimate IAK. This technique was based on an estimation of the IAK values from values of mAs, as recorded from the radiographic examination using the equation below (equation 1). $R^2 = 0.99$ was generated from **Figure (4-6)**:

$$Y = 86.3 * X \qquad (1)$$

Where:

Where: X represents the applied value of mAs; Y represents the resulted value of IAK; the value 86.3 represents the slope and means that for every increase of 1 in X (mAs), the value of Y (IAK) will increase by 86.3.

The extrapolation figure (**Figure (4-6)** was generated by performing the following experiment. Five PMMA slabs (26.5cm x 26.5 cm) of 50 mm thickness were used to represent a phantom as shown in **Figure (4-7).** The X-ray equipment was set to a manual exposure with a source to object distance (SOD) of 100 cm, kVp of 125 and additional filtration (0.1mm Cu+1mm Al). 8 exposures were performed using between 0.5 mAs and

5 mAs, in 0.5 increments. The values of IAK which correspond to their respective mAs values were recorded using the RaySafe X2 dosimeter. From the graphical representation of IAK against mAs **figure (4-6)**, the linear line of best fit and its equation (equation 1) were generated to provide a method for estimating the IAK.

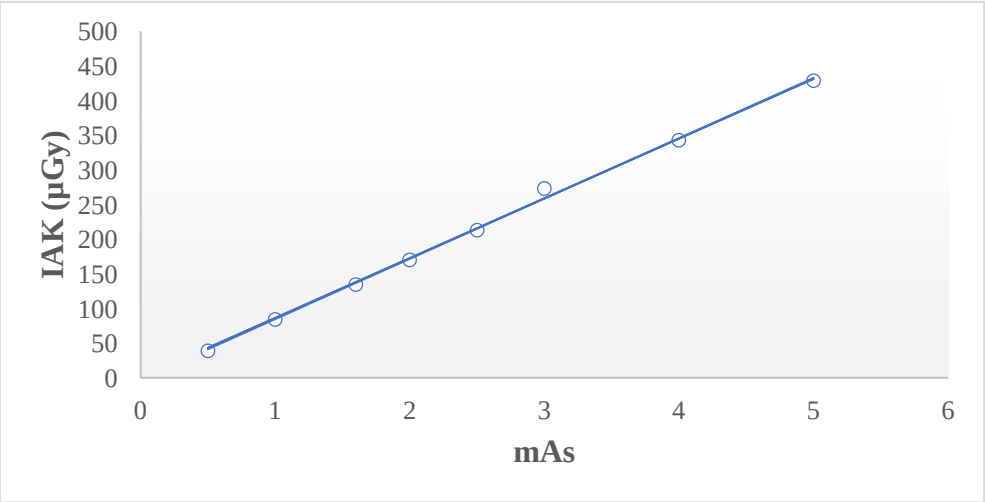

Figure 4-6 This figure demonstrates the extrapolation process for IAK from the mAs values using a linear relationship for the X–ray examination conducted using an AEC.

The line of best fit was used for estimating the IAK values from the values of mAs. The above equation (equation 1) is only valid when the acquisition parameters (kVp and filtration) are equal to 125 kVp and additional filtration of 0.1mm Cu+1mm Al. For each protocol and X-ray machine a relevant set of calibration data was obtained in order to allow the generation of IAK from post-exposure mAs values.

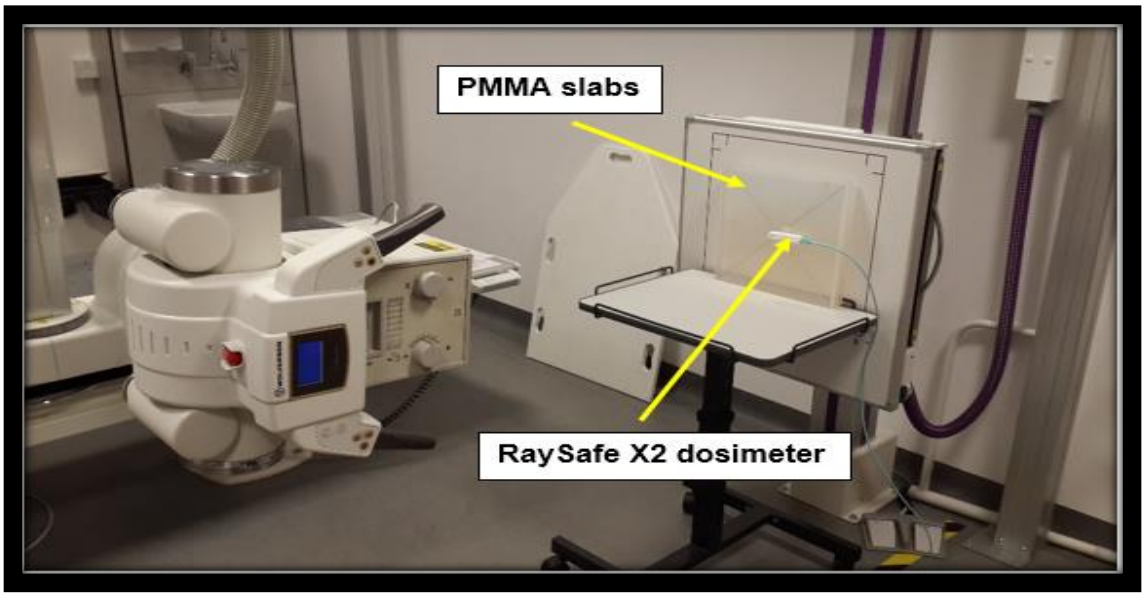

Figure 4-7 This figure demonstrates the experimental set up for the extrapolation of IAK values from mAs values for hospitals that used AEC in their CXR examinations.

4.2.4.1 Justification for using IAK as Indicator for Radiation Dose

The main purpose of the dose measurements was to allow the comparison of radiation dose values between the X-ray machines. In order to compare the IAK from the current study with other similar studies reported in the literature and in the UK DRL, conversion of IAK to ESD by multiplying the values of IAK with a backscatter factor (BSF) was required. However, the ESD has not been calculated in this thesis. The reason for this is that diagnostic radiography BSF depends on several factors: the technical parameters (the beam area and the X-ray spectrum), acquisition parameters used such as the tube voltage, filtration and SID, and the composition of the object to be imaged (Petoussi-Henss, Zankl, Drexler, Panzer, & Regulla, 1998). In this thesis, local protocols from the hospitals for the same age groups were different between the X-ray machines. They had different acquisition parameters including different values of kVp, filtration and SID which were used for the same age group among the different X-ray machines. Therefore, it is not valid to use the same value of BSF to convert IAK values, which resulted from the 17 X-ray machines that were obtained from different protocols and acquisition parameters, to ESD. Adding the same BSF will lead to an increase in the errors in dose values and make the comparison in radiation dose between the hospitals invalid. A more valid method would be to compare X-ray machines based on the IAK values that were recorded directly from the RaySafe X2 dosimeter without adding the BSF.

By way of example and based on a standard adult mathematical anthropomorphic phantom, a BSF equal to 1.47 was achieved for the PA CXRs of a male adult of 'standard size' when the following acquisition parameters were used: tube voltage = 125 kVp; filtration: 3.5 mm Al + 0.1 mm Cu; focus to skin distances (FSD) = 177.5 cm; and field size 35 cm x 40 cm (Schultz, Geleijns, & Zoetelief, 1994). This value of BSF (1.47) is valid only when the acquisition parameters used are similar to those just mentioned. This was also investigated by (Petoussi-Henss et al., 1998) who used different values of BSF resulting from a cuboid of PMMA slabs, water and ICRU tissue, based on the applied values of kVp, filtration, SID and beam area. Another reason for not converting the IAK values from the CDRAD 2.0 to ESD is because it is not a valid method for comparing the results of ESD from the CDRAD 2.0 phantom against those from a benchmark, such as DRLs. The reason behind this is because the CDRAD 2.0 phantom has a standard size (26.5 x 26.5 cm) which is used across all of the age groups. This will influence the resulting radiation dose values. Finally, a further reason for not calculating ESD from IAK by adding the BSF is that the Lungman phantom was used to represent two patient sizes- standard 'average' size and a larger sized adult patient, represented by the Lungman phantom with and without fat jacket. It is not valid to use the same BSF for both sizes because the BSF for the larger size Lungman phantom has not been reported in the literature.

4.2.5 Procedure for Image Acquisition

Three different phantoms were utilised to simulate the chest region of patients across a range of ages. This allowed the generation of images which could assess the variation in IQ, both within and between hospitals based on their local clinical acquisition protocols. Images were acquired for each X-ray machine at 4 stages, described as follows.

Before starting the imaging process, the characteristics of the X-ray equipment and the examination technique/protocol that would normally be utilised for each age group were collected. 2 data collection forms were used to record the information on the characteristics of the X-ray equipment and the examination technique / protocol. These are provided in **Appendix A-2** and **A-3**.

1. The CDRAD 2.0 phantom (Artinis Medical System, The Netherlands) (Burght et al., 2014) was used to acquire radiographic images for the 5 age groups for each X-ray machine using their existing protocol. The CDRAD 2.0 phantom was positioned at the centre of the primary radiation field, and the X-ray beam was collimated to

the edges of the CDRAD 2.0 phantom. The decision to use the erect or table Bucky was dependant on the local protocol as defined for each of the X-ray machines. The RaySafe X2 dosimeter was used to measure IAK for the phantom and it was placed above (tube facing side) the PMMA slabs in the centre of the X-ray field (at a 13.25 cm distance from the four phantom edges, the centre of was the X-ray intersection). Each of the CDRAD 2.0 acquisitions were repeated 3 times in order to acquire 3 duplicated images as recommended by the CDRAD 2.0 phantom manufacturer (Burght et al., 2014). For the measurement of IAK, each exposure was repeated 3 times and the average was calculated to minimise random error. The experimental setup is illustrated in **Figures (4-8)**.

2. The CDRAD 2.0 phantom was replaced (after finishing the image acquisitions) by the Lungman phantom and again imaged in line with local protocols. Further image acquisitions using the Lungman were obtained after attaching a fat jacket to simulate a larger sized patient. Image acquisitions were again obtained using local protocols. Similar to the CDRAD 2.0 phantom, the Lungman phantom was positioned at the centre of the X-ray beam and the collimation was set to the collimation borders (**Figure (4-8c)**). The RaySafe X2 dosimeter was also placed at the centring point (**Figure (4-8c)**).
3. The Lungman phantom was the replaced with the Gammex phantom and again imaged using local protocols.

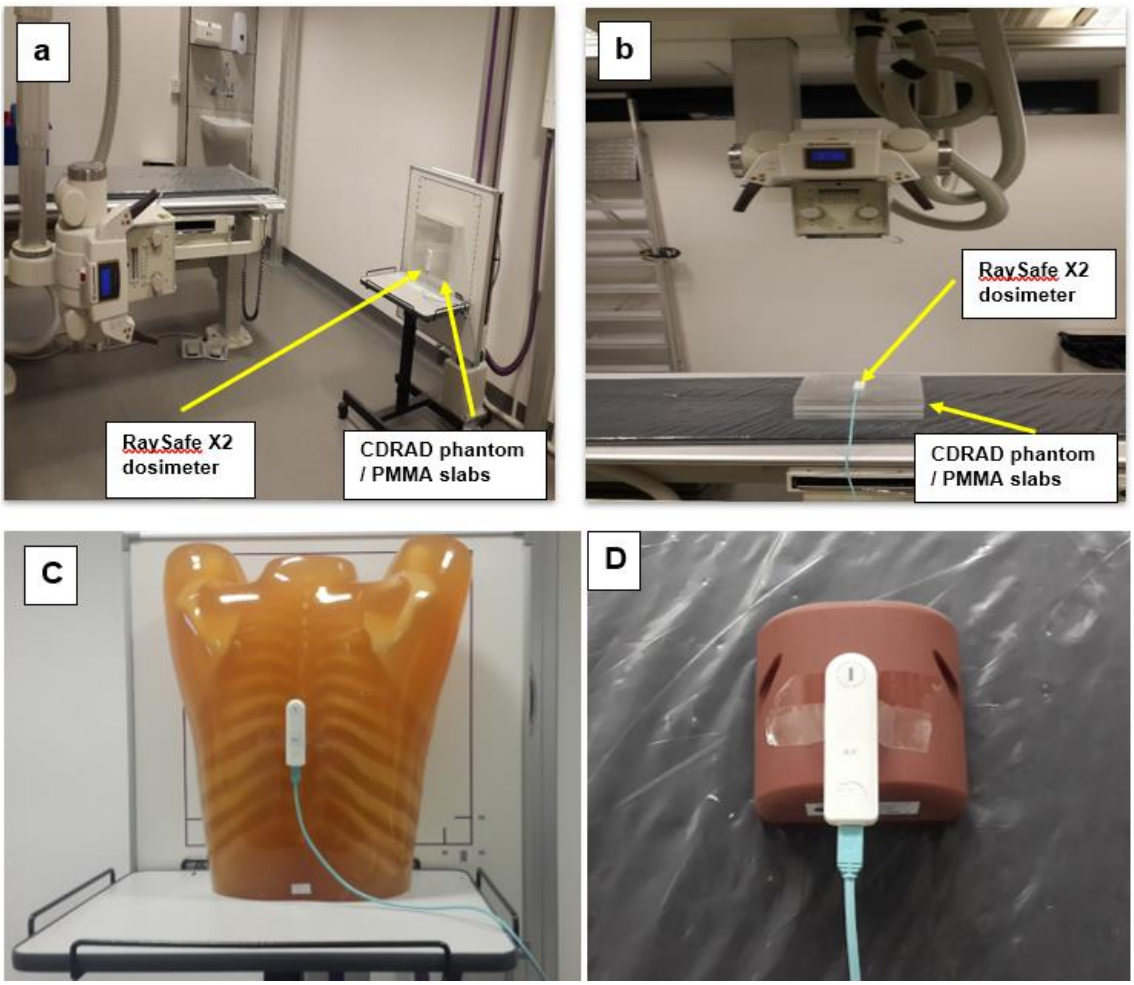

Figure 4-8 Illustrates the experimental set up of the phantoms used in in the hospitals. (a) Erect positioning of CDRAD 2.0 phantom. (b) Supine positioning CDRAD 2.0 phantom. (c) Positioned for PA CXR of Lungman phantom. (b) Supine positioning of Gammex phantom.

The collimation size of the X-ray beam was constant for each phantom, based on the size of the phantom used (i.e. three different constant collimation sizes were used according to the 3 sizes of the phantoms used: CDRAD 2.0, Lungman and Gammex), during all image acquisitions. This was necessary for increasing the reliability and to ensure that the collimation did not influence the IQ and radiation dose.

Images were collected from the hospitals according to the Digital Imaging and Communications in Medicine (DICOM) format and the images were coded using the following system: Coding used 3 numbers and 2 letters for ages 0, 1 and 5 years; and 4 numbers and 2 letters for ages 10 and for adults. The coding refers to hospital number, X-

ray machine number, patient age, X-ray device type (mobile or static) and phantom name, as shown in **Figure (4-9).**

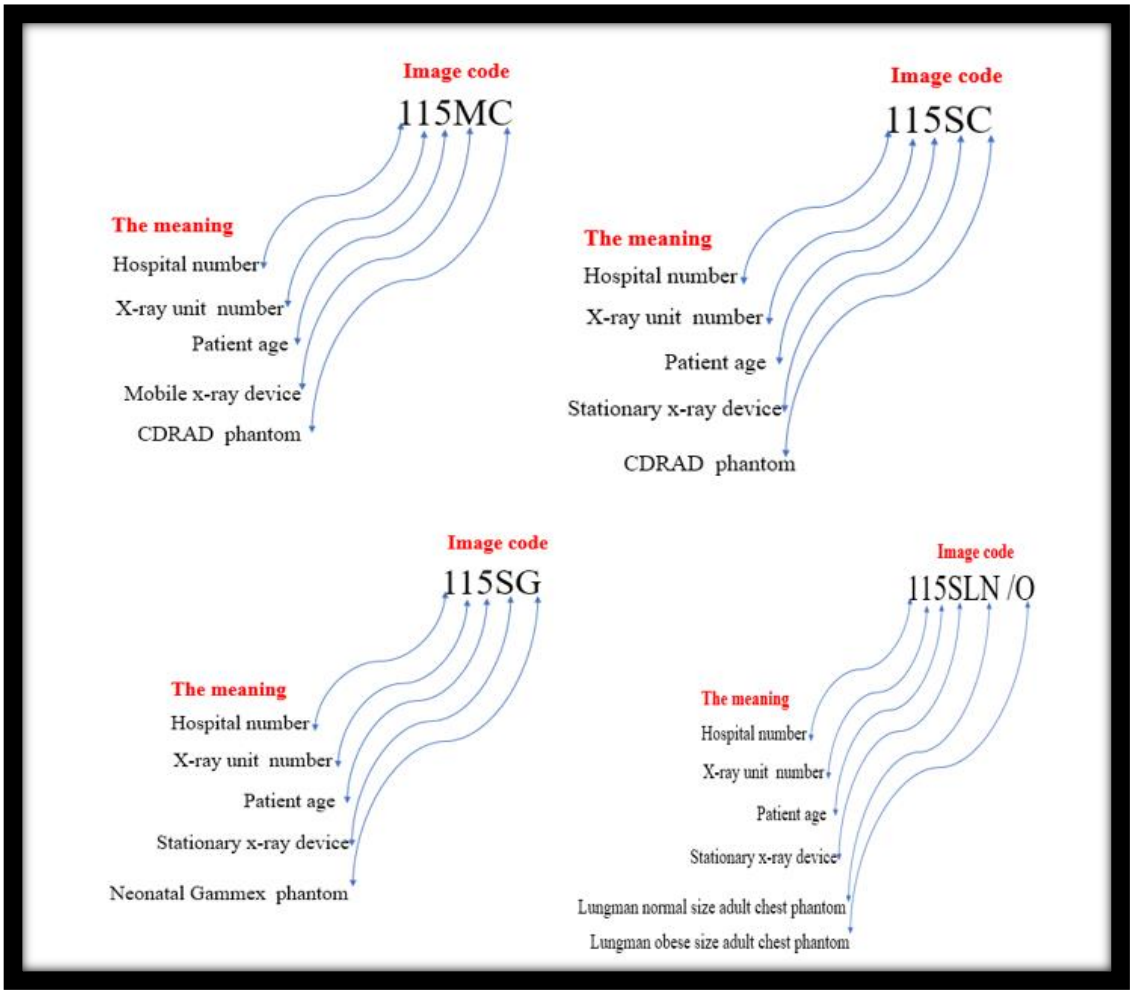

Figure 4-9 An illustration of the coding process for the phantom images.

4.2.6 Image Quality Evaluation

As indicated previously, the evaluation of IQ can be achieved either using physical measurements such as SNR and CNR or by visual evaluation. In this thesis, both physical and visual methods are used. The choice was based on the phantom type utilised and the required objectives. The relative VGA approach was utilised as a main method for IQ evaluation for the images acquired from the anthropomorphic phantoms (Lungman and Gammex), while the automated method using the CDRAD analyser software was used as a main method for IQ evaluation for the images generated from the CDRAD 2.0 phantom. There were no visual evaluations conducted using this phantom within this thesis. **Figure**

(4-10) demonstrates the different types of IQ evaluation that were performed on each type of phantom used in this study.

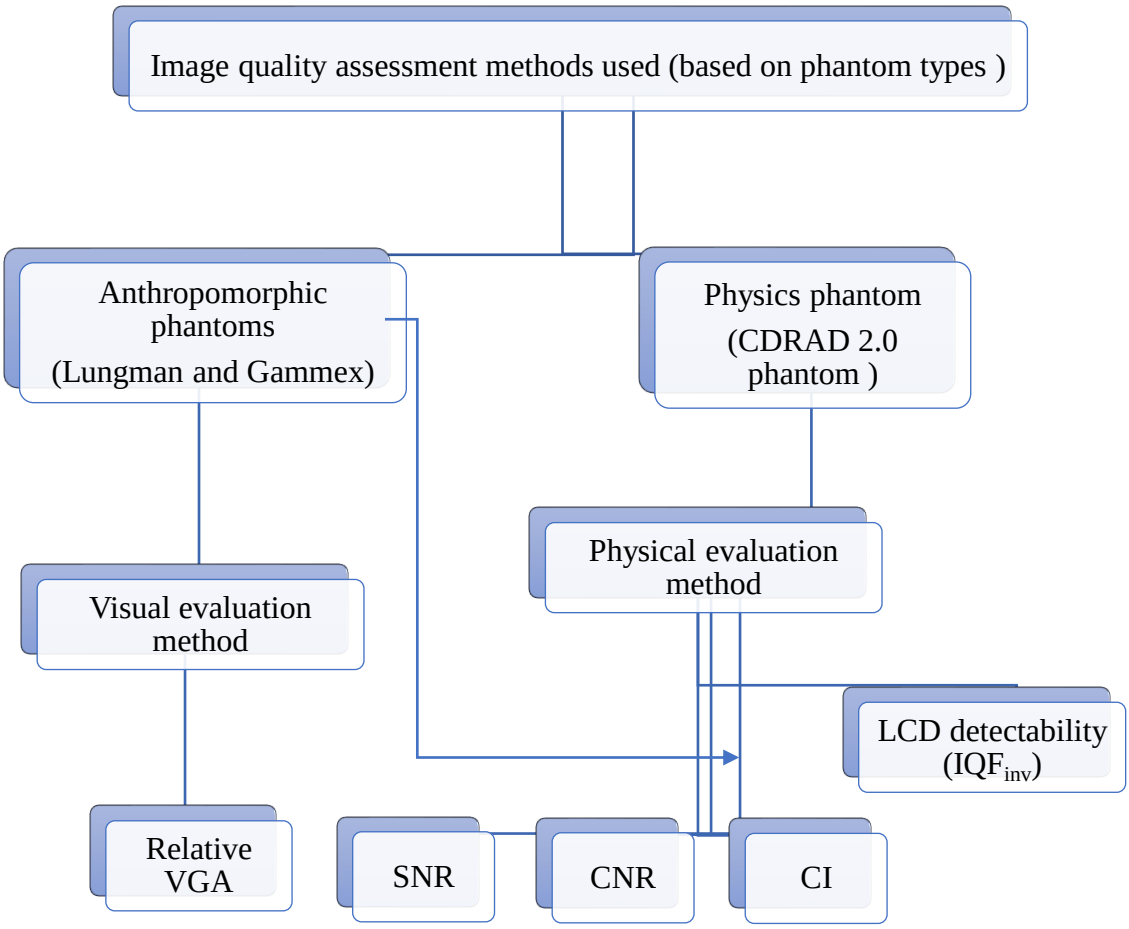

Figure 4-10 This figure demonstrates the different types of IQ evaluation methods (visual and physical) that have been used in this study (based on the type of the phantom).

4.2.6.1 Visual Image Quality Assessments

Ethical approval for this study is a requirement for visual IQ evaluations by human observers and was granted by the University of Salford (HSR1617-76) (**Appendix A-4**). A bank of the chest phantom images for both the Lungman phantom (n=17), with and without the fat jacket, and the Gammex phantom (n=17) were previously acquired from the different X-ray machines and were visually evaluated using relative VGA. All of the images were coded (**Figure** (**4-9)**) to ensure that the observers were blinded to the acquisition parameters and the hospital/X-ray machine. Images were evaluated visually using relative VGA by applying a 3 point Likert scale. The relative VGA method utilised the software is described by Hogg, P. and Blindell (2012). This is a Java-based software and it presents images to observers in a random sequence on dual screens. This software provides a method for allowing a reference image to be viewed on an adjacent monitor during the assessment. The relative VGA method has been chosen for 3 reasons: first, it is a sensitive method for detecting small variations in IQ; second, it shows minimal bias and variability; third, it is suitable for this study because of the type of questions that will be asked of the observers, e.g. 'whether there is a difference in LV and IQ', compared with a reference image.

For the relative VGA method, images were presented to observers on a computer monitor. On the left monitor there was a fixed reference image of median IQ and on the right monitor an image was randomly applied from the image bank (**Figure (4-11)**). The reference image was chosen by a consensus opinion of two experienced clinicians who interpret images as part of their normal clinical routine; the reference image had 'average' IQ in comparison with the other images; this was to ensure that all of the 3 Likert scale points were used in the evaluation (Allen, Hogg, Ma, & Szczepura, 2013; Ma et al., 2013). Each participant reviewed the images and responded with a set of statements/questions whilst making a comparison against the reference image.

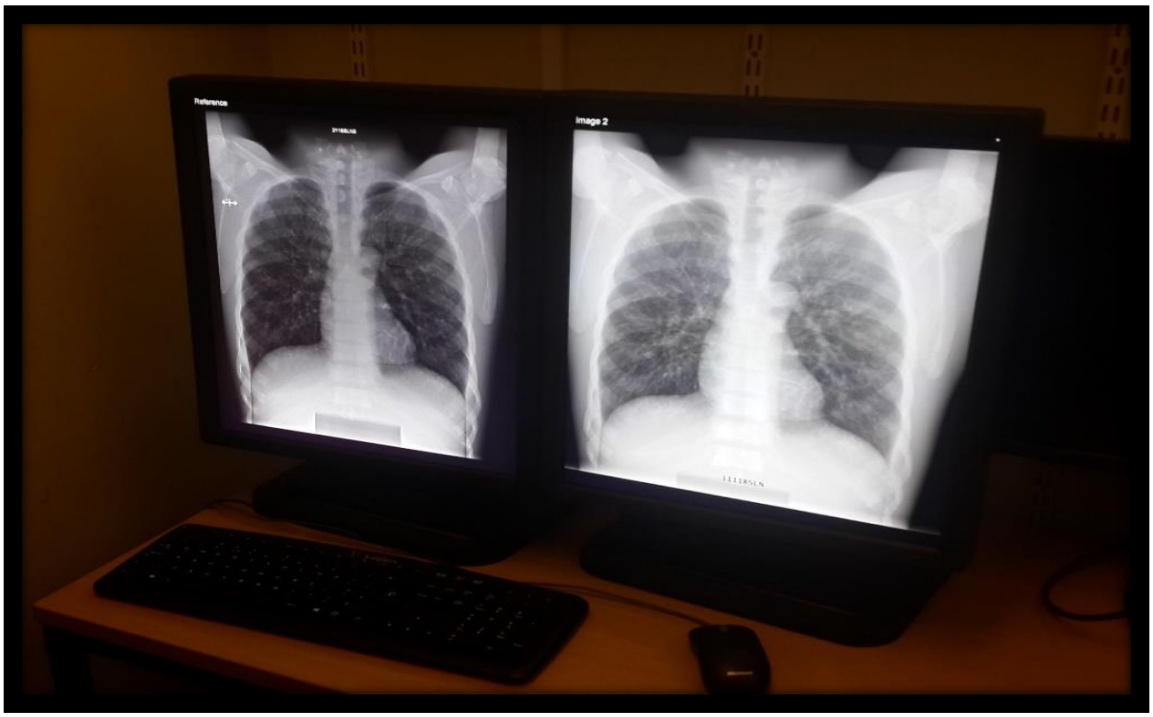

Figure 4-11 The setup for the image evaluations using the Lungman phantom images and a relative VGA method, 5 MP monochrome liquid crystal display monitors.

In terms of the Lungman phantom images, the observers decided whether the visual LV and IQ were either worse, equal to, or better than that of the reference image. 6 different types of criteria, based on the CEC guidelines (Commission of the European Communities (CEC), 1996a) and study by Ma et al. (2013), were selected to evaluate the IQ as illustrated in **Table (4-3)**. LV was assessed using 3 different criteria based on the study by Ma et al. (2013), as illustrated in **Table (4-4)**. For the Lungman images, the selected feature to be used as a reference for measuring LV across all of the acquired images was within the left hemi-thorax at the level of the 8th thoracic vertebra, immediately adjacent to the paraspinal tissues within the Lungman phantom (**Figure (4-12)**). The reasons for selecting this feature instead of the simulated phantom lesions were entirely practical since was deemed to be easier to recognise the position of this feature on all of the images (across a range of image qualities), when compared with the simulated lesions in other locations. Furthermore, it was difficult to control the fixation of the simulated lesions inside the phantom when moving the phantom between sites. This led to some simulated lesions moving around from their intended locations and could have generated limitations on the validity of study. Therefore, the feature described was selected as a vehicle for evaluating the LV as a result of its being fixed and located in the same location across all of the images. This represents a valid way for comparing LV between acquisitions/X-ray machines.

Regarding the Gammex phantom, the visibility of simulated pneumothorax and respiratory distress syndrome was selected as a criteria for assessing IQ based on the study by Quality et al. (2015).

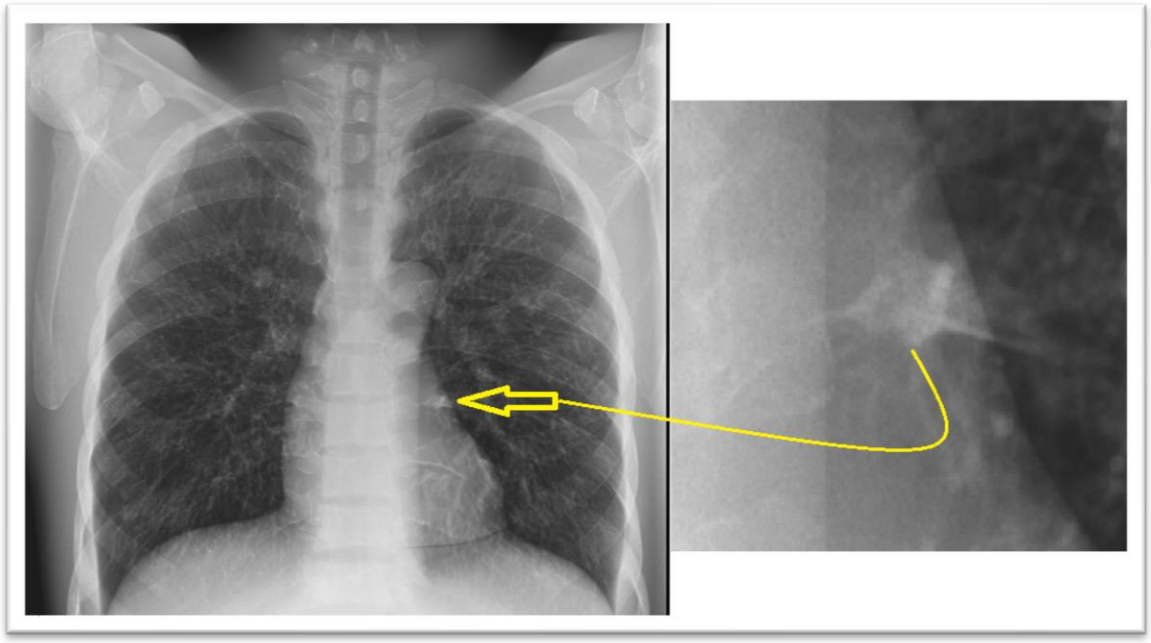

Figure 4-12 Lungman X-ray image illustrating the selected region of interest (feature) used for evaluation of LV.

Table 4-3 The six image criteria that were utilised for IQ assessments using a relative VGA method for the Lungman phantom images (Ma et al., 2013).

Item	Criteria
1	Demonstration of the vascular pattern in the whole lung, particularly the peripheral vessels.
2	Visually sharp demonstration of the trachea and proximal bronchi.
3	Visually sharp demonstration of the borders of the heart and aorta.
4	Visually sharp demonstration of the diaphragm and lateral costo-phrenic angles.
5	Visualisation of the retrocardiac lung and the mediastinum.
6	Visualisation of the spine through the heart shadow.

Table 4-4 The three simulated lesion criteria that were utilised for LV evaluations, using a relative VGA method, for Lungman phantom images (Ma et al., 2013).

Item	Criteria
1	Contrast of nodule, against background.
2	Brightness of nodule, against background.
3	Sharpness of nodule edge.

For the IQ evaluations, two 5 mega-pixel (MP) monochrome liquid crystal reporting monitors (DOME E5 (NDSsi, Santa Rosa, CA) monitors (2048 by 2560 pixels)) were used. The reasons behind selecting monitors with this technical specification were to simulate the clinical environment, because they are considered to be standard primary reporting monitors that are normally used for the interpretation of clinical images in most UK hospitals and to improve the viewing conditions recommended for better detectability and interpretation (Norweck et al., 2013). Furthermore, they are recommended by the Royal College of Radiologists (The Royal College of Radiologists, 2012).

The monitors were calibrated to DICOM grayscale standard display function (GSDF) (The Royal College of Radiologists, 2012) and the maximum luminance of the monitors was

set at 500 cd\cm^2 (**Figure (4-13)**), which represents the default value that is recommended by the manufacturer. This was achieved by utilising the quality-control software (RadiCS, Eizo Nanao Corporation) and a near range luminance meter (RadiCS UX1 Sensor, Eizo Nanao Corporation) without any ambient lighting as based on the user manual.

To simulate the clinical situation for IQ evaluation, the room light was switched off to prevent light reflecting onto the monitors and a calibrated Lux meter (RaySafe-X2 Light sensor) was utilised to check the ambient light level and to ensure that it was zero in front of the monitors. Furthermore, interruptions were not permitted during IQ evaluations in order to avoid distracting the observers. This was achieved by the room door being closed with a sign outside saying do not enter.

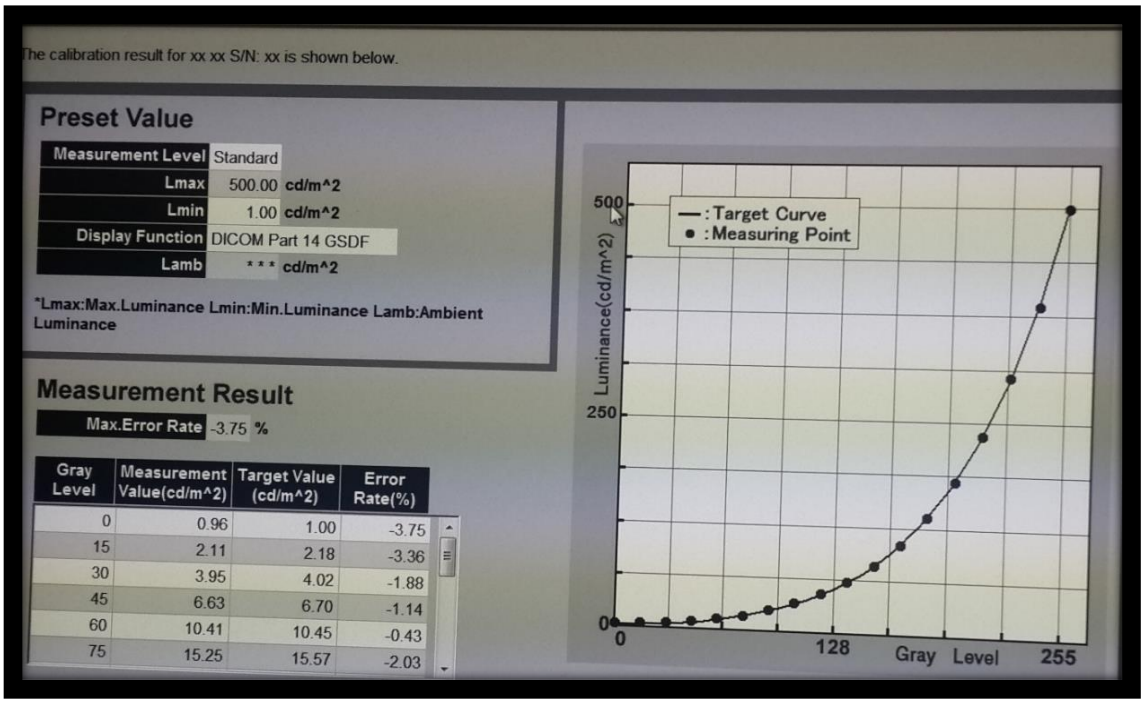

Figure 4-13 Calibration curves for the 5 MP monochrome liquid crystal display monitor, set with a maximum luminance = 500 cd/m2.

During image viewing the following conditions were applied: No time limitations were imposed on the observers during the scoring tasks; there were no restrictions on distance between the observer and the monitor; and the observers were not permitted to adjust the contrast, intensity or magnification of the image to ensure that any differences in visual

IQ were related to the quality of the image itself and not due to image manipulation/post-processing. By allowing the use of image viewing tools without restriction the image appearance could change and bring observational bias between observers (Allen, Hogg, Ma, & Szczepura, 2013; Ma et al., 2013; Jansson, Geijer, Persliden, & Andersson, 2006). The images were assessed by 7 qualified diagnostic radiographers (5 males and 2 females), range from 30 to 55 years. Their post-qualification experience ranged from 5 to 18 years. The observers in this study were included based on their levels of experience and having a minimum 5 years of experience was a condition in order to reduce the inter-observer variability and improve the validity of the results. To ensure that all the observers had a normal visual function, they were asked whether their eyesight was a typical 20/20 vision, the date of their latest eyesight test and if their eyesight was corrected with glasses or contact lenses. All the observers had a typical 20/20 vision, 3 of the observers utilised glasses and the rest (4) did not necessitate any eyesight correction. All observers had checked their eyesight within the last 12 months. Baseline data from the included observers is illustrated in **Table (4-5)**.

From the literature, there is no restriction on the number of the observers; however, compared with literature a relatively large number (7) of observers was used within this thesis. The overall score of the IQ for each image, by each observer, was calculated by adding together the scores from all the IQ criteria. The total scores represented the overall score of IQ for each image. This approach was similar for the calculations of overall LV for each image and by each observer. Next, the final score of IQ, for each image, was determined by taking the average IQ score across the 7 observers, and the same method was used for LV.

Table 4-5 This table demonstrates the characteristic of the observers involved in the IQ evaluations.

Observer number	Age, years	Qualifications	Time qualified, years	Experience looking at chest images, years
1	39	BSc, Masters and PhD degrees in Diagnostic Radiography	18	18
2	40	BSc in Diagnostic Radiography	16	16
3	35	BSc, Masters and PhD degrees in Diagnostic Radiography	13	13
4	55	BSc and Master's degree in Diagnostic Radiography	25	25
5	49	BSc and Master's degree in Diagnostic Radiography	19	19
6	36	BSc in Diagnostic Radiography and Master degree in Diagnostic Imaging	10	5
7	30	BSc in Diagnostic Radiography and Master's degree in Medical Physics	7	5

The method described above was utilised only for achieving the aim of the study relating to the investigation in variation of visual IQ, and LV, both within and between hospitals. On the other hand, a different technique for visual IQ evaluation was used to achieve the aim of the study. This related to the investigation of the influence of phantom thicknesses (represented by images acquired using the Lungman, with and without the fat jacket) on IQ for adults. This technique involved viewing two images from the Lungman phantom, with and without a fat jacket, acquired from the same X-ray machine and presented on 2 side-by-side monitors. Observers were asked to rate on a 2 -point scale, as follows: 1, the image with the better IQ (compared with the other image); 0, the image with the lower IQ (compared with the other image).

4.2.6.2 Physical Image Quality evaluation

4.2.6.2.1 IQF_{inv} Calculations

The CDRAD analyser software was used for the physical evaluation of LCD detectability and the output of the software is displayed as IQF_{inv} values (the average value of three repeated images). The input parameters for the CDRAD software analyser (Alpha, APD and the SID) are indicated below.

Alpha was selected to be equal to $1e^{-8}$. This value is equal to the default value of the CDRAD analyser software and is proposed by the manufacturer (Burght et al., 2014). The reason behind selecting this value is attributed from having the best correlation with perceptual IQ (Pascoal et al., 2005). APD should be considered and included in the calculation during the automated scoring method and it is set relative to the image bit depth. It was necessary to set the APD at 0 if the images had different bit depths to insure for a valid comparison between images (Brosi et al., 2011; Burght et al., 2014). APD was therefore set to 0 because the CDRAD 2.0 phantom images were acquired from different X-ray machines that have various numbers of bits stored per pixel. With respect to SID, this was set based on the local protocol values employed by the hospitals, and for each age group under investigation.

The major benefits of the physical evaluation method for CDRAD 2.0 phantom images using the CDRAD analyser software are the high reliability and consistency on the evaluation criteria utilised to assess the threshold CD detection, and that it does not suffer from the subjectivity of the human visual and cognition systems (Pascoal et al., 2005). In addition, many studies reported that there is good correlation between the visual and physical evaluation methods of LCD detectability using CDRAD 2.0 phantom images (De Crop et al., 2012; Norrman et al., 2005). Also, using the CDRAD software analyser in this thesis is extremely useful due to the large amount of data (CDRAD 2.0 phantom images) that was collected, since using visual assessment for CDRAD 2.0 image evaluation is very time consuming.

4.2.6.2.2 SNR and CNR Calculations

In terms of CDRAD 2.0 phantom images, the SNR and CNR values were computed for all the CDRAD 2.0 images in this study. The average value of SNR and CNR, for the 3 replicated images for each examination, were determined to represent the final values of SNR and CNR.

For practical reasons, SNR and CNR have been measured only from the first square in the CDRAD 2.0 image. This square has the highest contrast and diameter (0.8), rather than at different ROIs (square cells in CDRAD 2.0 images), and is similar to approaches used another similar study (Alzimami et al., 2009). Furthermore, comparing the values of SNR and CNR for the different CDRAD 2.0 images (based on the first square) is reasonable because the first square gives an indication as to the whole image.

The average pixel values of the central visible spot were considered as a signal, and the noise was measured from the SD of the background. The SD for the four regions in the background surrounding the central spot was measured to represent the noise. The computer software ImageJ was used to measure the SNR, and during the measurement process all of the ROIs had the same size and were positioned at the same place for all of the images. **Figure (4-14)** shows an example of the signal and the background, respectively. SNR was computed as the ratio between signal and noise ($SNR = S / \sigma_b$). This is considered to be a direct method for measuring SNR (Smans et al., 2010). CNR was computed as the ratio of the difference between signal and background, divided by the noise value ($CNR = S\text{-}B/\sigma_b$) (Hampel & Pascoal, 2018; Jiang, Baad, Reiser, Feinstein, & Lu, 2016).

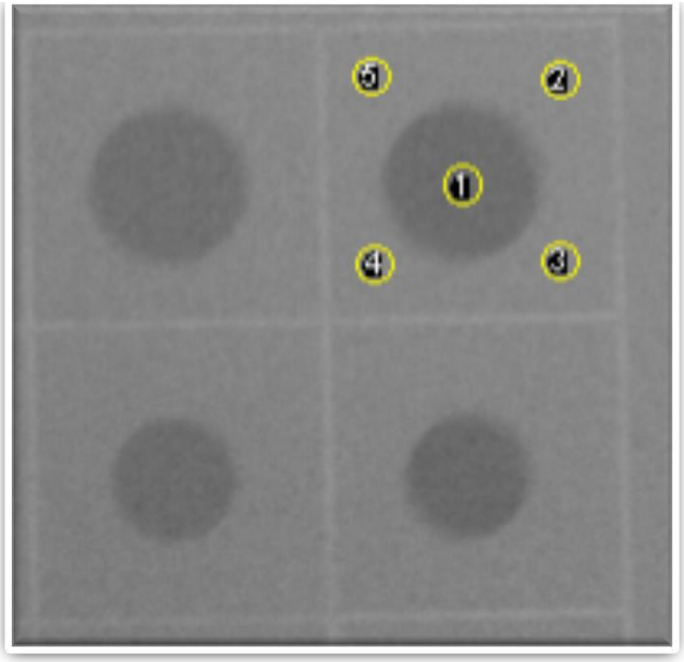

Figure 4-14 CDRAD 2.0 phantom image illustrating the selected ROI (circle 1) and the background (the circles 2,3,4 and 5).

Regarding the Lungman and Gammex phantom images, SNR and CNR were calculated in the same method that was used for the CDRAD 2.0 phantom images. The corresponding ROIs were selected as illustrated in **Figures (4-15)** and **(4-16)** and were based on those reported in the literature (Mraity et al., 2014; Lin et al., 2012). These 4 ROIs (circles: 1, 2, 3 and 4) were used only for the calculating the SNR. However, for CNR only one ROI, which is located in the lung (circle 3), was selected. The reason behind selecting this ROI was because it is more important clinically and any improvement in CNR for this region could influence the detection of lung disease.

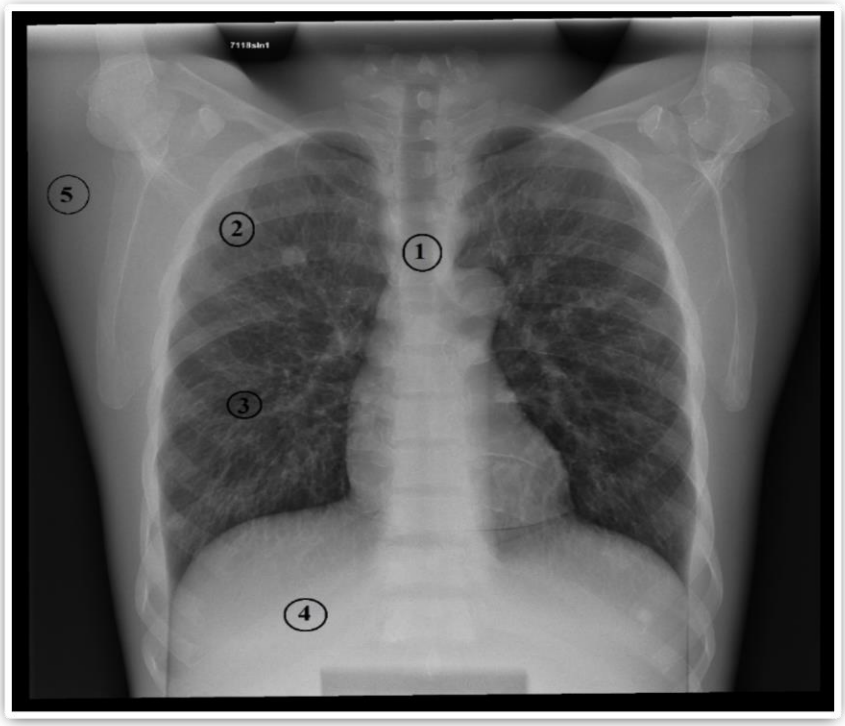

Figure 4-15 Lungman phantom image illustrating the selected ROIs (circles:1, 2, 3 and 4) and the background (circle: 5).

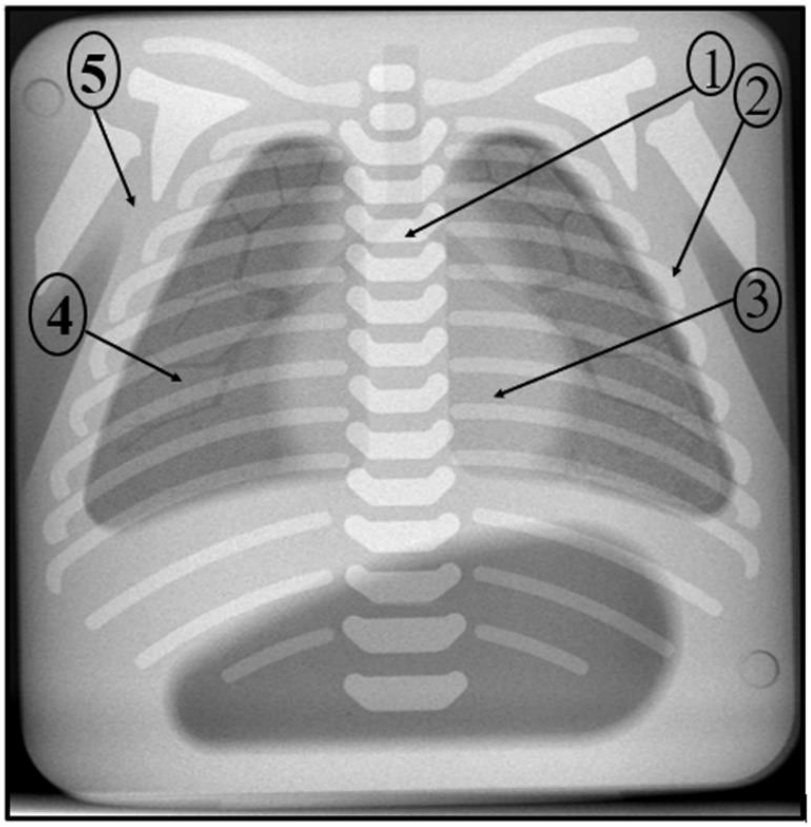

Figure 4-16 A Gammex phantom image illustrating the selected ROIs (circles:1, 2, 3 and 4) and the background (circle: 5).

4.2.6.2.3 Conspicuity Index (CI) Measurements

The CI was measured for both Lungman and CDRAD 2.0 phantom images using conspicuity software (Szczepura & Manning, 2016). Using this software a ROI is drawn around the required object (e.g. simulated lesion). This allows the identification of grey levels in the surrounding tissue that need to be compared against grey levels within the object to calculate conspicuity based on the object dimensions. Its use has been reported in the literature (Ohlmann-Knafo et al., 2016; Szczepura, Faqir, & Manning, 2017; Szczepura, Tomkinson, et al., 2017). It has been designed for measuring the CI for focal lesions and it developed and validated physically by Szczepura & Manning, (2016). It was developed to have high reliability. The software has been validated physically by K. R. Szczepura & Manning (2016) in two different ways: firstly, using the GAMMEX ACR CT accreditation phantom; secondly, using a Lungman phantom. Regarding the first method of validation, the GAMMEX ACR CT accreditation phantom was imaged using CT with a range of mAs values (50-400 mAs with 50 mAs increments), kVp values (80, 100, 120, 135 kVp) and slice thicknesses (1, 2, 3, 4 mm) and the CI was measured 4 times from the same selected

ROI to investigate the reliability of the CI calculation. A very low SD was observed from the 4 repeated values of CI from the same ROI and this is an indication to the high reliability of this software. Furthermore, the results showed a linear response between CI values and mAs and CI values behaved as expected with different kVp and slice thicknesses and this is an indication for the validity of the software in conspicuity measurements.

Regarding the second method of validation, a more clinically representative phantom represented by the Lungman phantom, loaded with 4 simulated focal lesions with different densities and with 2 diameters (8 and 4 mm), was imaged using the same CT scanner with different values of tube current time product (25-600 mAs) and the CI was measured for the lesions. The results showed that the CI increased alongside an increasing mAs up to a point where it plateaus. Also, it was found that for the low contrast simulated lesions, the CI was as low as expected and the simulated lesions with the greatest difference to the background reported the highest CI value. This is an indication about the validity of the software in CI measurements.

For the Lungman images, the same feature that was selected for visual LV evaluation tasks (**Figure (4-12)**) was selected for the CI measurements, as illustrated in **Figure (4-17)**. Regarding the CDRAD 2.0 phantom images, CI was measured only for the visible object (hole) in the first square of the image as illustrated in **Figure (4-18)**.

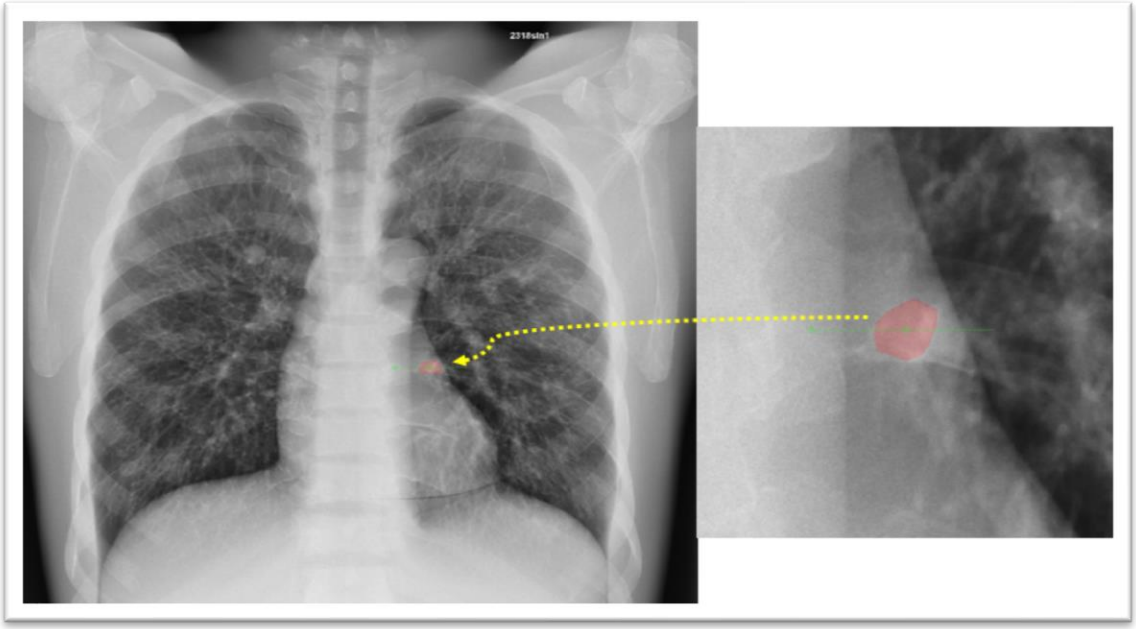

Figure 4-17 Illustrates ROI selecting in Lungman phantom images within the CI software.

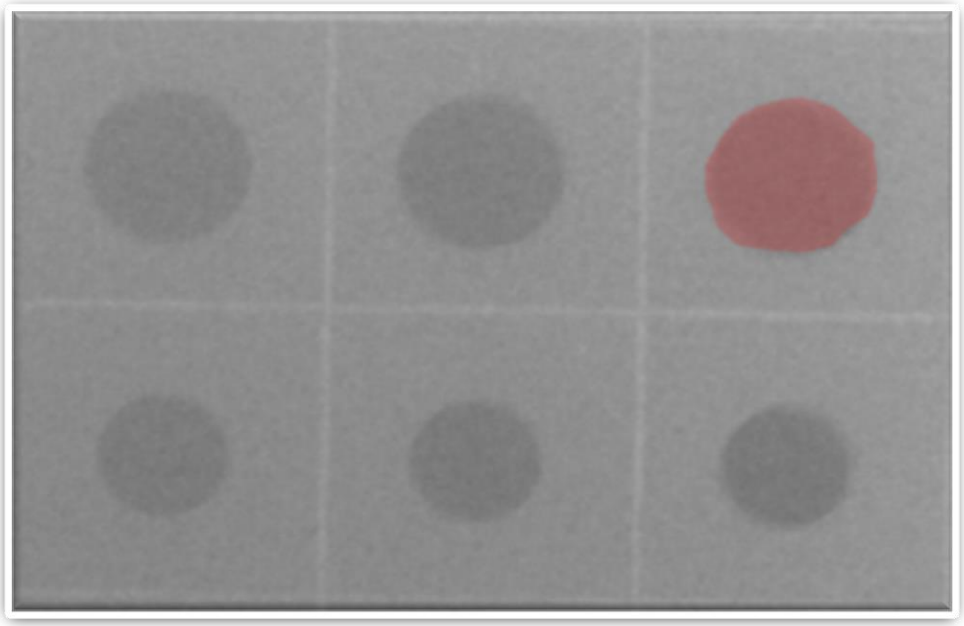

Figure 4-18 Illustrates ROI selecting in CDRAD 2.0 phantom images within the CI software.

4.2.7 Figure of Merit (FOM)

FOM calculations for the visual IQ evaluation of Lungman and Gammex phantoms images were done using the following equations:

$$FOM=(IQ)^2/\ IAK \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (2)$$

$$FOM=(LV)^2/IAK \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (3)$$

For the above equations, FOM calculations were used to describe the performance of the X-ray machines using the different visual metrics (IQ and LV). The FOM does not have an international system of units (SI), and for the FOM equations the unit of the FOM should be μGy^{-1} because the visual score of IQ does not have a unit. Within this thesis, FOM values will be presented as a numerical value without units; this approach is consistent to similar FOMs in the literature (Mraity et al., 2015).

A new FOM is also suggested and used within this thesis for the data obtained from the physical IQ measurements of CDRAD 2.0 phantom images. The new FOM equation is calculated from the LCD detectability that is represented by the IQF_{inv} based on the following:

$$FOM=(IQF_{inv})^2/\ IAK \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (4)$$

According to the equation (4), FOM units should be [mm^{-4}] [μGy^{-1}]. Nevertheless, and similar to the FOM generated from the visual evaluation, the results of FOM in equation (4) will be presented in this thesis as a numerical values without units. This is because the values of IQF_{inv} presented in the literature are calculated without a unit (De Crop et al., 2012; Geijer et al., 2001; Geuer, Norrman, & Persliden, 2009; McEntee et al., 2007). In comparison with the other FOMs that have been utilised in radiography in the previous studies, the new FOM has two distinctive features. In the previous studies, FOM uses SNR and CNR as IQ metrics (Bosmans et al., 2012; Dragusin, Smans, Jacobs, Inal, & Bosmans, 2008) while the new FOM uses the CD as a metric. This favours the detectability of lesions such as lung cancer, which is of paramount importance in CXR. Another feature of equation 4 is that it considers the size of the objects (lesions or anatomical features) in the image during the IQ evaluation. The above 2 features (utilising the CD as a metric whilst considering the size of the objects) are considered to be an advantage of the suggested FOM when compared with those previously published in the literature (Bosmans et al., 2012; Dragusin, Smans, Jacobs, Inal, & Bosmans, 2008). Generally, using CNR and SNR as IQ metrics for FOM calculations can be criticised because they do not consider the size of the object (Bath, 2010). Several studies (McAdams, Samei, Dobbins, Tourassi, & Ravin, 2006; Håkansson et al., 2005) have investigated lesion size and have reported that this is essential for visualisation and detection in CXR. Equation 4 ($(IQF_{inv})^2$/ IAK) helps to describe the performance of the medical imaging system in producing a high detection of LCD objects with low radiation dose. In other words, it examines how efficient the imaging system is for LCD detection. Therefore, proposing and utilising the new formula of FOM in this thesis, instead of using the existing FOM formulas, is appropriately justified.

4.3 Validation for using the CDRAD 2.0 Phantom for IQ and LV Evaluation and for CXR Optimisation Studies

This section of the methodology discusses the empirical work required for achieving the 2 objectives of this thesis: 1) to determine if a correlation exists between LCD detectability using CDRAD 2.0 phantom, visual IQ and LV; and 2) to investigate the validity of using LCD detectability using a CDRAD 2.0 phantom in CXR optimisation studies.

In order to achieve the 2 objectives, it was necessary to utilise 2 different phantoms- an anthropomorphic chest phantom and a CDRAD 2.0 phantom. The former is required for evaluating visual IQ and LV, while the latter is required for evaluating LCD detectability. |The former was represented by the Lungman phantom loaded with 2 spherical phantom lesions that simulate pulmonary nodules. A 10 mm +100 HU (Hounsefield unit) lesion was placed in left upper lobe; a 12 mm +100 HU lesion was placed in the right middle lobe (**Figure (4-19))**. Lesion type and placement was based on the need to simulate a clinically relevant scenario and also take into account the complexity of the anatomical surroundings. This approach has been previously reported in the literature (Jessop et al., 2015). A Wolverson Arcoma Arco Ceil general radiography system (Wolverson X-Ray Ltd, Willenhall, West Midlands, UK), with a Varian X-ray tube (Varian medical systems, Salt Lake City, UT, USA) was used to acquire the images. The X-ray tube has a Tungsten-Rhenium anode with an angle of 12º, and an inherent filtration of 3.0 mm Aluminium equivalent (for 75 kV). A fixed anti-scatter grid (10:1 ratio, 40 line/cm frequency) was used, as it is representative of the anti-scatter grids commonly used in clinical radiography departments (Fauber, 2016). A Cesium Iodide (CsI) image detector (Konica Minolta Medical Imaging USA INC, Wayne, NJ, USA) was used and this had an image area size of 35 cm x 43 cm with a 1,994 x 2,430-pixel matrix, and a pixel size=175 um.

42 images of the Lungman phantom were acquired at 180cm SID without using any additional filtration and using manual exposure control. Different settings of kVp (70, 80, 90, 100, 110,120 and 130 kVp) and mAs (1, 2 and 3.2 mAs), with and without anti-scatter grids were applied during image acquisition. The images were acquired as follows: at each kVp value six images were acquired but with different mAs levels, with and without an anti-scatter radiation grid as illustrated in **Table (4-6)**. For example, image number one was acquired with 70 kVp, 1 mAs and without a grid. 42 CDRAD 2.0 phantom images were acquired using the same acquisition parameters that were applied to the Lungman phantom.

Then, the CDRAD 2.0 phantom was placed between 10 cm of PMMA slabs (5 cm of PMMA above and 5 cm of PMMA below the CDRAD 2.0 phantom) to simulate an adult patient (Bacher, Smeets, Vereecken, et al., 2006; De Crop et al., 2012; De Hauwere et al., 2005). All the acquisition parameters (**Table (4-6**)) were selected since they were similar to those used for adult CXR in clinical practice. In addition, utilising a manual exposure control type with an extremely wide range of acquisition parameters allowed for the acquisition of a series of images with a wide range of qualities. This was felt to be important for investigating the correlation between the 2 phantoms across a wide range of image qualities. IQ and LV for the Lungman phantom images were evaluated visually using a relative VGA by 7 observers. Ethical approval for the visual IQ evaluations, by observers, was granted from the University of Salford (HSR1617-76) (**Appendix A-3**). The IQ and LV were evaluated using a relative VGA by applying a 5 point Likert scale (1 = much worse, 2 = worse, 3 = equal to, 4 = better, 5 = much better). The criteria used for IQ and LV evaluation were the observer's characteristics, the types of display monitor, the image viewing conditions and other evaluation issues- identical to those previously mentioned in subsection 4.2.6.1.

The reason behind selecting 2 simulated lesions was that one had relatively good visibility whilst the other had low visibility so that it was possible to determine if a correlation exists between LCD detectability and LV under 2 conditions.

LCD detectability represented by the IQF_{inv} for CDRAD 2.0 phantom images were evaluated physically using CDRAD phantom analyser software in the same way that was mentioned in subsection 4.2.6.2.1. The correlation between the IQF_{inv} from CDRAD 2.0 phantom images with IQ, LV for the specified simulated lesions independently was calculated. **Figure (4-2)** (subsection 4.1) demonstrates the stages of this method in a diagrammatical format.

Table 4-6 This table demonstrates the acquisition parameters used to acquire the images from the Lungman and the CDRAD 2.0 phantoms and the respective number (code) of each image and its corresponding acquisition parameters.

Image number	kVp	mAs			Grid	SID (cm)	Additional filtration
1,2,3	70	1	2	3.2	No	180	No
4,5,6	80	1	2	3.2	No	180	No
7,8,9	90	1	2	3.2	No	180	No
10,11,12	100	1	2	3.2	No	180	No
13,14,15	110	1	2	3.2	No	180	No
16,17,18	120	1	2	3.2	No	180	No
19,20,21	130	1	2	3.2	No	180	No
22,23,24	70	1	2	3.2	Yes	180	No
25,26,27	80	1	2	3.2	Yes	180	No
28,29,30	90	1	2	3.2	Yes	180	No
31,32,33	100	1	2	3.2	Yes	180	No
34,35,36	110	1	2	3.2	Yes	180	No
37,38,39	120	1	2	3.2	Yes	180	No
40,41,42	130	1	2	3.2	Yes	180	No

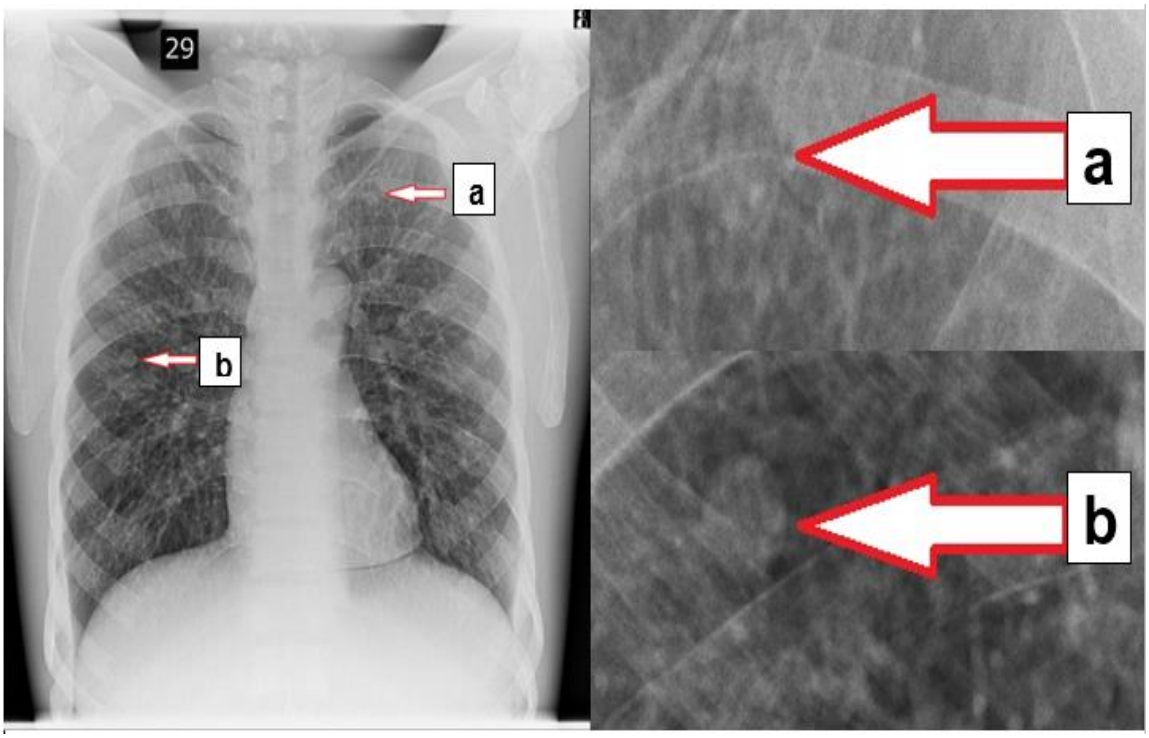

Figure 4-19 Lungman chest phantom image illustrating the 2 simulated lesions for the LV evaluation. (a) The 10 mm +100 HU lesion was placed in left upper lobe. (b) The 12 mm +100 HU lesion was placed in the right middle lobe.

4.4 Statistical Analysis

4.4.1 Normality tests

The normality of the data was examined visually by a frequency distribution (histogram), a boxplot, and a Q-Q plot (quantile-quantile plot). In addition, an objective analysis using a Shapiro-Wilk test was conducted in conjunction with the visual examination to obtain a precise estimation of the normality as recommended by Filed (2009). Evaluating the normality of the data is an important statistical procedure because it determines the path of the statistical tests that need to be used, and also whether parametric or non-parametric statistics are needed.

4.4.2 Statistical Analyses

For the purposes of this thesis, graphical illustration and descriptive statistics were utilised for describing the results that have been collected from the X-ray machines. Graphical illustration was chosen because it is extremely beneficial for summarising data and illustrating patterns. In addition, it helps interpret the data in a straightforward form, and

this enables the reader to understand it easily (Pallant, 2010). Another reason behind using graphical illustration is related to a lack of control group for the collected data. This is because, depending on the type of the X-ray machines used, the detectors, protocols, acquisition parameters such as kVp, mAs, SID, and filters can be different. Therefore, the data of the variations in IQ, LV, IQF_{inv}, SNR, CNR, CI, FOM and radiation dose between and within hospitals was analysed graphically and presented as a series of bar charts. The mean, the first quartile and the third quartile values are presented on bar charts. Data is also presented as mean ± SD, and the SD represents the following: 1) visual IQ and LV data, wherein SD represents the variation in image scoring between observers for the same image; 2) physical IQ values from CDRAD 2.0 phantom images data (IQF_{inv}, CNR, SNR and CI), wherein the SD represents the measured values of the physical IQ from three replicated images.; and 3) radiation dose data, wherein the SD represents the measured values of three repeated exposures for each examination.

On the other hand, Pearson's correlation was used to investigate the relationship between IQ and LV evaluation from the normal adult sized and larger sized Lungman phantom images with their corresponding IAK values. Furthermore, Pearson's correlation was utilised for investigating the correlation between the IQF_{inv} from the CDRAD 2.0 phantom images represented the adult patient and their corresponding IAK values. Pearson's correlation, for normal distribution data, and the Spearman's correlation, for non-normal data, were used together to investigate the correlation between the IQF_{inv}, IQ and LV and their corresponding IAK values. They also helped to find the CNR, SNR and CI values for all the age groups under study. The IQ data for the experiment that investigated the influence of phantom thickness on IQ (represented by Lungman chest phantom with and without fat jacket) was analysed statistically using the Mann-Whitney test for non-normal distribution data, and the significance level was determined when $p<0.05$. The Mann-Whitney was used to test the hypotheses that phantom thickness has no statistically significant impact on IQ. The unpaired t-test for parametric data was utilised for investigating the effect of phantom thickness on physical IQ metrics (SNR, CNR and CI) and IAK. The interpretation of the strength of the correlation (r) was conducted based on the studies of Cohen, 1988; Filed (2009), at which r=0.1-0.29 (small), r=0.30-0.49 (medium), and r=0.50-1.0 (large).

To investigate the inter-observer variability scores between the observers during IQ and LV evaluation, the interclass correlation coefficients (ICC) using SPSS was utilised and

the 95% confidence levels are also reported. ICC is often utilised as a statistic for measuring the reliability index in test-retest, intra-rater, and inter-rater reliability analyses (Koo & Li, 2016). A more eligible measurement of reliability must consider both the degree of correlation and the agreement between measurements, and the ICC is an appropriate index for this (Koo & Li, 2016). According to (Koo & Li, 2016;Portney & Watkins, 2000), ICC values of less than 0.5 indicate poor reliability, values ranging between 0.5 and 0.75 indicate moderate reliability, values ranging between 0.75 and 0.9 indicate good reliability and values greater than 0.90 indicate excellent reliability. However, a level of reliability of either good or excellent is generally more preferable (Koo & Li, 2016).

Regarding the data generated for investigating the validity of LCD detectability using a CDRAD 2.0 phantom in IQ and LV evaluations and its validity for use in optimisation studies, data here was analysed by investigating the correlation (Pearson's correlation) between the IQ from the Lungman phantom and the IQF_{inv} values from the CDRAD 2.0 phantom. Spearman's correlation was utilised to investigate the correlation between both the LV for the first simulated lesion and the second simulated lesion (resulting from the Lungman phantom) with the IQF_{inv} values (resulting from the CDRAD 2.0 phantom).

4.5 Chapter Summary

In this chapter, the experimental work used for investigating the radiation dose and IQ variation, between and within hospitals, is presented. Within this framework, a performance evaluation of the included X-ray machines, regarding their imaging protocols, machines types, radiation dose and different types of IQ from the different phantoms types were investigated for both paediatric and adult CXR examinations. In addition, the FOM formulas used for evaluating and comparing IQ and radiation dose, between and within hospitals, is reported in this chapter. This also includes reporting the new FOM formula (from the CDRAD 2.0 phantom data) that utilised the IQF_{inv} as a metric for IQ evaluation. The advantages of the new FOM formula with a CDRAD 2.0 phantom is highlighted in this chapter.

This chapter also reports the experimental work designed to test a novel method used for validating the use of LCD detectability, using CDRAD 2.0 phantom, for IQ and LV evaluation and for its more wider use in optimisation studies. Within this framework the relationship between the LCD detectability, represented by the IQF_{inv}, and IQ and LV from the Lungman phantom was investigated.

Chapter 5: Results

5.1 Overview of Chapter

This chapter presents the results of this PhD thesis. Data is represented in 3 sections in order to aid understanding. The first section (section 5.2) presents data from the validation of the LCD detectability methodology. The aim of this section is to assess and report the validity of utilising LCD detectability, using the CDRAD 2.0 phantom, for evaluating visual IQ and LV, as well as to determine its validity in CXR optimisation studies. The second section (section 5.3) presents the radiation dose, IQ, FOM and clinical protocols collected from the different hospitals and X-ray machines. The aim of this section is to evaluate and report the variability in radiation dose, IQ, FOM and clinical protocols, both between and within hospitals. The third section (section 5.4) presents a comparison between the visual IQ data, physical IQ data and radiation dose data for the standard sized Lungman phantom and the larger sized Lungman phantom, between the different hospitals and X-ray machines. Here, the aim was to investigate the influence of phantom thickness on physical IQ, visual IQ and radiation dose, both between and within hospitals.

In Section 5.2 of the results chapter, the data for the LCD detectability validation experiment is presented in two ways: visually, as represented by linear regression and by Pearson/Spearman correlation tests.

Section 5.3 is divided into 6 subsections. Subsection 5.3.1 provides a summary of the main variation results and is presented as a series of tables (**Tables (5-2)** - (**5-10)**). These tables present the minimum and maximum values as well as the values of the 3rd quartile for radiation dose, IQ metrics and the FOM, both between and within hospitals for all of the age groups and for the phantoms.

The next subsections (5.3.2. to 5.3.6.) are organised by age (neonate, 1 year, 5 years, 10 years and adult). Within each group, the results are presented under 2 subheadings: 1) IQ, radiation dose and FOM variations; and 2) clinical protocol variations. To avoid any confusion, and to consider any additional complexity introduced within the neonate and adult age groups (because two different phantoms have been utilised for data collection), the results for IQ, radiation dose and FOM variation are further separated based by phantom type (i.e. physics-based and anthropomorphic).

The first subheading (IQ, radiation dose and FOM variations) compares the values of visual IQ, LV, IQF_{inv} and their corresponding IAK. Also, the FOM graphs from the physical and visual IQ evaluations are presented under these subheadings.

Due to the large number of bar charts generated, the results of the variation of the physical IQ metrics (CNR, SNR, CI) for acquisitions in the five age groups obtained using the CDRAD 2.0 phantom, and that for neonate and adult age groups obtained from anthropomorphic phantoms (Gammex and Lungman) and their corresponding radiation dose (IAK), are presented in **Appendix B** and **Appendix C**, respectively. The error bars for CNR, SNR and CI resulting from the CDRAD 2.0 phantom represent their respective SD obtained from assessing the three repeat images. The error bar for IAK represents the SD obtained from measuring the three repeat radiation exposures. It should be noted that the CNR, SNR and CI results from the anthropomorphic phantoms do not have error bars because these physical measurements were conducted for only one image, while the error bars for IAK represent the SD obtained from measuring 3 repeat radiation exposures.

The second subheading of section 2 (clinical protocols variation) identifies differences in the clinical protocols used between X-ray machines and hospitals across all of the age groups.

The third section (5.4) presents visual IQ, SNR, CNR and CI data alongside IAK, using the standard and larger sized Lungman phantom. The aim of this subsection is to evaluate and report the influence of phantom thickness on visual IQ, SNR, CNR, CI and IAK, both between and within hospitals.

In this chapter, the data on IQ and FOM are presented as a series of bar charts, whereas the IAK values are presented as a dashed line against the corresponding IQ values. The colour coding system of the bars within these graphs allocates each hospital a different coloured bar. The X-ray machines within a hospital are given a code but not a different colour. As previously stated, hospitals and X-ray machines are coded using two letters and one number: the letter (H) refers to the hospital and the letter (X) refers to the X-ray machine i.e. Hospital 1, X-ray machine 1 would be H1X1. It is important to stress that no hospitals are individually identifiable within this thesis.

5.2 Results of the LCD Detectability Method Validation for IQ and LV and for Optimisation Studies in CXR

As mentioned in the methods chapter (section 4.3.), 42 images from both CDRAD 20 and Lungman phantoms, with different image qualities, were acquired using a range of acquisition parameters. In order to investigate the correlation between LCD detectability, visual IQ and LV with the anthropomorphic Lungman phantom, the CDRAD 2.0 phantom images were evaluated physically using a CDRAD phantom software analyser. The Lungman phantom images were evaluated visually by seven observers using a relative VGA method. The IQF_{inv} values and the visual (IQ and LV) data together with the acquisition parameters are illustrated in **Table (5-1)**. These data were first examined to investigate the normality of the data using the Shapiro-Wilk test and also visually to determine the most appropriate correlation statistic. It was concluded that the IQ and the IQF_{inv} were distributed normally. Conversely, LV for both simulated lesions showed a non-normal distribution.

Correlation analysis using a Pearson correlation coefficient (r) was done to determine the level of the correlation between the visual IQ and IQF_{inv} and demonstrated a strong positive correlation r=0.91 ($p<0.001$). Spearman correlation was used to compare IQF_{inv} and LV. A good correlation was observed for both simulated lesions; the first simulated lesion (the simulated lesion with the lower visibility located in the left upper lobe) showed a r value of 0.79 (p=0.001); the second simulated lesion (the simulated lesion with the higher visibility that located in the right middle lobe) showed an value of 0.68 (p=0.001). **Figure (5-1)** demonstrates the relationship between the IQ as a function of the IQF_{inv} as a linear regression curve. **Figure (5-2)** and **Figure (5-3)** demonstrate the relationship between the LV for the first selected simulated lesion and the second selected simulated lesion respectively, again as a function of the IQF_{inv}, using linear regression.

Excellent inter-observer agreement between the observers was calculated (ICC=0.90; 95% confidence interval: 0. 84-0. 94) during the IQ evaluations. A good inter-observer agreement was calculated against the first simulated lesions (ICC= 0.84; 95% confidence interval: 0. 68-0. 90) and second simulated lesions (ICC=0.83; 95% confidence interval: 0. 714-0. 91)).

Table 5-1 A summary of the acquisition parameters used to acquire images and the resultant IQFinv , IQ and LV values.

	Mean (SD) IQF_{inv} score			**Mean (SD) Relative VGA score**								
				LV						**IQ**		
				Left upper lobe			**Right middle lobe**					
mAs / kVp	1	2	3.2	1	2	3.2	1	2	3.2	1	2	3.2
70	0.28(0.02)	0.40(0.07)	0.54(0.06)	3.00(0.00)	3.43(1.13)	4.00(1.73)	4.42(2.15)	4.57(2.44)	5.71(2.36)	6.86(1.57)	8.00(3.42)	9.00(4.00)
80	0.79(0.07)	0.86(0.03)	1.00(0.09)	5.43(1.81)	5.14(2.12)	7.14(1.95)	7.28(1.50)	5.85(2.27)	7.42(2.30)	9.29(3.50)	10.71(3.45)	12.00(3.32)
90	1.00(0.04)	0.59(0.06)	0.64(0.08)	7.29(1.25)	4.43(1.81)	5.86(1.57)	7.00(2.16)	5.57(2.70)	8.00(2.65)	12.75(3.55)	8.43(2.82)	11.43(3.26)
100	0.89(0.05)	1.24(0.06)	1.21(0.08)	6.57(2.07)	8.71(0.76)	9.00(0.00)	7.00(3.27)	8.71(1.25)	8.28(1.89)	11.43(4.58)	14.14(3.02)	15.71(2.57)
110	1.41(0.08)	1.57(0.10)	0.82(0.12)	7.43(1.51)	9.00(0.00)	5.86(1.77)	8.14(1.46)	8.28(1.89)	8.14(2.79)	17.43(1.51)	17.14(2.27)	10.71(3.82)
120	1.17(0.03)	1.40(0.05)	1.45(0.11)	7.57(2.15)	9.14(0.83)	9.00(0.00)	8.14(1.46)	8.57(1.27)	8.57(1.13)	13.71(2.21)	17.14(2.34)	17.43(1.51)
130	1.59(0.08)	1.84(0.18)	2.10(0.16)	8.57(1.13)	8.00(1.41)	9.00(1.83)	9.00(1.00)	8.71(0.76)	8.85(0.90)	17.86(0.38)	19.29(1.89)	18.57(1.62)
70*	0.60(0.09)	0.74(0.04)	0.88(0.06)	4.43(1.81)	8.00(1.53)	6.57(2.15)	9.71(2.63)	10.00(2.38)	8.85(3.76)	9.43(3.26)	10.57(4.08)	11.71(3.64)
80*	0.94(0.01)	1.15(0.11)	1.12(0.03)	7.43(1.51)	7.86(2.91)	8.00(2.58)	9.85(1.77)	10.00(1.73)	9.71(2.06)	12.57(3.41)	14.71(1.98)	14.86(1.86)
90*	1.21(0.10)	0.97(0.04)	1.17(0.04)	8.71(2.14)	8.00(2.45)	9.86(2.12)	10.57(1.62)	10.28(2.14)	10.28(1.70)	14.71(1.98)	14.00(2.16)	14.86(1.68)
100*	1.39(0.13)	1.68(0.13)	1.65(0.01)	8.86(2.79)	9.43(1.72)	9.00(1.53)	10.42(2.51)	10.28(1.80)	10.14(2.04)	15.86(1.35)	16.14(2.54)	18.00(2.83)
110*	1.89(0.16)	1.75(0.05)	1.35(0.15)	9.71(1.98)	9.00(3.00)	9.14(2.27)	10.42(2.70)	10.85(1.46)	10.14(1.86)	16.14(2.04)	16.57(1.51)	15.43(2.51)
120*	1.57(0.03)	1.88(0.12)	2.18(0.13)	10.29(2.50)	8.86(1.68)	8.71(3.40)	10.57(1.72)	11.00(1.15)	11.00(2.00)	16.57(0.98)	16.43(3.41)	16.43(2.57)
130*	2.24(0.13)	2.39(0.12)	2.38(0.30)	8.71(2.69)	9.00(3.87)	9.14(3.48)	10.42(1.40)	10.14(1.57)	10.42(1.40)	19.14(3.24)	18.43(3.78)	19.00(2.89)
Protocols indicated by a * were undertaken using an anti-scatter radiation grid.												

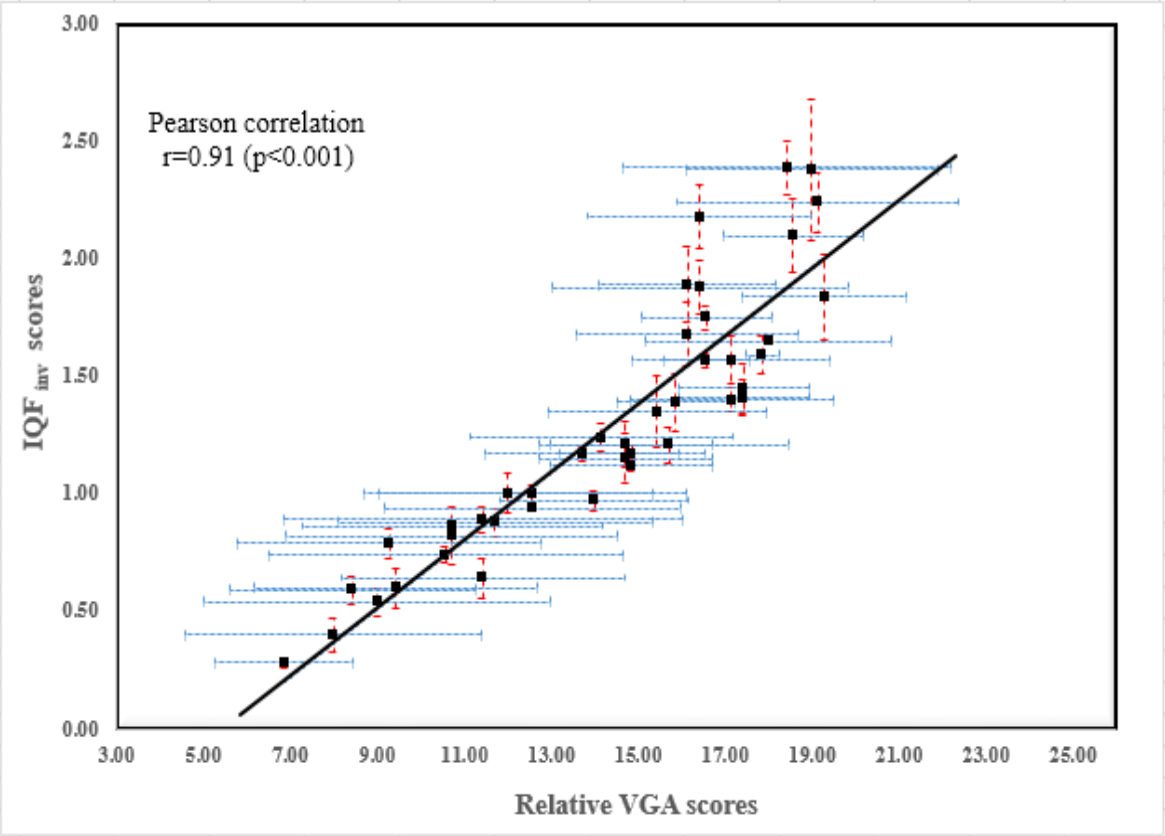

Figure 5-1 Linear regression curve between the mean IQFinv scores against the mean IQ scores. Error bars across the x-axis (the blue dotted lines) represent the SD between the scores of observers, while error bars on the y-axis (the red dashed lines) represent the SD of the scores of the three replicated CDRAD 2.0 images.

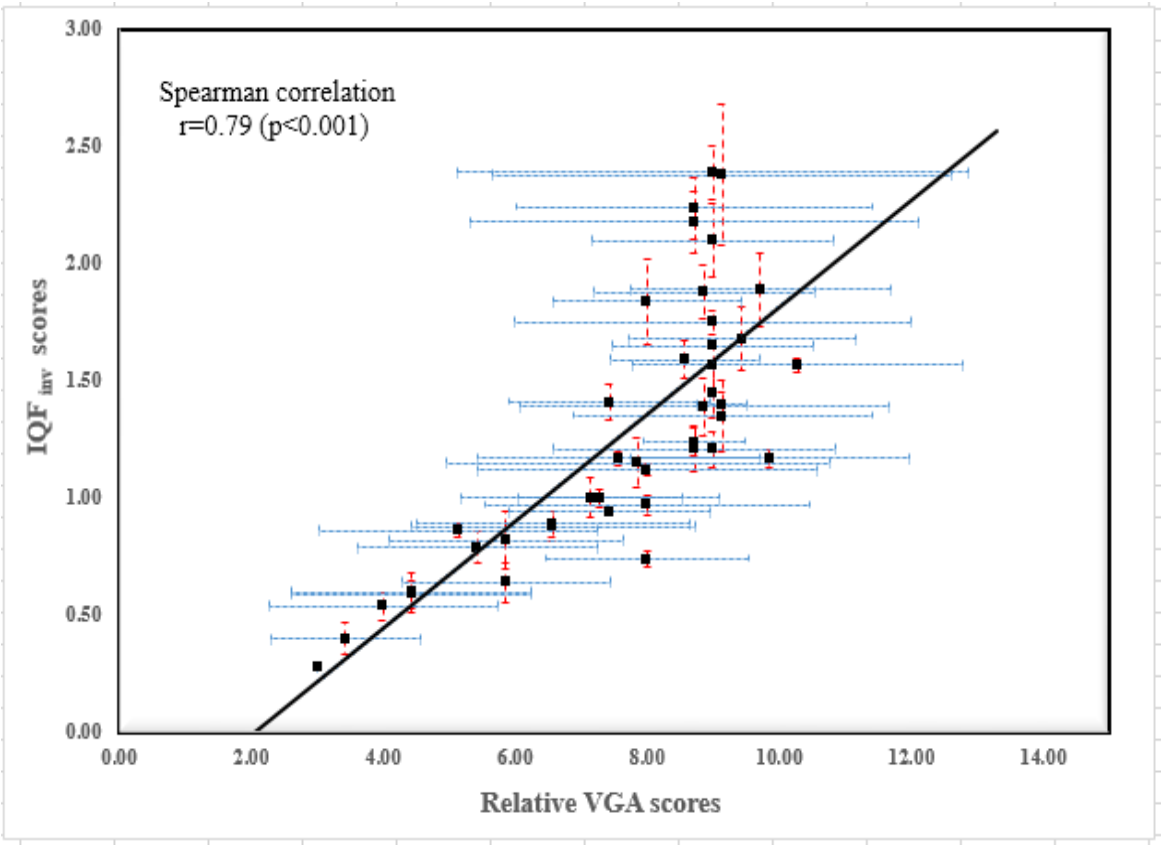

Figure 5-2 Linear regression curve between the mean IQFinv scores against the mean LV scores for the first simulated lesion (the simulated lesion with the lower visibility that located in the left upper lobe). Error bars in the x direction (the blue dotted lines) represent the SD between the scores of observers, while error bars in the y direction (the red dashed lines) represent the SD of the scores of the three replicated CDRAD 2.0 images.

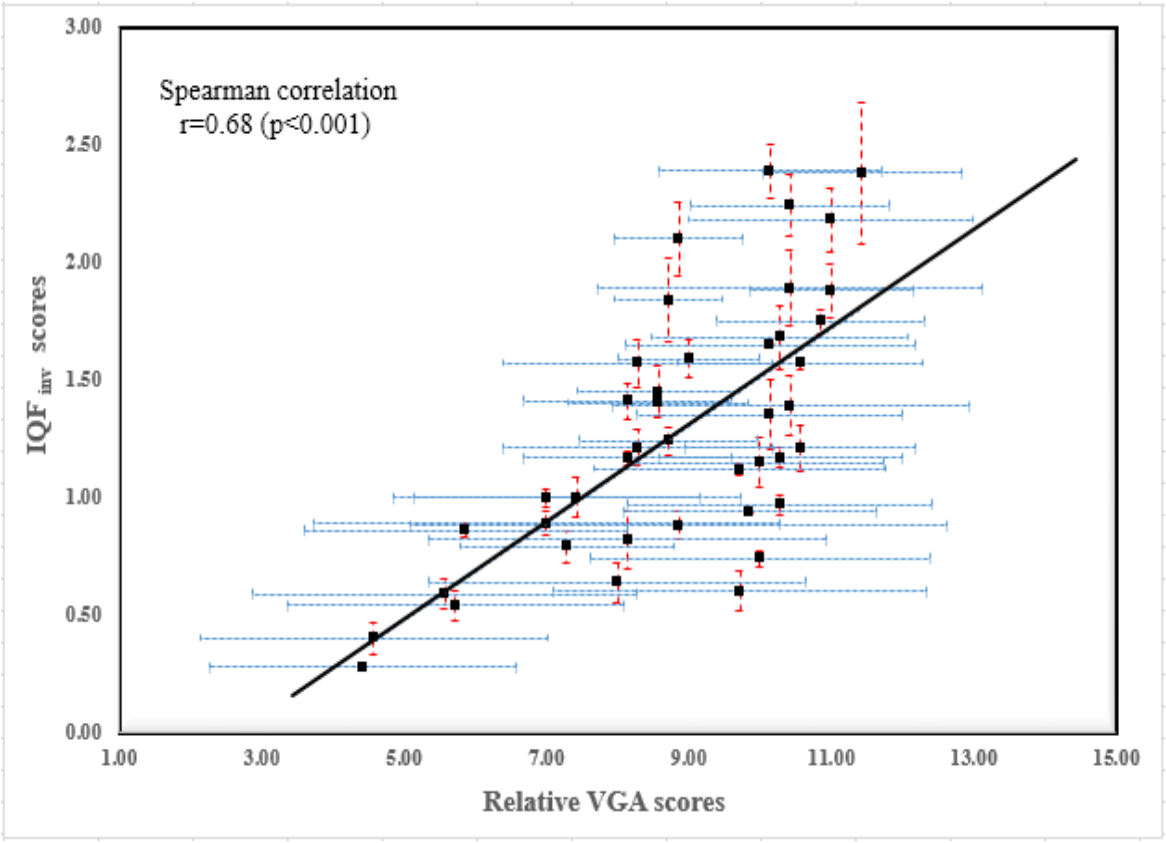

Figure 5-3 Linear regression curve between the mean IQFinv scores against the mean LV scores for the second simulated lesion (the simulated lesion with higher visibility that located in the right middle lobe). Error bars in the x direction (the blue dotted lines) represent the SD between the scores of observers, while error bars in the y direction (the red dashed lines) represent the SD of the scores of the three replicated CDRAD 2.0 images.

5.3 Results of Image Quality, Radiation Dose, FOM and Clinical Protocols for the Range of X-ray Machines

5.3.1 Summary of Key Findings

Nine tables have been used to provide a summary of the key results (**Tables (5-2)-(5-10)**). **Tables (5-2)**, (**5-3)** and (**5-4)** present the minimum, maximum and 3rd quartile values of the IAK, IQF_{inv} and FOM obtained using the CDRAD 2.0 phantom, between and within the hospitals; **Table (5-3)** also presents the correlation between the IQF_{inv} and the respective IAK values for all 5 age groups; **Tables (5-5)**, (**5-6)** and (**5-7)** present the minimum, maximum and 3rd quartile values of the SNR, CNR and CI using the CDRAD 2.0 phantom.

Table 5-2 Lists the variation in IAK (µGy) obtained using the CDRAD 2.0 phantom for the 5 age groups, both between and within the hospitals.

Age	Between the hospitals		Within the hospitals		3rd quartile
	Minimum	Maximum	Minimum	Maximum	
Neonate	8.56	52.62	21.79	52.62	34.45
1-year	5.44	82.82	36.78	82.82	53.37
5-year	10.97	59.22	11.75	59.22	45.52
10-year	13.97	100.77	35.72	100.77	43.61
Adult	17.26	239.15	122.58	239.15	132.32

Table 5-3 Lists the variation in mean IQF_{inv} score for the 5 age groups, both between and within the hospitals and the correlation between IQF_{inv} scores and the respective IAK.

Age	Between the hospitals		Within the hospitals		3rd quartile between hospitals	Correlation (r) between IQF_{inv} and IAK
	Minimum	Maximum	Minimum	Maximum		
Neonate	1.40	4.44	2.45	4.44	3.18	0.54
1-year	0.96	4.73	2.33	4.73	2.84	0.20
5-year	0.87	1.81	0.98	1.46	1.51	0.24
10-year	0.90	2.39	1.27	2.39	1.73	0.10
Adults	0.83	2.18	1.52	2.18	1.61	0.45

Table 5-4 Lists the variation in mean FOM obtained using CDRAD 2.0 phantom for the 5 age groups, both between and within the hospitals.

Age	Between the hospitals		Within the hospitals		3rd quartile
	Minimum	Maximum	Minimum	Maximum	
Neonate	0.08	0.41	0.14	0.41	0.33
1-year	0.05	0.51	012	0.51	0.17
5-year	0.04	0.22	0.06	0.22	0.18
10-year	0.01	0.14	0.02	0.13	0.13
Adults	0.01	0.14	0.02	0.05	0.11

Table 5-5 Lists the variation in mean SNR obtained using CDRAD 2.0 phantom for the 5 age groups, both between and within the hospitals.

Age	Between the hospitals		Within the hospitals		3rd quartile
	Minimum	Maximum	Minimum	Maximum	
Neonate	16.74	84.56	27.89	84.56	45.71
1-year	14.48	90.15	38.05	90.15	49.28
5-year	10.28	74.93	19.20	38.69	39.90
10-year	13.07	55.72	13.07	55.72	50.16
Adults	15.39	58.88	15.39	58.88	38.98

Table 5-6 Lists the variation in mean CNR obtained using CDRAD 2.0 phantom for the 5 age groups, both between and within the hospitals.

Age	Between the hospitals		Within the hospitals		3rd quartile
	Minimum	Maximum	Minimum	Maximum	
Neonate	3.50	13.94	4.31	13.94	9.18
1-year	2.41	10.43	5.82	10.43	9.47
5-year	2.22	5.33	3.14	5.33	4.78
10-year	2.10	7.72	2.10	7.72	5.61
Adults	2.26	6.92	3.84	6.92	4.78

Table 5-7 Lists the variation in mean CI obtained using CDRAD 2.0 phantom for the 5 age groups, both between and within the hospitals.

Age	Between the hospitals		Within the hospitals		3rd quartile
	Minimum	Maximum	Minimum	Maximum	
Neonate	49.00	156.95	101.74	156.95	127.00
1-year	14.48	90.15	38.05	90.15	49.28
5-year	40.57	83.74	60.78	83.74	73.26
10-year	50.66	122.64	30.08	122.64	80.82
Adults	22.12	197.88	55.61	197.88	69.01

Table (5-8) and **Table (5-9)** present the minimum, maximum and 3rd quartile values of IAK, as well as the visual IQ for adults and neonates, between and within the hospitals. In addition, **Table (5-9)** presents the correlation between the visual IQ for adults and neonates with their IAK values. Finally, **Table (5-10)** presents the FOM values from the visual IQ assessments for adults and neonates and their 3rd quartile values.

Table 5-8 Lists the variation in mean IAK (µGy) obtained using the anthropomorphic phantoms (Lungman and Gammex), both between and within the hospitals.

Age	Between the hospitals		Within the hospitals		3rd quartile
	Minimum	Maximum	Minimum	Maximum	
Neonate (Gammex)	8.11	49.94	22.26	49.94	34.44
Standard size adult (Lungman)	19.28	136.29	61.24	136.29	89.05
Larger size adult (Lungman)	27.43	384.73	195.76	384.73	213.21

Table 5-9 Lists the variation in mean visual IQ score for adults and neonate, both between and within the hospitals, and the correlation between the IQ and IAK is presented.

Age	Between the hospitals		Within the hospitals		3rd quartile	Correlation IQ and IAK
	Minimum	Maximum	Minimum	Maximum		
Neonate (Gammex)	2.17	3.50	2.33	3.33	3.00	0.46
Standard sized adult (Lungman)	12.00	21.33	15.00	21.33	17.00	0.34
Larger size adult (Lungman)	14.17	23.67	16.83	21.50	23.00	0.60

Table 5-10 Lists the variation in mean FOM obtained using the anthropomorphic phantom (Lungman and Gammex), both between and within the hospitals.

Age	Between the hospitals		Within the hospitals		3rd quartile
	Minimum	Maximum	Minimum	Maximum	
Neonate (Gammex)	0.13	0.55	0.16	0.55	0.31
Standard size adult (Lungman)	0.26	1.76	1.12	1.76	1.12
Larger size adult (Lungman)	0.12	1.57	1.05	1.57	0.71

5.3.2 Neonate Age Group

5.3.2.1 CDRAD 2.0 Phantom

5.3.2.1.1 Image Quality, Radiation Dose and FOM variations

Figure (5-4) demonstrates the IQF_{inv} values against the corresponding IAK values, between and within the hospitals. The IQF_{inv} scores between the hospitals ranged from 1.40 (H6X2) to 4.44 (H2X4) (mean 2.58) with a difference of 104.1%, while within the hospitals the range was lower: 2.45 (H2X3) to 4.44 (H2X4) (mean 3.37) with a difference of 57.8%. The third quartile value between the hospitals was 3.18. IAK values between the hospitals ranged from 8.56 (H6X1) to 52.62 µGy (H2X4) (mean 24.93 µGy) with a difference of 144.0%, whereas the range within the hospitals was smaller - 21.79 (H2X1) to 52.62 µGy (H2X4) (mean 34.28 µGy) with a difference of 82.9%. The third quartile value between the hospitals was 34.45µGy. The calculated FOM values for the same acquisitions, between hospitals, ranged from 0.08 (H8X1) to 0.14 (H4X1) (mean 0.02) with difference of 54.5%, while the FOM within the hospitals ranged from 0.14 (H4X2) to 0.41 (H4X1) (mean 0.21) with a difference of 98.2% (the 3rd quartile value between the hospitals was 0.33; **Figure (5-5)**).

Physical IQ parameters (CNR, SNR and CI) also showed wide variation, both between and within hospitals. The CNR values (between the hospitals) ranged from 3.50 (H6X2) to 13.94 (H6X5) (mean 7.17) with a difference of 119.7%. The within hospitals range was 4.31 (H3X4) to 13.94 (H6X5) (mean 9.38) with a difference of 105.5% (the 3rd quartile value between the hospitals was 9.18). SNR ranged from 16.74 (H6X2) to 84.56 (H2X4) (mean 40.80) with a difference of 133.9% between hospitals. The within hospital variation in SNR was 27.89 (H2X1) to 84.56 (H2X4) (mean 50.79) with difference of 100.8%. The third quartile between the hospitals was 45.71. CI varied from 49.00 (H6X1) to 156.95 (H2X1) (mean 99.56) with a difference of 104.8% between hospitals. The within hospital CI variation was 57.95 (H3X3) to 143.12 (H3X1) (mean 12.64) with a difference of 84.7% (3rd quartile value between the hospitals 123.48).

The Shapiro-Wilk test shows that IQF_{inv}, IAK, SNR, CNR and CI are normally distributed ($p>0.05$; **Appendix D-1**). A correlation analysis using the Pearson's correlation coefficient between the IQF_{inv} and the IAK resulted in a good positive correlation ($r=0.54$; $P=0.02$). Pearson's correlations between the IQF_{inv} and SNR, CNR and CI were $r=0.67$ ($P=0.003$), $r=0.69$ ($P=0.002$) and $r=0.80$ ($P=0.001$), respectively.

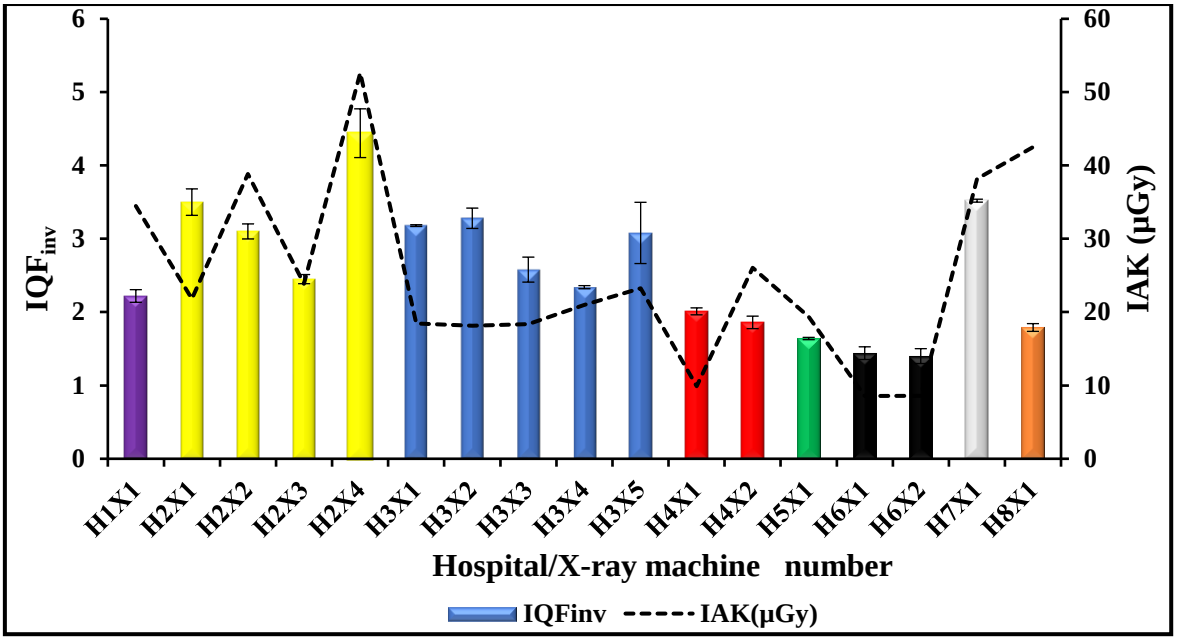

Figure 5-4 A comparison of IQFinv against the IAK for the neonate age group, between/within hospitals. The error bars in this chart represent the SD in IQFinv obtained from measuring three repeated images. The dashed line represents the average value of IAK resulting from measuring 3 repeated radiation exposures.

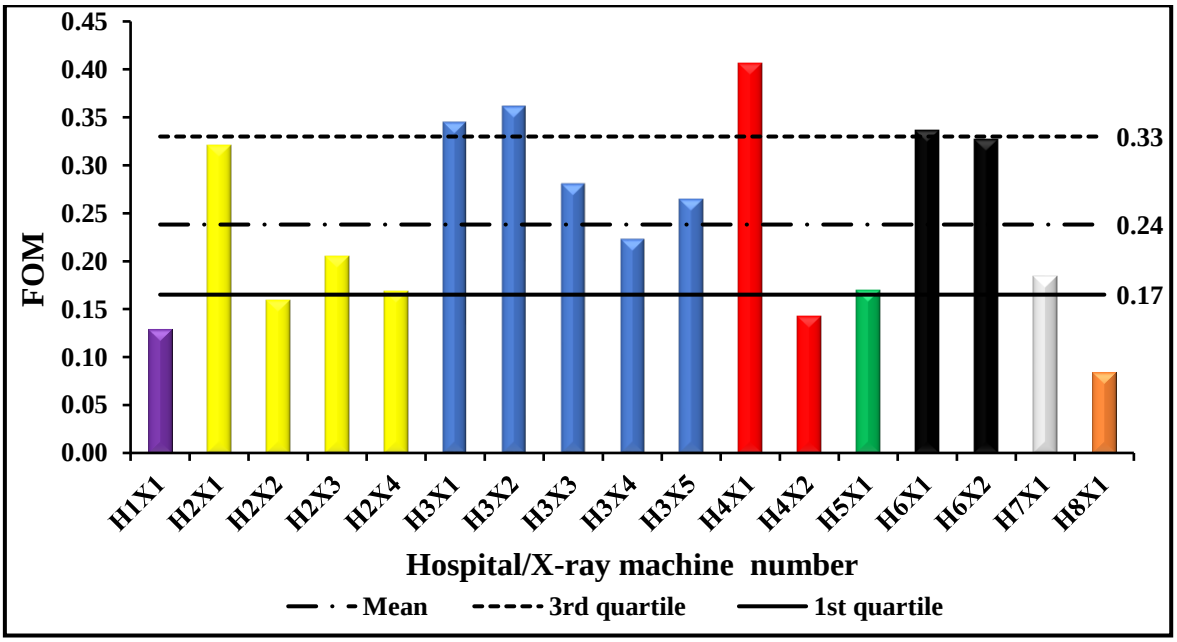

Figure 5-5 Bar chart displaying the distribution of FOM for the neonate age group. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) of the FOM values, respectively.

5.3.2.2 Gammex Phantom:

5.3.2.2.1 Image Quality, Radiation Dose and FOM Variations

The bar chart below (**Figure (5-6)**) compares the visual IQ scores against the corresponding IAK values for the neonate age group. **Figure (5-7)** presents the variation in FOM using the same values. The IQ scores between the hospitals ranged from 2.00 (H1X1) to 3.50 (H2X4) (mean 2.69) with a difference of 54.5% (3rd quartile between the hospitals was 3.00), while within the hospitals the range was lower 2.33 (H3X3) to 3.33 (H3X5) (mean 2.73) with a difference of 35.3%. IAK values between the hospitals ranged from 8.11 (H6X1) to 49.94 µGy (H2X4) (mean 24.45 µGy), with a difference of 144.1% percentage, whereas the range within the hospitals was smaller: 22.26 (H2X3) to 49.94 µGy (H2X4) (mean 34.86 µGy) with a difference of 76.7%. The 3rd quartile IAK value between the hospitals was 34.44µGy. The FOM values for the same acquisitions between hospitals ranged from 0.13 (H1X1) to 0.55 (H4X1) (mean 0.28) with a difference of 123.5%. The FOM range within the hospitals ranged from 0.16 (H4X2) to 0.55 (H4X1) (mean 0.36) with a difference of 109.9%; the third quartile between the hospitals was 0.31.

Physical IQ parameters (CNR and SNR) showed a large variation, both between and within hospitals. The CNR values between the hospitals ranged from 7.65 (H1X1) to 33.18 (H3X3) (mean 17.53) with a difference of 125.1%. The within hospitals range was smaller: 10 (H3X4) to 33.18 (H3X3) (mean 19.85) with difference of a 107.4%. The 3rd quartile CNR value between the hospitals was 21.44. SNR ranged from 31.48 (H6X2) to 97.99 (H3X3) (mean 53.40) with a difference of 102.7% between hospitals. Within hospitals, CNR varied from 44.61 (H3X1) to 97.99 (H3X3) (mean 61.22) with a smaller difference of 74.9%. The 3rd quartile CNR value between the hospitals was 62.29.

Shapiro-Wilk tests showed that IQ, IAK, SNR and CNR were normally distributed ($p>0.05$; **Appendix D-2**). A correlation analysis using the Pearson's correlation coefficient compared IQ and IAK and gave a moderate positive correlation ($r=0.46$) which was not statistically significant ($P=0.06$). A moderate inter-observer agreement between the observers was observed with an ICC of 0.54 (ICC=0.0.54; 95%CI: 0. 17-0. 80) during the IQ evaluations. Pearson's correlations between the IAK and SNR and CNR were $r = 0.08$ ($P=0.74$) and $r = 0.002$ ($P=0.99$), respectively. Pearson's correlations between the IQ and SNR and CNR were observed to be $r = 0.27$ ($P=0.29$) and $r = 0.13$ ($P=0.61$) respectively.

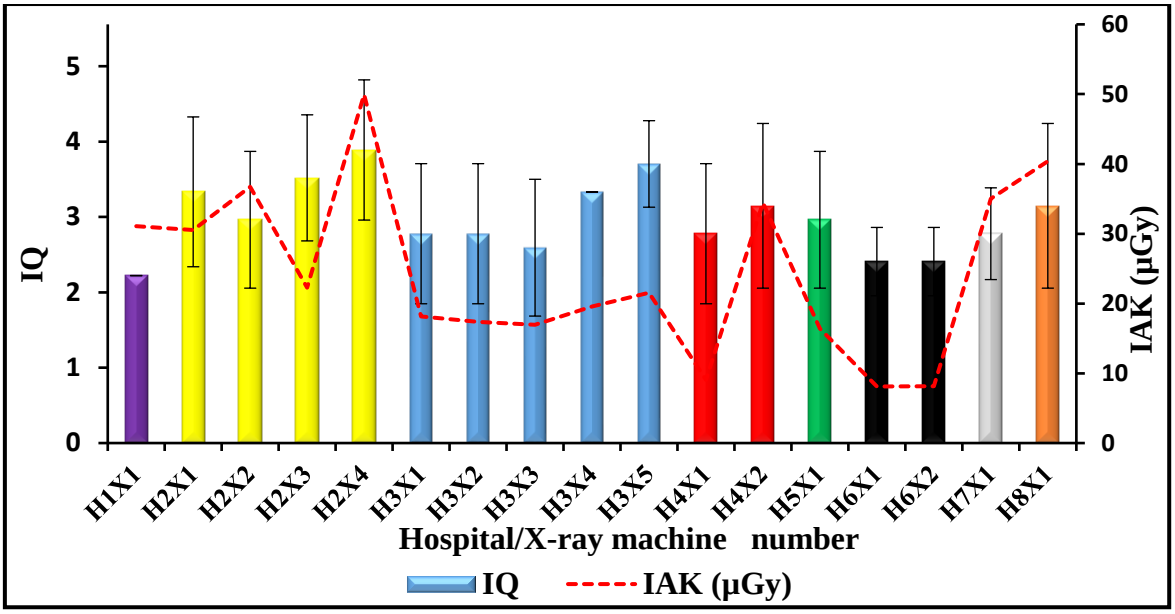

Figure 5-6 A comparison of measured IQ and IAK for the neonate age group. The error bars in this chart represent the SD in IQ obtained from the visual evaluations by observers. The dashed line represents the average value of IAK resulting from measuring 3 repeated radiation exposures.

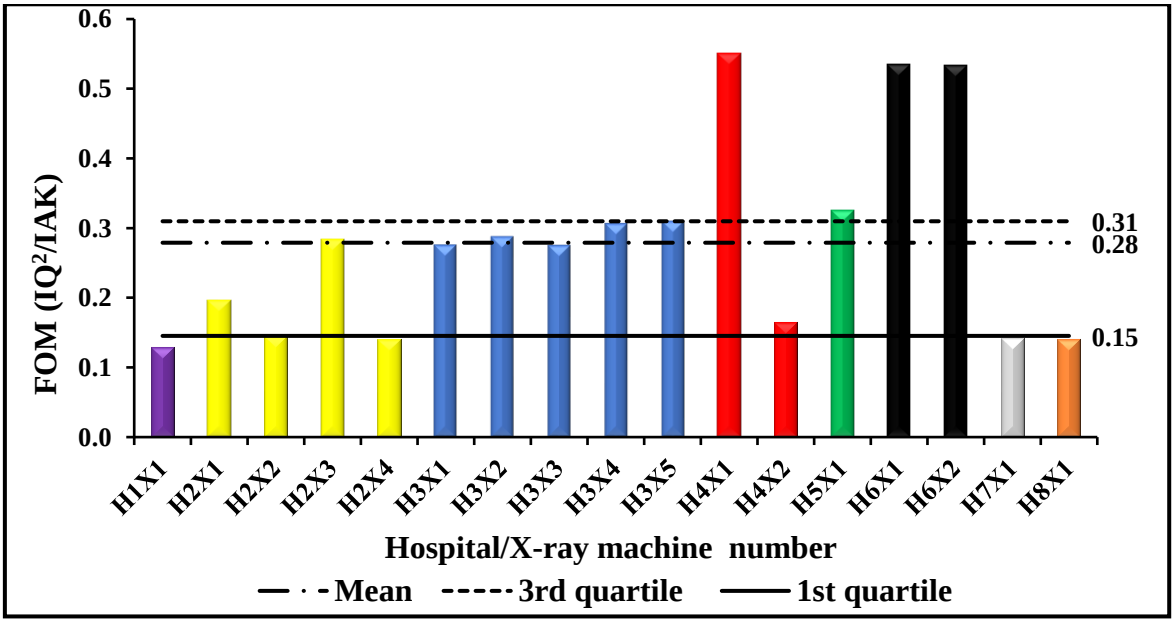

Figure 5-7 Bar chart displaying the distribution FOM calculated for the neonate age group. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) of the FOM values are displayed, respectively.

5.3.2.3 Clinical Protocol Variations:

Table (5-11) describes the examination techniques and acquisition parameters used for the neonate age group, both between and within the hospitals. All X-ray machines used manual exposure control and no anti-scatter radiation grids; additional filtration was not utilised except for the 5 X-ray machines in hospitals 3 (H3) and the X-ray machine H4X1 (additional filtration equal to 0.1 mm Cu and 1mm Al+0.1 mm Cu, respectively). The variations in the range of kVp and SID were 60 to 64.5 kVp and 110 to 135 cm, respectively. A large variation was observed in mAs, from 0.63 to 2 mAs. Finally, there was no consistency between the X-ray machines in the focal spot type used: eight of the X-ray machines (H1X1, H2X1, H2X2, H2X3, H2X4, H6X1, H6X2, H7X1 and H8X1) used broad focus while the other 9 X-ray machines used fine.

Table 5-11 A summary of X-ray machine types, examination techniques and acquisition parameters used for **neonate** age group.

H^A No.	X^B No.	X type	D^C type	X Manufacturer	D Manufacturer	Position of phantom	Type of exposure control	AEC	Grid	Additional filtration (mm)	Focal spot type	SID (cm)	kVp	mAs
1	1	Static	IDR	Philips	Philips	Supine AP	Manual	None	No	No	Broad	135	60	1.25
2	1	Mobile	IDR	Carestream	Carestream	Supine AP	Manual	None	No	No	Broad	115	64	0.80
	2	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	No	Broad	115	68	1.00
	3	Static	DDR	Carestream	Carestream	Supine AP	Manual	None	No	No	Broad	130	60	1.20
	4	Static	IDR	Samsung	Samsung	Supine AP	Manual	None	No	No	Broad	115	63	2.00
3	1	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	0.1 Cu	Fine	115	64.5	1.40
	2	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	0.1 Cu	Fine	115	64.5	1.40
	3	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	0.1 Cu	Fine	115	64.5	1.40
	4	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	0.1 Cu	Fine	115	64.5	1.40
	5	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	0.1 Cu	Fine	115	64.5	1.40
4	1	Static	DDR	Philips	Philips	Supine AP	Manual	None	No	1 Al + 0.1 Cu	Fine	110	60	1.00
	2	Static	CR	Philips	Carestream	Supine AP	Manual	None	No	No	Fine	110	60	1.00
5	1	Static	CR	Siemens	Carestream	Supine AP	Manual	None	No	No	Fine	120	60	0.63
6	1	Static	DDR	Philips	Philips	Supine AP	Manual	None	No	No	Broad	135	60	1.20
	2	Static	DDR	Philips	Philips	Supine AP	Manual	None	No	No	Broad	135	60	1.20
7	1	Static	DDR	Carestream	Carestream	Supine AP	Manual	None	No	No	Broad	115	60	1.20
8	1	Static	CR	Siemens	Carestream	Supine AP	Manual	None	No	No	Broad	110	63	1.25

H^A: Hospital; X^B: X-ray machine; D^C: Detector.

5.3.3 One Year Age Group

5.3.3.1 Image Quality, Radiation Dose and FOM Variations

Figure (5-8) demonstrates the IQF_{inv} values against the corresponding IAK values, between and within hospitals. **Figure (5-9)** demonstrates the FOM values, both between and within hospitals. The IQF_{inv} scores between the hospitals ranged from 0.96 (H4X2) to 4.73 (H2X4) (mean 2.39) with difference of 132.5%. Within the hospitals the range was lower: 2.33 (H2X3) to 4.73 (H2X4) (mean 3.32) with difference of a 68.0%. The 3rd quartile IQF_{inv} value between the hospitals was 2.84. IAK values between the hospitals ranged from 5.44 (H4X1) to 82.82 µGy (H2X4) (mean 36.84 µGy) with a difference of 175.3%. The range within hospitals was smaller 36.78 (H2X3) to 82.82 µGy (H2X4) (mean 54.73 µGy) with a difference of 77.0%. The 3rd quartile IAK value between the hospitals was 53.37µGy. The FOM values for the same acquisitions, between hospitals, ranged from 0.05 (H5X1) to 0.51 (H4X1) (mean 0.17) with a difference of 164.3%. The FOM within the hospitals ranged from 0.12 (H4X2) to 0.51 (H4X1) (mean 0.32) with a difference of 123.8%. The third quartile between the hospitals was 0.17. Physical IQ parameters (CNR, SNR and CI) also showed wide variation, both between and within hospitals. The CNR values between the hospitals ranged from 2.41 (H4X2) to 10.43 (H2X4) (mean 7.02) with a difference of 124.9%. The within hospital range was 4.67 (H3X4) to 9.82 (H3X3) (mean 8.42) with a smaller difference of 71.1%. The 3rd quartile CNR value between the hospitals was 9.47. SNR values varied from 14.48 (H8X1) to 90.15 (H2X4) (mean 42.28) with a difference of 45.1% between hospitals. Within the hospitals the SNR varied from 31.59 (H2X3) to 90.15 (H2X4) (mean 50.58) with a difference of 96.2%. The 3rd quartile SNR value between the hospitals was 49.28. CI ranged from 41.58 (H4X2) to 199.95 (H2X4) (median 81.55) between the hospitals with a difference of 131.1%. Within the hospitals the range was smaller: 93.28 (H2X3) to 199.95 (H2X4) (median 117.21), which represented a difference of 72.8%. The 3rd quartile CI value between the hospitals was 98.03. The Shapiro-Wilk tests showed that IQF_{inv}, IAK, SNR and CNR were normally distributed ($p>0.05$; **Appendix D-1**), while CI values had a non-parametric distribution ($p<0.05$; **Appendix D-1**). Correlation analysis using the Pearson's correlation coefficient compared IQF_{inv} and IAK and resulted in a weak positive correlation ($r=0.20$) which was not statistically significant ($P=0.44$). However, Pearson's correlations between the IQF_{inv} and SNR, CNR and Spearman's correlation between the IQF_{inv} and the CI were $r=0.80$ ($P=0.001$), CNR $r=0.83$ ($P=0.001$) and $r=0.90$ ($P=0.001$), respectively.

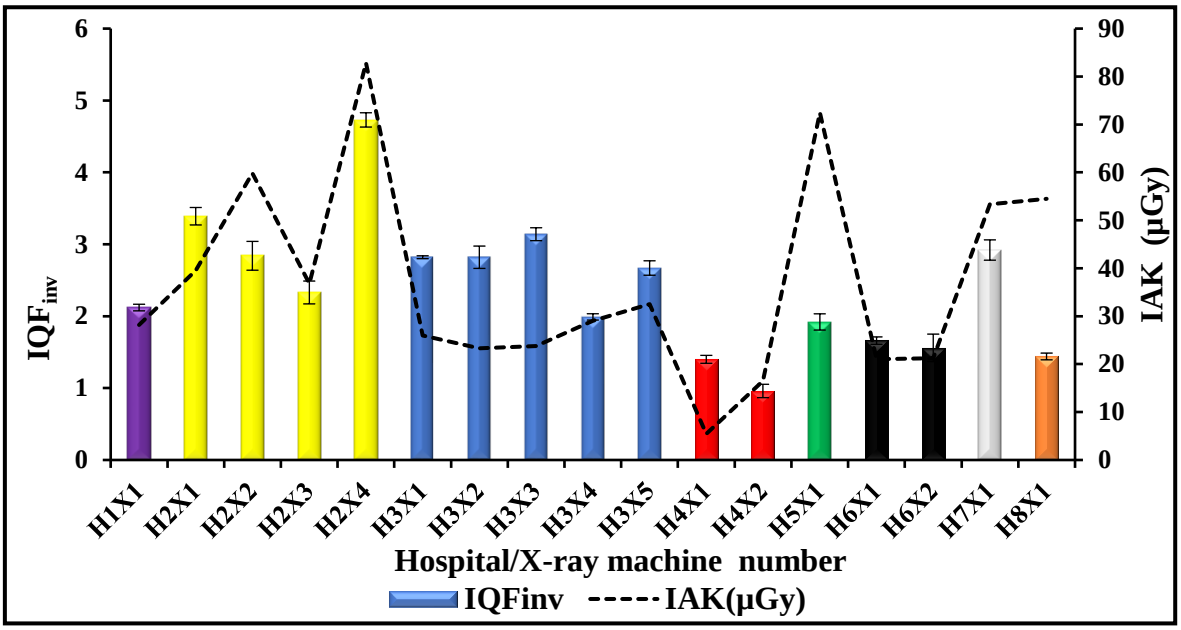

Figure 5-8 Illustrates the comparison of IQFinv scores against IAK for the 1-year age group. The error bars in this chart represent the SD in IQFinv obtained from measuring 3 repeated images. The dashed line represents the average value of IAK resulting from measuring 3 repeated radiation exposures.

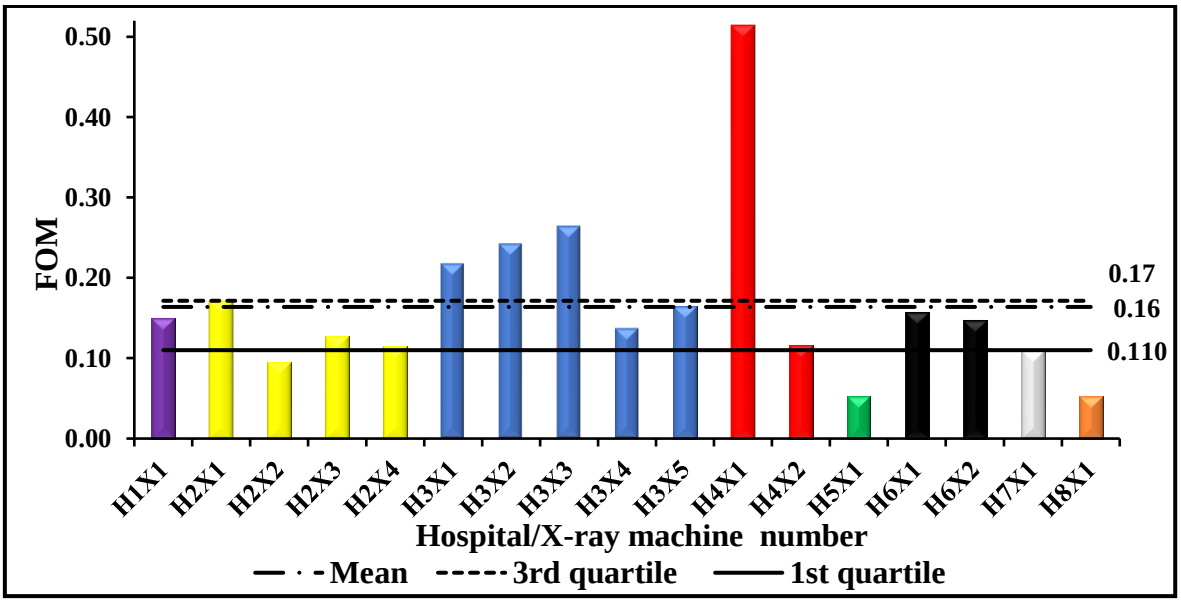

Figure 5-9 Bar chart displaying the distribution of FOM values obtained for the 1-year age group. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) of the FOM values, respectively.

5.3.3.2. Clinical Protocol Variations:

Table (5-12) describes the examination techniques and acquisition parameters used for the one-year age group, between and within the hospitals. All of the X-ray machines used manual exposure control without the presence of an anti-scatter radiation grid (except one X-ray machine (H1X1) which used the AEC. The kVp and mAs values ranged from 60 to 73 kVp, and 1.00 to 2.5 mAs, respectively. SID varied more widely with a range from 110 to 180 cm. There was no consistency between the X-ray machines in the focal spot type used: eight of the X-ray machines (H1X1, H2X1, H2X2, H2X3, H2X4, H6X1, H6X2 H7X1and H8X1) used broad focus while the other nine X-ray machines used fine focus. Additional filtration (0.1 mm Cu) was used in the 5 X-ray machines in hospitals 3 (H3) and the X-ray machine (H4X1).

Table 5-12 A summary of X-ray machine types, examination techniques and acquisition parameters used for 1-year age group.

H[A] No.	X[B] No.	X type	D[C] type	X Manufacturer	D Manufacturer	Position of phantom	Type of exposure control	AEC	Grid	Additional filtration	Focal spot type	SID (cm)	kVp	mAs
1	1	Static	IDR	Philips	Philips	Supine AP	AEC	Central	No	No	Broad	135	73	1.55
2	1	Mobile	IDR	Carestream	Carestream	Supine AP	Manual	None	No	No	Broad	115	68	1.20
	2	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	No	Broad	115	73	1.25
	3	Static	DDR	Carestream	Carestream	Supine AP	Manual	None	No	No	Broad	130	63	1.60
	4	Static	IDR	Samsung	Samsung	Supine AP	Manual	None	No	No	Broad	115	68	2.50
3	1	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	0.1 mm Cu	Fine	115	66	1.60
	2	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	0.1 mm Cu	Fine	115	68	1.60
	3	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	0.1 mm Cu	Fine	115	66	1.60
	4	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	0.1 mm Cu	Fine	115	68	1.60
	5	Static	IDR	Siemens	Siemens	Supine AP	Manual	None	No	0.1 mm Cu	Fine	115	68	1.60
4	1	Static	DDR	Philips	Philips	Erect AP	Manual	None	No	1 mm Al + 0.1 mm Cu	Fine	180	70	1.00
	2	Static	CR	Philips	Carestream	Erect AP	Manual	None	No	No	Fine	180	70	1.00
5	1	Static	CR	Siemens	Carestream	Supine AP	Manual	None	No	No	Fine	120	68	1.80
6	1	Static	DDR	Philips	Philips	Supine AP	Manual	None	No	No	Broad	135	73	1.30
	2	Static	DDR	Philips	Philips	Supine AP	Manual	None	No	No	Broad	135	73	1.60
7	1	Static	DDR	Carestream	Carestream	Supine AP	Manual	None	No	No	Broad	115	60	1.60
8	1	Static	CR	Siemens	Carestream	Supine AP	Manual	None	No	No	Broad	110	70	1.25

H[A]: Hospital; X[B]: X-ray machine; D[C]: Detector.

5.3.4 Five Year Age Group

5.3.4.1 Image Quality, Radiation Dose and FOM Variations

Figure (5-10) demonstrates the IQF_{inv} values against the corresponding IAK values, between and within the hospitals. **Figure (5-11)** demonstrates the FOM values between and within the hospitals. The IQF_{inv} scores between the hospitals ranged from 0.87 (H8X1) to 1.81 (H2X2) (mean 1.35) with a difference of 70.1%. Within the hospitals the IQF_{inv} range was lower: 0.98 (H3X3) to 1.46 (H3X4) (mean 1.24) with a difference of 39.3%. The 3rd quartile IQF_{inv} value between the hospitals was 1.50. IAK values between the hospitals ranged from 10.97 (H3X3) to 59.22 µGy (H6X1) (median 21.43 µGy) with a difference of 137.5%. The range in variability for X-ray machines within the same hospitals was smaller: 11.75 (H4X1) to 52.94 µGy (H4X2) (median 32.35 µGy) with a difference of 127.3%. The 3rd quartile IAK value between the hospitals was 45.52µGy. The FOM values for the same acquisitions between hospitals ranged from 0.04 (H4X2 and H6X1) to 0.22 (H2X1 and H3X4) (mean 0.13) with a difference of 138.5%, while the FOM within the hospitals ranged from 0.07 (H2X4)- 0.22 (H2X1) (mean 0.12) with a smaller difference of 103.4%. The third quartile between the hospitals was 0.18.

Physical IQ measures (CNR, SNR and CI) also showed wide variation, both between and within hospitals. The SNR values between the hospitals ranged from 10.28 (H8X1) to 47.93 (H5X1) (mean 31.36) with a difference of 129.4%. Within the hospitals the range was smaller: 13.22 (H3X3) to 39.68 (H3X4) (mean 27.22) with a difference of 100.0%. The 3rd quartile SNR value between the hospitals was 39.90. CNR varied from 2.22 (H8X1) to 5.33 (H2X2) (mean 3.89) with a difference of 82.4% between hospitals. Within the hospitals the range was slightly smaller, 2.44 (H3X2 and H3X3) to 5.21 (H3X4) (mean 3.83), with a difference of 72.4%. The 3rd quartile SNR value between the hospitals was 4.78. CI varied from 40.57 (H4X2) to 83.74 (H2X2) (mean 59.77) with a difference of 69.5% between hospitals. Within the hospitals the variability was smaller: from 51.42 (H3X3) to 77.34 (H3X5) (mean 66.07) with difference of 40.4%. The 3rd quartile CI value between the hospitals was 73.26.

Shapiro-Wilk tests showed that IQF_{inv}, CI, SNR and CNR were normally distributed ($P>0.05$; **Appendix D-1**), while IAK had a non-normal distribution ($P<0.05$; **Appendix D-1**). Analysis using the Spearman's correlation coefficient compared IQF_{inv} and IAK resulting in a weak non-statistically significant positive correlation ($r=0.24$; $P=0.35$).

Pearson's correlations between the IQF_{inv} and SNR, CNR and CI included a moderate non-statistically significant positive correlation (r = 0.42; P=0.10), a good statistically significant positive correlation (r =0.81; P=0.001) and r =0.51/P=0.05, respectively.

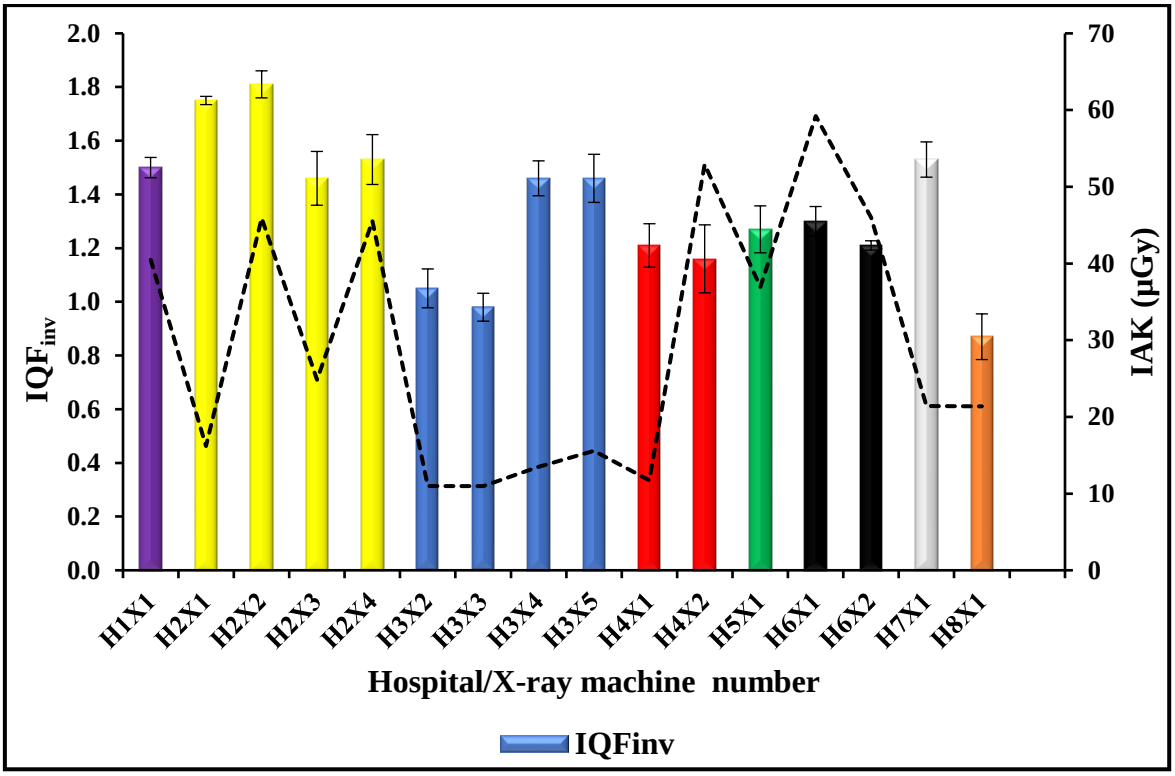

Figure 5-10 Graphical comparison of IQFinv values against IAK for the 5-year age group. The error bars in this chart represent the SD in IQFinv obtained from measuring 3 repeated images. The dashed line represents the average value of IAK resulted from measuring 3 repeated radiation exposures.

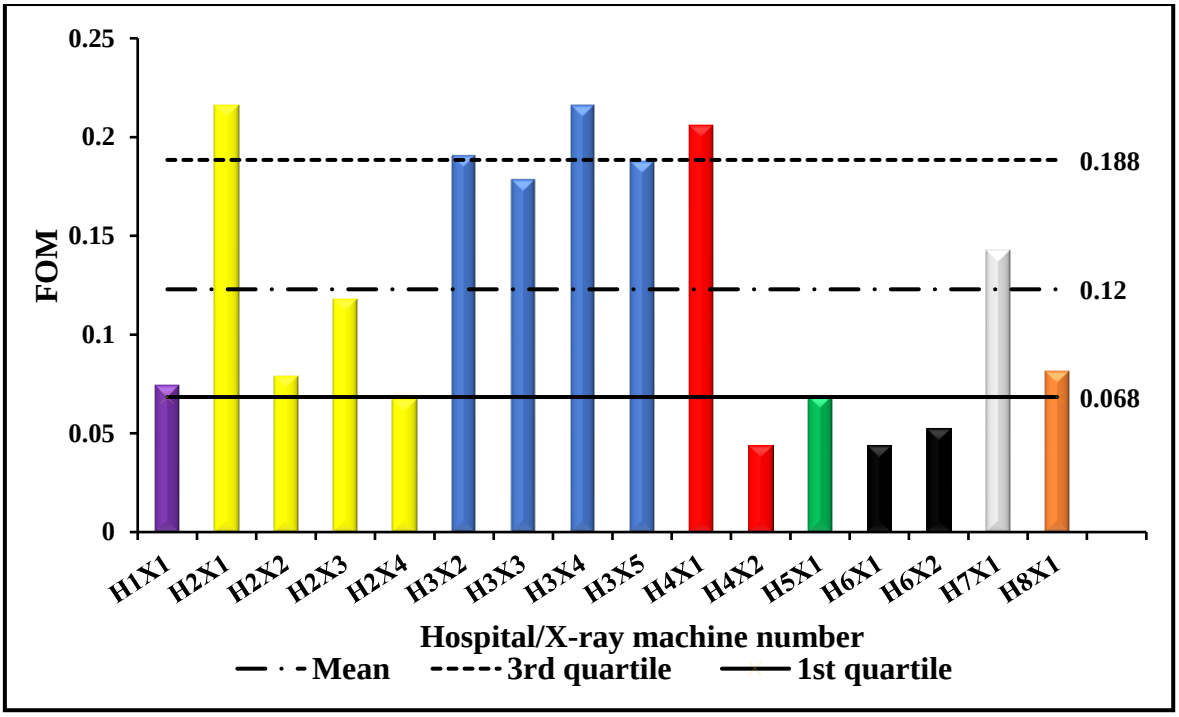

Figure 5-11 Bar chart displaying the FOM values calculated for the 5-year age group. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) of the FOM values, respectively.

5.3.4.2. Clinical Protocol Variations:

Table (5-13) displays the examination techniques and acquisition parameters used for the 5-year age group, both between and within the hospitals. An anti-scatter radiation grid was not used and the SID was set to 180 cm, except for with 1 X-ray machine (H8X1) which used 200 cm. However, there was no consistency between and within the hospitals on the other acquisition parameters used. Types of exposure control varied with H1X1, H2X3, H6X1, and H6X2 (24%) using an AEC, while the remainder of the X-ray machines (H2X1, H2X2, H2X4, H3X1, H3X2, H3X4, H3X5, H4X1, H4X2, H4X5, H7X1 and H8X1) (76%) used manual exposure termination. The kVp and mAs values varied considerably, ranging from 63 to 120 kVp and 1.00 to 3.5 mAs, respectively. Additional filtration (0.1 mm Cu) was used in 5 X-ray machines in hospital 3 (H3) and the X-ray machine, H4X1. Finally, there was no consistency between the hospitals when selecting the focal spot type used. 8 of the X-ray machines (H1X1, H2X1, H2X2, H2X3, H2X4, H6X1, H6X2, H7X1 and H8X1) used broad focus while the other 9 X-ray machines used fine focus.

Table 5-13 A summary of X-ray machine types, examination techniques and acquisition parameters used for **5-years** age group between/within hospitals.

H^A No.	X^B No.	X type	D^C type	X Manufacturer	D Manufacturer	Position of phantom	Type of exposure control	AEC	Grid	Additional filtration	Focal spot type	SID (cm)	kVp	mAs
1	1	Static	IDR	Philips	Philips	Erect	AEC	R+L	No	No	Broad	180	73	2.16
2	1	Mobile	IDR	Carestream	Carestream	Erect	Manual	None	No	No	Broad	180	80	1.00
	2	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	No	Broad	180	77	2.50
	3	Static	DDR	Carestream	Carestream	Erect	AEC	R+L	No	No	Broad	180	120	1.20
	4	Static	IDR	Samsung	Samsung	Erect	Manual	None	No	No	Broad	180	72	3.50
3	1	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	0.1 mm Cu	Fine	180	70	1.60
	2	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	0.1 mm Cu	Fine	180	70	2.00
	3	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	0.1 mm Cu	Fine	180	70	1.60
	4	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	0.1 mm Cu	Fine	180	70	2.00
	5	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	0.1 mm Cu	Fine	180	70	2.00
4	1	Static	DDR	Philips	Philips	Erect	Manual	None	No	1mm Al + 0.1 mm Cu	Fine	180	70	2.00
	2	Static	CR	Philips	Carestream	Erect	Manual	None	No	No	Fine	180	70	2.00
5	1	Static	CR	Siemens	Carestream	Erect	Manual	None	No	No	Fine	180	70	2.00
6	1	Static	DDR	Philips	Philips	Erect	AEC	R+L	No	No	Broad	180	81	2.50
	2	Static	DDR	Philips	Philips	Erect	AEC	R+L	No	No	Broad	180	73	2.50
7	1	Static	DDR	Carestream	Carestream	Erect	Manual	None	No	No	Broad	180	63	2.00
8	1	Static	CR	Siemens	Carestream	Erect	Manual	None	No	No	Fine	200	85	1.25

H^A: Hospital; X^B: X-ray machine; D^C: Detector.

5.3.5 Ten Years Age Group:

5.3.5.1 Image Quality, Radiation Dose and FOM variations

Figure (5-12) demonstrates the IQF_{inv} values against the corresponding IAK values, both between and within hospitals. **Figure (5-13)** demonstrates the FOM values between and within hospitals. The IQF_{inv} scores between the hospitals ranged from 0.90 (H8X1) to 2.39 (H2X1) (median 1.23) with a difference of 90.6%. For the within hospital comparison, the IQF_{inv} range was lower (1.27 (H2X4) to 2.39 (H2X1); median 38.33 which was a difference of 61.2%. The 3rd quartile IQF_{inv} value between the hospitals was 1.73. IAK values, between the hospitals, ranged from 13.97 (H3X3) to 100.77 µGy (H2X2) (median 35.72 µGy). This was a difference of 151.3%. However, the range for X-ray machines within the hospitals was smaller 35.72 (H2X3) to 100.77 µGy (H2X2) (median 38.34 µGy), being a 95.3% difference. The 3rd quartile IAK value between the hospitals was 43.61µGy.

FOM values for the same acquisitions, between hospitals, ranged from 0.03 (H4X2 and H6X1) to 0.15 (H3X3) (mean 0.09) with a difference of 133.3%. The FOM values for the within hospital comparisons ranged from 0.04 (H2X2) to 0.13 (H2X1) (mean 0.08) and had a lower (105.9%) percentage difference. The 3rd quartile FOM value between the hospitals was 0.13. Physical IQ measures (CNR, SNR and CI) also showed wide variation, both between and within hospitals. The CNR values between the hospitals ranged from 2.10 (H3X2)-7.72 (H3X5) (mean 4.27) with a 114.5% percentage difference, which was the same within the hospitals. The 3rd quartile CNR value between the hospitals was 5.61. SNR varied from 13.07 (H3X2) to 55.72 (H3X4) (mean 35.80) with a difference of 124.0%, and again was the same within the hospitals. The 3rd quartile SNR between the hospitals was 50.16. CI, ranging from 30.09 (H4X2) to 122.64 (H2X2) (mean 59.91) with a difference of 121.2 %. Between the hospitals, this ranged from 50.67 (H2X4) to 122.64 (H2X2) (mean 87.53) with a smaller (83.1%) difference. The 3rd quartile CI values between the hospitals was 80.82. Shapiro-Wilk tests showed that CI, SNR and CNR were approximately normally distributed ($p>0.05$; **Appendix D-1**), while IQF_{inv} and IAK values had a non-normal distribution ($p<0.05$; **Appendix D-1**). Analysis using Spearman's correlation coefficient resulted in a positive weak, non-statistically significant correlation between IQF_{inv} and the IAK ($r=0.10$; $P=0.71$). While, correlations between IQF_{inv} and SNR, CNR and CI were $r=0.22$ ($P=0.39$), $r = 0.85$ ($P=0.001$) and $r=0.68$ ($P=0.001$), respectively.

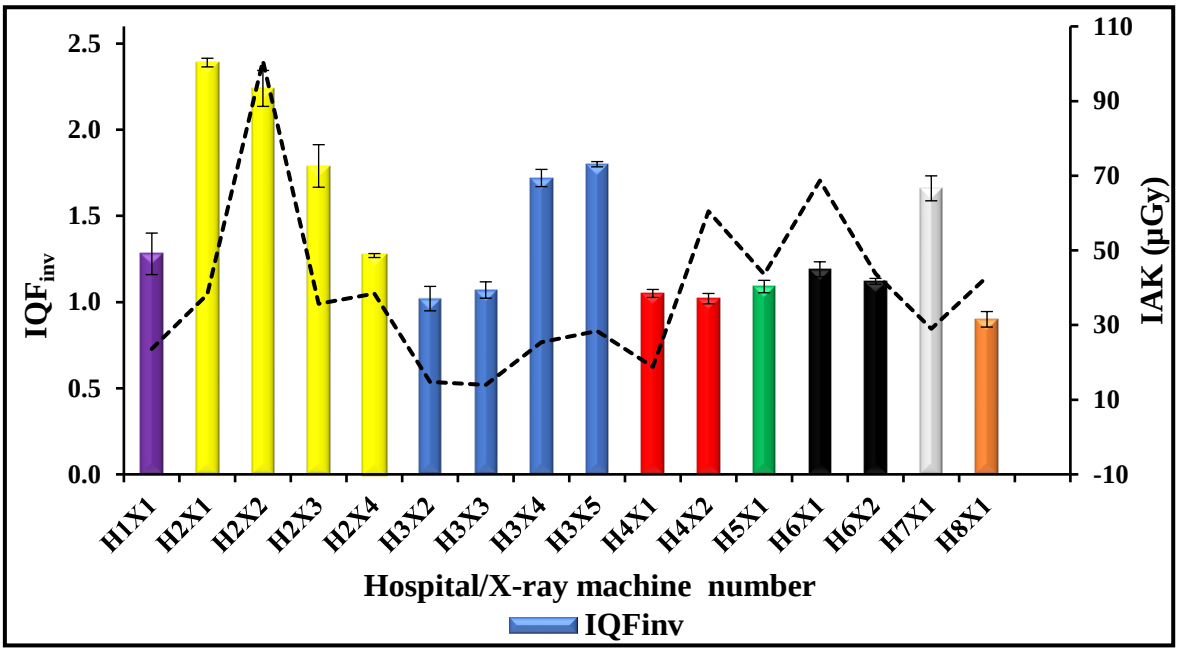

Figure 5-12 A comparison of IQFinv values against IAK for the 10-year age group. The error bars in this chart represent the SD in IQFinv obtained from measuring 3 repeated images. The dashed line represents the average value of IAK resulted from measuring 3 repeated radiation exposures.

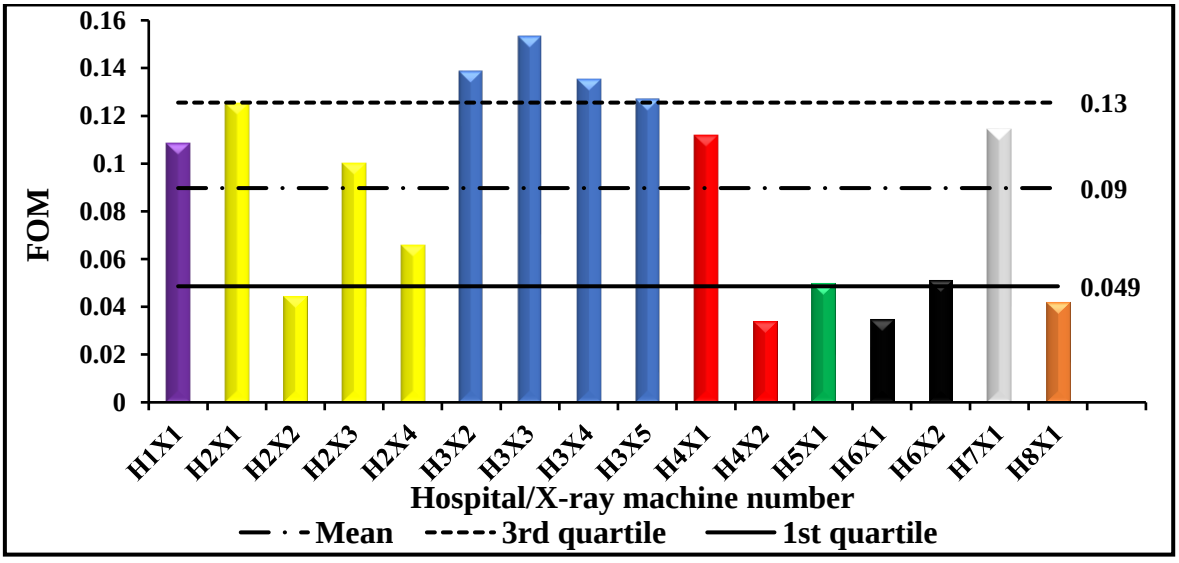

Figure 5-13 Bar chart displaying the range of FOM values for the 10 -year age group. Dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) of the FOM values, respectively.

5.3.5.2 Clinical Protocols Variations:

Table (5-14) describes the examination techniques and acquisition parameters used for the 10-year age group, both between and within the hospitals. All the hospitals did not use an anti-scatter radiation grid, and SID was set at 180 cm (except one hospital (H8X1) which used 200 cm). However, there was no consistency in the other acquisition parameters used. The type of exposure control varied. H1X1, H2X3, H2X4, H4X1, H4X2, H6X1 and H6X2 (41%) used the AEC, while the remaining X-ray machines utilised a manual exposure control. The kVp and mAs ranged from 65 to 125 kVp and 0.53 to 5.0 mAs, respectively. Additional filtration was used in three hospitals, H1X1, H3 and H4X1(equal to 1.0 Al+0.1 mm Cu and 0.1 mm Cu). Finally, there was no consistency between the hospitals in the choice of focal spot type used; ten of the X-ray machines (H1X1, H2X1, H2X2, H2X3, H2X4, H3X4, H5X1, H6X1, H6X2 and H7X1) used broad focus while the other seven X-ray machines used fine focus.

Table 5-14 A summary of X-ray machine types, examination techniques and acquisition parameters used for the **10-year** age group.

H[A] No.	X[B] No.	X type	D[C] type	X Manufacturer	D Manufacturer	Position of phantom	Type of exposure control	AEC	Grid	Additional filtration	Focal spot type	SID (cm)	kVp	mAs
1	1	Static	IDR	Philips	Philips	Erect	AEC	R+L	No	1.0 Al+0.1 mm Cu	Broad	180	81	2.33
2	1	Mobile	IDR	Carestream	Carestream	Erect	Manual	None	No	No	Broad	180	80	2.20
	2	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	No	Broad	180	79	5.00
	3	Static	DDR	Carestream	Carestream	Erect	AEC	R+L	No	No	Broad	180	120	0.85
	4	Static	IDR	Samsung	Samsung	Erect	AEC	R+L	No	No	Broad	180	120	1.10
3	1	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	0.1 mm Cu	Fine	180	73	1.80
	2	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	0.1 mm Cu	Fine	180	81	2.50
	3	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	0.1 mm Cu	Fine	180	73	1.30
	4	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	0.1 mm Cu	Broad	180	81	2.50
	5	Static	IDR	Siemens	Siemens	Erect	Manual	None	No	0.1 mm Cu	Fine	180	81	2.50
4	1	Static	DDR	Philips	Philips	Erect	AEC	R+L	No	1 mm Al + 0.1mm Cu	Fine	180	125	0.53
	2	Static	CR	Philips	Carestream	Erect	AEC	R+L	No	No	Fine	180	125	1.42
5	1	Static	CR	Siemens	Carestream	Erect	Manual	None	No	No	Broad	180	75	2.00
6	1	Static	DDR	Philips	Philips	Erect	AEC	R+L	No	No	Broad	180	81	2.86
	2	Static	DDR	Philips	Philips	Erect	AEC	R+L	No	No	Broad	180	81	2.50
7	1	Static	DDR	Carestream	Carestream	Erect	Manual	None	No	No	Broad	180	65	2.50
8	1	Static	CR	Siemens	Carestream	Erect	Manual	None	No	No	Fine	200	96	2.00

H[A]: Hospital; X[B]: X-ray machine; D[C]: Detector.

5.3.6 Adult Group

5.3.6.1 CDRAD 2.0 Phantom

5.3.6.1.1 Image Quality, Radiation Dose and FOM Variations

Figure (5-14) illustrates the IQF_{inv} values against the corresponding IAK values, both between and within hospitals. The IQF_{inv} scores between the hospitals ranged from 0.83 (H8X1) to 2.18 (H2X1) (mean 1.42) with a difference of 89.7%. Within the hospitals the range was lower: 1.52 (H2X2) to 2.18 (H8X1) (mean 1.87) with only a 35.7% difference. The third quartile between the hospitals was 1.61. IAK values between the hospitals ranged from 17.26 (H3X1) to 239.15 µGy (H4X2) (mean 93.56 µGy) with a difference of 173.1%. The IAK range for X-ray machines within the hospitals was smaller- 122.58 (H4X1) to 239.15 µGy (H4X2) (mean 180.85 µGy) with a difference of 64.5%. The third quartile between the hospitals was 132.32µGy. The FOM values for the same acquisitions between hospitals ranged from 0.01 (H4X2) to 0.14 (H3X1) (mean 0.05) with a difference of 173.3%, while the FOM within the hospitals ranged from 0.02 (H2X4)-0.05 (H2X1) (mean 0.03) with a small (85.7%) difference. The third quartile between the hospitals was 0.11, as illustrated in **Figure (5-15)**.

Shapiro-Wilk tests showed that IQF_{inv}, IAK, SNR and CNR were normally distributed ($p>0.05$; **Appendix D-1**), while CI values had a non-normal distribution ($p<0.05$; **Appendix D-1**). Analysis using the Pearson correlation coefficient resulted in a moderate positive correlation ($r=0.45$; $p=0.02$) between IQF_{inv} and the IAK. The physical image quality measures (CNR, SNR and CI) also showed wide variation, both between and within hospitals. The CNR values between the hospitals ranged from 2.26 (H8X1) to 6.92 (H2X3) (mean 4.16) with a difference of 101.5%. The within hospital range was 3.84 (H2X4) to 6.92 (H2X3) (mean 5.26) with a smaller difference of 57.ere. The 3rd quartile CNR value between the hospitals was 4.63. CI ranged 22.12 (H8X1) to 197.88 (H2X3) (median 53.26) with a difference of 159.8% between hospitals. Within the hospitals, the range was smaller: 55.61(H2X2) to 197.88 (H2X3) (median 111.43) with a difference of 112.2%. The 3rd quartile CI value between the hospitals was 69.01. SNR ranged from 15.39 (H3X3) to 58.88 (H3X5) (mean 30.45) with a difference of 117.1%, which was the same both between and within the hospitals. The third quartile SNR between the hospitals was 38.98.

Good correlations were observed between IQF_{inv} and both CNR ($r = 0.87$; $P=0.001$) and CI ($r=0.72$; $P=0.001$), while a weak non-significant correlation was observed between IQF_{inv} and SNR ($r =0.05$; $P=0.82$).

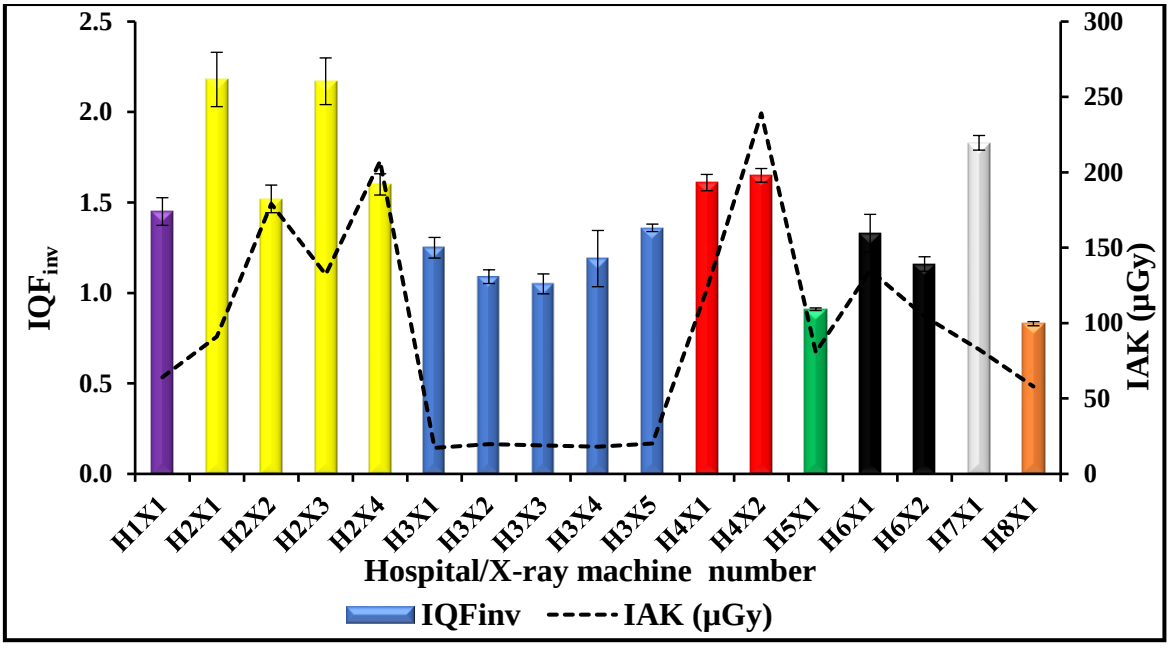

Figure 5-14 Comparison of IQFinv scores against the IAK values for the adult age group. The error bars in this chart represent the SD in IQFinv obtained from measuring 3 repeated images. The dashed line represents the average value of IAK resulted from measuring 3 repeated radiation exposures.

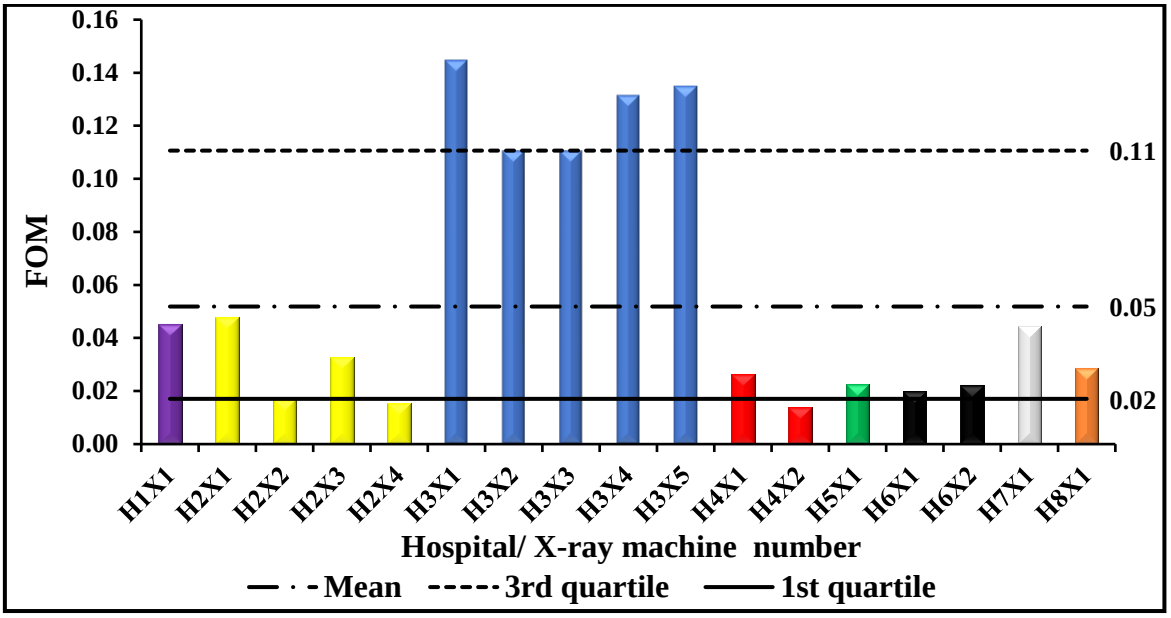

Figure 5-15 Bar chart displaying the range of FOM scores calculated for the adult age group. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) of the FOM values, respectively.

5.3.6.2 Lungman Phantom (Standard Size):

5.3.6.2.1 Image Quality, Radiation Dose and FOM Variations

Figure (5-16) compares the scores for the visual IQ against the corresponding IAK values, between and within hospitals, acquired using the standard size Lungman phantom (without the fat jacket). **Figure (5-17)** presents the variation in FOM values obtained using the same data, between and within hospitals. **Figure (5-18)** compares the scores for the LV evaluations against the corresponding IAK data, between and within hospitals. **Figure (5-19)** presents the variation in FOM values obtained using the same data, for between and within hospitals.

The visual IQ scores between the hospitals ranged from 12.00 (H8X1) to 21.33 (H2X3) (mean 16.26) with a difference of 56.0%. For the within hospital comparisons, the range was lower: 15.00 (H2X1) to 21.33 (H2X3) (mean 18.13) with only a 34.8% difference. The 3rd quartile IQ between the hospitals was 17.00. IAK values between the hospitals ranged from 19.28 (H3X1) to 136.29 µGy (H4X2) (mean 63.38 µGy) with a difference of 150.4%. The IAK range within the hospitals was smaller: 61.24 (H4X1) to 136.29 µGy (H4X2) (mean 98.77 µGy) with a difference of 76.0%. The third quartile IAK value between the hospitals was 89.05 µGy. The FOM values obtained using the same data (**Figure (5-17)**) between the hospitals ranged from 0.26 (H4X2) to 1.76 (H3X1) (mean 1.49) with a difference of 148.5%. The FOM for the within hospital comparison ranged from 1.12 (H3X5) to 1.76 (H3X1) (mean 2.67) with a difference of 44.4%. The third quartile FOM value was 1.12 between the hospitals.

Visual LV scores between the hospitals ranged from 6.00 (H8X1) to 11.67 (H2X3) (mean 8.09), with a difference of 64.2%. For the within hospital comparisons, the range was lower: 7.33 (H2X1) to 11.67 (H2X3) (mean 9.71) with a difference of 45.7%. The third quartile LV score (between the hospitals) was 9.00. FOM values obtained using the same data (**Figure (5-19)**), between hospitals, ranged from 0.45 (H4X2) to 4.20 (H3X1) (mean 1.49) with a difference of 161.3%, while the FOM variations within the hospitals ranged from 2.16 (H3X5) to 4.20 (H3X1) (mean 2.67) with a smaller (64.2%) difference. Between the hospitals the 3rd quartile FOM was 2.0.

Physical IQ parameters (CNR, SNR and CI) also showed wide variation, between and within hospitals. The CNR values between hospitals ranged from 9.60 (H3X3) to 45.51 (H2X1) (mean 22.67) showing a 130.3% difference. Within hospitals the range was 15.04 (H2X2)

to 45.51 (H2X1) (mean 27.13) with a smaller (100.6%) difference. The 3rd quartile CNR value between the hospitals was 27.09. SNR varied from 13.39 (H3X3) to 67.93 (H2X1) (mean 39.32) with a difference of 134.1%; and within the hospitals it ranged from 34.21 (H2X2) to 90.15 (H2X4) (mean 44.05). This translated into a 90.0% difference. The 3rd quartile SNR value between the hospitals was 45.57. CI ranged from 15.85 (H8X1) to 63.35 (H2X3) (mean 44.14) with a difference of 119.9%. Within the hospitals this ranged from 36.83 (H2X2) to 63.35 (H2X3) (mean 47.66) with a difference of 52.9%. The 3rd quartile CI value between the hospitals was 52.37. Shapiro-Wilk tests showed that IAK, IQ, LV, SNR, CNR and CI were normally distributed ($p>0.05$; **Appendix D-2**). Analysis using the Pearson correlation coefficient resulted in a moderate positive, non-significant, correlation $r=0.34$ ($p=0.17$) between IAK and IQ. Correlation between LV and IAK also produced a moderate positive, non-significant, correlation ($r=0.36$; $p=0.15$). Correlations between the IQ and SNR, CNR and CI were $r=0.08$ ($P=0.75$), $r=-0.02$ ($P=0.91$) and $r=0.35$ ($P=0.17$), respectively.

Good inter-observer agreement between the observers was observed in the IQ (ICC=0.97; 95% confidence interval: 0. 61-0. 91) and LV (ICC=0.83; 95% confidence interval: 0. 68-0. 93).

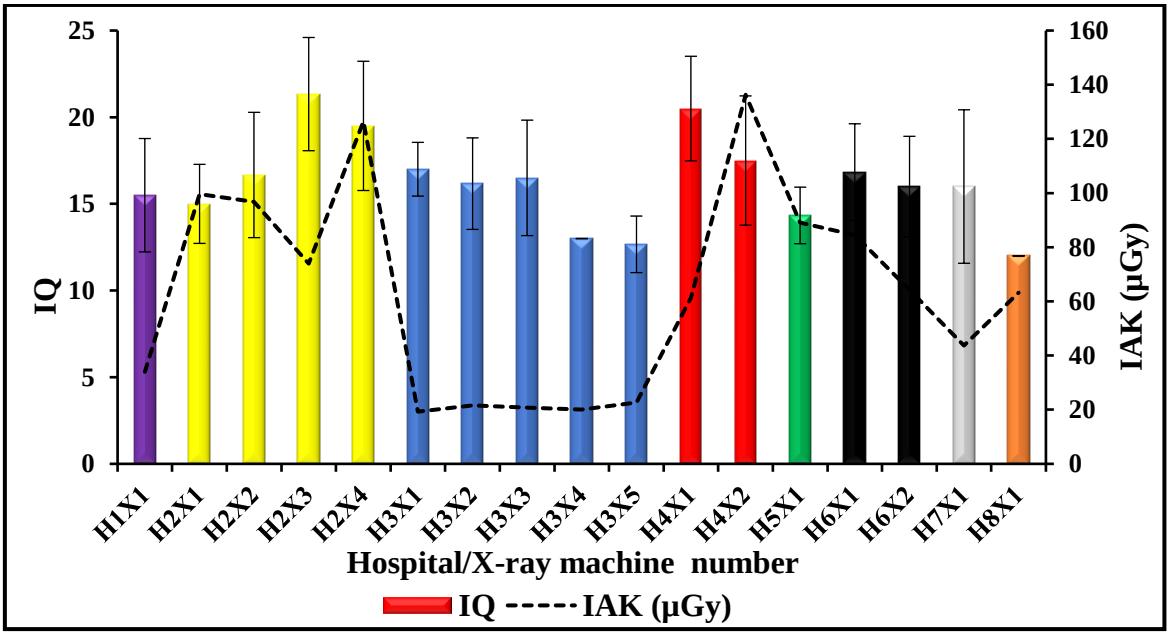

Figure 5-16 A comparison of the measured visual IQ values against the IAK values for the standard size adult group. The error bars in this chart represent the SD in IQ obtained from the visual evaluations by observers. The dashed line represents the average value of IAK resulted from measuring 3 repeated radiation exposures.

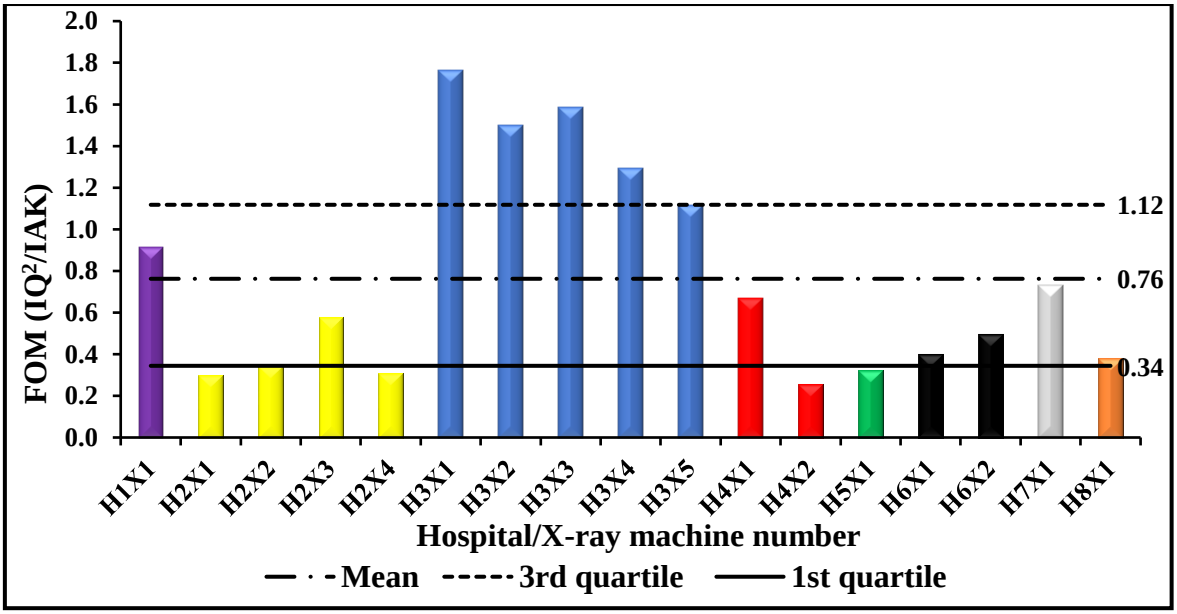

Figure 5-17 Bar chart displaying the range of FOM values, based on the visual IQ evaluation. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) of the FOM values, respectively.

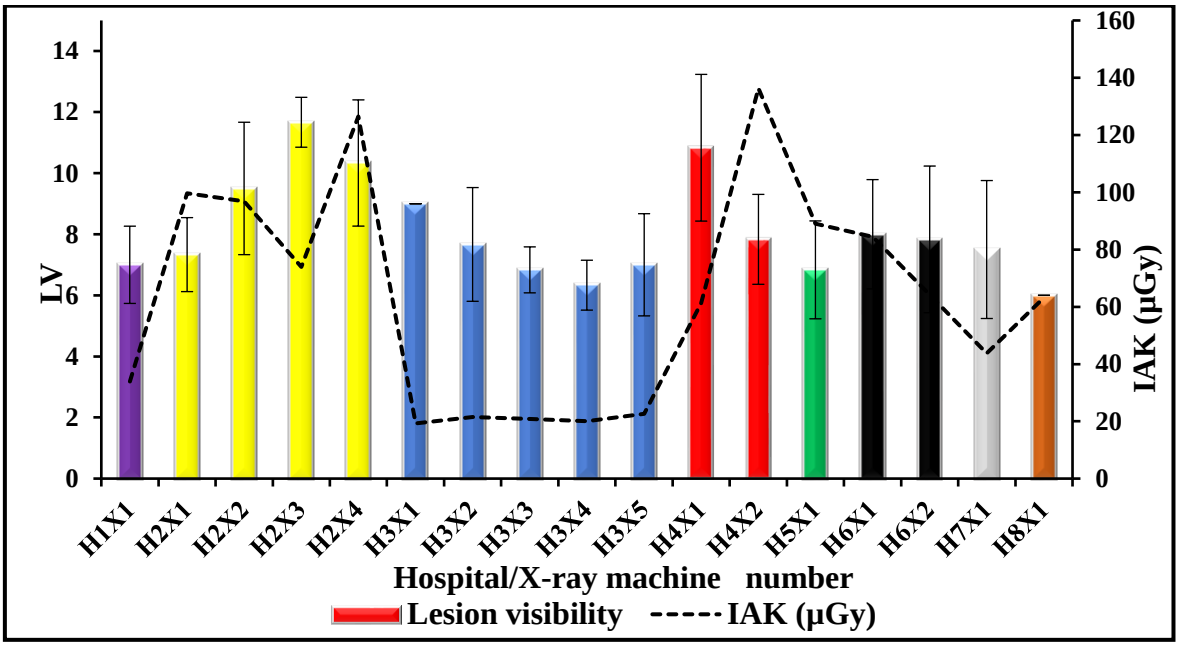

Figure 5-18 A comparison of visual LV values against their respective IAK values for the standard size adult group. The error bars in this chart represent the SD in LV obtained from the visual evaluations by observers. The dashed line represents the average value of IAK resulted from measuring 3 repeated radiation exposures.

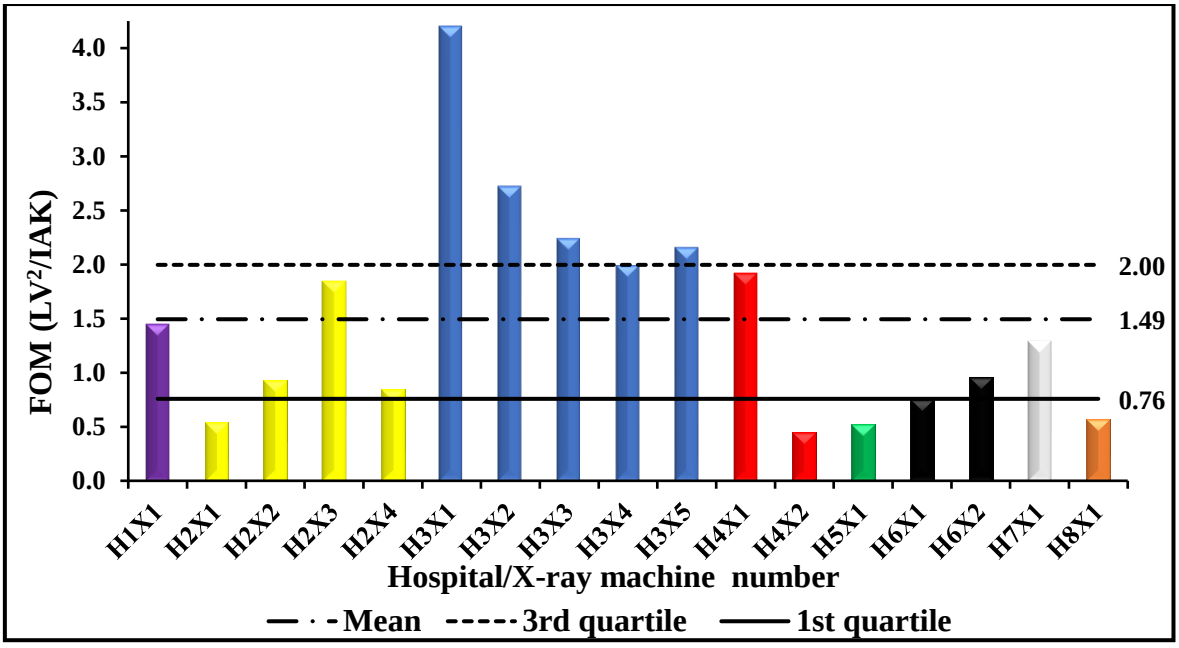

Figure 5-19 Bar chart displaying the range of FOM values, based on visual LV evaluations. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) of the FOM values, respectively.

5.3.6.3 Larger size Lungman Phantom:

5.3.6.3.1 Image Quality, Radiation Dose and FOM Variations

Figures (5-20) compares the scores for the visual evaluations of IQ against the corresponding IAK data, between and within hospitals, acquired using the larger sized Lungman phantom (with a fat jacket). **Figure (5-21)** presents the variation in FOM values obtained using the same data, between and within hospitals. **Figure (5-22)** compares the scores for the visual evaluations of LV against the corresponding IAK data, between and within hospitals. **Figure (5-23)** presents the variation in FOM values obtained using the same data, for between and within hospitals.

The IQ scores between the hospitals ranged from 14.17 (H8X1) to 23.67 (H2X3) (mean 20.90) with a difference of 50.2%. For within the hospitals comparisons, the range was lower: from 16.83 (H3X2) to 21.50 (H3X1) (mean 18.63) with a smaller (24.4%) difference. The 3rd quartile IQ score between the hospitals was 23.00. IAK values between the hospitals ranged from 27.43 (H3X1) to 384.73 µGy (H4X2) (mean 159.20 µGy) with a difference of 173.4%. The range of IAK values within the hospitals was smaller: 195.76 (H4X1) to 384.73 µGy (H4X2) (mean 290.25 µGy) with a difference of 65.1%. The 3rd quartile IAK value between the hospitals was 213.21 µGy. The FOM values obtained using the same data (**Figure (5-17)**), between the hospitals, ranged from 0.12 (H4X2) to 1.57 (H3X1) (mean

1.42) with a difference of 171.6%. The FOM values within the hospitals ranged from 1.05 (H3X2) to 1.57 (H3X1) (mean 3.23) with a lower (39.7%) difference. The third quartile between the hospitals was 0.71. The LV scores between the hospitals ranged from 5.83 (H8X1) to 13.67 (H4X1) (mean 10.81), with a difference of 80.4%; while, within the hospitals the range was lower: 11.17 (H4X2) to 13.67 (H4X1) with a mean of 12.42 and a total difference of 20.1%. The 3rd quartile LV score between the hospitals was 11.67. The FOM values obtained using the same data (**Figure (5-19)**), between the hospitals, ranged from 0.28 (H8X1) to 4.41 (H3X1) (mean 2.52) with a difference of 176.1%. The FOM within the hospitals ranged from 2.52 (H3X2) to 4.41 (H3X1) (mean 3.23) with a difference of 54.5%. The third quartile between the hospitals was 2.52.

Physical IQ measures (CNR, SNR and CI) also showed wide variation, both between and within hospitals. The CNR values between the hospitals ranged from 7.48 (H3X3) to 30.31 (H2X1) (mean 20.23) with a difference of 120.8%. Within the hospitals the range was lower: from 12.27 (H2X2) to 30.31 (H2X1) (mean 21.22), which was a difference of 84.7%. Between the hospitals the 3rd quartile CNR was 26.88. SNR ranged from 11.33 (H3X3) to 66.25 (H5X1) (mean 35.42) with a difference of 141.6%. Within the hospitals the SNR ranged from 11.33 (H3X3) to 38.55 (H3X4) (mean 23.71) with a smaller (109.1%) difference (3rd quartile between the hospitals was 41.83). CI ranged from 5.59 (H8X1) to 67.64 (H2X3) (mean 39.93), with a 169.5% difference between hospitals. Within the hospitals the difference (83.3%) ranged from 27.87 (H2X1) to 67.64 (H2X3) (mean 46.38). The 3rd quartile CI between the hospitals was 46.33.

Shapiro-Wilk tests showed that IAK, IQ, LV, SNR, CNR and CI were normally distributed ($p>0.05$; **Appendix D-2**). Correlation analysis (Pearson's correlation) demonstrated large positive correlation between IAK and IQ ($r=0.60$; $p=0.01$). There was a moderate positive correlation between the LV and IAK ($r=0.49$) ($p=0.04$). Correlations (Pearson's correlation) between the IQ and SNR, CNR and CI were $r=0.52$ ($P=0.03$), CNR $r=0.64$ ($P=0.006$) and $r=0.54$ ($P=0.02$), respectively. Good inter-observer agreement was reported for the IQ and LV evaluations ((ICC=0.87; 95% confidence interval: 0. 74-0. 94) and (ICC=0.84; 95% confidence interval: 0. 69-0. 93), respectively). **Table (5-15)** summarises the inter-observer variability and their confidence interval values of the observers during IQ and LV evaluation of the anthropomorphic phantoms (Lungman and Gammex) images.

Table 5-15 Analysis of inter-observer variability of seven observers during IQ and LV evaluations for the Lungman and Gammex phantoms.

Phantom type	Test type	Intraclass Correlation (ICC)	95% confidence interval	
			Lower	Upper
Lungman phantom without the fat jacket	IQ	0.79	0.61	0.91
Lungman phantom with the fat jacket	IQ	0.87	0.74	0.94
Lungman phantom without the fat jacket	LV	0.83	0.68	0.93
Lungman phantom with the fat jacket	LV	0.84	0.69	0.93
Gammex phantom	IQ	0.54	0.17	0.80

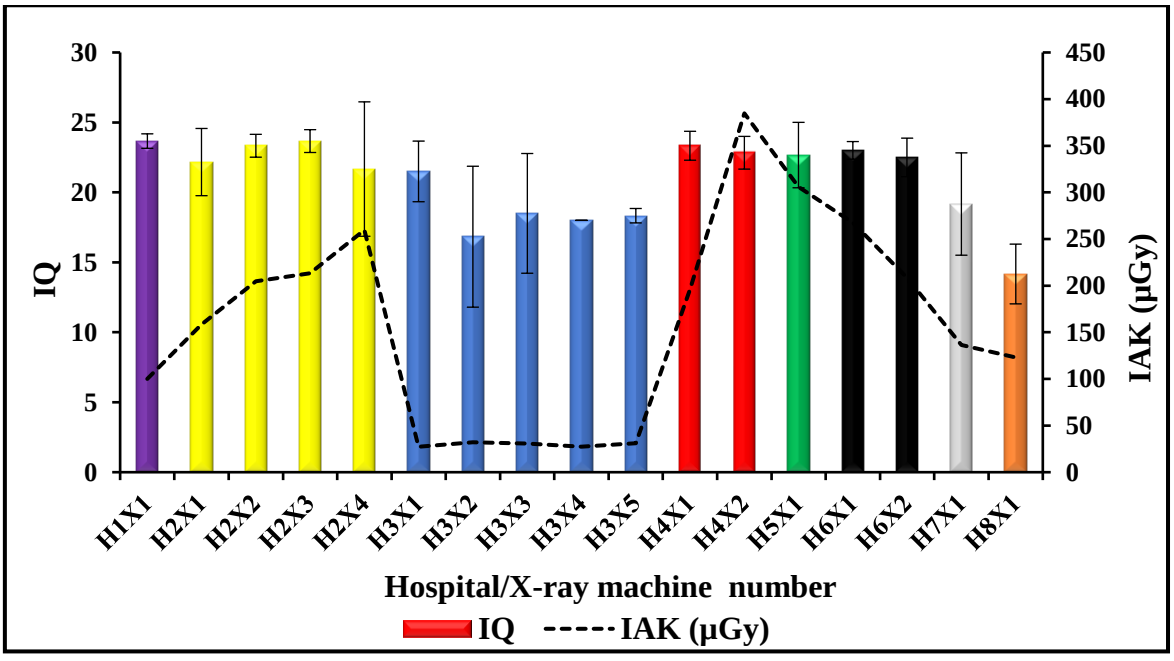

Figure 5-20 A comparison of measured IQ against IAK for the larger size adult protocols. The error bars in this chart represent the SD in IQ obtained from the visual evaluations by observers. The dashed line represents the average value of IAK resulted from measuring 3 repeated radiation exposures.

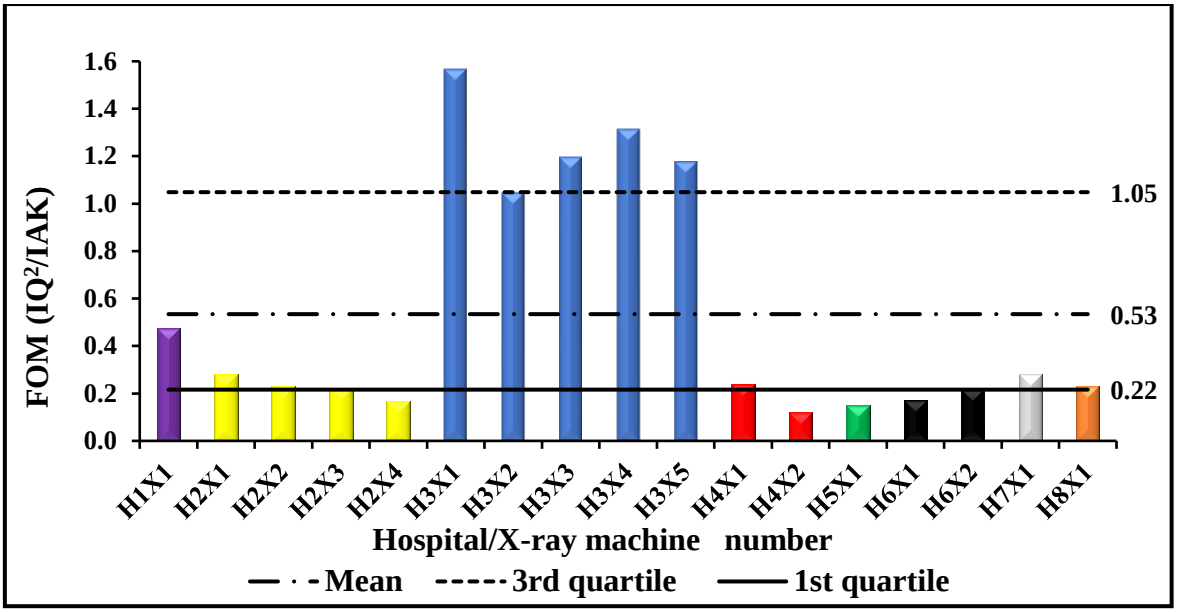

Figure 5-21 Bar chart displaying the range of FOM values, based on IQ evaluations and calculated for the larger size adult group. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) of the FOM values, respectively.

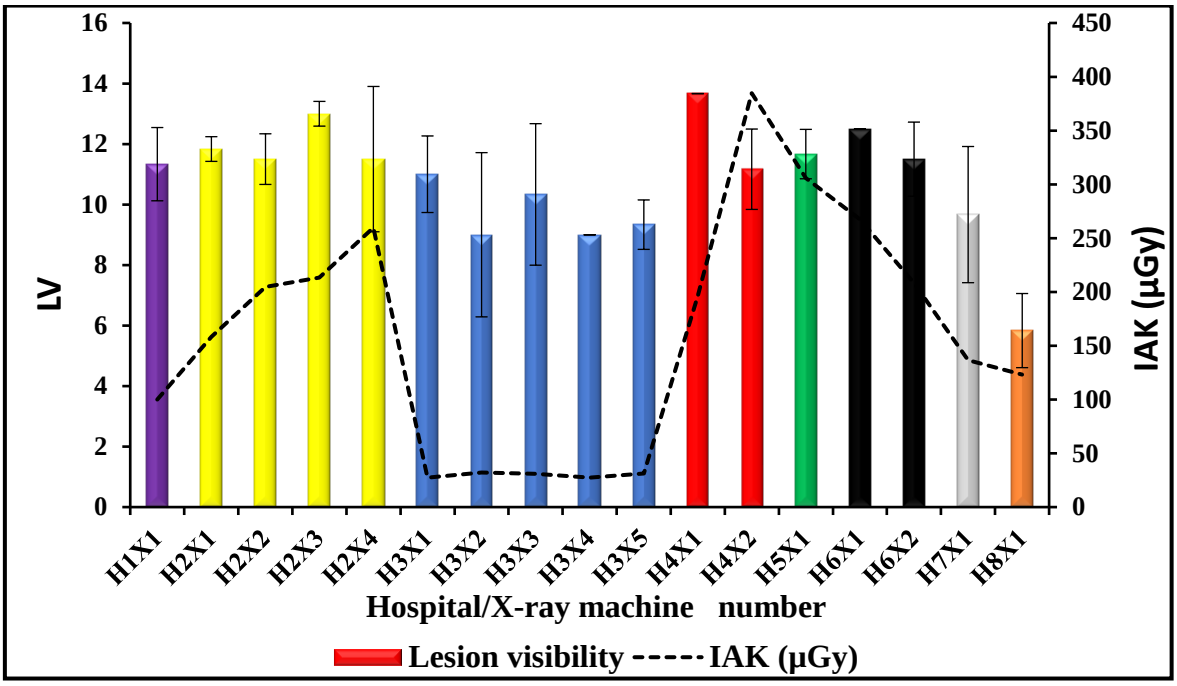

Figure 5-22 A comparison of measured LV against IAK for the larger size adult group. The error bars in this chart represents the SD in LV obtained from the visual evaluations by observers. The dashed line represents the average value of IAK resulted from measuring 3 epeated radiation exposures.

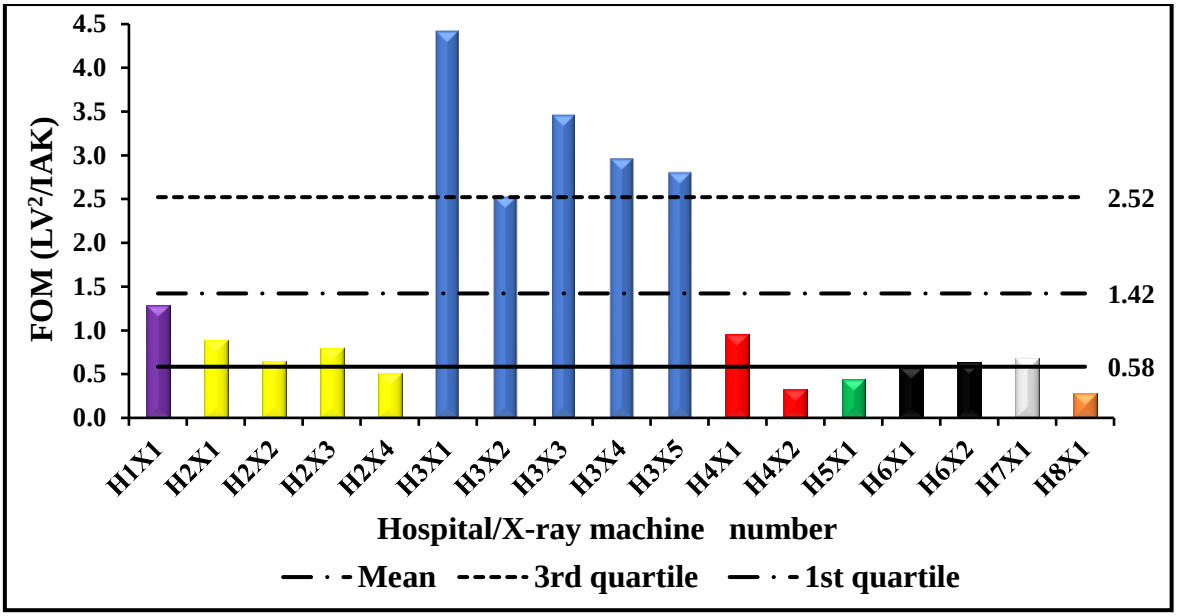

Figure 5-23 Bar chart displaying the range of FOM values, based on LV evaluations and calculated for the larger size adult group. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) of the FOM values, respectively.

5.3.6.4. Clinical Protocols Variations:

Table (5-16) describes the examination techniques and acquisition parameters used for adults using s CDRAD 2.0 phantom, while **Table (5-17)** describes those of the normal size and larger size adult protocols using a Lungman phantom, between and within the hospitals. The only agreement between hospitals was the use of 180 cm SID (except for one X-ray machine (H8X1) which used 200 cm). However, there was no consistency between and within the hospitals on the other acquisition parameters used. The type of exposure control varied, however H1X1, H2X2, H2X3, H2X4, H4X1, H4X2, H6X1, H6X2 and H7X1 (41%) all used AEC. All of the X-ray machines which made use of the AEC also a used an anti-scatter radiation grid. X-ray machines which did not use an anti-scatter radiation grid primarily used a manual exposure mode. The kVp and tube mAs ranged from 85 to 125 kVp and 1.6 to 4.4 mAs using CDRAD 2.0 phantom, and 0.80 to 2.80 mAs for a standard size Lungman phantom. Finally, there was no consistency between the hospitals on using the additional filtration; all the X-ray machines used a broad focal spot type except 3 X-ray machines (H4X1, H4X2 and H8X1) which used a fine focal spot. Regarding the larger size adult protocols, most of the hospitals used the same protocols except for 4 of the X-ray machines (H2X1, H2X4, H5X1, H8X1) that used different kVp values which were higher than that of the standard size phantom, and 3 X-ray machines (H1X1, H3 and H8X1) that

used higher values of mAs compared to that of the standard size phantom (**Tables (5-17)**). Finally, X-ray machine H5X1 used AEC for the larger size phantom as opposed to manual exposure control for the standard size phantom.

Table 5-16 A summary of X-ray machine types, examination techniques and acquisition parameters used for adult CDRAD 2.0 phantom.

H[A] No.	X[B] No.	X type	D[C] type	X Manufacturer	D Manufacturer	Type of exposure control	AEC	Grid	Additional filtration	Focal spot type	SID (cm)	kVp	mAs
1	1	Static	IDR	Philips	Philips	AEC	R+L	yes	1mm Al +0.1 mm Cu	Broad	180	125	2.04
2	1	Mobile	IDR	Carestream	Carestream	Manual	None	Yes	No	Broad	180	110	2.80
	2	Static	IDR	Siemens	Siemens	AEC	R+L	Yes	0.1 mm Cu	Broad	180	125	2.10
	3	Static	DDR	Carestream	Carestream	AEC	R+L	Yes	No	Broad	180	125	2.88
	4	Static	IDR	Samsung	Samsung	AEC	R+L	Yes	No	Broad	180	125	4.40
3	1	Static	IDR	Siemens	Siemens	Manual	None	No	0.2 mm Cu	Broad	180	96	1.60
	2	Static	IDR	Siemens	Siemens	Manual	None	No	0.2 mm Cu	Broad	180	96	1.60
	3	Static	IDR	Siemens	Siemens	Manual	None	No	0.2 mm Cu	Broad	180	96	1.60
	4	Static	IDR	Siemens	Siemens	Manual	None	No	0.2 mm Cu	Broad	180	96	1.60
	5	Static	IDR	Siemens	Siemens	Manual	None	No	0.2 mm Cu	Broad	180	96	1.60
4	1	Static	DDR	Philips	Philips	AEC	R+L	Yes	1mm Al +0.1 mm Cu	Fine	180	125	3.13
	2	Static	CR	Philips	Carestream	AEC	R+L	Yes	No	Fine	180	125	5.50
5	1	Static	CR	Siemens	Carestream	Manual	None	No	No	Broad	180	85	2.80
6	1	Static	DDR	Philips	Philips	AEC	R+L	Yes	No	Broad	180	125	2.50
	2	Static	DDR	Philips	Philips	AEC	R+L	Yes	No	Broad	180	125	2.00
7	1	Static	DDR	Carestream	Carestream	AEC	R+L	Yes	No	Broad	180	125	1.80
8	1	Static	CR	Siemens	Carestream	Manual	None	No	No	Fine	200	113	2.00

H[A]: Hospital; X[B]: X-ray machine; D[C]: Detector.

Table 5-17 A summary of X-ray machine types, examination techniques and acquisition parameters used for standard and larger size Lungman phantoms.

H[A]	X[B]	X type	X manufacturer	D[C] type	D manufacturer	Exposure control (S[D]/L[E])	Grid (S/L)	Additional filtration (S=L)	kVp p (S/L)	kVp increase (%)	mAs (S/L)	mAs Increase (%)
1	1	Static	Philips	IDR	Philips	AEC/AEC	Yes/yes	0.1mm Cu + 1mm Al	125/125	0.0	0.97/2.64	172.16
2	1	Mobile	Carestream	IDR	Carestream	Manual/Manual	Yes/yes	No	110/120	9.1	2.80/4.00	42.86
	2	Static	Siemens	IDR	Siemens	AEC/AEC	Yes/yes	0.1 mm Cu	125/125	0.0	1.20/2.40	100.00
	3	Static	Carestream	DDR	Carestream	AEC/AEC	Yes/yes	No	125/125	0.00	1.44/3.84	166.67
	4	Static	Samsung	IDR	Samsung	AEC/AEC	Yes/yes	No	125/130	4.00	2.40/5.20	116.67
3	1	Static	Siemens	IDR	Siemens	Manual/Manual	No/No	0.2 mmCu	96/96	0.00	1.60/2.00	25.00
	2	Static	Siemens	IDR	Siemens	Manual/Manual	No/No	0.2 mm Cu	96/96	0.00	1.60/2.00	25.00
	3	Static	Siemens	IDR	Siemens	Manual/Manual	No/No	0.2 mm Cu	96/96	0.00	1.60/2.20	37.50
	4	Static	Siemens	IDR	Siemens	Manual/Manual	No/No	0.2 mm Cu	96/96	0.00	1.60/2.00	25.00
	5	Static	Siemens	IDR	Siemens	Manual/Manual	No/No	0.2 mm Cu	96/96	0.00	1.60/2.00	25.00
4	1	Static	Philips	DDR	Philips	AEC/AEC	Yes/Yes	1 mm Al+ 0.1 mm Cu	125/125	0.00	1.40/4.13	195.00
	2	Static	Philips	CR	Carestream	AEC/AEC	Yes/yes	No	125/125	0.00	2.80/7.32	161.43
5	1	Static	Siemens	CR	Carestream	Manual/AEC	No/Yes	No	85/96	12.94	2.80/6.08	117.14
6	1	Static	Philips	DDR	Philips	AEC/AEC	Yes/yes	No	125/125	0.00	1.40/4.10	192.86
	2	Static	Philips	DDR	Philips	AEC/AEC	Yes/yes	No	125/125	0.00	1.10/3.30	200.00
7	1	Static	Carestream	DDR	Carestream	AEC/AEC	Yes/yes	No	125/125	0.00	0.80/2.50	212.50
8	1	Static	Siemens	CR	Carestream	Manual/Manual	No/No	No	113/121	7.08	2.00/3.20	60.00

H[A]: Hospital; X[B]: X-ray machine; C detector; S[D]: standard size phantom; L[E]: larger size phantom

5.4 Influence of Phantom Thickness on Image Quality and Radiation Dose

As mentioned in the methods chapter (section 4.2.6.1), 7 qualified diagnostic radiographers were involved in the visual evaluation tasks designed for investigating the influence of phantom thickness on visual IQ. The observers rated on a 2-point scale, wherein the ratings signified the following: 1, for the image which has superior IQ; and 0, for the image which has the lower IQ. The bar chart below (**Figure (5-24)**) presents the scores of the IQ, as given by the observers, for images generated from a Lungman phantom, with and without the fat jacket. From this figure, it can be seen that the additional phantom thickness has a negative impact on IQ for most of the X-ray machines, except for H2X2, H6X1 and H6X2. These three X-ray machines were found to have superior IQ for the larger sized adult images when compared with that of the standard sized**. Figures (5-25)** to **(5-27)** highlight the influence of phantom thickness on the different physical IQ parameters (SNR, CNR and CI) which were generated from a Lungman phantom. Similar to the visual IQ evaluation, it was found that the additional phantom thickness led to a degradation of physical IQ. Regarding SNR, 4 X-ray machines (H3X1, H5X1, H6X1, H6X2) were found to have higher SNR for larger sized phantom; 3 X-ray machines (H1X1, H2X2, H2X4) had approximately equal SNR values; and the remainder (H2X1, H2X3, H3X2, H3X3, H3X4, H3X5, H4X1, H4X2, H7X1, H8X1) were found to have higher SNR for the 'standard size' when compared with the 'larger size' acquisitions. In terms of CNR, only 4 X-ray machines (H4X1, H5X1, H6X1, H6H2) were found to have higher CNR for the larger sized phantom, while the additional thickness caused a reduction in the CNR values for the other 13 X-ray machines. With respect to CI, only 4 X-ray machines (H2X2, H2X3, H5X1, H6X1) were found to have higher CI for the larger sized phantom, while the additional thickness caused a reduction of the CI values in the other thirteen X-ray machines.

Figure (5-28) illustrates the influence of phantom thickness on radiation dose, between and within X-ray machines. The radiation dose for the larger sized phantom was higher than that of the standard size for all of the seventeen X-ray machines studied. It is clear from **Figure (5-28)** that the lowest difference in radiation dose values between the larger sized and standard size phantom were observed in Hospital 3 and in all of its X-ray machines (H3X1, H3X2, H3X3, H3X4, H3X5).

The Mann-Whitney test, for non-parametric data (Shapiro-Wilk test: $p=0.001$), was utilised to test the hypotheses that phantom thickness has no statistically significant difference on

visual IQ. The unpaired t-test, for parametric data (Shapiro-Wilk test: p>0.05; **Appendix D-2**), was used to test the hypothesis that phantom thickness has no statistically significant difference on SNR, CNR, CI and IAK.

The Mann-Whitney test results showed that there was a significant difference (P=0.001) in visual IQ evaluations between the larger sized and standard size phantom images. The unpaired t-test results demonstrated that there was no significant difference in SNR, CNR and CI (P=0.40, P=0.37 and P=0.37 between larger size and standard size phantom images, respectively). Finally, the results from the unpaired t-test demonstrated that there was a significant difference in IAK between the larger and standard sized phantom acquisitions (P=0.002).

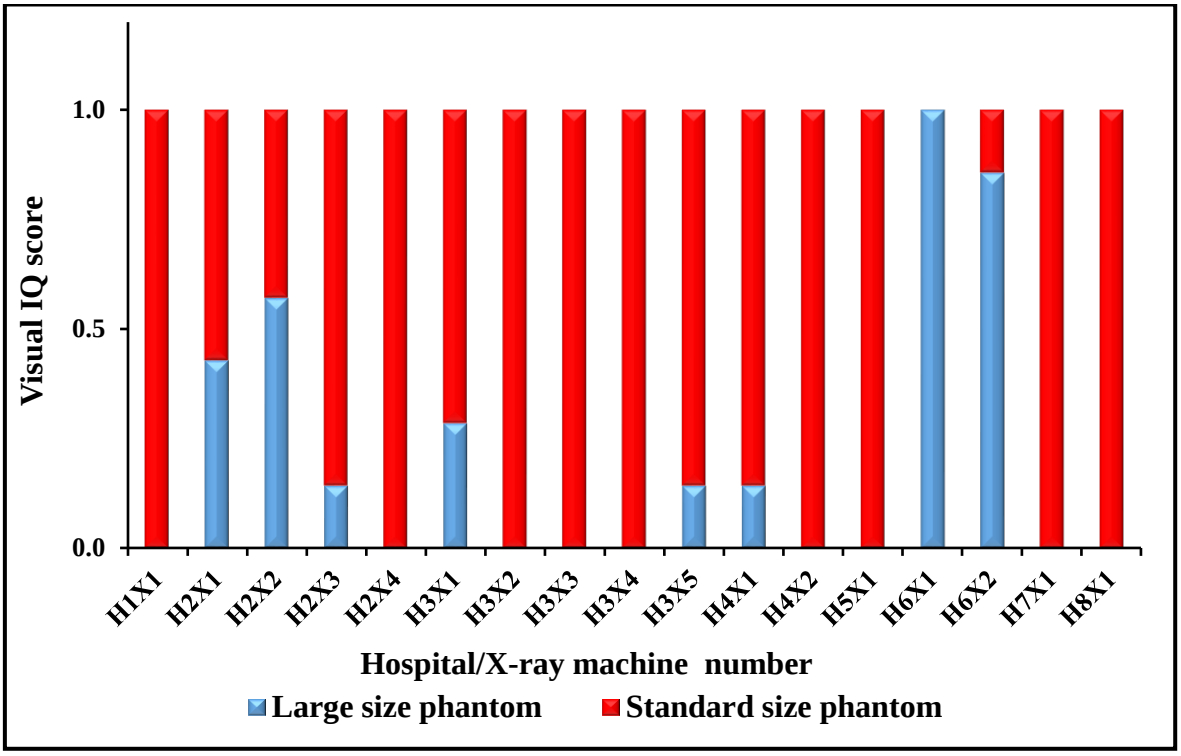

Figure 5-24 Bar chart displaying the scores of the visual IQ evaluations for the seven qualified diagnostic radiographers on images generated using the larger sized and standard size Lungman phantoms.

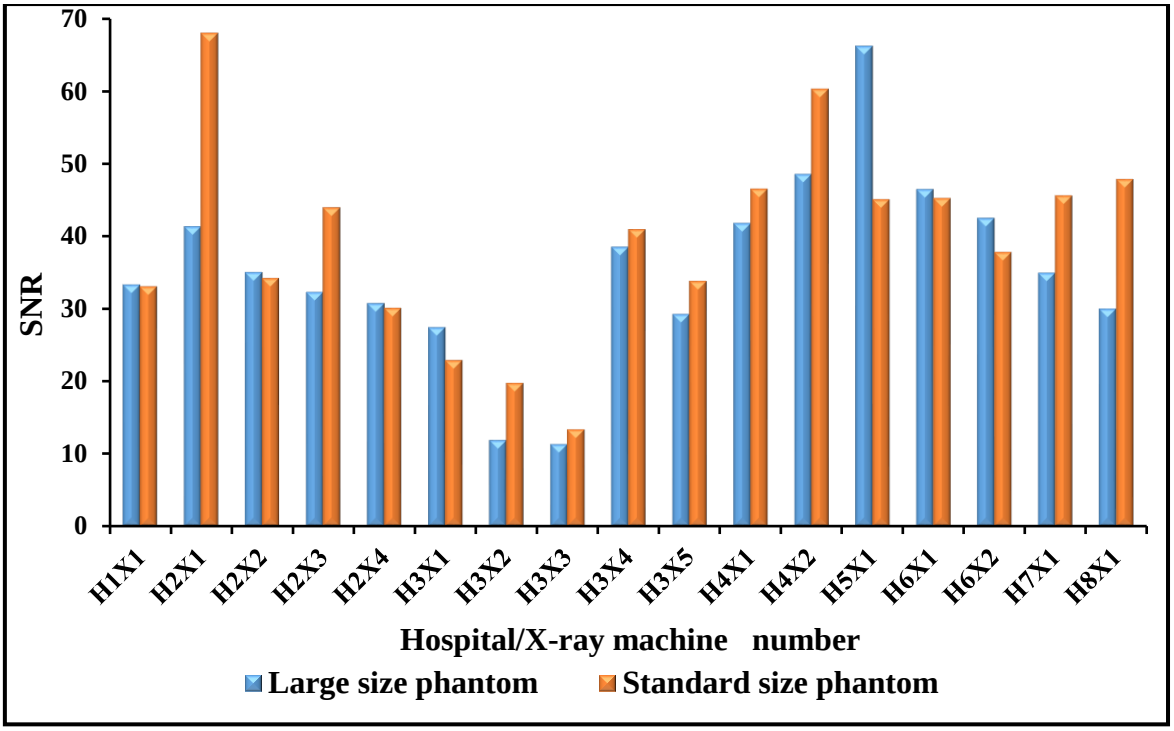

Figure 5-25 Bar chart displaying SNR values from images generated from the larger sized and standard size Lungman phantoms.

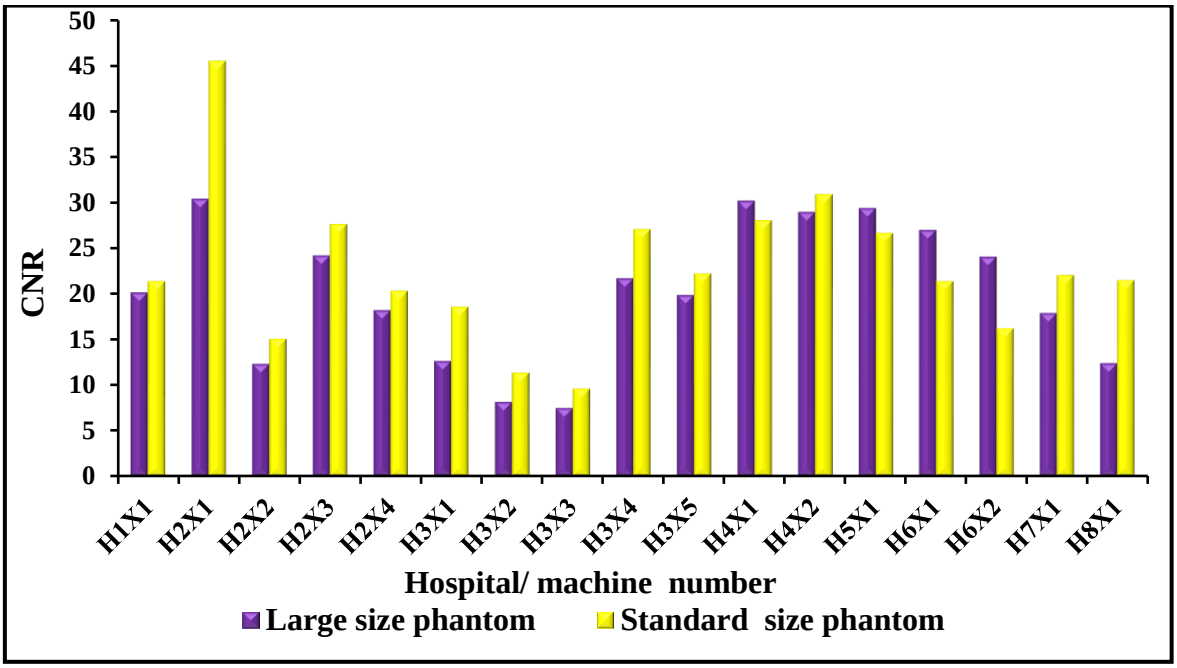

Figure 5-26 Bar chart displaying CNR values from images generated from the larger sized and standard size Lungman phantoms.

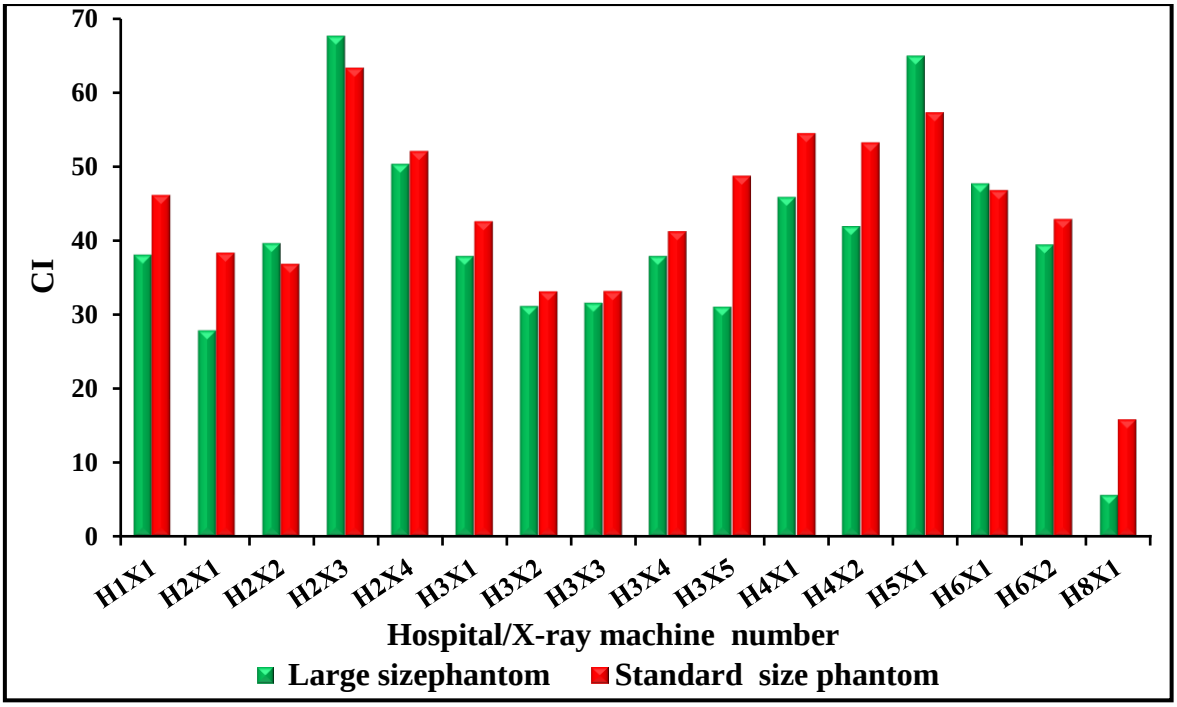

Figure 5-27 Bar chart displaying CI values from images generated from the larger sized and standard size Lungman phantoms.

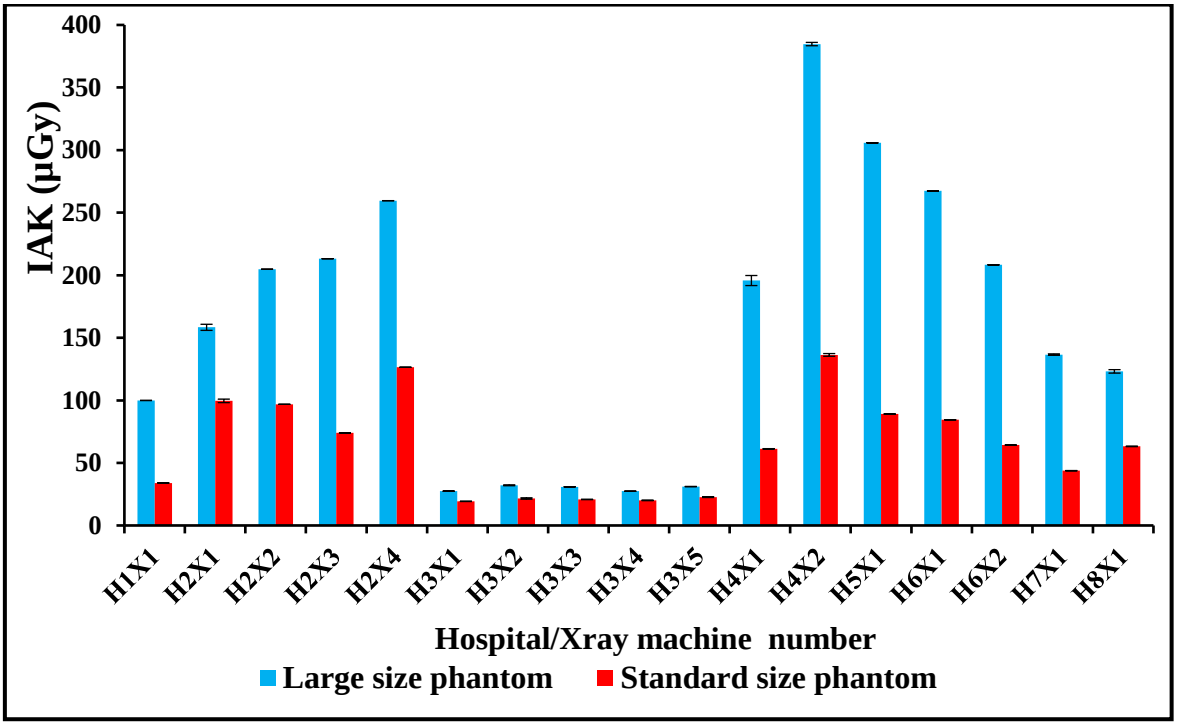

Figure 5-28 Bar chart displaying IAK values from images generated from the larger sized and standard size Lungman phantoms.

5.5. Chapter Summary

The first section of this chapter presents the results of the experimental work designed to investigate the validity of utilising the LCD detectability method in this thesis. It was found that LCD detectability, using CDRAD 2.0 phantom, was a valid method for visual IQ and LV evaluation and also for use in optimisation studies. This latter point was proven from the good positive correlations for both the IQ and LV evaluations. The results of IQ variation, radiation dose and FOM between and within hospitals are presented in the second section. A wide difference is evident in IQ, radiation dose and FOM, between and within hospitals, and correlation is low for IQ and radiation dose in most of the age groups. In addition, there is a considerable variation in the clinical protocols used between and within hospitals for the same age group. The results of the third section of this chapter show that the additional thickness of the phantom has a negative impact on both the physical and visual IQ, and that image quality is degraded in most of the X-ray machines as a result of the additional thickness. Furthermore, the results also demonstrate that the additional thickness increases the radiation dose by approximately 50%, irrespective of the standard phantom size, across all of the X-ray machines.

Chapter 6: Discussion and Conclusions

6.1 Chapter Overview

This chapter discusses the results (Chapter 5) and is divided into 7 sections to aid understanding. First is the validation of utilising LCD detectability, using the CDRAD 2.0 phantom, for visual IQ and LV evaluation and for digital CXR optimisation studies (6.2.). This section will discuss the results of the validity of utilising LCD detectability for digital CXR optimisation studies. Within this section the results of the experimental work will be compared against those from the available literature. In addition, the methodological advantages used for the validation of CDRAD 2.0 phantom will be compared with that from previous studies. The second section (6.3.) focuses on the new method that was used for evaluating IQ and radiation dose, based on the FOM concept. This section also considers the advantages of the new FOM concept as compared with existing methods which are based on DRLs or which evaluate IQ and radiation dose separately. Additionally, the advantages of the new FOM formula are discussed and compared with existing FOM formulas. The third section (6.4.) concentrates on discussing the variability in radiation dose, IQ and FOM for the X-ray machines studied in this thesis. In Chapter 5, the results of the variation in IQ, radiation dose and FOM were presented based on the different age groups (neonate, one year, five years, ten years and adult). This chapter will use the same structure. Within each group, the results will be discussed under two subheadings: 1) the analysis of IQ; and 2) the analysis of radiation dose and FOM. Within this section the possible reasons for variation in IQ and dose will be discussed and linked to imaging equipment and the clinical protocols used. The forth section (6.5) will discuss the influence of patient size, represented by the Lungman chest phantom with and without a fat jacket, on IQ and radiation dose. Within this section, the influence of fat thickness (fat jacket) added to the standard size Lungman phantom on visual and physical IQ, and also radiation dose, will be considered. Also, this section will highlight and discuss the differences between local clinical protocols for standard size and larger sized adult patients. The fifth section (6.7) will consider limitations, future work and recommendations. The seventh section (6.8) will presents the conclusion of this thesis.

6.2 Validation for utilising LCD detectability with CDRAD 2.0 phantom in CXR optimisation studies

The CDRAD 2.0 phantom has been widely used within the literature for optimisation studies and comparing the performance of imaging systems (Alzimami et al., 2009; Bacher et al., 2003; Geijer et al., 2001; Rong et al., 2001; Veldkamp et al., 2006). It must be acknowledged that there are potential limitations when using the CDRAD 2.0 phantom for visual IQ, LV evaluation and CXR optimisation studies because it has a uniform background in which it does not consider the impact of anatomical noise from human anatomy. The chest region is classified as a quantum-saturated region wherein the anatomical structures are the main factors that influence lesion detection and IQ when compared with quantum noise (Samei et al., 2000). Quantum noise is a limiting factor for the LCD detectability measured from CDRAD 2.0 phantom images, whereas the IQ for images acquired on patients would be limited by anatomical noise. It would be extremely useful if the CDRAD 2.0 phantom could predict visual (perceptual) IQ and LV in the clinical setting when undertaking routine quality assurance and comparing IQ between systems, as these parameters are more useful in optimisation studies. IQ and LV studies have a number of applications within medical imaging but are notoriously labour intensive. Automated methods for evaluating IQ are attractive but historically there has been a lack of data on their relationships with perceptual (visual) methods and clinical reality. One of the aims of this thesis is to investigate the potential of using LCD detectability, with the CDRAD 2.0 phantom, for digital CXR optimisation studies by investigating its correlation with IQ and LV. As mentioned in the methods chapter (section 4.3.), 42 images from both the CDRAD 2.0 and Lungman phantoms, with different levels of image qualities, were acquired using a range of acquisition parameters. To investigate the correlation between LCD detectability, IQ and LV with the Lungman phantom, the CDRAD 2.0 phantom images were assessed physically using a CDRAD phantom software analyser. The Lungman phantom images were assessed visually by seven observers (radiographers) using a relative VGA method.

Results from this thesis demonstrate a strong positive correlation ($r=0.91$; $p<0.001$) between the physical evaluation of LCD detectability (IQF_{inv} from the CDRAD 2.0 phantom) and IQ evaluations from the Lungman phantom. Based on IQ alone, the CDRAD 2.0 phantom appears to be a valid and a well justified approach and. It can therefore be used as a surrogate for IQ tasks using an anthropomorphic phantom when evaluating the performance of radiography systems or during optimisation studies. For LV, it is important to know to what

extent there is correlation between the IQF_{inv} values and LV from the Lungman phantom images, since the CDRAD 2.0 phantom was designed to examine the performance of imaging systems in visualising LCD objects.

The findings of this thesis demonstrate a good positive correlation between IQF_{inv} and LV for the two simulated lesions; the first lesion showed r=0.79 ($p<0.001$) and the second lesion showed r=0.68 ($p<0.001$). It was useful to examine the correlation between the IQF_{inv} and the LV when the lesions were located in different positions and with different levels of visibility (high and low). This was necessary to understand how the variations in the location and visualisation of the lesions affect IQF_{inv} correlation values. Although the two simulated lesions have the same HU value (+100 HU), the radiographic visibility of the first simulated lesion (in the left upper lobe) was lower than that of the second one (in the right middle lobe). This is related to the complexity of the anatomical background surrounding the first simulated lesion. A relatively large part of this simulated lesion is covered by the first anterior rib and third posterior rib. The HU values for the surrounding lung, the first anterior rib and the third posterior rib were -890, 353 and 575, respectively. In contrast, a small part of the second simulated lesion was covered by the fourth anterior rib. The HU values for the surrounding lung and the fourth anterior rib were -1027 and 467, respectively (**Figure (4-20))**. In addition, the first simulated lesion had a smaller diameter (10 mm) compared with that of the second simulated lesion (12 mm). It was found that the lesion with lower visibility (left upper lobe) had higher correlation (r=0.79) when compared with that of the higher visibility lesion (right middle lobe, r=0.68). A possible reason behind this could be because of the small level of improvement and degradation of general IQ. This would have a higher impact on increasing or decreasing the visibility of the lower visibility lesion when compared with that of the higher visibility lesion (acquisition parameters had more of an effect on the lower visibility lesion). The good (positive) correlation observed under these 2 conditions provides a useful indication on the utility of the CDRAD 2.0 phantom for investigating the performance of an imaging system in LCD detection.

Following an extensive literature review, only one study attempted to investigate the correlation between the LCD detectability using a CDRAD 2.0 phantom and IQ from cadaveric images using an absolute VGA method (De Crop et al., 2012). However, the correlation between LV and LCD detectability when using the CDRAD 2.0 has not been investigated. Within the work of De Crop et al. (2012), comparable results were reported with a strong positive correlation (r=0.91; p=0.001) between the physical evaluation of IQF_{inv} (CDRAD 2.0) and IQ (cadavers). The work in this thesis builds upon the study by De

Crop et al. (2012) in which a Lungman phantom loaded with simulated pulmonary lesions was used, instead of three normal cadavers with no inherent pathology. The inclusion of a pathology provides the opportunity for a more accurate evaluation of IQ in the detection task, and it is an important factor in optimisation studies. This thesis therefore has some methodological benefits when compared with the work of De Crop et al. Within this thesis a wider range of image qualities were generated and tested compared with that by De Crop et al.. This can be seen from the lowest value of IQF_{inv} in this thesis which was equal to 0.28, compared with 2.56 in study by De Crop et al. The range of image qualities generated in this thesis closely reflected those acquired in clinical practice but also included both high (IQF_{inv}=2.39) and low (IQF_{inv}=0.28) qualities; this was necessary to ensure that the correlations between the IQF_{inv} and IQ/LV were tested across a full range of different levels of IQ. Image qualities in this thesis were deemed similar to those acquired in clinical practice for several reasons. Firstly, a group of experienced radiographers were asked to provide acquisition parameters routinely used within their clinical practice. Secondly, the acquisition parameters were compared with those cited within the literature. Thirdly, the resultant images were reviewed by a group of experts and were deemed to demonstrate a range of image qualities which may be expected within the clinical environment. Relative VGA method was also used in this thesis, without allowing windowing and magnification. This was different to the absolute VGA system, with windowing and magnification, as used in the study by De Crop et al.. Utilising a relative VGA approach is justified because it is likely to be less biased and more sensitive in the detection of subtle variations in IQ. This is due to the presence of a reference image. Although the free adjustment of image viewing parameters, such as brightness, contrast and magnification are considered part of a clinical review of an image, it could, however, influence the correlation between the IQF_{inv} and the IQ. LV and was not permitted within this thesis. A consistent image display is necessary to ensure that any measured variation in IQ results only from the imaged object and not the adjustment of image viewing parameters such as zooming or windowing.

Geijer et al. (Geijer & Persliden, 2005) investigated the influence of different tube potentials (by adjusting the mAs values to keep the effective dose constant) on IQ for lumbar spine radiography, using two different phantoms, CDRAD 2.0 and the adult anthropomorphic Alderson phantom. Findings of the study by Geijer et al. (Geijer & Persliden, 2005) demonstrated that the IQ was higher for high kVp values (≥96 kVp) for both phantoms. Bacher et al. compared the performance of an amorphous silicon and amorphous selenium flat-panel detector using two different approaches: clinical images generated from patients

and using CDRAD 2.0 phantom images (Bacher, Smeets, Vereecken, et al., 2006). Both methods showed that the amorphous silicon flat-panel system requires a lower effective dose to produce images with the same level of quality when compared with that of amorphous selenium flat-panel system. However, in these two studies the correlation between the IQ from the Alderson phantom, images from patients and the IQF_{inv} from the CDRAD 2.0 phantom were not calculated.

Within mammography, the use of the CDMAM phantom is well established in practice (Fausto et al., 2017; Mackenzie et al., 2016; Warren et al., 2012) and there is some similarity in the evaluation of IQ with CDMAM phantom and the methods described in this thesis. Following an extensive literature review, no study was identified which investigated the correlation between the IQF_{inv} and the LV. As a result there is no data available in the literature to compare this thesis' findings with. The good positive correlations reported in this thesis (IQF_{inv} and IQ and LV) might also have a value in indicating the detectability of abnormalities and possibly could extend to lesion detection performance. Several studies have found that there is a correlation between pathology detectability and the visibility of normal anatomical structures (Sund et al., 2000; Tingberg, 2000; Tingberg, Båth, et al., 2000; Tingberg, Herrmann, et al., 2000). To further understand these relationships, additional studies are warranted and will be described in the later section about future work.

Based on the findings in this thesis, there is an excellent correlation between LCD detectability using a CDRAD 2.0 phantom and IQ. A good correlation existed between LCD detectability and LV. It can be confirmed that the CDRAD 2.0 phantom is valid for evaluating IQ and LV and could be of use in digital CXR optimisation studies. Also, this thesis' findings suggest it is possible that the CDRAD 2.0 phantom can be utilised along with the automated evaluation method for LCD detectability within routine quality assurance and optimisation studies. Again, further work to understand the utility of this is warranted.

After investigating the validity of utilising a CDRAD 2.0 phantom for IQ and LV assessment and its validity for use in optimisation studies, the CDRAD 2.0 phantom was used in this thesis. It was used with a very high degree of confidence as a main method to evaluate the variation in IQ for different paediatric age groups and adults, between and within the hospitals.

6.3 New Method for Evaluating Image Quality and Radiation Dose

DRLs are defined in legislation as a tool for monitoring radiation dose for X-ray imaging procedures (International Atomic Energy Agency, 2012; Martin & Vano, 2018). However, DRLs are limited because they only take into account radiation dose and not IQ. Practically, DRLs can be beneficial for identifying variations in radiation dose, within and between hospitals, however they cannot identify or predict variations in IQ. The reason behind this is that reports in the literature have found no straightforward correlation between IQ and radiation dose between hospitals in different countries (Almén et al., 1996; Geleijns et al., 1993; Struelens et al., 2008; Veldkamp et al., 2006). On the other hand, several studies have attempted to assess the variation in IQ and the radiation dose administered to the patient between hospitals (Almén et al., 1996; Geleijns et al., 1993; Veldkamp et al., 2006). These studies presented variations in IQ and radiation dose separately and did not combine them as a single metric. This makes comparisons between X-ray units difficult and restricts the determination of optimal parameters. Within radiography, there is always a trade-off between the radiation dose and the IQ by which no single factor should be analysed on its own.

The major purpose of this thesis was to develop a method for evaluating and comparing IQ and radiation dose, between and within hospitals, and then utilise it to determine the likely variation in IQ and radiation dose for paediatric and adult CXRs. From the reviewed literature (chapter 3, section 3.4.), it was observed that there is no standardised method that can be used for evaluating IQ and radiation dose, between and within hospitals. In this thesis a new method based on the FOM concept was used. It is presented as viable for use in the optimisation and evaluation of IQ and radiation dose in clinical practice, between and within hospitals. An FOM concept was used for establishing a standardised approach for comparing IQ and dose between and within hospitals. The FOM concept has been widely reported in the literature in optimisation studies and the selection of optimal protocols (Doyle et al., 2006; Lee, Wang, Liu, & Jiang, 2007; Vano, Geiger, Schreiner, Back, & Beissel, 2005). However, following an extensive literature review, no studies were found that used the FOM concept for evaluating and comparing IQ and radiation dose between X-ray systems, nor between hospitals or even in the same hospital. Utilising an FOM offers an attractive way of jointly comparing IQ and radiation dose. FOM values are commonly offered as a single number that reflects the efficiency of an imaging system under set conditions. It can offer a useful, fast and descriptive index that summarise the overall performance of the imaging

systems concerning dose and IQ between and within hospitals. In addition, it is a useful way to determine the optimal protocols among hospitals because it considers both IQ and radiation dose.

The new FOM formula used in this thesis assesses IQ using the CDRAD 2.0 phantom and the IQF_{inv} metric. This approach has numerous advantages. IQF_{inv} overcomes the limitations of using other physical IQ measures such as SNR and CNR by considering the object size. IQF_{inv} has been validated within this thesis (see section 6.2.) and was found to have a correlation with the visual measures of IQ and LV. Based on this, the FOM utilised in this thesis was considered to be an effective, valid and reliable approach for comparing and optimising imaging protocols, both between and within hospitals. Another important feature for the new FOM is that it provides a standardised and reliable method for future comparisons of FOM values. Justification for this comes from the IQF_{inv} values generated from the CDRAD 2.0 phantom, which are exactly the same as when the automated CDRAD phantom analyser software is used (greater reliability). In contrast, the other existing FOM formulas that utilise CNR or SNR as a metric for IQ evaluation could be difficult to compare between sites and machines. This is because there are many different techniques for measuring these physical measures, including differences in the selection of the positions and sizes of the ROIs for the measured objects and the background (noise). In addition, there are different mathematical formulas for measuring the same physical measure of IQ, such as CNR. Based on this, the final calculated value of FOM can be influenced by the different techniques used for measuring the SNR or CNR, and it might then be difficult to compare the FOM values between X-ray units. The new suggested FOM is based on using IQF_{inv} and can be compared between the different studies when using the same PMMA thicknesses. Recently, a study by Fausto et al. (2017) proposed using the IQF_{inv} values from CDMAM phantom images as a metric for IQ. Within their work, the calculation of FOM was FOM= $(IQF_{inv})^2$/AGD, where AGD is the average glandular dose. This was used for the optimisation of exposure parameters of digital mammography for three simulated breast thicknesses. The results demonstrated that the use of the suggested FOM based on the IQF_{inv} permits the optimisation of exposure parameters for digital mammography whilst considering the breast thicknesses. A similar study was undertaken by Rojas et al. (2017) who also used the same proposed FOM reported by Fausto et al. (2017), again using the CDMAM phantom for optimising the exposure factors in digital mammography for 2 simulated breast thicknesses. It was concluded that their proposed FOM is an efficient tool

for the optimisation of the digital mammography units. These 2 studies in mammography provide further indication to the value of the proposed FOM formula reported within this thesis.

Applying the method for evaluating and comparing IQ and dose between and within hospitals proposed in this thesis, based on the FOM concept, will be extremely beneficial for optimising the protocols used for the same radiography examinations. For instance, it would be beneficial if hospitals within specific countries had access to the CDRAD 2.0 phantom, had the same thicknesses of medical PMMA slabs, had a solid-state dosimeter, used the automated method for LCD detectability evaluation using phantom analyser software, and used the new proposed FOM formula (that uses IQF_{inv} as a metrics for IQ evaluation). Then the values of FOM resulting from the X-ray units in each hospital, for each age group and for different anatomical areas under radiographic examination, can be sent to the NHS institute. This will provide an up-to-date database on the level of the variation in IQ and dose between hospitals and even in the same hospital. Then the hospitals will be able to receive feedback about their protocols, and this can help direct optimisation as needed. This would be an improvement on current methods which rely solely on DRLs. It is therefore recommended that FOM data should be published for each hospital/X-ray machine and clinical procedure.

6.4 Evaluation of Image Quality, Radiation Dose, FOM and Clinical Protocols among X-ray Machines

6.4.1 Neonate Age Group

6.4.1.1 Analysis of Image Quality

Regarding the CDRAD 2.0 phantom, this thesis identified wide variation in LCD detectability performance (**Figure (5-4)**), both between and within hospitals. Only 4 out of 17 X-ray machines had IQF_{inv} scores higher than the 75^{th} percentile (3.18). 4 X-ray machines (H5X1, H6X1, H6X2 and H8X1) had IQF_{inv} scores lower than the 25^{th} percentile (1.86). 2 points need to be discussed in relation to **Figure (5-4)**. First, the high outlier value for X-ray machine H2X4 and second, the lowest 2 scores which were lower than the first quartile (H6X1, H6X2).

In terms of the first point, the reasons for the high IQF_{inv} score for X-ray machine H2X4 could be related to the high value of mAs employed when using this X-ray machine (H2X4). As can be seen from **Table (5-12)**, this X-ray machine employed the highest mAs value (2

mAs) when compared with other X-ray machines. A higher setting of mAs increases the number of the photons that reach the detector and this leads to an increase in the SNR and a subsequent increase in the detection of the objects (visible holes of CDRAD 2.0 phantom image) (H. Alsleem, U, Mong, & Davidson, 2014). It was found from previous CDRAD 2.0 LCD detectability studies that there is a direct relationship between mAs and the detection of LCD (when increasing mAs, the detection of LCD increases and vice versa) (Alsleem et al., 2014; De Hauwere et al., 2005; Norrman et al., 2005; Precht et al., 2014; Saito et al., 2008). Regarding the second point, the low IQF_{inv} scores observed for the X-ray machines H6X1 and H6X2 might be due to the relatively high SID and the low mAs employed, as can be seen from **Table (5-12)**. Both the high SID (due to the inverse square law) and low mAs will decrease the number of the photons reaching the detector, and this decreases the SNR (De Hauwere et al., 2005; Tugwell et al., 2014) with a subsequent reduction in the score of IQF_{inv}.

Although it not plausible to attribute IQF_{inv} scores to a single factor, it appears that in general the type of image detector also has an impact on the IQF_{inv} scores. This can be justified by observing that all X-ray machines which used a CR image detector (H4X2, H5X1and H8X1) showed IQF_{inv} values which do not exceed the first quartile limit. Several CDRAD 2.0 phantom studies have shown that the performance of DR systems for LCD detectability is significantly better than that of CR systems, for the same radiation dose level (Alsleem et al., 2014; Bacher et al., 2003; Hamer et al., 2003; McEntee et al., 2007). The reason behind the excellent performance of DR when compared with CR could be related to the high DQE for DR systems when compared with that of CR (Bertolini et al., 2012; Schaefer-Prokop et al., 2008). The reason behind the low DQE value for CR systems is related to increased noise sources such as plate granularity and noise from the readout stage. Low DQE increases the visibility of noise on CR images compared with that of DR for the same employed exposure factors (McEntee et al., 2007).

Regarding the Gammex phantom, it was observed that there is wide variation in the visual IQ evaluation, both between and within the hospitals. From **Figure (5-6)** it can be seen that there is a similar trend between the IQ evaluation and the IQF_{inv} scores when using the CDRAD 2.0 phantom for neonatal CXR examinations (**Figure (5-4)**). There were also 3 outliers in **Figure (5-6)**. They are as follows: 1), the high outlier value of IQ score for the X-ray machine H2X4; 2), the lowest 2 scores of the IQ in for the X-ray machines H6X1,

H6X2. The expected reasons for these 2 listed points regarding **Figure (5-6)** are similar to what has been discussed previously for the CDRAD 2.0 phantom (**Figure (5-4)**).

6.4.1.2 Analysis of Radiation Dose and FOM

IAK (**Figure (5-4)**) was observed to vary both between and within hospitals. The majority of the X-ray machines (12) had IAK values lower than the 3rd quartile (34.45 µGy), and 4 X-ray machines (H2X2, H2X4, H7X1and H8X1) had IAK values higher than the 3rd quartile. 4 X-ray machines (H3X2, H4X1, H6X1and H6X2) had IQF_{inv} scores lower than the 25th percentile (18.36 µGy). 2 points need to be discussed in relation to **Figure (5-4)**. Firstly, the high outlier value of IAK for X-ray machine H2X4 and secondly, the lowest two IAK values observed in X-ray machines H6X1 and H6X2. In terms of the first point, the reason for, the high IAK (52.62 µGy) for the X-ray machine H2X4 is attributed to the high value of mAs employed in this X-ray machine (H2X4). As can be seen from **Table (5-12**), this X-ray machine employed the highest mAs value (2.0) when compared with other X-ray machines, and from the literature it can be seen that there is a direct relationship between the mAs values, the number of X-ray photons generated and the radiation dose received (Seeram, Davidson, Bushong, & Swan, 2016; Sun, Lin, Tyan, & Ng, 2012). Regarding the second point, H6X1 and H6X2 X-ray machines were observed to have the lowest IAK values- equal 8.56 and 8.57 µGy, respectively. Possible reasons could include the relatively high SID and low mAs and kVp employed as can be seen from **Table (5-12**), and both that the high SID (due to the inverse square law) and low mAs will decrease IAK values.

Although the 2 X-ray machines H4X1 and H4X2 have the same exposure parameters and SID values, H4X1 X-ray machine has a very low IAK value (9.89 µGy) compared with that in X-ray machine H4X2 (26.05 µGy). This could be attributed to the type of detector used- CR in X-ray machine H4X2 compared with DR in X-ray machine H4X1. It is accepted that CR needs more radiation to obtain a similar IQ to that from DR. Several studies have shown that the performance of the DR system is significantly better than that of CR in terms of dose reduction, with possible dose reduction of up to 75% in comparison with the CR (Fischbach et al., 2002; Hamer et al., 2003; Jansson et al., 2006; McEntee et al., 2007). Furthermore, additional filtration (1mm Al+0.1 mm Cu) was used in X-ray machine H4X1, and this is also considered to be an influential factor for reducing the IAK value. This is especially important since the ICRP has recommended the use of additional filtration in regards to paediatric imaging, in publication 121 (Zoetelief, 2013). Employing additional beam filtration removes the low X-ray energy and increases the penetration energy. This then leads

a reduction in the absorbed dose for the imaged object. Several studies have also recommended utilising additional filtration for both paediatric and adult acquisitions and have concluded that employing additional filtration can reduce radiation dose but without a reduction in IQ (Brosi et al., 2011; Ekpo et al., 2014; Hamer et al., 2004). Regarding the neonatal Gammex phantom, there was a similar trend for IAK distribution that resulted from the CDRAD 2.0 phantom, between and within hospitals.

It was observed from **Figure (5-5)** that the FOM values varied considerably, both between and within hospitals; only four X-ray machines (H1X1, H2X2, H4X2, H8X1) had FOM values lower than the 1st quartile (0.17), while four X-ray machines (H3X1, H3X2, H4X1, H6X1) had FOM values higher than the 3rd quartile (0.33). The main reason for the highest value of FOM observed in the X-ray machine (H4X1) may relate to additional beam filtration (1 Al+0.1 mm Cu) (see **Table (5-12)**).

6.4.2 One Year Age Group

6.4.2.1 Analysis of Image Quality

A large difference in LCD detectability performance (**Figure (5-8)**) was identified, between and within hospitals. 4 X-ray machines (H2X1, H2X4, H3X3 and H7X1) had IQF_{inv} scores higher than the 75th percentile (2.84). 4 X-ray machines (H4X1, H4X2, H6X2 and H8X1) had IQF_{inv} scores lower than the 25th percentile (1.66). The most important scores that need to be discussed were the highest IQF_{inv} score (H2X4) and the 3 lowest IQF_{inv} scores (H4X1, H4X2 and H8X1), all of which were lower than the first quartile limit (1.66).

The X-ray machine H2X4 has an outlier value of IQF_{inv}, which is 4 times larger than that of the X-ray machine H4X2. **Table (5-13)** shows that one of the main reasons for the observed high value of IQF_{inv} (H2X4) could be the value of the mAs employed. This X-ray machine used 2.5 mAs which was the highest value among all of the X-ray machines. In fact, this was approximately more than double the values of that of the other X-ray machines. Furthermore, this X-ray machine employed a short SID value (115 cm) compared with some of the other X-ray machines, which typically used either 135 or 180 cm. These 2 reasons (mAs and SID) would have a large influence on the values of IQF_{inv} (as discussed in subsection 6.4.1.1) since they are directly responsible for an increase in X-ray photons at the image detector surface. The lowest two values of IQF_{inv} were observed in the 2 X-ray machines: H4X1 and H4X2. Possible reasons for this could relate to the fact that these 2 X-ray machines used the lowest mAs values (1) and the largest SID (180 cm), resulting in the opposite- a lower quantity of X-ray photons at the image receptor (**Table (5-13)**). Although

these 2 X-ray machines (H4X1 and H4X2) used relatively similar acquisition parameters and techniques (except the additional filtration in H4X1), the X-ray machine H4X1 has higher score of IQF_{inv} (1.40) compared with that from H4X2 (0.96). These differences can be attributed to the machine type used, where the X-ray machine H4X1 used a DR detector compared with a CR detector used for the H4X2 X-ray machine **(Table (5-13))** (the influence of detector type was discussed in subsection 6.4.1.1). The third lowest value of IQF_{inv} (1.44) was observed in the X-ray machine H8X1 and this could again be related to the type of the detector used (CR) and its relatively low mAs value (1.25).

6.4.2.2 Analysis of Radiation Dose and FOM

It is clear from the distribution of IAK values in **Figure (5-8)** that there is a large variation both between and within hospitals. Only 4 X-ray machines (H2X2, H2X4, H5X1 and H8X1) had IAK values higher than the 3rd quartile (53.37 µGy). 4 X-ray machines (H4X1, H4X2, H6X1 and H6X2) had IAK scores lower than the 25th percentile (23.30 µGy).

X-ray machine H2X4 had a high IAK outlier value of 82.82µGy. **Table (5-13)** shows that one of the main reasons for this might be the mAs value used. This X-ray machine used 2.5 mAs which was the highest value among all of the other X-ray machines. Furthermore, this X-ray machine (H2X4) employed a short SID value (115 cm). These are likely to be the reason for the IAK high value (as discussed in subsection 6.4.1.2.). Similarly, the X-ray machine H5X1 had the second highest value of IAK (72.47µGy). This was the result of its using the second highest mAs value (1.8) among hospitals and a relatively short SID (120 cm). Furthermore, the type of the detector used in this X-ray machine was CR which is not recommended for radiation dose saving (as discussed in subsection 6.4.1.2.). The lowest value of IAK (5.44 µGy) was observed in X-ray machine H4X1 and, as can be seen from **Table (5-13)**, this might be attributed to its using the lowest mAs value (1.0) and the highest SID value (180 cm). These 2 factors have a big influence on IAK values (as discussed in subsection 6.4.1.2.). Similar to the neonatal age group, the 2 X-ray machines H4X1 and H4X2 have the same exposure parameters and SID values, however the H4X1 X-ray machine had a very low IAK (5.44 µGy) compared with that of H4X2 (16.48 µGy). This might be related to the detector type (CR) used with H4X2, which is not recommended for dose saving, and also that H4R1 X-ray machine utilised additional filtration (1 mmAl and 0.1cu mm) (as discussed in subsection 6.4.1.2.).

The FOM (**Figure (5-9)**) varied between and within hospitals. The highest value of FOM (0.51) was observed in X-ray machine H4X1, while the lowest value of FOM (0.05) was observed in X-ray machine H5X1. The main reason for the highest value of FOM in X-ray machine H4X1 is related to its utilising additional beam filtration (1 Al+0.1 mm Cu) and high value of SID (180) (see **Table (5-13)**).

6.4.3 Five Years Age Group

6.4.3.1 Analysis of Image Quality

There was a large variation in LCD detectability **(Figure (5-10)**), between and within hospitals. 4 of the X-ray machines (H2X1, H2X2, H2X4, H7X1) were found to meet the third quartile limit of IQF_{inv} score (1.51), while the other X-ray machines (12 X-ray machines) were lower than this limit with 4 X-ray machines (H3X2, H3X3, H4X2 and H8X1) having IQF_{inv} scores lower than the 25^{th} percentile (1.20). The lowest IQF_{inv} score was observed in the X-ray machine H8X1 (0.87) (**Figure (5-10))**. This low score could have resulted from 3 possible factors. Firstly, the relatively low mAs value employed when compared with other X-ray machines; secondly, the high SID value (200 cm), which is the highest value reported; and finally, the detector type used, which was CR (the influence of these 3 factors were discussed in subsection 6.4.1.1) **(**see **Table (5-14))**. Here, the technical factor (detector type) has a considerable influence on IQF_{inv} scores in the 2 X-ray machines H4X1 and H4X2, since many studies (as reported in subsection 6.4.1.1) have concluded that the IQF_{inv} resulting from DR systems is significantly higher than that of the CR systems for the same radiation dose level.

Although these 2 X-ray machines used similar acquisition parameters (except for utilising additional beam filtration in H4X1), H4X1 had a higher IQF_{inv} score compared with X-ray machine H4X2 and this is related to detector type. A DR type was used in H4X1 whereas a CR was used in H4X2. From **Figure (5-10)**, X-ray machine H2X2 has the highest IQF_{inv} score (1.81) among the other X-ray machines and this is likely to be related to the relatively high mAs and kVp values used in this X-ray machine, as illustrated in **Table (5-14)**. It must be acknowledged that the IQF_{inv} threshold that leads to a change in visual IQ has not been investigated within this thesis, nor has it been within the literature. Future work should be conducted to determine the IQF_{inv} threshold that leads to a change in visual IQ in clinical practice. It is worth highlighting that by utilising a high SID value (e.g. 200 cm) for X-ray machine H8X7, for this age group and also for the 10-year age groups and adults, this thesis was in conflict with common clinical practice and the literature (Commission of the

European Communities (CEC), 1996a). However, utilising a 200 cm SID for CXR examinations has been suggested in a study by Hirose et al. (Hirose, Ikeda, Ito, Ishigaki, & Sakuma, 1993) for bringing about a further reduction in radiation dose. Another possible reason for using the high value of SID (200 cm) in X-ray machine H8X7 might be to reduce image magnification.

6.4.3.2 Analysis of Radiation Dose and FOM

From the distribution of the IAK values in **Figure (5-10)**, there is a wide variation that can be seen in IAK, both between and within hospitals. Only 5 X-ray machines (H2X2, H2X4, H4X2 H6X1 and H6X2) had IAK values higher than the 3rd quartile (45.52 µGy), while the other X-ray machines were lower than the 3rd quartile. 4 X-ray machines (H3X2, H3X3, H3X4 and H4X1) had IAK scores lower than the 25th percentile (14 µGy). The X-ray machine H6X1 was observed to have the highest IAK (59.22µGy). **Table (5-14)** shows that the main reasons for this value are the inclusion of the AEC and the relatively high kVp and mAs recorded from using AEC. The second highest IAK (52.94 µGy) was observed in H4X2 and this also could be related to AEC use, CR technology and recording a relatively high mAs post-exposure (the influence of these factors on IAK was discussed in subsection 6.4.1.2.).

In this age group (and in the 10-year old group) there is also an increase in AEC usage. This may reflect suggestions made by the CEC (Commission of the European Communities (CEC), 1996b), who do not recommend using AEC for infants and young children. It should be noted that implementing the use of the AEC could lead to an increase in dose, as noted in X-ray machines H6X1 and H4X2. This is in agreement with previous studies which conclude that AEC is associated with higher radiation dose for paediatrics when compared with the manual control (Almén et al., 1996; Hintenlang et al., 2002). The likely reason for its delivering higher a radiation dose with the AEC for paediatrics when compared with that of manual exposure control may be because most AEC systems are not designed specifically for paediatric patients. These systems have comparatively large and fixed ionisation chambers and their size, shape, and location do not reflect the differences in body size in paediatric patients (Hintenlang et al., 2002). In addition, the differences in the calibration of the AEC device amongst the different X-ray machines might be considered another factor that can lead to the variation in radiation dose between units.

The lowest IAK values were observed in H3X2 and H3X3 and were equal to 11.00 and 10.97 µGy, respectively. These values might be attributed to using manual mAs (1.6), low

kVp (70) and additional beam filtration (0.1 mm Cu). These parameters have been reported to have an influence on radiation dose reduction (as discussed in subsection 6.4.1.2.).

FOM (**Figure (5-11)**) was observed to differ both between and within hospitals. Only 3 X-ray machines (H4X2, H6X1, H6X2) had FOM values lower than the 1st quartile (0.06), while the remainder of the X-ray machines were higher than this limit. The high values of FOM in all the X-ray machines in H3 (0.19,0.18,0.22 and 0.19) could be explained by the utilisation of additional beam filtration (0.1 mm Cu) in all of these X-ray machines. The low radiation dose resulting from addition beam filtration is not always linked with low IQ, as discussed in subsection 6.4.1.2. Several studies have also found that utilising additional filtration for both paediatric and adult CXR imaging can reduce radiation dose but without a significant reduction in IQ (Brosi et al., 2011; Ekpo et al., 2014; Hamer et al., 2004).

6.4.4 Ten Years Age Group

6.4.4.1 Analysis of Image Quality

It is obvious from the bar chart (**Figure (5-12)**) that there is considerable variation in the IQF_{inv} values between X-ray machines. Only 4 X-ray machines (H2X1, H2X2, H2X3 and H3X5) had IQF_{inv} scores higher than the 75th percentile (1.73). Another 4 X-ray machines (H3X2, H4X1, H4X2 and H8X1) had IQF_{inv} scores lower than the 25th percentile (1.06). An outlier IQF_{inv} value was observed in X-ray machine H2X1, with a value equal to 2.39 which might be related to the highest value of mAs being used with this X-ray machine compared with the other X-ray machines (**Table (5-15)**) (the influence of this factor was discussed in subsection 6.4.1.1). The X-ray machine H8X1 was observed to have the lowest score of IQF_{inv} (0.90), and this might be attributed to the high SID value (200 cm), which is the highest value reported, and the detector type being CR (the influence of these two factors were discussed in subsection 6.4.1.**1).**

6.4.4.2 Analysis of Radiation Dose and FOM

It is clear from the distribution of the IAK values in **Figure (5-12)** that there is a large variation in IAK values amongst the X-ray machines. 4 X-ray machines (H2X2, H4X2, H5X1, H6X1 and H6X2) had IAK values higher than the 3rd quartile (43.61 µGy), while the other X-ray machines were lower than the 3rd quartile. 4 X-ray machines (H1X1, H3X2, H3X3 and H4X1) had IAK scores lower than the 25th percentile (25.40 µGy). 2 points need to be discussed in relation to **Figure (5-12)**. Firstly, the outlier value of IAK for X-ray machine H2X2; and secondly, the lowest IAK which was reported from H3X3. In terms of the first point, the reason for the high IAK (100.77 µGy) can be attributed to the use of the

highest value of mAs (5) (**see Table (5-15)**) among the hospitals (as discussed in subsection 6.4.1.2.). Regarding the second point, as can be seen from **Table (5-15)**, the X-ray machine H3X3 employed a manual technique setting with the lowest value of mAs (1.3), a low kVp (73) and additional beam filtration (0.1 mm Cu). These parameters have influence on dose reduction (as discussed in subsection 6.4.1.2.). In general, all 5 X-ray machines in hospital number three (H3) were observed to have low IAK when compared with the other hospitals. This might be attributed to its using a manual exposure technique together with low kVp, mAs and additional beam filtration.

On the other hand, FOM (**Figure (5-13)**) varied both between and within hospitals. Only 5 X-ray machines (H2X1, H3X2, H3X3, H3X4, H3X5) were found to meet the 3rd quartile (0.13), and 4 X-ray machines (H2X2, H4X2, H6X1, H8X1) had FOM scores lower than the 1st quartile (004). The highest values of FOM observed in all of the X-ray machines in H3 might be related to their utilisation of additional beam filtration (0.1 mm Cu).

6.4.5 Adult Group

6.4.5.1 Analysis of Image Quality

Regarding the CDRAD 2.0 phantom, our study identified considerable variations in LCD performance (**Figure (5-14)**), between and within hospitals. Only 5 out of 17 X-ray machines had IQF_{inv} scores higher than the 75th percentile (1.61). 3 X-ray machines had IQF_{inv} scores lower than the 25th percentile (1.16). The lower IQF_{inv} observed for X-ray machine H8X1 had a value of 0.83, this could have been caused by the type of the image detector (CR) and its using a higher SID value (200 cm) when compared with that of the other X-ray machines (**Table (5-17)**) (the influence of this factor was discussed in subsection 6.4.1.1). In addition, this X-ray machine (H8X1) did not make use of an anti-scatter grid as opposed to the other X-ray machines. Utilising an anti-scatter grid can lead to a significant improvement of IQF_{inv} values when compared with not doing so, as demonstrated in study by Ween et al (2009).

The type of detector (CR), together with the relatively low kVp, is likely to be the main reasons for the low IQF_{inv} score in X-ray machine H5X1 (lower than the 1st quartile limit). Manual exposure control was used in H5X1, in which any reduction in kVp values can lead to a reduction in the final number of X-ray photons which reach the image detector and as such lower values of IQF_{inv},. If AEC was used then a lower kVp is more optimal for obtaining high IQF_{inv} values, because the mAs is automatically compensating for the

numbers of X-ray photons resulting from the reduction of kVp, and the greater numbers of X-ray photons reaching image detector.

From **Table (5-17)**, it appears that the high IQF_{inv} score for X-ray machine H2X1 is related to using a manual exposure technique (with an anti-scatter grid) and high mAs values when compared with other X-ray machines that used a manual exposure technique (the influence of mAs was discussed in subsection 6.4.1.1). The use of AEC was observed to have an influence on the IQF_{inv} scores. The mean (SD) of the IQF_{inv} scores, with and without AEC, were 1.59 (SD=0.29) and 1.23 (SD=0.42), respectively. This might be explained by considering that all of the hospitals which used an AEC also used an anti-scatter grid which improved the image contrast by reducing the scattered radiation that reached the detector. A phantom study was undertaken by Ween et al. (using a CDRAD 2.0 phantom), who investigated anti-scatter grids and concluded that its use produces statistically significant differences in LCD detectability improvement (Ween et al., 2009). It must be acknowledged that the AEC is not the only factor that can lead to differences in the mean IQF_{inv} values for the 2 groups of X-ray machines, with and without AEC. This is because variations in the other acquisition factors such as SID, detector types, and the inclusion of additional filtrations also exist.

In terms of the standard and larger sized (with added fat jacket) Lungman phantoms, as can be seen from **Figure (5-16)** and **Figure (5-20),** the lowest IQ scores were observed for X-ray machine H8X1. This is similar to the evaluation of the IQF_{inv} results from the CDRAD 2.0 phantom for the adult group (**Figure (5-14)**), since the lowest score of IQF_{inv} was also observed in the X-ray machine H8X1. In addition, the highest IQ score for both the standard and larger sized Lungman phantoms was observed in the X-ray machine H2X3 and this also was approximately similar to the highest score of IQF_{inv} resulting from CDRAD 2.0 phantom for the adult group. On the other hand, there are relatively similar trends between the LV evaluation for each of the standard (**Figure (5-18)**) and larger sized (**Figure (5-22)**) Lungman phantoms and their corresponding IQ evaluation.

The influence of the additional phantom thickness represented by the fat jacket (larger size phantom) on IQ and radiation dose across the X-ray machines is discussed in section 6.5. This also includes the discussion of the variations in routine clinical protocols for the larger and standard sized adult CXR examination.

6.4.5.2 Analysis of Radiation Dose and FOM

IAK (**Figure (5-14)**) was observed to vary both between and within hospitals. The majority of X-ray machines (12) had IAK values lower than the 3rd quartile (132.32µGy); and all X-ray machines in hospital 3 (H3) had IAK values lower than the 1rd quartile (20.16µGy). X-ray machine H4X2 was observed to have the highest IAK value (239.15 µGy). **Table (5-17)** shows that the main reasons for the higher value in this X-ray machine can be attributed to the use of an AEC, an anti-scatter grid and its not using any additional filtration. A CR system was also used in this X-ray machine and this has been shown to be an influencing factor on increasing the radiation exposure compared with DR systems (the influence of system type was discussed in subsection 6.4.1.2.. The lowest IAK values were observed in all 5 X-ray machines in one hospital (H3) (17.26, 19.70, 18.98, 18.09 and 20.16 µGy) and this can be attributed to using a manual technique together with additional filtration (0.2 mm Cu) and no anti-scatter grid. Several LCD detectability studies (using CDRAD 2.0 phantom) (Brosi et al., 2011; Hamer et al., 2004) concluded that it is possible to reduce radiation dose by around 30% in CXR examinations by utilising additional filtration. Brosi et al. (2011) concluded that utilising additional Cu filtration equal to 0.1, 0.2, and 0.3 mm thickness reduced ESD by 25–32%, 32–39%, and 40–44%, respectively. Dose reduction ranges depend on the kVp values used and in their study they recommended utilising 0.3 mm Cu addition filtration without any significant reduction in IQ.

The results show that using AEC has a high influence on IAK. The mean IAKs, with and without using AEC, were 140.72 (SD=57.68) and 40.50 (SD=31.25) µGy, respectively. This difference in the delivered IAK from both the manual technique setting and from AEC can be explained by considering that all the hospitals that used AEC also used an anti-scatter grid. Several studies report that using an anti-scatter grid increases the radiation dose significantly because a number of X-ray photons are absorbed by the grid. This leads to an increase in the radiation dose to the patient by way of compensation (Jessen, 2004; Ween et al., 2009). It must be acknowledged that the AEC is not the only factor that caused the difference in the mean values of IAK for the 2 groups of the X-ray machines, with and without utilising AEC. Other factors such as kVp, the use of an anti-scatter grid and additional filtration also exist.

On the other hand, FOM (**Figure (5-15)**) was observed to vary both between and within hospitals. Most of the FOM values of X-ray machines (11) were between the mean and the 1st quartile value (0.02). FOM results presented in **Figure (5-15)** do not essentially mean

that the image is better than the other, however the image has the best optimal LCD detectability with the lowest radiation dose. It should be noted that there are many variables in the existing protocols when acquiring CXR images and this could make it hard to determine the true factors causing these performance differences. However, it is clear from **Figure (5-15)** that using DR with manual exposure control, without an anti-scatter radiation grid and with additional filtration appears (protocol used in H3) to be the optimal technique for reducing dose and maximising LCD detectability. It is suggested that other X-ray machines which have low values of FOM, especially those lower than the 1st quartile (i.e. H4X2), should consider reviewing their protocols for CXR examinations as this could lead to radiation dose reductions and protocol optimisation.

Regarding the standard sized Lungman phantom, the mean IAK, with and without AEC, was equal to 80.12 and 44.53 µGy, respectively. With respect to the larger Lungman phantom, the mean IAK, with and without using AEC, was 218.91 and 92.03 µGy, respectively. The high difference in the delivered radiation dose from the manual technique setting and from AEC control can be explained by considering that all the hospitals that used AEC controls also used an anti-scatter grid. Several studies found that using anti-scatter grid causes a statistically significant increase in the radiation dose (Ween et al., 2009; Jessen, 2004).

6.4.6 General Discussion on Radiation Dose and Image Quality Findings across all Age Groups

4 points are worthy of discussion from the above findings: 1) the correlation between radiation dose and IQ, both between and within hospitals; 2) the changes in IQF_{inv} values across the different age groups; 3) a comparison of IQ, radiation dose and FOM with the literature and guidelines; 4) the implications of the IQ and radiation dose results to clinical practice and their applicability to radiology departments in general.

6.4.6.1 The Correlation between Radiation Dose and Image Quality, both between and within Hospitals

It was observed that there was a weak correlation between IQ and radiation dose in most of the age groups. This might be attributed to the high variability of the technical characteristics of image detectors and their manufacturers (the influence of this factor on IQ and radiation dose was discussed in subsection (6.4.1.)), or to the difference X-ray generator types and manufacturers used between and within hospitals. Furthermore, there was a considerable variation in the exposure parameters and techniques utilised for the same examination,

between and within hospitals. In addition, the variability in IQ and radiation dose among X-ray machines could also be attributed to the X-ray machine's age, since the ageing of the equipment can influence radiation output and consequently IQ. The weak correlation between IQ and dose found in this thesis is confirmed by the literature. Several survey studies that were conducted to investigate the variation in IQ and radiation dose among hospitals of different examinations using digital and analogue systems showed that there was no straightforward correlation between radiation dose and IQ (Almén et al., 1996; Geleijns et al., 1993; Struelens et al., 2008; Ween et al., 2009).

6.4.6.2 Changes in the IQF_{inv} Values across the different Age Groups

In general, it is noticeable from the bar charts of IQF_{inv} across the different age groups that there is a tendency for IQF_{inv} to decline as age/size increases. For example, the mean values for IQF_{inv} for neonates, 1-year and 5-year groups were 2.58, 2.39 and 1.24, respectively. Such changes in IQF_{inv} can be related to increasing the PMMA thickness (simulating increased patient thickness) with increasing the phantom age groups. For instance, the neonatal age group has higher values of IQF_{inv} compared to that of the other age groups because it has the lowest thickness of PMMA which is equal to 6.5 cm compared with 8.5 and 10 cm for the 1-year and 5-years group, respectively. Increasing the PMMA thickness leads to an increase in the attenuation of the incident X-ray beam through the PMMA slabs, and increases the photon scatter (Carucci, 2013). The scattered radiation decreases the dynamic range of the X-ray intensities that exit from the patient/phantom. It reaches the detector, leading to a decrease in the contrast of the image and also a reduction in SNR. This is as result of an increase in the quantum noise in the image without any additional signal (Williams et al., 2007). Therefore, it is not valid to compare the numerical values of the IQF_{inv} between the different age groups because their PMMA thickness is different. The valid comparison needs be achieved within each group independently.

6.4.6.3 Comparing image quality, radiation dose and FOM with the literature and guidelines

There was no possibility of comparing the individual numerical values of IQF_{inv} acquired within this thesis against those published in literature. The reason for this was that publications which included a CDRAD 2.0 phantom were obtained using different thicknesses of PMMA slabs and as such comparison would not be valid. This is also true for FOM values, which depend on IQF_{inv} values as a metric for the evaluation of IQ.

On the other hand, the radiation dose in this thesis was represented by the IAK instead of the ESD. The major aim of dose measurements in this thesis was to compare radiation dose values among the X-ray machines. However, to compare radiation dose values from the current study with guidelines such as those from the National Radiological Protection Board (NRPB) (Hart, D. W. B. F., Wall, B., & Shrimpton, 2000) and UK NPDD 2010 (Hart, Wall, Hillier, & Shrimpton, 2010) would require the conversion of IAK values to ESD, by multiplying the values of IAK with a suitable BSF. The ESD has not been calculated in this thesis because, in diagnostic radiography, BSF depends on numerous factors such as the tube voltage, filtration, SID and the composition of the imaged object (Petoussi-Henss et al., 1998). In this thesis, standard clinical protocols for the same age groups were different between the X-ray machines and different values of kVp, additional filtration and SID were utilised for the same age group among the X-ray machines. Consequently, it is not valid to utilise the same value of BSF to convert IAK values that have resulted from the 17 X-ray machines to ESD. Utilising the same BSF leads to an increase in errors in radiation dose values and makes the comparison in radiation dose between the X-ray machines not valid. A valid approach would be to compare X-ray machines based on IAK values that were recorded directly from the RaySafe X2 dosimeter without adding the BSF. Another reason for not calculating ESD from the CDRAD 2.0 phantom data, especially when compared with anthropomorphic phantoms, is because it is not a valid to compare the results of ESD from the CDRAD 2.0 phantom against those from a benchmark. This is because the CDRAD 2.0 phantom does not simulate a patient's size and is instead a standard size that is utilised across all of the age groups.

To give an indication about variation in radiation dose between the results from the participating hospitals in this thesis and that from the benchmark, the IAK results from the Gammex phantom and the standard sized Lungman phantom were converted to ESD by using suitable BSFs. These 2 phantoms were selected rather than the CDRAD 2.0 phantom because they reflect a good a simulation of the size and attenuation of their corresponding patient age groups when compared with that of the CDRAD 2.0 phantom. However, the ESD values for the larger sized Lungman phantom (with a fat jacket) were not calculated because the BSF for the larger sized adult phantom/patient has not been previously reported in the literature. It would not be valid to use the same value of BSF for that of the standard sized phantom.

A BSF equal to 1.1 was applied for the Gammex phantom. This BSF value was designed for neonates, based on a study by Wall et al. (Wall, B. F., Harrison, R. M., & Spiers, 1988). Several studies have also applied this BSF value for neonates (Olgar et al., 2008; Armpilia, C. I., Fife, I. A. J., & Croasdale, 2001; Kostova-Lefterova, Taseva, Ingilizova, Hristova-Popova, & Vassileva, 2011). The BSF for the standard sized Lungman phantom was taken in this thesis to be equal to 1.47, based on the study by Schultz et al. (1994).

Regarding Gammex phantom ESD data, only 1 X-ray machine (H2X4) had an ESD value (54.93 μGy) higher than the recommended reference dose values from the NRPB (50 μGy), while the other X-ray machines were observed to have ESD values lower than the reference value. In terms of the average sized Lungman phantom ESD data, all the X-ray machines had ESD values lower than the reference dose values from the NRPB, which is 300 μGy. However, two X-ray machines (H2X4 and H4X2) were observed to have ESD values (185.8 and 200.3 μGy, respectively) higher than the recommended reference dose values of UK NPDD 2010 (150 μGy).

6.4.6.4 Implications of the Image Quality and Radiation Dose Results in Clinical Practice and their Applicability to Radiology Departments in General

The visual IQ evaluation for the neonatal age group in this thesis, indicated by the variation in visibility of the simulated pneumothorax and respiratory distress syndrome, did not include a detection task (simulated pathology detection). The clinical influence of the observed differences in visual IQ should be considered in future work to help maximise pathology detection performance. This is especially true for the X-ray machines observed with low IQ. The variation in IQ might influence the detectability of subtle pathologies in clinical practice. Further studies are required to investigate to what extent this variation in IQ might influence pathology detectability. If the low values of IQ observed in this thesis do not influence the diagnostic accuracy regarding pathology detection, the radiation dose can be minimised considerably, both between and within institutions.

On the other hand, in this thesis the IQ for all the paediatric age groups and adults is represented by the LCD detectability using the CDRAD 2.0 phantom. This is a widely used and accepted method for evaluating IQ, assessing and comparing imaging systems and for optimisation studies. (Alzimami et al., 2009; Bacher et al., 2003; Geijer et al., 2001; Rong et al., 2001; Veldkamp et al., 2006). Clinically, improvement of LCD detectability could be beneficial for increasing the detectability of lesions and lung disease, particularly in chest radiography. Also, the findings of this thesis have confirmed the validity of using LCD

detectability, using a CDRAD 2.0 phantom, for evaluating visual IQ and LV and confirmed its potential use in CXR optimisation studies. Within this thesis, it was concluded that there is a correlation between the LCD detectability, the visual IQ and the general visual LV within CXR. Therefore, the clinical effects of the observed differences in LCD detectability between and within hospitals in this work, especially for the X-ray machines with the lowest values of LCD detectability, should be taken into account in future works to help increase the pathology detectability performance. These differences in LCD detectability could influence the detection of subtle pathologies in clinical practice. More research is necessary to investigate how these differences in LCD detectability can influence lesion detection. Similarly to the neonate age group, if the observed variations in LCD detectability reported in this thesis do not have a negative impact on the diagnostic accuracy of pathology, then further reduction of the radiation dose can be achieved.

The findings of this thesis demonstrate a weak correlation between radiation dose and IQ for the majority of the age groups investigated. This indicates that high IQ is not necessarily related to a high radiation dose and that radiographic imaging protocols are not well optimised and need further optimisation. This is especially necessary for the X-ray machines with low IQ and high radiation dose. Optimisation of their imaging protocols could increase the pathology detection performance and also minimise radiation exposure and risk to patients. The acquisition parameters for each age group should be selected based on the clinical indications for achieving acceptable IQ for diagnosis. It could be that an imaging protocol and its resultant radiation dose is optimal for general IQ evaluation, but it might not be acceptable for the detectability of a small subtle lung pathology.

Work within this thesis was conducted on a small sample of X-ray machines, however it is reasonable to speculate that our findings would be applicable to paediatric and adult CXR examinations in general. The reason relates to IQ and dose optimisation being difficult to achieve when undertaking chest radiography due to the wide range of tissue types and medical indications within the thorax which makes it a difficult examination.

Furthermore, such differences in IQ and radiation dose are likely to exist in other imaging centres/hospitals which conduct paediatric and adult CXR examinations, since variations in standard clinical imaging protocols and imaging system performance still exist. With this in mind, it is suggested that paediatric and adult CXR examinations should be investigated in all institutions to ensure conditions are optimal, such as that the IQ is maximised and the dose/risk is minimised.

6.5 Influence of Phantom size on Image Quality and Radiation Dose among X-ray Machines

The number of overweight and obese people has increased rapidly over recent years in many countries. In 2014, it was reported that around 51.6% of the European Union (EU's) residents were overweight (Eurostat Statistics Explained, 2017). Obesity is classified as a global epidemic according to the World Health Organization (WHO)(World Health Organization., 2000) and it increases a person's risk of developing diseases which can result in more medical procedures being require. This includes X-ray imaging. In comparison to people of standard weight, obese people tend to absorb larger radiation doses for the same examinations (Hofmann, 2016).

In radiography, practical difficulties often arise with imaging overweight and obese patients and there is an associated increase in radiation dose and IQ reduction (Carucci, 2013). Such difficulties can include X-ray beam attenuation, scatter radiation and long exposure times, resulting in motion artefacts. With increases in body part thickness comes an increase in X-ray beam attenuation, which in turn increases in the exposure time and the quantity of scattered radiation. This increases patient dose (Carucci, 2013; Uppot, Sahani, Hahn, Gervais, & Mueller, 2007; Yanch, Behrman, Hendricks, & McCall, 2009). X-ray images from patients with a large body habitus are often of lower IQ and practitioners need research evidence regarding adjustments to both pre-acquisition and post-acquisition imaging parameters (Lazar, Plocher, & Egol, 2010).

The CEC (Commission of the European Communities (CEC), 1996a) has published guidelines on adult radiographic imaging, which includes the CXR examination. These guidelines provide information on the selection of optimal acquisition parameters, but it was designed only for standard sized patients. Currently, no guidelines or recommendations are available for undertaking CXRs on overweight or obese patients. In addition, a further limitation of the CEC guidelines is that they were developed in an era of analogue film/screen systems and are not specifically for digital acquisitions.

One of the aims of this thesis was to evaluate the influence of patient size on IQ, radiation dose and acquisition parameters when undertaking adult CXR examinations using routine acquisition protocols, between and within hospitals. The findings show that patient size has an influence on IQ and IAK. Regarding IQ, phantoms with a fat jacket had reduced IQ compared with that of the standard size phantom for the majority of the X-ray machines (thirteen X-ray machines) (**Figure (5-24)**). This achieved a statistical significant difference (P=0.001). This IQ decrease is expected because of the additional soft tissue thickness,

which leads to an increase in the attenuation of the incident X-ray beam along with the production of additional scatter (Carucci, 2013).

Another possible reason for the low IQ obtained from a larger size phantom might be attributed to the majority of the hospitals using relatively similar protocols (except mAs values) for both the standard and larger size phantom/patients. This is surprising because, for many, no adaptation has been made to their routine imaging protocols as can be seen from **Tables (5-17)**. Using a standard size PA chest protocol for a larger size patient could lead to insufficient exposure reaching the image detector and therefore deliver a suboptimal IQ which might influence pathology identification (Audun et al., 2017). This is especially true when the examination is undertaken using manual exposure control compared with that using AEC (Buckley et al., 2009).

Regarding the nine X-ray machines that utilised AEC (H1XI, H2X2, H2X3, H2X4, H4X1, H4X2, H6X1, H6X2, H7X1) (**Table (5-17))**, they all used exactly the same protocols and acquisition parameters (except mAs values; see **Table (5-17)**) for both the standard and larger sized phantoms. The percentage of the increase in mAs values when the larger size phantom was imaged compared to the standard size. It varied from 100% to 212.5 % (**Table (5-17)**). The kVp values were the same for both standard and larger sized phantoms, except for in only one X-ray machine (H2X4) that changed the kVp. Here, kVp increased from 125 for standard sized phantom to 130 for the larger sized phantom. With respect to the 8 X-ray machines that used manual exposure (H2X1, H3X1, H3X2, H3X3, H3X4, H3X5, H5X1, H8X1) (**Table (5-17))**, the percentage of the increase in kVp and mAs values when moving from imaging the standard sized phantom to the larger sized phantom ranged from 0% to 9.09% and from 25% to 60%, respectively (**Table (5-17)**). The X-ray machine H5X1 used the AEC and anti-scatter grid with the larger sized phantom instead of manual exposure control without anti-scatter grid, which was used for imaging the standard size phantom. This resulted in a 12.9% and 117.1% increase in kVp and mAs for the larger sized phantom, respectively (**Table (5-17)**). From **Figure (5-24)**, 2 X-ray machines (H6X1 and H6X2) had better IQ for the images obtained from the phantom with a fat jacket, compared with that of the standard sized phantom. However, these images were had a higher IAK (216.6% and 224.32% percentage increase, H6X1 and H6X2, respectively). This could be due to the high mAs values recorded from the AEC with the larger sized phantom compared with that of the standard sized phantom- a 192.86% and 200% increase for X-ray machines H6X1 and H6X2, respectively (**Table (5-17)**). These high mAs values associated with the larger sized phantom led to an increase in the photon density and facilitated an adequate X-ray

penetration of the X-ray beam. This increased the number of the photons that reached the image detector which in turn increased the SNR, decreased the noise and resulted in good IQ. It has been demonstrated from the literature that there is a direct relationship between mAs and the IQ (De Hauwere et al., 2005; Precht et al., 2014; Saito et al., 2008). Two X-ray machines, H2X1 and H2X2 (**Figure (5-24)**), were found to have a relatively comparable IQ for the standard and larger sized phantom images. For the X-ray machine H2X1, this system used manual exposure control during the imaging of both the standard and larger sized phantoms. The percentage increases for kVp and mAs for standard sized to the larger sized phantom were 9.1% and 42.9%, respectively (**Table (5-17)**).

The combined radiation dose values of X-ray machine H2X1 for the standard sized and larger sized phantoms were 99.6 µGy and 158.4 µGy, with 59.0% percentage increase. Regarding X-ray machine H2X2, it's comparable IQ might be related to the value of mAs recorded from the AEC during imaging the larger sized phantom. This was equal to 2.4 mAs compared with 1.2 mAs that was recorded from imaging the standard sized phantom, with a 100% percentage increase (**Table (5-7)**).

From the results, utilising the AEC was not always helpful in obtaining satisfactory IQ when the larger sized phantom was used. This observation concurs with Audun et al. (2017). For adult CXR, Audun et al. (2017) investigated the influence of kVp and mAs on SNR, CNR and visual IQ and radiation dose for both standard sized and larger sized Lungman phantoms. Study findings demonstrate that the AEC does not always result in optimal visual IQ or a lower radiation dose and that it is possible to obtain higher SNR, CNR and visual IQ scores for a larger sized phantom at lower mAs values than that given by the AEC, because of post-processing. The study concluded that the manual exposure parameters were superior for controlling IQ for the larger size Lungman phantom than using an AEC technique, and it is required to propose protocols tailored for the larger size Lungman phantom.

More studies are needed to investigate how different acquisition parameters can influence IQ when larger sized patients are being imaged. Several studies have sought to establish strategies for achieving dose optimisation in adult CXR (Dobbins III James T. et al., 2003; Grewal, Young, Colins, Karunnaratne, & Sabharwal, 2012; Hamer et al., 2005). All of these studies focused on average sized patients and did not investigate optimisation for overweight and obese patients. Only a limited number of phantom studies have been conducted to investigate the impact of acquisition parameters on IQ and radiation dose for the larger body sized patients in CXR (Audun et al., 2017; Hauge et al., 2017; Otto et al., 2000). However,

these studies investigated the influence of only some parameters, such as kVp and mAs and did not cover all the acquisition parameters.

In general, it was noticed that the physical IQ represented by SNR, CNR and CI for the larger sized phantom images was lower than that of the standard sized phantom images (**Figure (5-25) - (5-27))**. The results showed that there is no statistically significant difference in SNR, CNR and CI measurements between images for the larger sized phantom and for that of the standard sized phantom images: (P=0.40), (P=0.37) and (P=0.37), respectively.

The general reduction in physical IQ (SNR, CNR, CI) for the larger sized phantom, compared with that of the standard sized, is also related to the influence of the additional fat thickness. This leads to an increase in the scattered radiation and a reduced number of the photons reaching the image detector, which in turn decreases the SNR and increases the noise in the image. This results in a lower IQ (Williams et al., 2007).

Regarding radiation dose, this thesis' results (**Figure (5-28)**) demonstrate that the additional thickness increases the radiation dose by 151.2 % compared with that of a standard sized phantom across all of the X-ray machines, with a statistical significant difference between them (P=0.002). The mean (range) values of IAK obtained from the standard sized and larger sized phantoms between the hospitals were 63.38 (19.28 to136.29) µGy and 159.20 (27.43 to 384.73) µGy, respectively. Comparable results regarding the influence of the additional fat thickness on radiation dose were reported in a study by Otto et al. (2000). In this thesis, the larger sized phantom was observed to have both a higher IAK and a lower IQ compared with that of the standard sized phantom. Interestingly, the higher IAK received by the larger sized phantom did not considerably improve IQ in most of the X-ray machines compared with standard sized phantom. This emphasises that the protocols used for larger sized phantom may need further optimisation. With growing levels of obesity, it is ever more important to develop and establish imaging protocols designed for patients who are overweight and obese and this should be applied across a greater range of anatomical areas.

6.6 Statement of Novelty

It is now possible to indicate the novel contributions arising from this thesis, and these can be seen below:

1. A novel method for comparing radiation dose and IQ between X-ray units based on the use an FOM concept has been established. This new method is also likely to be beneficial for the optimisation of X-ray units, between and within hospitals. Establishing this novel method was necessary since it was observed that there is no standardised method which considers both IQ and dose, between and within hospitals. The existing methods either depend on DRLs. These could be considered self-limiting approaches because they only take into account radiation dose and not IQ, or depend on separate IQ and dose evaluations without a common link. This makes the comparison between hospitals difficult and complex.
2. Use of a novel FOM formula {FOM= $(IQF_{inv})^2$ /IAK} which can be used with CD phantoms and utilises the LCD detectability as metric for IQ evaluation. This novel formula is useful because existing FOM formulas depend on SNR and CNR as metrics for IQ evaluation. These are considered to be a limitation since they do not take into account the object size during IQ evaluation. Also, they do not provide a standardised method for IQ evaluations as a result of the variations in the selected sizes and locations for ROIs.
3. The provision of new information regarding how routine standard sized paediatric and adult CXR protocols result in variation in IQ and radiation dose, between different X-ray machines and between and within hospitals.
4. The provision of new information regarding how routine larger sized adult CXR protocols result in variations in IQ and radiation dose, between and within hospitals.
5. The provision of new information on the relationship between body habitus and IQ and radiation dose when undertaking adult CXR examinations using routine clinical protocols.
6. The provision of new information on the validity of utilising LCD detectability, using a CDRAD 2.0 phantom, for evaluating visual IQ and a visual LV, and for CXR optimisation studies.

6.7 Limitations and Future Work

6.7.1 Limitations of Validation CDRAD 2.0 Phantom for use in CXR optimisation

A limitation of this thesis is that the correlations between the LCD detectability, IQ and LV were only carried out using a standard sized Lungman phantom, and that variations in phantom size were not taken into account. A further limitation was that the correlation between the LCD detectability and the LV was carried out using only one type of simulated lesion (that simulating pulmonary nodules). The other types of lesion (such as micro nodules, lines, and reticular opacities) were not considered. Future work should be undertaken taking into account these limitations by using different anthropomorphic phantom sizes, loaded with different types of simulated lesions.

Another limitation in the validation of the CDRAD 2.0 phantom was that the Lungman phantom was utilised for general visual IQ. However, this phantom is limited by its lack of anatomical variation compared with that of the real patients. In 2011, Craig S. et al. (Moore, Liney, Beavis, & Saunderson, 2011) created and validated a computer algorithm which was able to simulate CR chest images from digitally reconstructed radiographs (DRRs), generated by ray tracing virtual X-rays through clinical patient CT datasets. Several studies have been conducted by Moore et al. (Moore et al., 2012; Moore et al., 2015) that focused on the optimisation of CR CXR imaging using a DRR computer simulation. The simulated clinical experiment (simulated DRR) used by Moore et al. has been successfully utilised for deriving an optimal exposure parameter for CXR examinations. Therefore, further studies should be conducted using DRRs or clinical images from patients. These should apply the same acquisition parameters used in this thesis to investigate the validity of using the CDRAD 2.0 phantom as a measure of clinical IQ in optimisation studies. This could provide further evidence of the validity of using the CDRAD 2.0 phantom for optimisation studies.

Also, based on the findings, future work could focus on establishing a baseline IQF_{inv} value for CXR that identifies the threshold of LCD detectability that is required for obtaining an adequate IQ for diagnosis. The availability of such baseline measures would be highly beneficial for routine quality assurance programs to ensure that the IQ is within accepted limits. It would also be beneficial for optimisation studies as it could facilitate the evaluation of how different techniques can influence IQ in clinical practice.

6.7.2 Limitations of Evaluating Image Quality, Radiation Dose and FOM among X-ray Machines

Within the work described in this thesis a number of limitations have been identified, including:

1. As a result of access restrictions in hospitals, this thesis investigated the variation in IQ and radiation dose only for the standard sized age groups (except for in the adult group). Variation for simulated patients with different sizes, and for different anatomical area examinations, has not been included, e.g. for larger sized or for those with serious malnutrition/metabolic conditions. However, the influence of patient size on IQ and radiation dose was evaluated for the adult group represented by the Lungman phantom (with and without the fat jacket). This provided valuable information about the influence of patient size on IQ and radiation dose. Future work could be conducted using the CDRAD 2.0 phantom for different age groups and for larger phantom sizes that simulate overweight and obese patients, and also for different anatomical area examinations, e.g. for pelvis examinations.
2. The visual IQ evaluation was undertaken only using a relative VGA method and there were no detection tasks. Using an ROC method, for example, could be an option for investigating the variability in lesion detection performance between the hospitals/systems. Future work could involve conducting an investigation into the variability in lesion detection performance for adult CXRs, between and within hospitals, using an ROC method. This could be achieved by including simulated lesions with different sizes, contrast and locations in Lungman phantoms, and then imaging the phantom on different X-ray machines in different hospitals. This could investigate how their local protocols influence lesion detection performance.
3. The number of the hospitals involved in this thesis could be considered relatively limited, and only a limited number of X-ray machines were included. The reason behind this is that they were the only hospitals that granted access for the purpose of the study, and that could be included within the time available for this thesis.
4. Not all of the radiographic positions were covered; the lateral chest position, for instance, was not covered. This is because of the limited time available and also because this is a relatively infrequent examination.
5. Only the variation in IQ and radiation dose among X-ray machines was investigated, without performing optimisation for the protocols of the participating hospitals. Future work could be conducted to focus on IQ and dose optimisation for the

participating hospitals. This could be achieved by going back to the hospitals and optimising their local protocols to investigate to what extent the radiation dose could be reduced without reduction IQ. It would be advantageous to establish whether all X-ray units could achieve the same IQ for the same radiation dose.

6. Utilising the phantoms to obtain the images in this thesis raises some issues since they are limited by their lack of movement, anatomical variation and that the simulated pathologies (spherical shapes) inserted in these phantoms do not always represent clinical reality.
7. The fat deposits (fat jacket) added to the Lungman phantom to simulate the larger sized patients only include subcutaneous fat. Visceral fat deposits were not considered. This is because it was difficult to add such fat to the Lungman phantom, however future research could determine its impact on IQ and radiation dose.
8. A questionnaire was utilised for collecting information regarding the eyesight of the observers; however, further research could include formal eyesight tests prior to starting the data collection. This could increase both the reliability and validity of the study.
9. The type of imaging equipment used for neonatal examinations could have been different. With this work they were predominantly stationary X-ray machines, while neonatal imaging is often undertaken using mobile X-ray machines. Further work using the method described could be undertaken for mobile systems.
10. The evaluation of the simulated LV, between and within hospitals, was based on an anatomical feature within the left hemi-thorax at the level of the 8^{th} thoracic vertebra, immediately adjacent to the paraspinal tissues within the Lungman phantom instead of based on a simulated lesion. This because it was not possible to control the fixation of the simulated lesions inside the phantom when moving it between and within the hospitals. This caused all of the simulated lesions to move around from their intended locations and generated limitations on the validity of this study. Therefore, the feature described was selected as a vehicle for assessing the LV as a result of its being fixed and located in the same position across all of the images.
11. The threshold value of IQF_{inv} that can lead to a change in visual IQ was not investigated within this thesis. Future work should be conducted to determine the threshold of IQF_{inv} that would lead to a change in visual IQ in clinical practice.

6.8 Conclusions

In this thesis, a novel method has been established to evaluate and compare IQ and radiation dose between and within hospitals. This novel method utilises an FOM concept which combines IQ and radiation dose into a single factor. This is considered to be a new method for IQ and radiation dose assessment. In comparison with other existing methods, it takes into account both the IQ and radiation dose and produces a single figure which can allow comparisons between systems and also the monitoring of performance. In addition, the new FOM formula proposed in this thesis uses the IQF_{inv} as a metric for IQ. This is instead of using existing metrics such as SNR and CNR, which again brings in additional advantages. The major objective of this thesis was to answer the question: "to what extent do standard protocols for undertaking paediatric and adult CXR vary, between and within hospitals, and what is the resultant impact on IQ and radiation dose?". The findings from this thesis clearly demonstrate a wide variation in radiation dose, visual and physical IQ metrics and FOM, between and within the participating hospitals. A weak correlation between the radiation dose and IQ was observed for most of the age groups studied. It was concluded that good IQ is not necessarily related to a high radiation dose. Likely causes of this can be attributed to differences in:

- Acquisition parameters and the radiographic techniques used
- Technical characteristics of the machines used

This may prove that there is opportunity for optimising the radiographic technique with a significant potential for radiation dose reduction in these hospitals, without degrading IQ. Data from this thesis suggests utilising additional beam filtration for all paediatric age groups and for adults, without utilising an anti-scatter radiation grid, should be considered. This was found to have an influence on dose reduction, but without necessarily degrading IQ.

When considering the influence of patient size on IQ and radiation dose, the findings demonstrate that CXRs for the larger sized Lungman phantom had a lower IQ compared with that of the standard sized phantom. This was true for the majority of the X-ray machines, between and within the participating hospitals. The IAK was found to be higher for the larger sized phantom and double that of the standard sized phantom, for all the X-ray machines between and within the participating hospitals. The data suggests that, when using routine clinical CXR protocols for larger patients, IQ may be compromised (possibly resulting in reduced pathology detection) and radiation dose will be increased. Routine protocols may therefore require greater optimisation for this subcategory of patients.

The validity of using physical evaluations of LCD detectability, with a CDRAD 2.0 phantom, as a main method for IQ evaluation in this thesis was investigated prior to collecting data from the hospitals. The results showed that this technique is valid for IQ, LV assessment and for CXR optimisation studies since it has a good correlation with the visual IQ and visual LV from the Lungman chest phantom.

Data collected within this thesis provides valuable, new and up-to-date information on IQ, radiation dose and standard clinical protocol variations among hospitals. These findings can be used as a baseline for any future local and national reference doses, wherein such data, especially for paediatric age groups, is currently not available. Furthermore, the data from this study could be useful as a baseline for any new surveys and studies on the optimisation of IQ and radiation does within UK.

Appendices

Appendix A:

A-1. The QC test form sheet.

	Exposure number	**kV**	**mA**	**time (ms)**	**focus**	**kV**	**µGy**	**ms**
Output variation with kV	1	60	200	100	Broad			
	2	70	200	100	Broad			
	3	80	200	100	Broad			
	4	90	200	100	Broad			
	5	100	200	100	Broad			
	6	110	200	100	Broad			
	7	120	200	100	Broad			
Output variation with mA	8	80	25	100	Broad			
	9	80	50	100	Broad			
	10	80	100	100	Broad			
	11	80	200	100	Broad			
	12	80	320	100	Broad			
	13	80	500	100	Broad			
	14	80	800	100	Broad			
Output variation with time	15	80	200	20	Broad			
	16	80	200	50	Broad			
	17	80	200	100	Broad			
	18	80	200	200	Broad			
	19	80	200	500	Broad			
	20	80	200	1000	Broad			
Reproducibility	21	80	200	100	Broad			
	22	80	200	100	Broad			
	23	80	200	100	Broad			
	24	80	200	100	Broad			
	25	80	200	100	Broad			

A-2. Form sheet for collecting the information of the X-ray machines for the hospitals.	
Hospital name: -	X-Ray machine number: -
Equipment type	Mobile x-ray ☐ Static x-ray ☐
Company	
Product name	
Country	
Year of purchase	
Tube type	
Generator type	
Target angle	
Inherent filtration	
Focal spot size (mm)	
Anti-scatter grid type (stationary /removable) and ratio	
System type (chest stand or general purpose)	
System type (Detector)	CR ☐ IDR ☐ DDR ☐
Detector manufacturer	

A-3. Form sheet for collecting the information regarding the acquisition parameters/techniques used for the CXR in the hospitals.	
Hospital name: - X-ray machine number: - Patients age group: -	
Type of exposure control	AEC mode ☐ Manual mode ☐
Patient position type	Supine AP ☐ Stand Bucky PA ☐
kVp	
mAs	
Additional filtration	
Focal spot type	
Grid in place	
SID	
ODD	
Numbers and positions of selected chambers	

A-4. The ethical approval letter for this thesis granted from the University of Salford (HSR1617-76).

University of Salford MANCHESTER

Research, Innovation and Academic Engagement Ethical Approval Panel

Research Centres Support Team
G0.3 Joule House
University of Salford
M5 4WT

T +44(0)161 295 2280

www.salford.ac.uk/

17 March 2017

Dear Sadeq,

RE: ETHICS APPLICATION–HSR1617-76–'A comparative study evaluating the performance of diagnostic radiography units and protocols for paediatric and adult chest radiography examinations in medical imaging centres across the North-west of England.'

Based on the information you provided I am pleased to inform you that application HSR1617-76 has been approved.

If there are any changes to the project and/or its methodology, then please inform the Panel as soon as possible by contacting Health-ResearchEthics@salford.ac.uk

Yours sincerely,

Sue McAndrew
Chair of the Research Ethics Panel

Appendix B:

Appendix B-1:

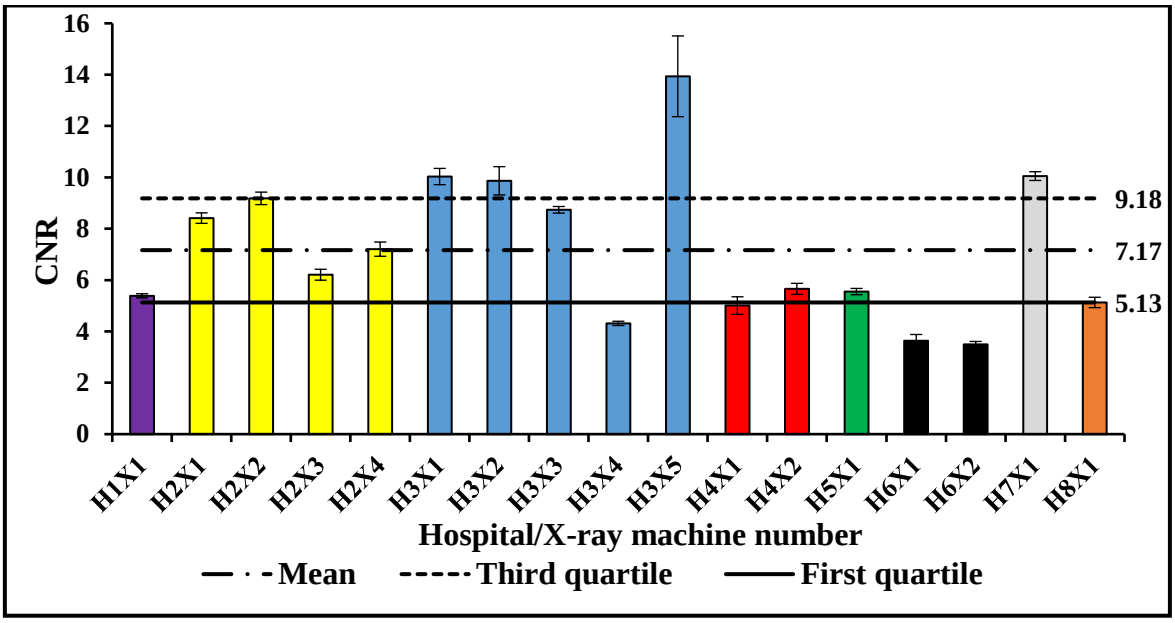

Figure B-1.1. Bar chart displaying the distribution of CNR for the neonate age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CNR values, respectively. The error bars in this chart represent the SD in CNR obtained from the measurements of three repeated images.

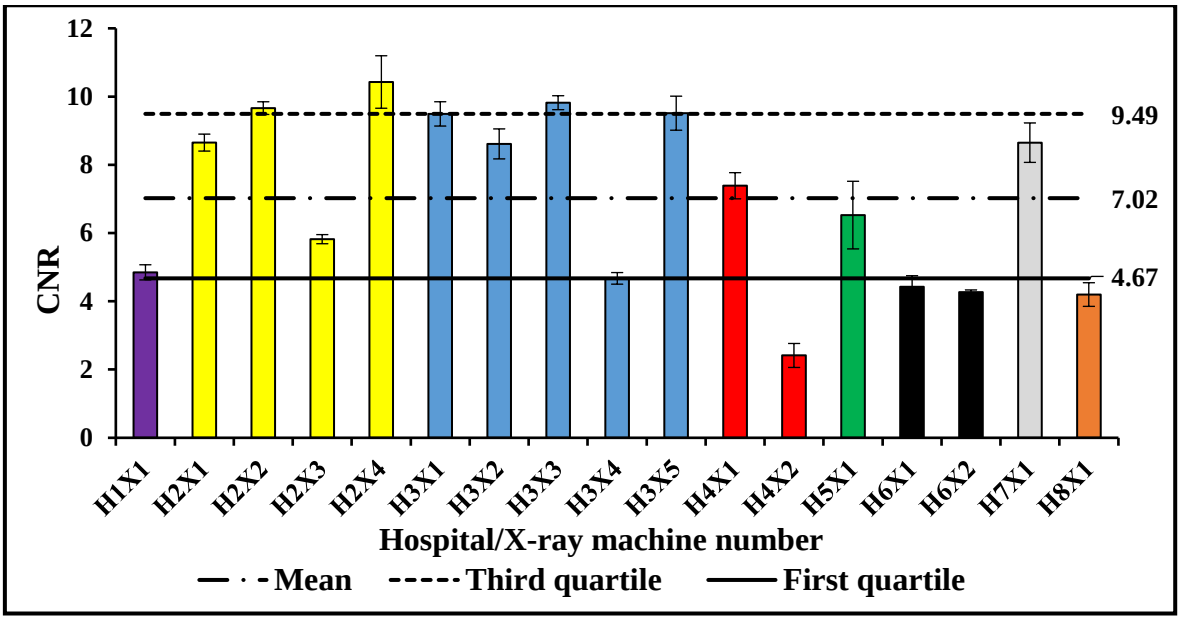

Figure B-1.2. Bar chart displaying the distribution of CNR for the one year age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CNR values, respectively. The error bars in this chart represent the SD in CNR obtained from the measurements of three repeated images.

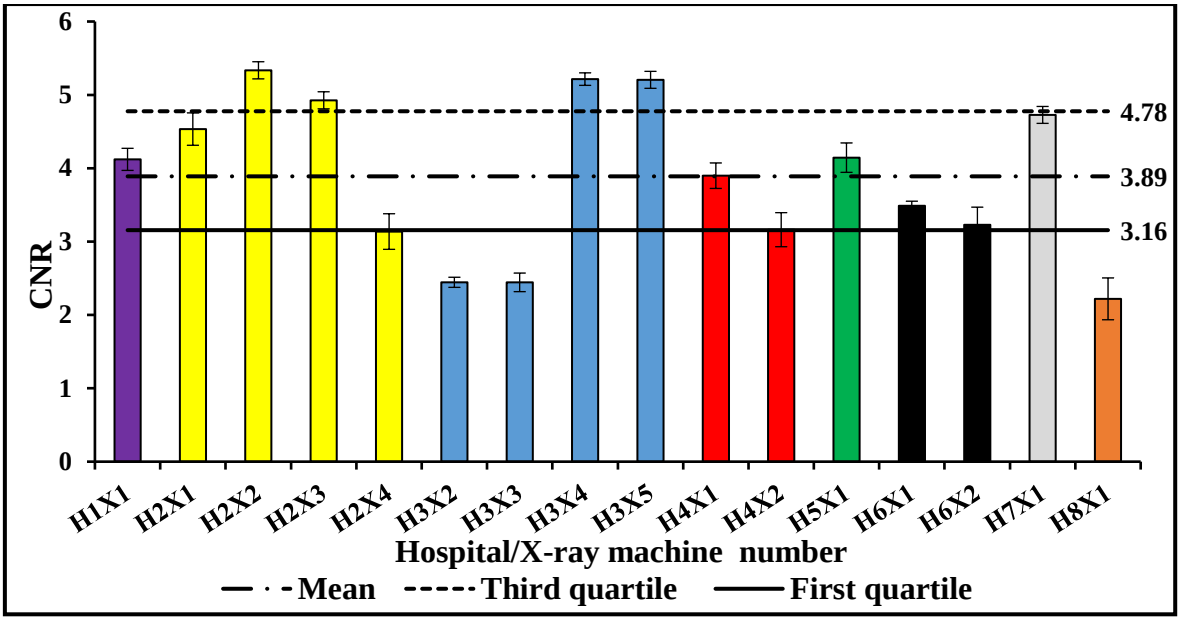

Figure B-1.3. Bar chart displaying the distribution of CNR for the five years age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CNR values, respectively. The error bars in this chart represent the SD in CNR obtained from the measurements of three repeated images.

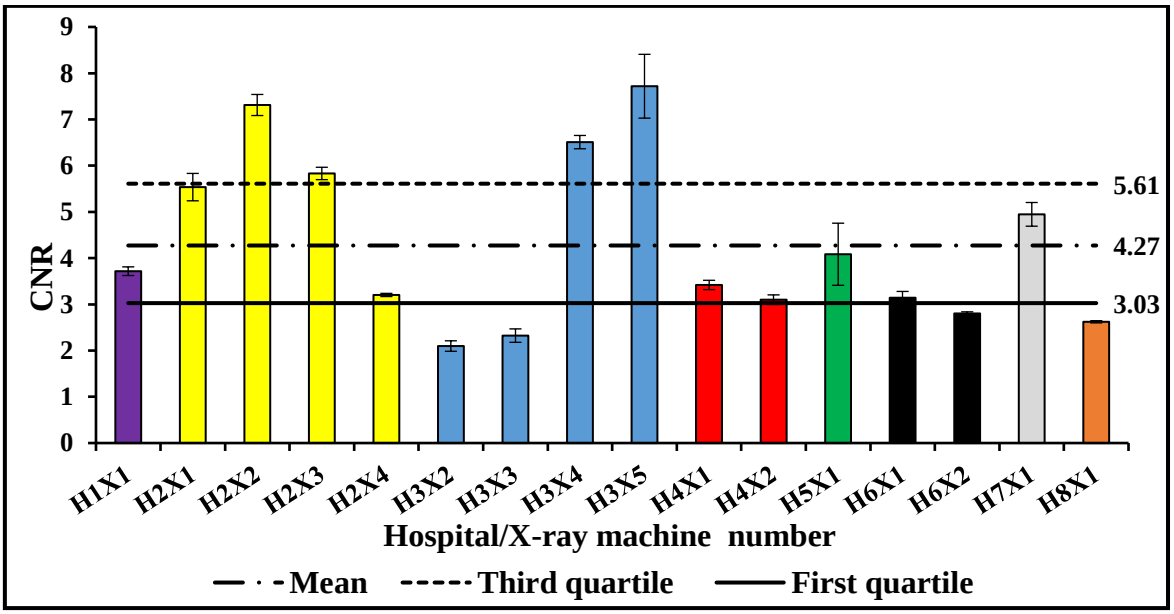

Figure B-1.4. Bar chart displaying the distribution of CNR for the ten years age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CNR values, respectively. The error bars in this chart represent the SD in CNR obtained from the measurements of three repeated images.

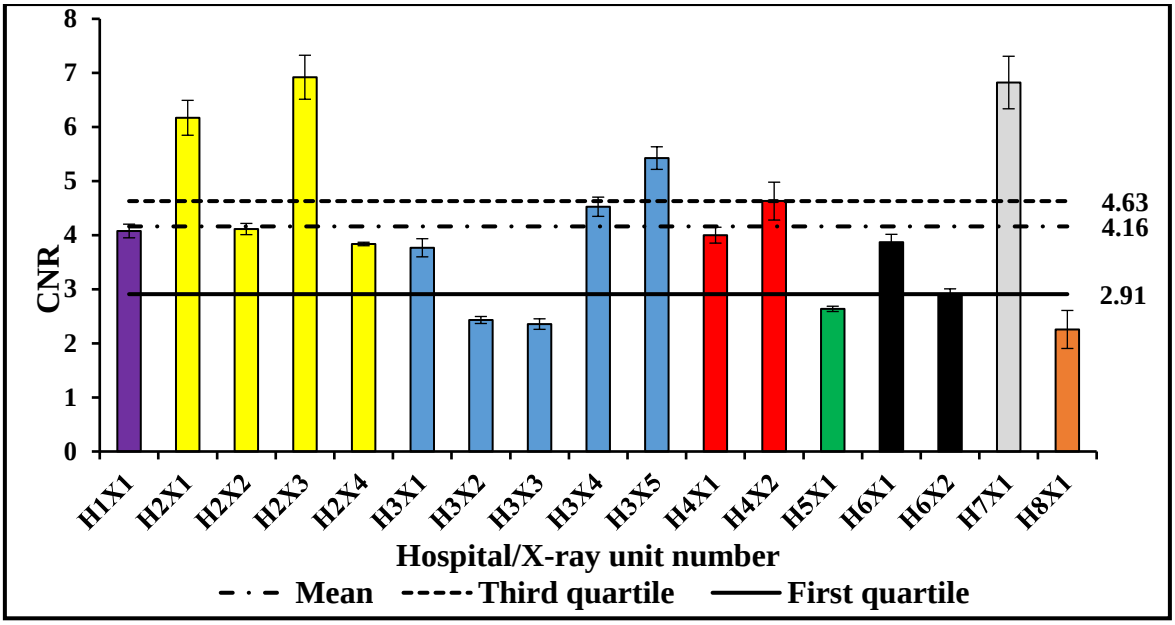

Figure B-1.5. Bar chart displaying the distribution of CNR for the adult age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CNR values, respectively. The error bars in this chart represent the SD in CNR obtained from the measurements of three repeated images.

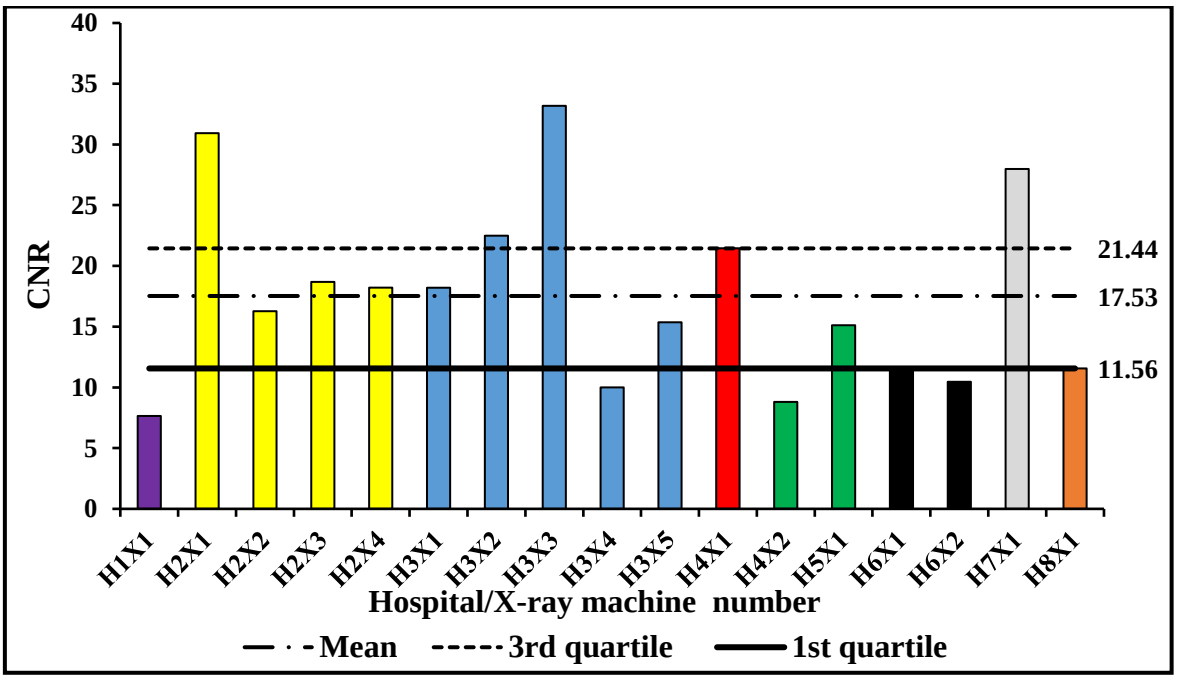

Figure B-1.6. Bar chart displaying the distribution of CNR for the neonate age group using the Gammex phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CNR values, respectively.

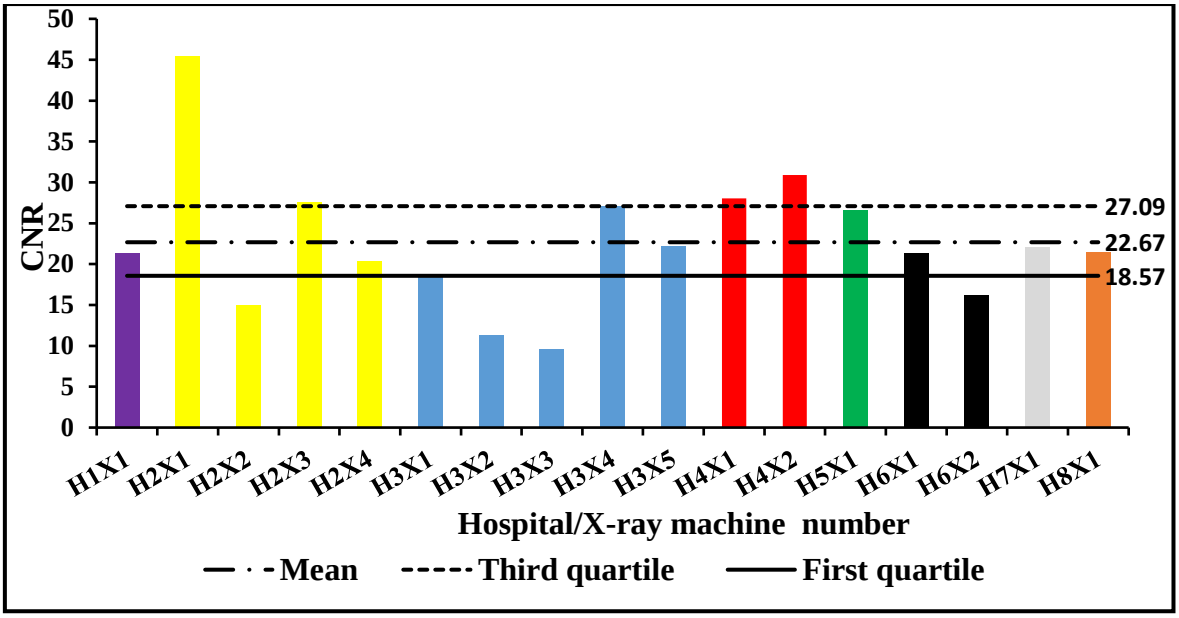

Figure B-1.7. Bar chart displaying the distribution of CNR for the standard size adult age group using the standard size Lungman phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CNR values, respectively.

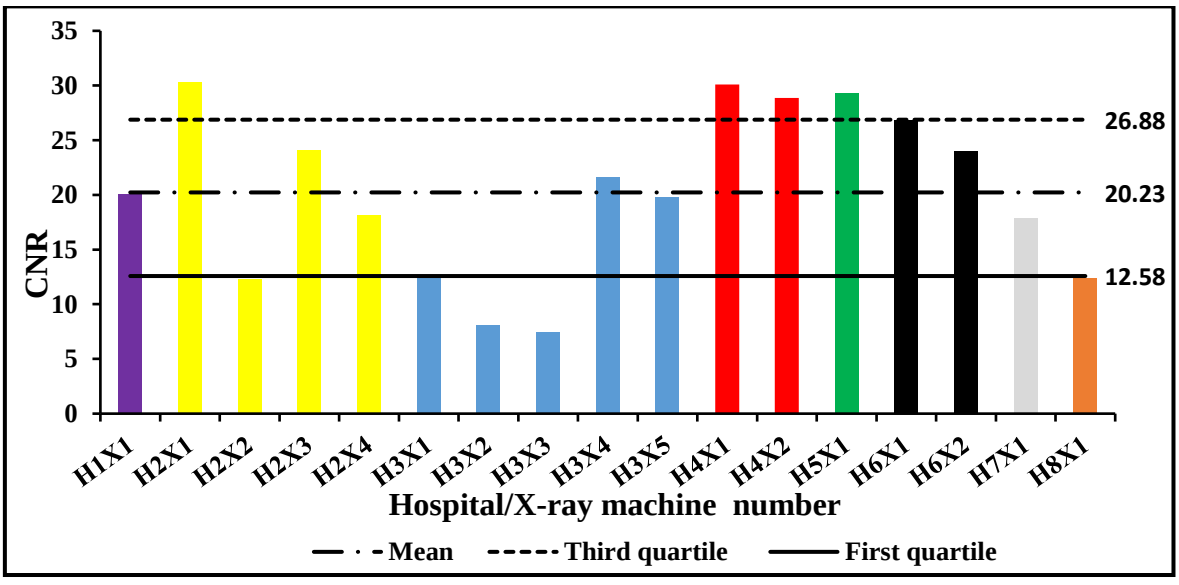

Figure B-1.8. Bar chart displaying the distribution of CNR for the larger size adult age group using the larger size Lungman phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CNR values, respectively.

Appendix B-2:

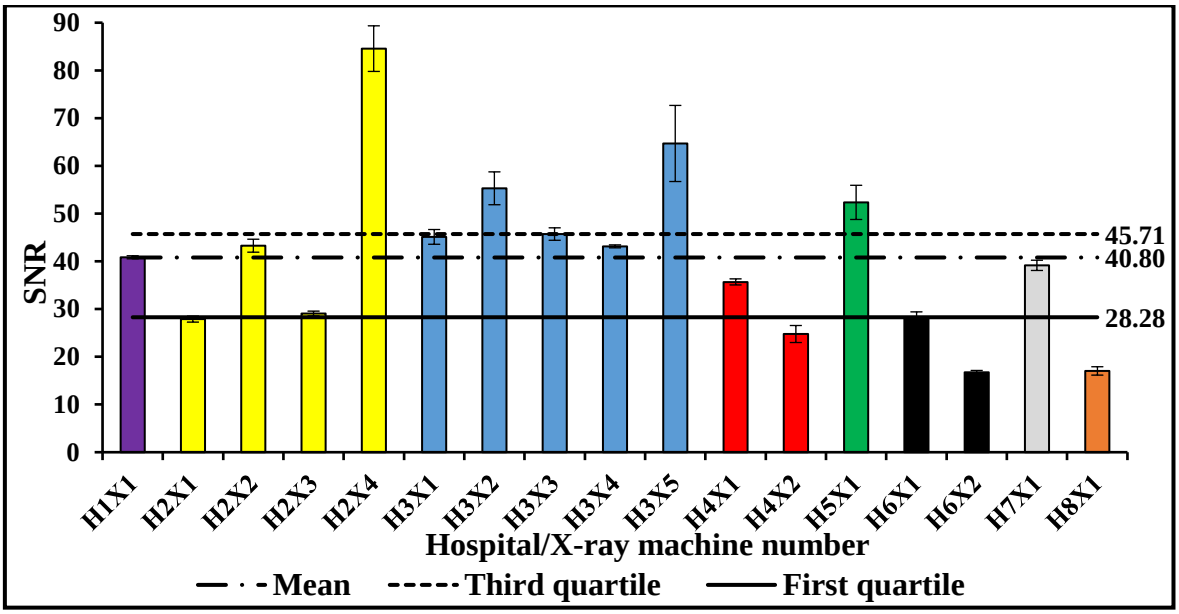

Figure B-2.1. Bar chart displaying the distribution of SNR for the neonate age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the SNR values, respectively. The error bars in this chart represent the SD in SNR obtained from the measurements of three repeated images.

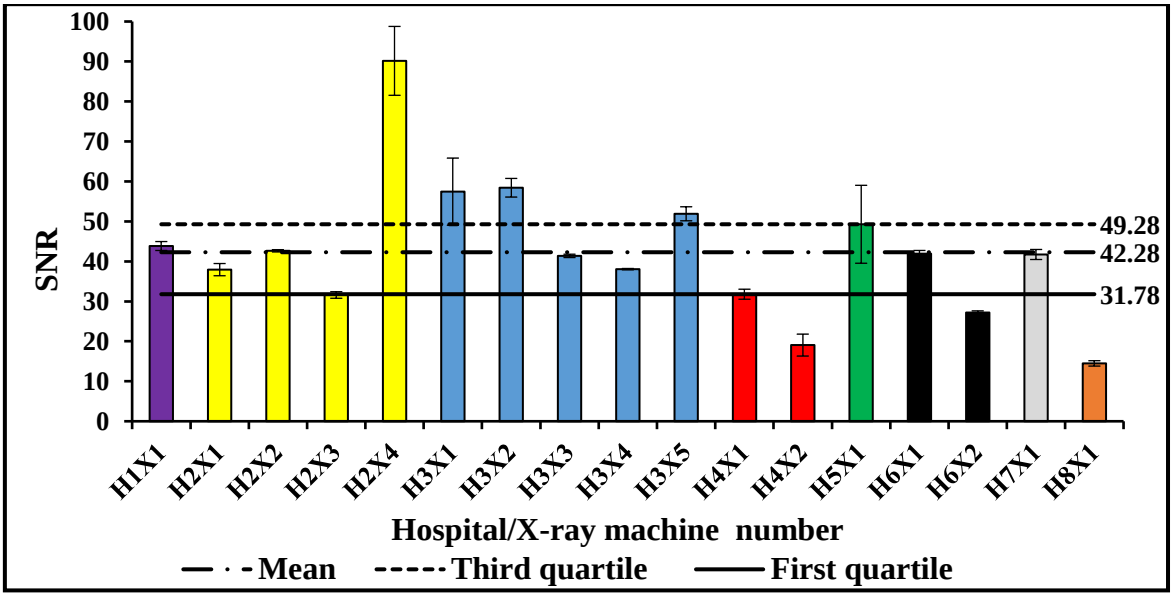

Figure B-2.2. Bar chart displaying the distribution of SNR for the one year age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the SNR values, respectively. The error bars in this chart represent the SD in SNR obtained from the measurements of three repeated images.

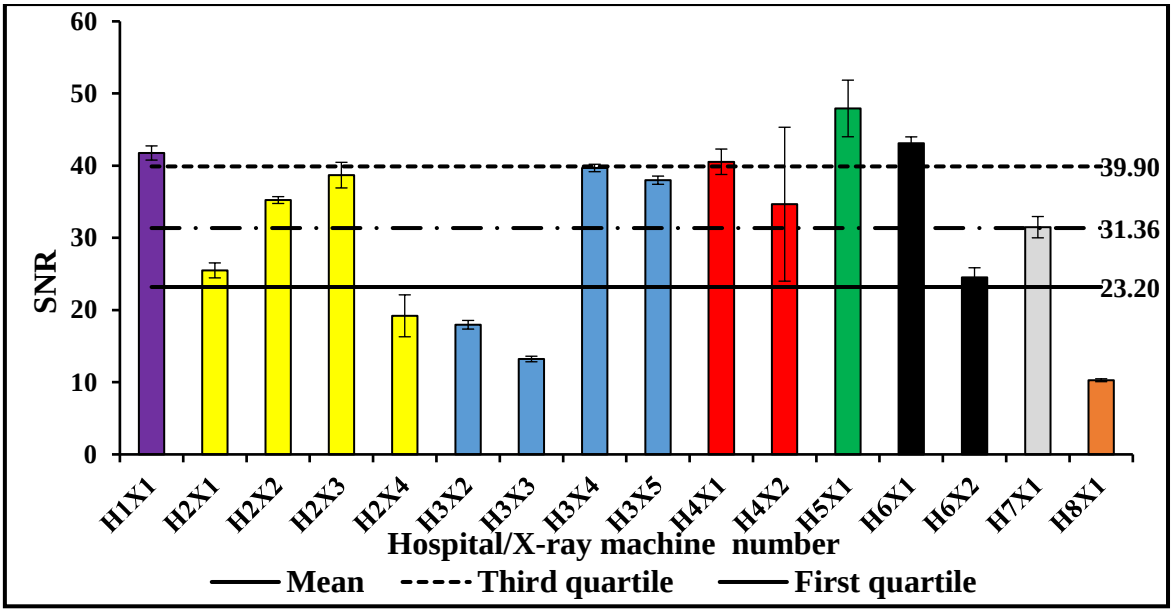

Figure B-2.3. Bar chart displaying the distribution of SNR for the five years age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the SNR values, respectively. The error bars in this chart represent the SD in SNR obtained from the measurements of three repeated images.

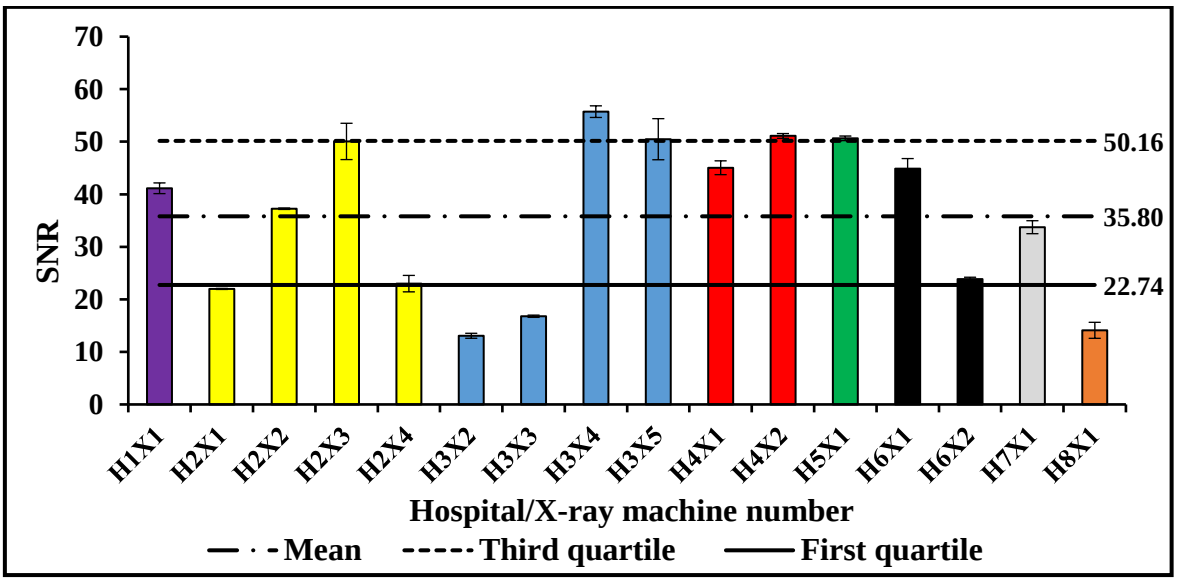

Figure B-2.4. Bar chart displaying the distribution of SNR for the ten years age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the SNR values, respectively. The error bars in this chart represent the SD in SNR obtained from the measurements of three repeated images.

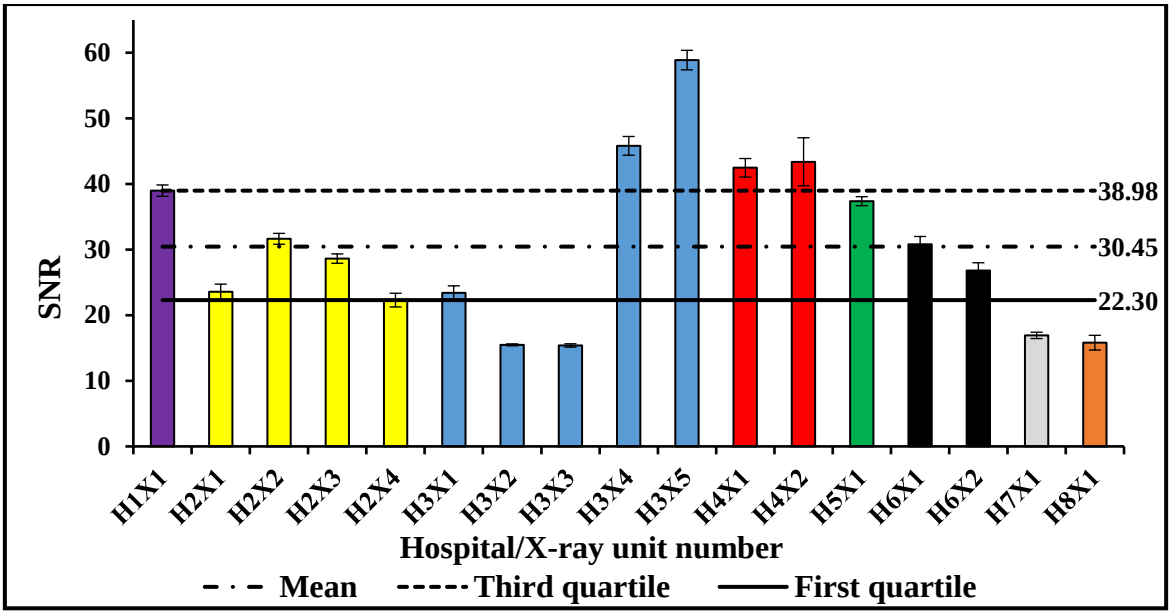

Figure B-2.5. Bar chart displaying the distribution of SNR for the adult age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the SNR values, respectively. The error bars in this chart represent the SD in SNR obtained from the measurements of three repeated images.

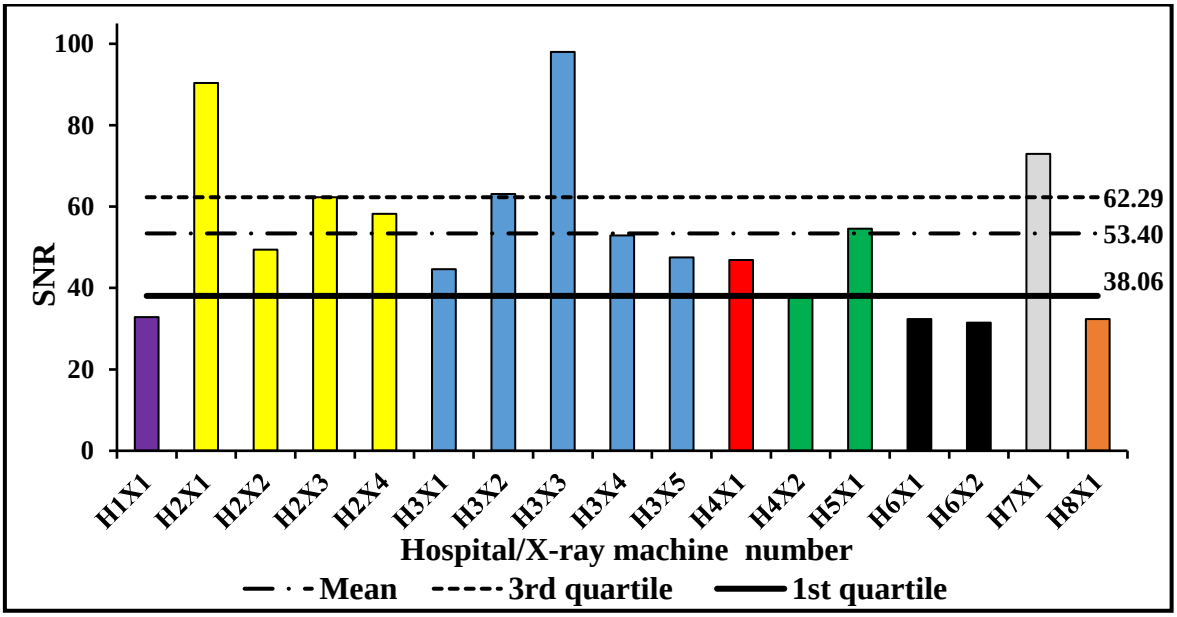

Figure B-2.6. Bar chart displaying the distribution of SNR for the neonate age group using the Gammex phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the SNR values, respectively.

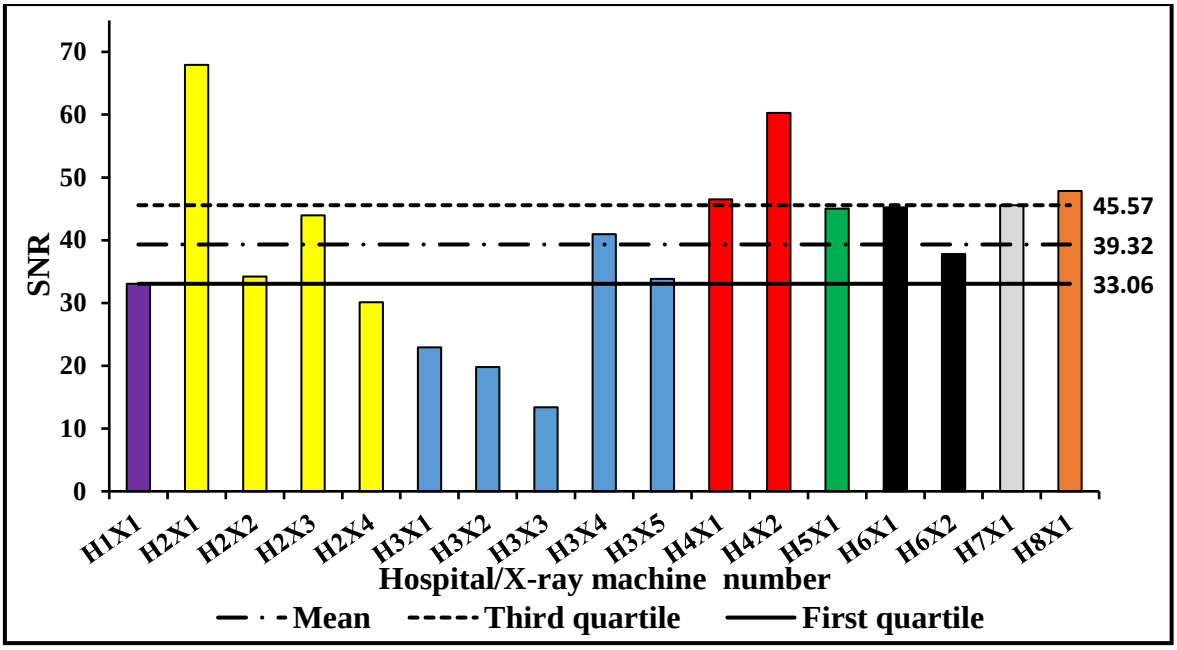

Figure B-2.7. Bar chart displaying the distribution of SNR for the standard size adult age group using the standard size Lungman phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the SNR values, respectively.

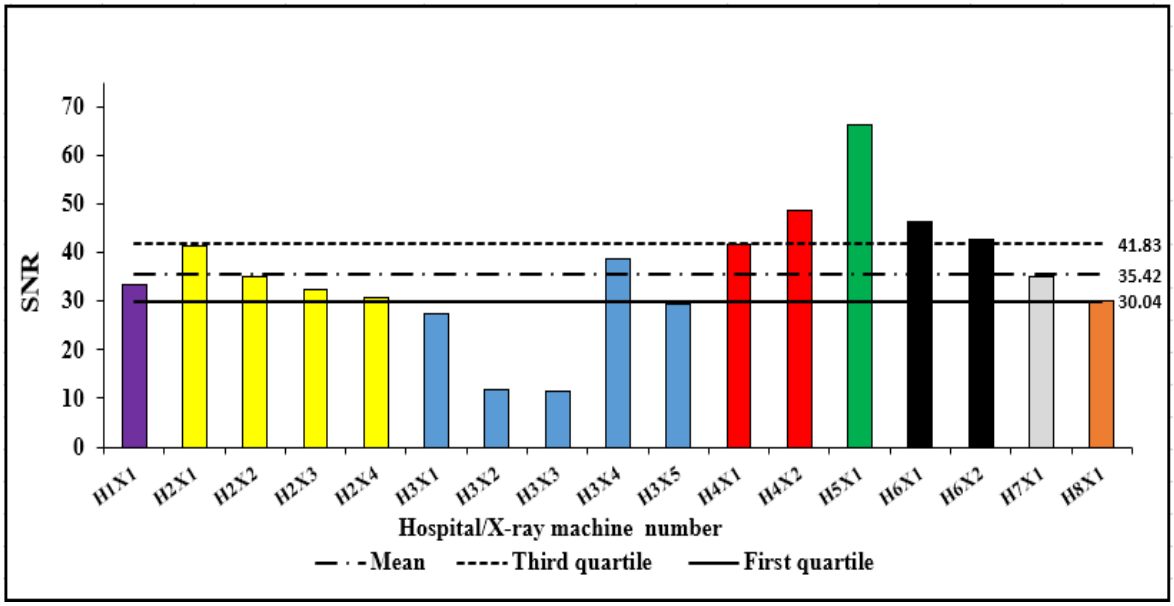

Figure B-2.8. Bar chart displaying the distribution of SNR for the larger size adult age group using the larger size Lungman phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the SNR values, respectively.

Appendix B-3:

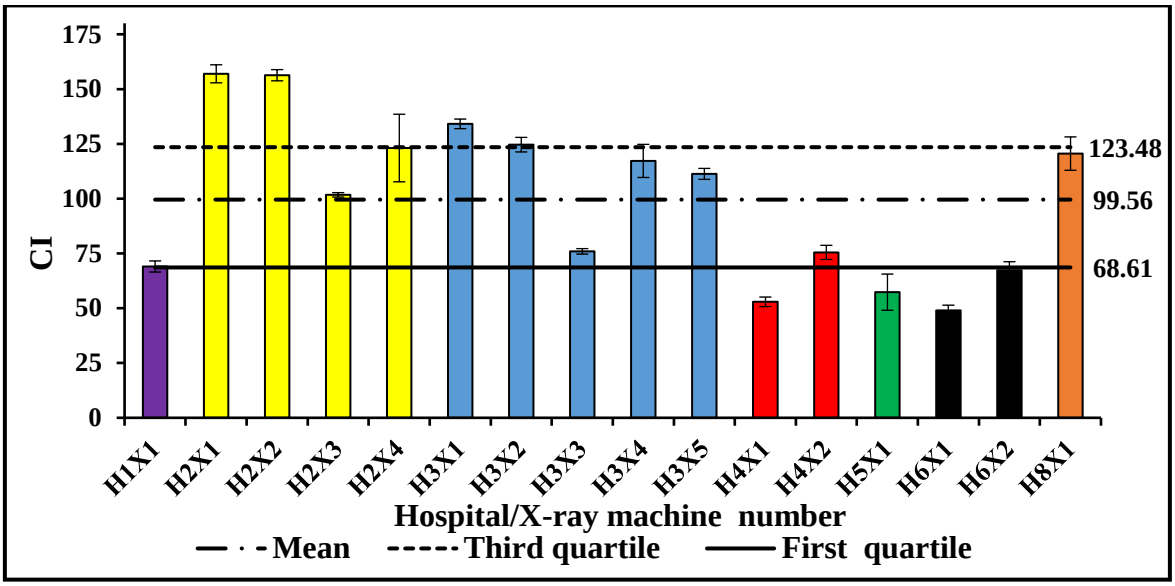

Figure B-3.1. Bar chart displaying the distribution of CI for the neonate age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CI values, respectively. The error bars in this chart represent the SD in CI obtained from the measurements of three repeated images.

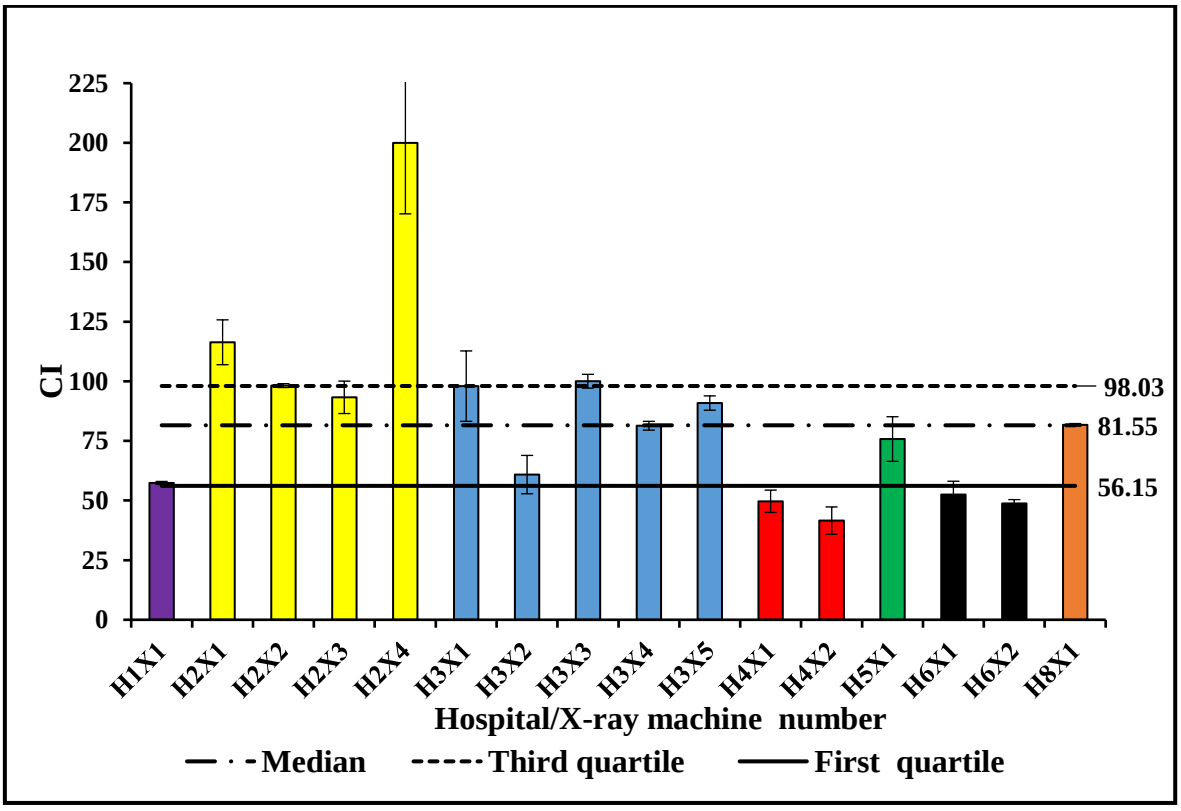

Figure B-3.2. Bar chart displaying the distribution of CI for the one year age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (median) and the solid line (1st quartile) display the CI values, respectively. The error bars in this chart represent the SD in CI obtained from the measurements of three repeated images.

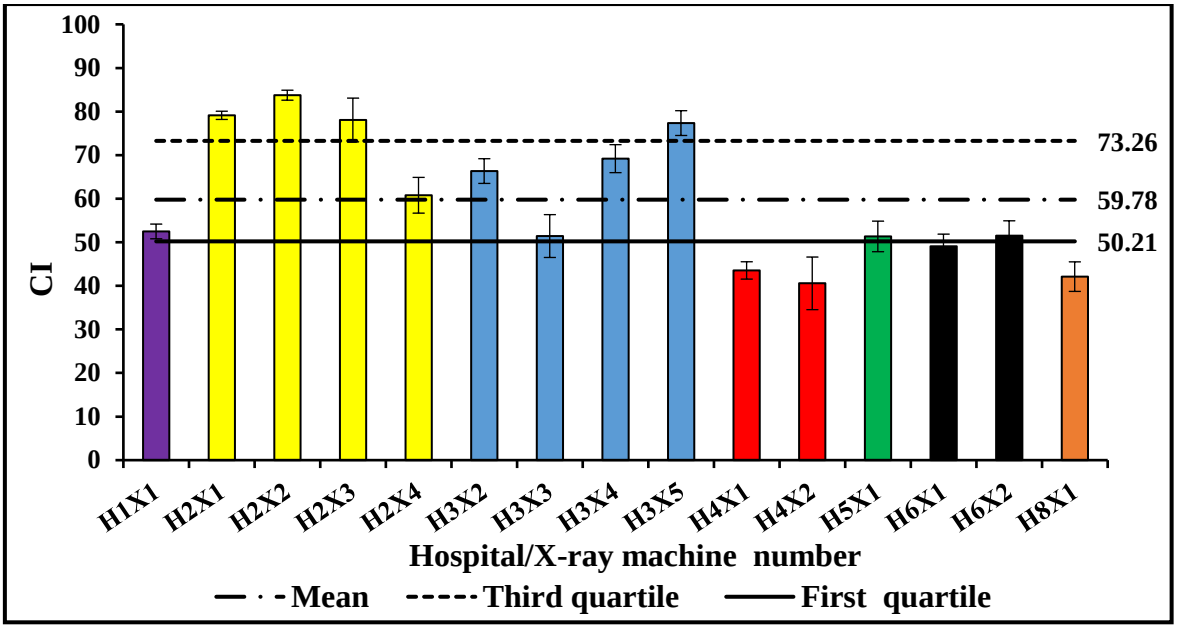

Figure B-3.3. Bar chart displaying the distribution of CI for the five years age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CI values, respectively. The error bars in this chart represent the SD in CI obtained from the measurements of three repeated images.

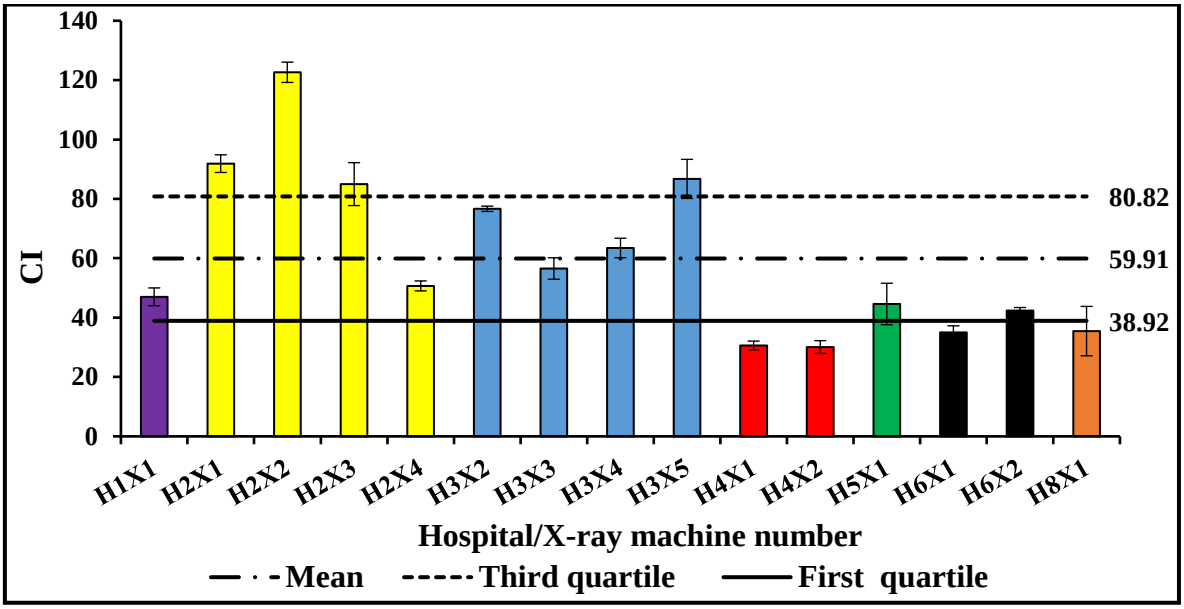

Figure B-3.4. Bar chart displaying the distribution of CI for the ten years age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CI values, respectively. The error bars in this chart represent the SD in CI obtained from the measurements of three repeated images.

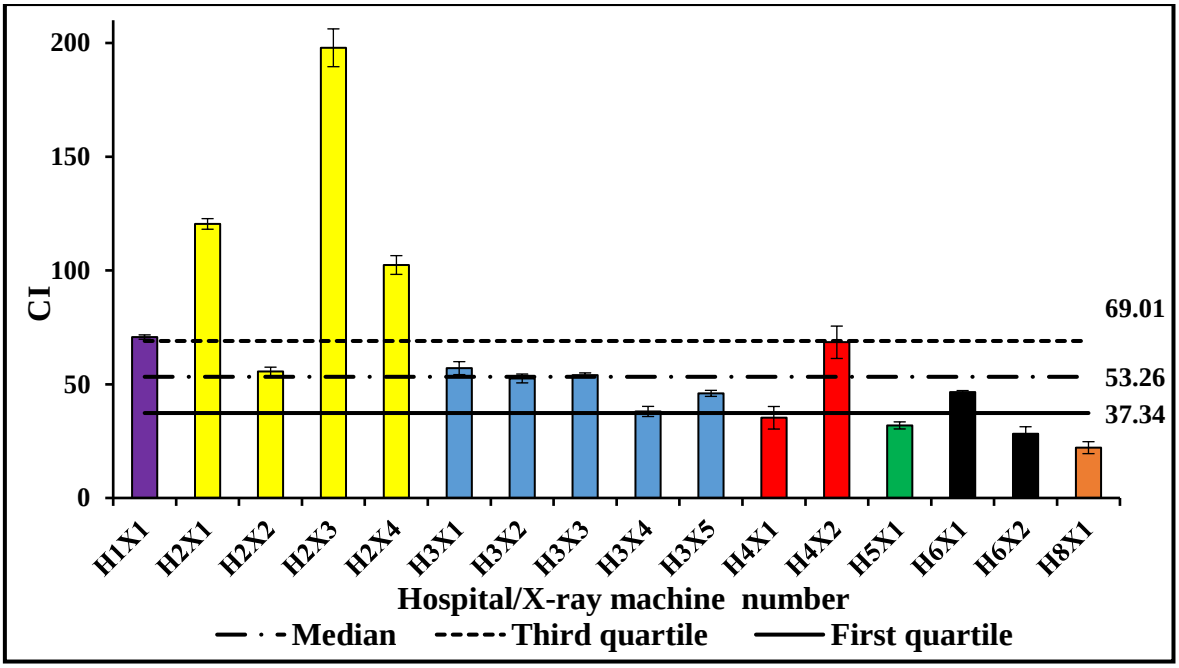

Figure B-3.5. Bar chart displaying the distribution of CI for the adult age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (median) and the solid line (1st quartile) display the CI values, respectively. The error bars in this chart represent the SD in CI obtained from the measurements of three repeated images.

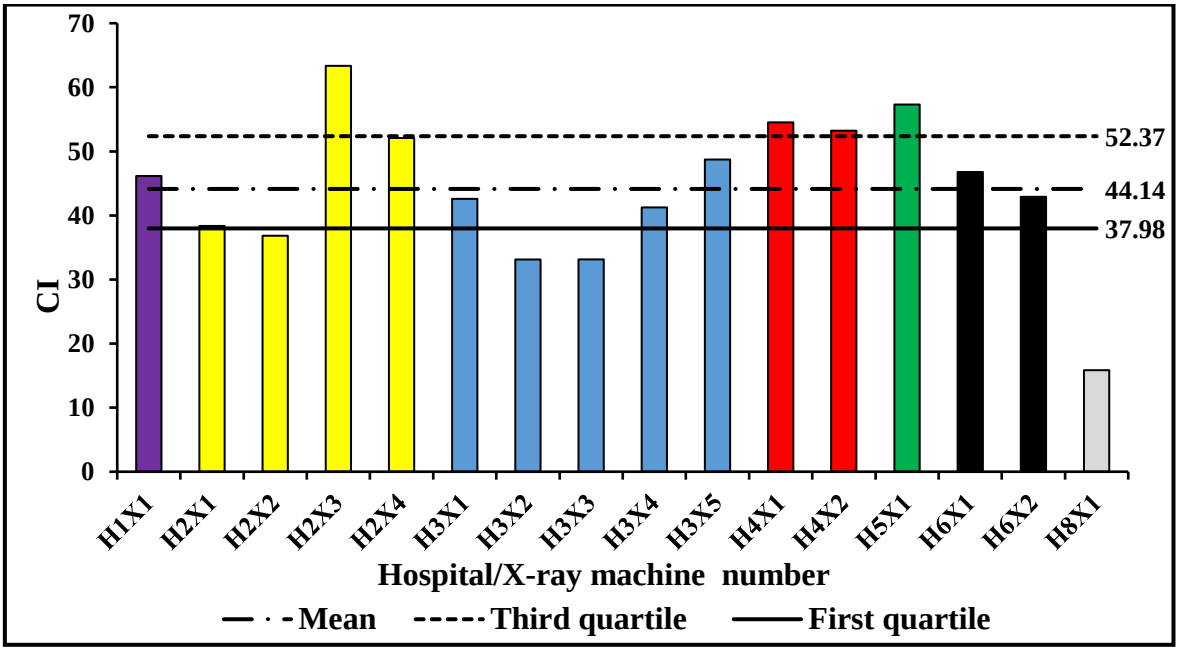

Figure B-3.7. Bar chart displaying the distribution of CI for the standard size adult age group using the standard size Lungman phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CI values, respectively.

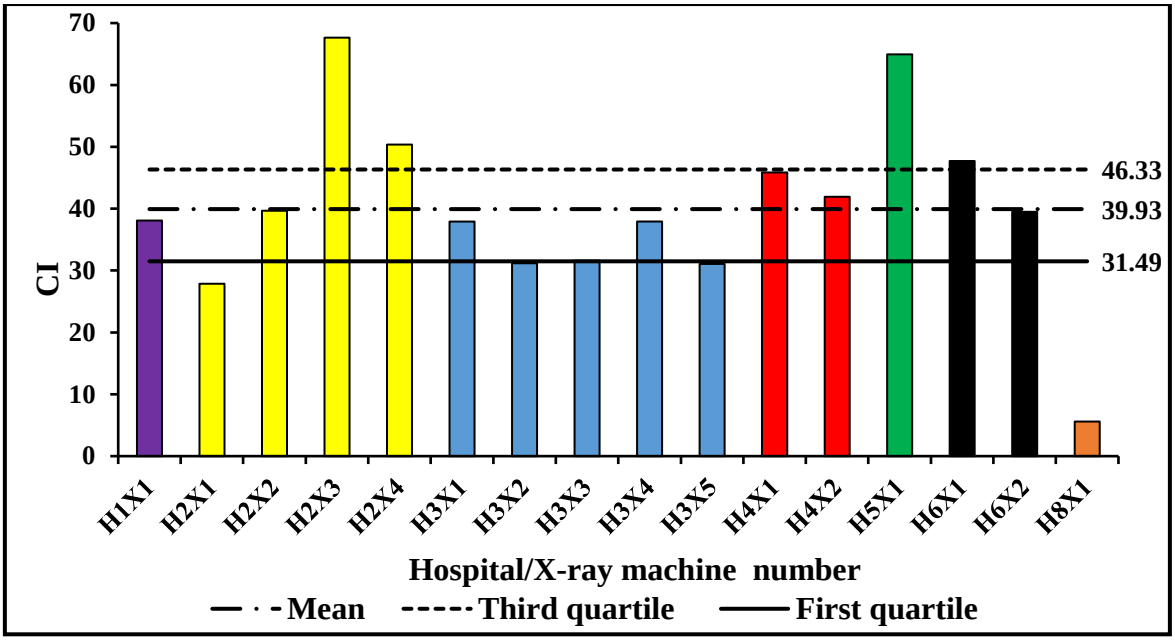

Figure B-3.8. Bar chart displaying the distribution of CI for the larger size adult age group using the larger size Lungman phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the CI values, respectively.

Appendix C:

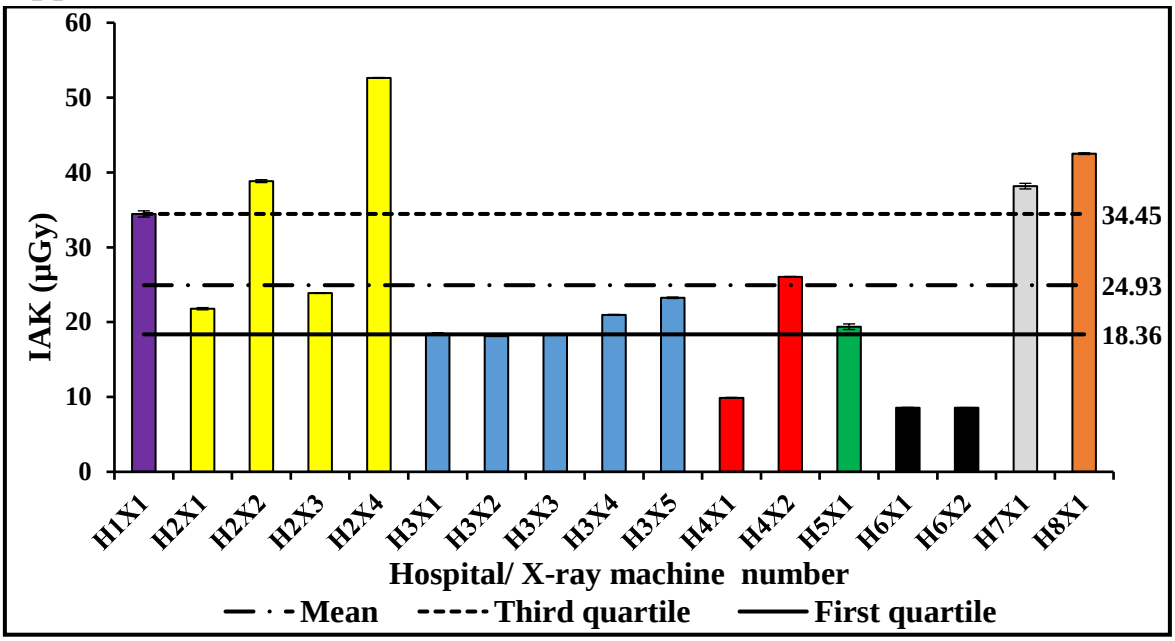

Figure 1. Bar chart displaying the distribution of IAK for the neonate age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the IAK values, respectively. The error bars in this chart represents the SD in IAK obtained from measuring the three repeat radiation exposures.

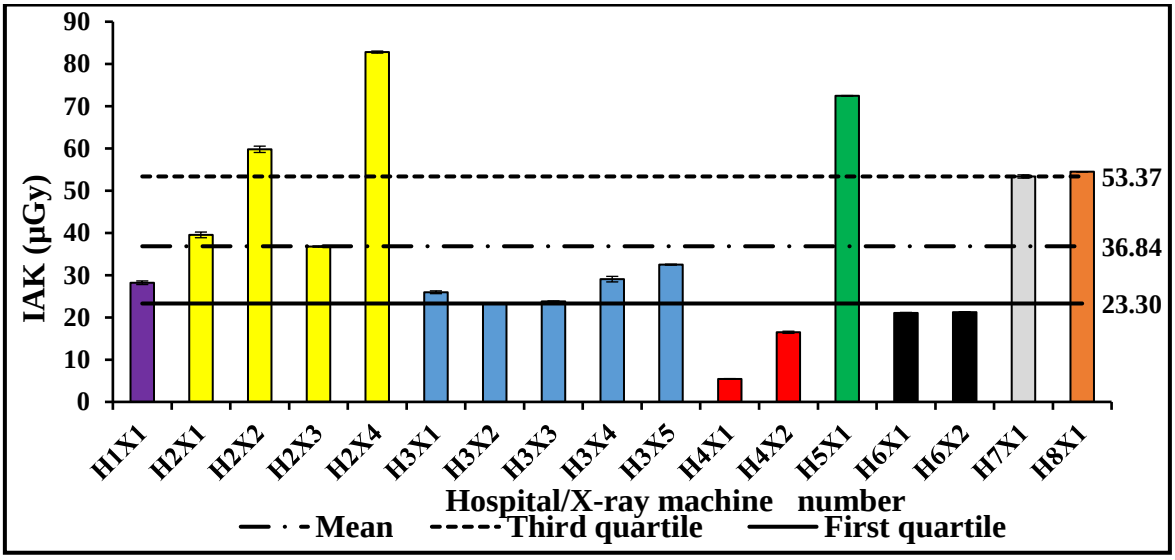

Figure 2. Bar chart displaying the distribution of IAK for the one year age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the IAK values, respectively. The error bars in this chart represents the SD in IAK obtained from measuring the three repeat radiation exposures.

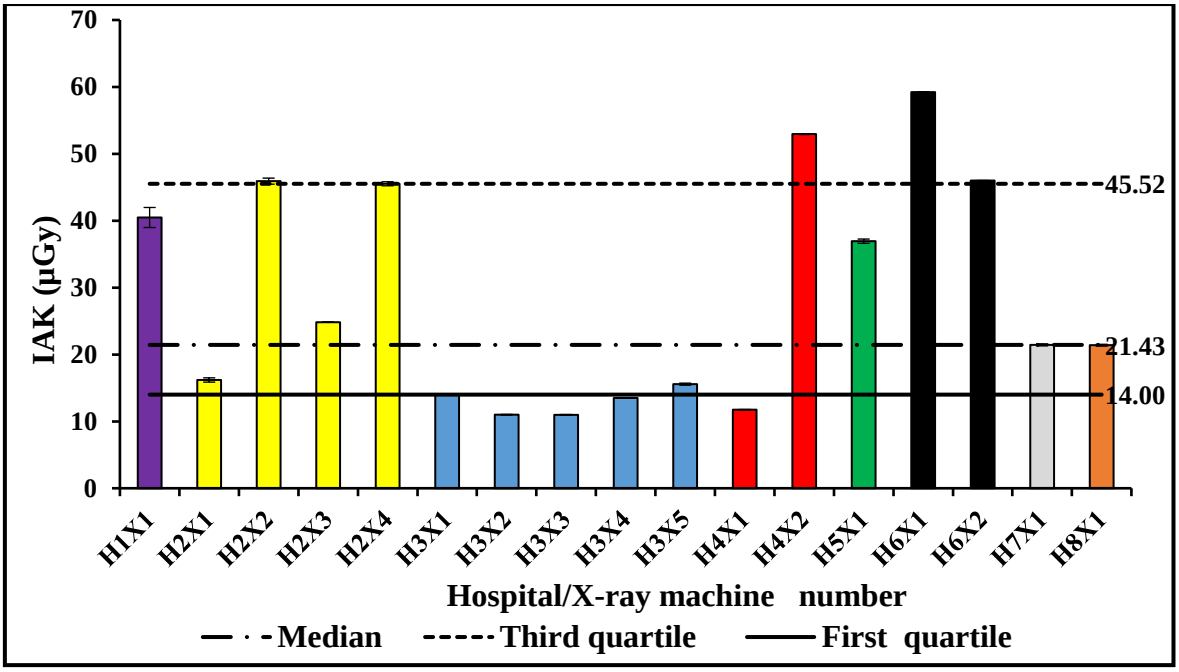

Figure 3. Bar chart displaying the distribution of IAK for the five years age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (median) and the solid line (1st quartile) display the IAK values, respectively. The error bars in this chart represents the SD in IAK obtained from measuring the three repeat radiation exposures.

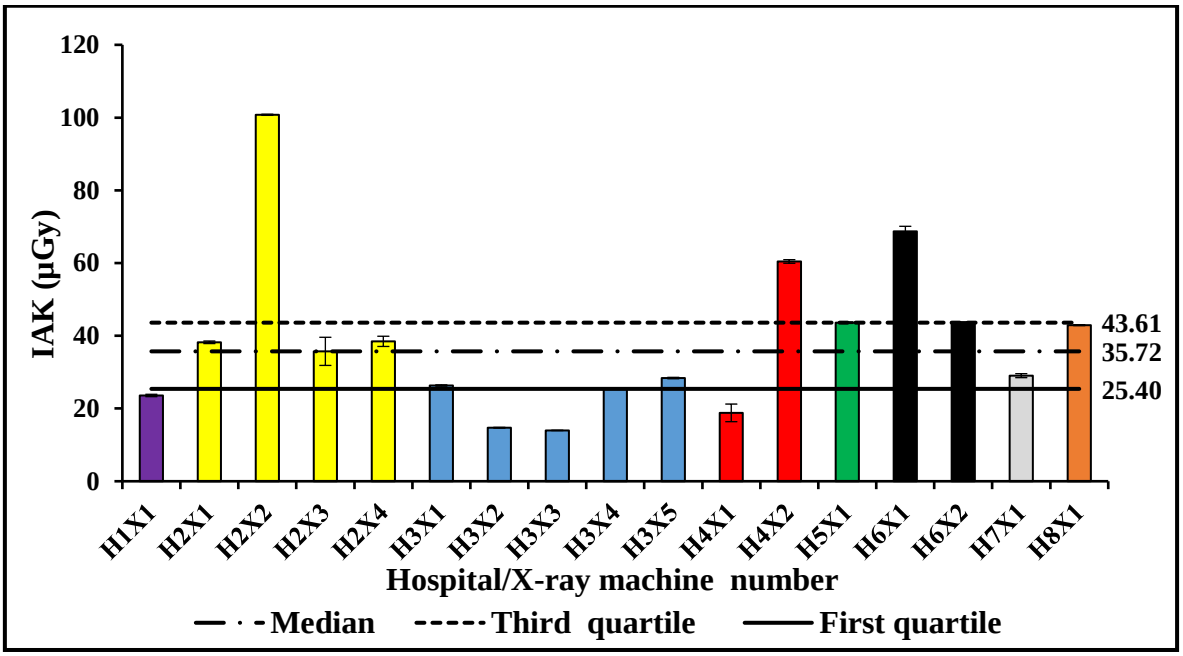

Figure 4. Bar chart displaying the distribution of IAK for the ten years age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (median) and the solid line (1st quartile) display the IAK values, respectively. The error bars in this chart represents the SD in IAK obtained from measuring the three repeat radiation exposures.

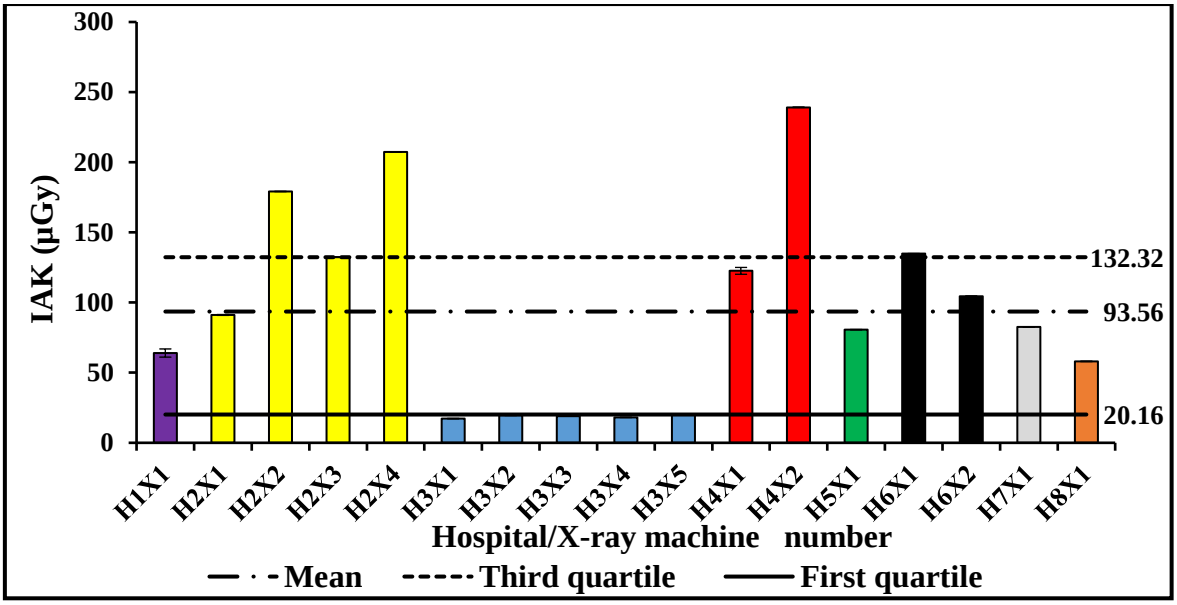

Figure 5. Bar chart displaying the distribution of IAK for the adult age group using the CDRAD 2.0 phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the IAK values, respectively. The error bars in this chart represents the SD in IAK obtained from measuring the three repeat radiation exposures.

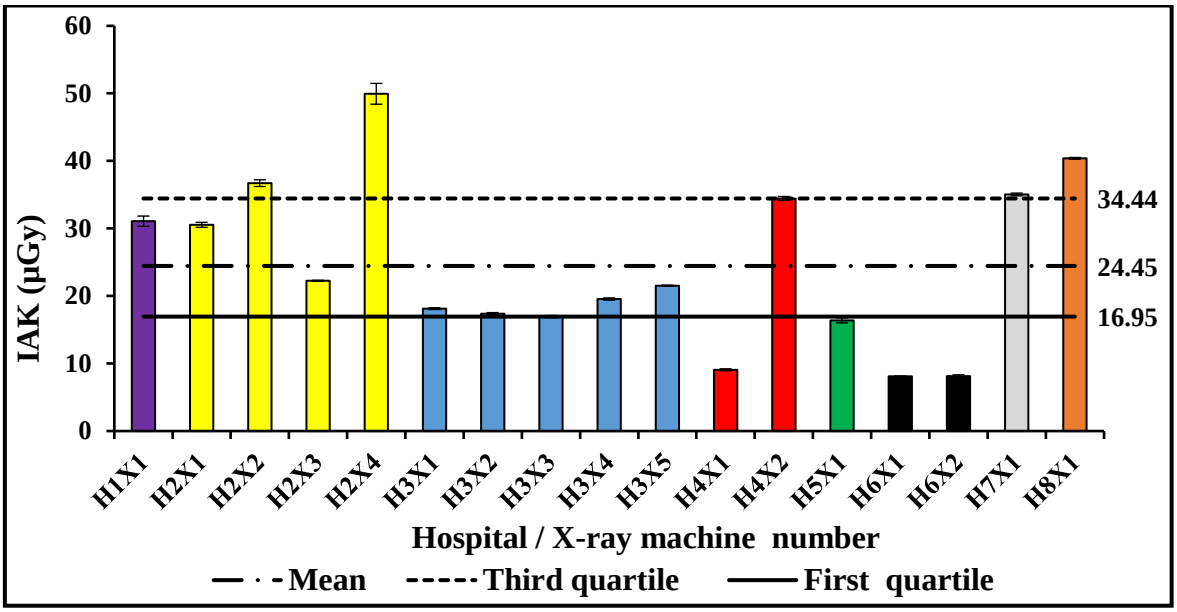

Figure 6. Bar chart displaying the distribution of IAK for the neonate age group using the Gammex phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the IAK values, respectively. The error bars in this chart represents the SD in IAK obtained from measuring the three repeat radiation exposures.

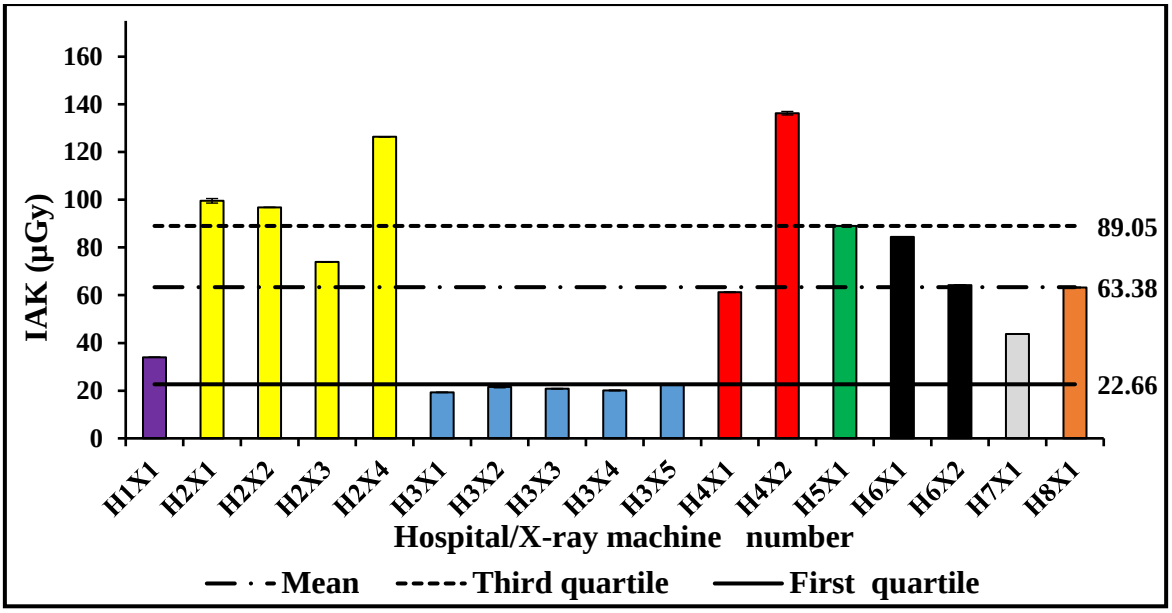

Figure 7. Bar chart displaying the distribution of IAK for the standard size adult age group using the standard size Lungman phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the IAK values, respectively. The error bars in this chart represents the SD in IAK obtained from measuring the three repeat radiation exposures.

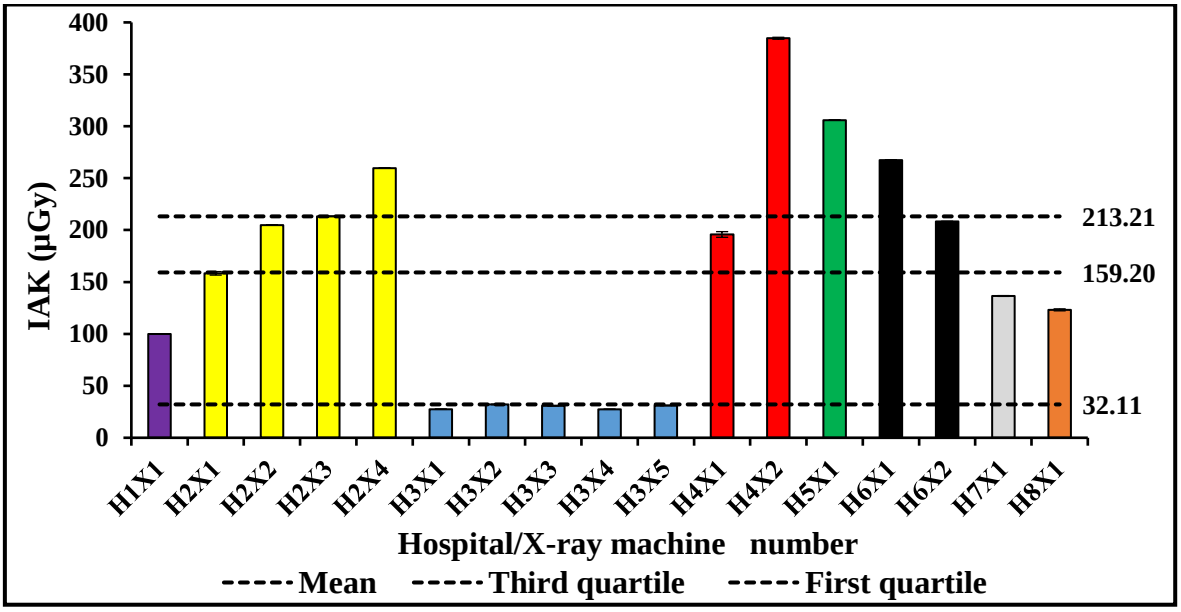

Figure 8. Bar chart displaying the distribution of IAK for the larger size adult age group using the larger size Lungman phantom. The dashed line (3rd quartile), the dashed dotted line (average, mean) and the solid line (1st quartile) display the IAK values, respectively. The error bars in this chart represents the SD in IAK obtained from measuring the three repeat radiation exposures.

Appendix D:

D-1. This table shows the test of normality values of radiation dose and different types of physical IQ resulted from CDRAD 2.0 phantom based on the Shapiro-Wilk test.

Age group	The normality (p) values based on the Shapiro-Wilk test				
	IAK	IQF_{inv}	SNR	CNR	CI
Neonate	0.13	0.62	0.42	0.26	0.21
1-year	0.65	0.57	0.09	0.16	0.008
5-year	0.04	0.74	0.32	0.29	0.116
10-year	0.02	0.01	0.11	0.16	0.111
Adult	0.10	0.27	0.44	0.24	0.001

D-2. This table shows the normality values of radiation dose and different types of IQ resulted from Gammex phantom and the standard size and larger size Lungman phantom based on the Shapiro-Wilk test.

Phantom type	The normality (p) values based on the Shapiro-Wilk test					
	IAK	IQ	LV	SNR	CNR	CI
Gammex	0.28	0.61	-	0.12	0.29	-
Standard size Lungman	0.10	0.92	0.10	0.81	0.23	0.74
Large size Lungman	0.15	0.23	0.43	0.40	0.23	0.24

References

Agarwal, V., & Newbery, S. (2016). Paediatric radiation dose measurements undergoing X-ray examination in Tasmania. *Radiation Protection Dosimetry*, 1–5. http://doi.org/10.1093/rpd/ncw143

Aichinger, H., Dierker, J., Joite-Barfuß, S., & Säbel, M. (2012). *Radiation Exposure and Image Quality in X-Ray Diagnostic Radiology*. *Berlin: Springer- Verlag*. Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-11241-6

Akpochafor, M., Omojola, A., Adeneye, S., Aweda, M., & Ajayi, H. (2016). Determination of reference dose levels among selected X-ray centers in Lagos State, South-West Nigeria. *Journal of Clinical Sciences*, *13*(4), 167. http://doi.org/10.4103/2468-6859.192275

Alghamdi, S. S. (2016). *Radiologic Image Assessment Using Information Loss Theory by Specially Designed Low Contrast Detail Phantoms and Extending it to CT.*

Allen, E., Hogg, P., Ma, W. K., & Szczepura, K. (2013). Fact or fiction: An analysis of the 10 kVp 'rule' in computed radiography. *Radiography*, *19*(3), 223–227. http://doi.org/10.1016/j.radi.2013.05.003

Almén, A., Lööf, M., & Mattsson, S. (1996). Examination Technique, Image Quality, and Patient Dose in Paediatric Radiology. *Acta Radiologica*, *37*(3), 337–342. http://doi.org/10.3109/02841859609177662

Almen, A., Tingberg, A., Mattsson, S., Besjakov, J., Kheddache, S., Lanhede, B., … Zankl, M. (2000). The influence of different technique factors on image quality of lumbar spine radiographs as evaluated by established CEC image criteria. *British Journal of Radiology*, *73*(875), 1192–1199.

Alsleem, H. A. (2014). *Evaluation of Factors that Affect Contrast-Detail in Digital X-Ray and Computed Tomography A thesis submitted in fulfilment of the requirements for the degree of PhD of Medical Radiations Science*.

Alsleem, H., U, P., Mong, K. S., & Davidson, R. (2014). Effects of radiographic techniques on the low-contrast detail detectability performance of digital radiography systems. *Radiologic Technology*, *85*(6), 614–22.

Alzimami, K., Sassi, S., Alkhorayef, M., Britten, A. J., & Spyrou, N. M. (2009). Optimisation of computed radiography systems for chest imaging. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *600*(2), 513–518. http://doi.org/10.1016/j.nima.2008.12.023

American College of Radiology. (2014). *ACR–SPR practice parameter for the performance of chest radiography*.

Armpilia, C. I., Fife, I. A. J., & Croasdale, P. L. (2001). Radiation doses to neonates and issues of radiation protection in a special care baby unit. *Adiological Protection of Patients in Diagnostic and Interventional Radiology, Nuclear Medicine and Radiotherapy*.

Aroua, A., Rickli, H., Stauffer, J. C., Schnyder, P., Trueb, P. R., Valley, J. F., … Verdun, F. R. (2007). How to set up and apply reference levels in fluoroscopy at a national level. *European Radiology*, *17*(6), 1621–1633. http://doi.org/10.1007/s00330-006-0463-3

Audun, S., Emmanuel, A., Burke, A., Susanne, D., Julie, F., Sabine, S., … Soares Flavio. (2017). Analysis of image quality and effective dose in adult chest phantom radiography with high BMI. In *(Eds.) A. Meijer, C. Buissink and P. Hogg .Radiation dose and image quality optimisation in medical imaging, Erasmus Intensive Programme OPTIMAX#2017*.

Aufrichtig, R. (1999). Comparison of low contrast detectability between a digital amorphous silicon and a screen-film based imaging system for thoracic radiography. *Medical Physics*, *26*(7), 1349–1358. http://doi.org/10.1118/1.598630

Aufrichtig, R., & Xue, P. (2000). Dose efficiency and low-contrast detectability of an amorphous silicon x-ray detector for digital radiography. *Physics in Medicine and Biology*, *45*(9), 2653–2669. http://doi.org/10.1088/0031-9155/45/9/316

Bacher, K., Smeets, P., Bonnarens, K., De Hauwere, A., Verstraete, K., & Thierens, H. (2003). Dose Reduction in Patients Undergoing Chest Imaging: Digital Amorphous Silicon Flat-Panel Detector Radiography Versus Conventional Film-Screen Radiography and Phosphor-Based Computed Radiography. *American Journal of Roentgenology*, *181*(4), 923–929. http://doi.org/10.2214/ajr.181.4.1810923

Bacher, K., Smeets, P., De Hauwere, A., Voet, T., Duyck, P., Verstraete, K., & Thierens, H. (2006). Image quality performance of liquid crystal display systems: Influence of display resolution, magnification and window settings on contrast-detail detection. *European Journal of Radiology*, *58*(3), 471–479. http://doi.org/10.1016/j.ejrad.2005.12.016

Bacher, K., Smeets, P., Vereecken, L., De Hauwere, A., Duyck, P., De Man, R., … Thierens, H. (2006). Image Quality and Radiation Dose on Digital Chest Imaging: Comparison of Amorphous Silicon and Amorphous Selenium Flat-Panel Systems. *American Journal of Roentgenology*, *187*(3), 630–637. http://doi.org/10.2214/AJR.05.0400

Barrett, H. H., Myers, K. J., & Rathee, S. (2004). Foundations of Image Science. *Medical Physics*, *31*(4), 953. http://doi.org/10.1118/1.1677252

Bath, M. (2010). Evaluating imaging systems: practical applications. *Radiation Protection Dosimetry*, *139*(1–3), 26–36. http://doi.org/10.1093/rpd/ncq007

Båth, M., Håkansson, M., Börjesson, S., Kheddache, S., Grahn, A., Bochud, F. O., … Månsson, L. G. (2005). Nodule detection in digital chest radiography: part of image background acting as pure noise. *Radiation Protection Dosimetry*, *114*(1–3), 102–108. http://doi.org/10.1093/rpd/nch523

Båth, M., Håkansson, M., Börjesson, S., Kheddache, S., Grahn, A., Ruschin, M., … Månsson, L. G. (2005). Nodule detection in digital chest radiography: Introduction to the radius chest trial. *Radiation Protection Dosimetry*, *114*(1–3), 85–91. http://doi.org/10.1093/rpd/nch575

Båth, M., & Månsson, L. G. (2007). Visual grading characteristics (VGC) analysis: A non-parametric rank-invariant statistical method for image quality evaluation. *British Journal of Radiology*, *80*(951), 169–176. http://doi.org/10.1259/bjr/35012658

Berbaum, K. S., Dorfman, D. D., Franken, E. A., & Caldwell, R. T. (2000). Proper ROC analysis and joint ROC analysis of the satisfaction of search effect in chest radiology. *Academic Radiology*, *7*(11), 945–958. http://doi.org/10.1016/S1076-6332(00)80176-2

Berbaum, K. S., Franken, E. A., Dorfman, D. D., Caldwell, R. T., & Krupinski, E. A. (2000). Role of faulty decision making in the satisfaction of search effect in chest radiography. *Academic Radiology*, *7*(12), 1098–1106. http://doi.org/10.1016/S1076-

6332(00)80063-X

Bertolini, M., Nitrosi, A., Rivetti, S., Lanconelli, N., Pattacini, P., Ginocchi, V., & Iori, M. (2012). A comparison of digital radiography systems in terms of effective detective quantum efficiency. *Medical Physics*, *39*(5), 2617. http://doi.org/10.1118/1.4704500

Billinger, J., Nowotny, R., & Homolka, P. (2010a). Diagnostic reference levels in pediatric radiology in Austria. *European Radiology*, *20*(7), 1572–1579. http://doi.org/10.1007/s00330-009-1697-7

Billinger, J., Nowotny, R., & Homolka, P. (2010b). Diagnostic reference levels in pediatric radiology in Austria. *European Radiology*, *20*(7), 1572–1579. http://doi.org/10.1007/s00330-009-1697-7

Borg, M., Badr, I., & Royle, G. J. (2012). The use of a figure-of-merit (FOM) for optimisation in digital mammography: a literature review. *Radiation Protection Dosimetry*, *151*(1), 81–88. http://doi.org/10.1093/rpd/ncr465

Bosmans, H., Carton, A. K., Rogge, F., Zanca, F., Jacobs, J., Van Ongeval, C., … Marchal, G. (2006). Image quality measurements and metrics in full field digital mammography: An overview. *Radiation Protection Dosimetry*, *117*(1–3), 120–130. http://doi.org/10.1093/rpd/nci711

Bosmans, H., Nens, J., Delzenne, L., Marshall, N., Pauwels, H., De Wever, W., & Oyen, R. (2012a). Exploration of exposure conditions with a novel wireless detector for bedside digital radiography. In N. J. Pelc, R. M. Nishikawa, & B. R. Whiting (Eds.), *Proceedings of International Conference "Medical Physics in the Baltic States"* (Vol. 8313, p. 83135K). http://doi.org/10.1117/12.911588

Bosmans, H., Nens, J., Delzenne, L., Marshall, N., Pauwels, H., De Wever, W., & Oyen, R. (2012b). Exploration of exposure conditions with a novel wireless detector for bedside digital radiography. In N. J. Pelc, R. M. Nishikawa, & B. R. Whiting (Eds.), *Physics in Medicine and Biology* (Vol. 53, p. 83135K). http://doi.org/10.1117/12.911588

Bourne, R. (2010). *Fundamentals of Digital Imaging in Medicine*. London: Springer London. http://doi.org/10.1007/978-1-84882-087-6

Bourne, R., & Kagadis, G. C. (2010). *Fundamentals of Digital Imaging in Medicine*.

Springer Science & Business Media. http://doi.org/10.1118/1.3456439

Brennan, P. C., & Johnston, D. (2002). Irish X-ray departments demonstrate varying levels of adherence to European guidelines on good radiographic technique. *British Journal of Radiology*, *75*(891), 243–248.

Britain, G. and H. and S. C. (2000). *The Ionising Radiation (Medical Exposure) Regulations 2000*.

Brosi, P., Stuessi, A., Verdun, F. R., Vock, P., & Wolf, R. (2011). Copper filtration in pediatric digital X-ray imaging: its impact on image quality and dose. *Radiological Physics and Technology*, *4*(2), 148–155. http://doi.org/10.1007/s12194-011-0115-4

Buckley, O., Ward, E., Ryan, A., Colin, W., Snow, A., & Torreggiani, W. C. (2009). European obesity and the radiology department. What can we do to help? *European Radiology*, *19*(2), 298–309. http://doi.org/10.1007/s00330-008-1154-z

Bunch, P. C., Hamilton, J. F., Sanderson, G. K., & Simmons, A. H. (1977). A Free Response Approach To The Measurement And Characterization Of Radiographic-Observer Performance. In J. E. Gray & W. R. Hendee (Eds.), *Application of Optical Instrumentation in Medicine VI* (pp. 124–135). http://doi.org/10.1117/12.955926

Burgess, A. (1995). Image quality, the ideal observer, and human performance of radiologic decision tasks. *Academic Radiology*, *2*(6), 522–526. http://doi.org/10.1016/S1076-6332(05)80411-8

Burght, R. van der, Floor, M., Thijssen, M., & Bijkerk, R. (2014). *Manual CDRAD 2.0 phantom and analyser software version 2.1*. *Artinis Medical Systems, The Netherlands*. Einsteinweg 17 ,The Netherlands: Artinis Medical Systems.

Busch, H. P., & Faulkner, K. (2005). Image quality and dose management in digital radiography: a new paradigm for optimisation. *Radiation Protection Dosimetry*, *117*(1–3), 143–147. http://doi.org/10.1093/rpd/nci728

Bushong, S. C. (2013). *Radiologic Science for Technologists-E-Book: Physics, Biology, and Protection. Elsevier Health Sciences.*

Carroll, E. M., & Brennan, P. C. (2003). Radiation doses for barium enema and barium meal examinations in Ireland: Potential diagnostic reference levels. *British Journal of*

Radiology, *76*(906), 393–397. http://doi.org/10.1259/bjr/13457134

Carucci, L. R. (2013). Imaging obese patients: problems and solutions. *Abdominal Imaging*, *38*(4), 630–646. http://doi.org/10.1007/s00261-012-9959-2

Chakraborty, D. P. (2005). Recent advances in observer performance methodology: Jackknife free-response ROC (JAFROC). *Radiation Protection Dosimetry*, *114*(1–3), 26–31. http://doi.org/10.1093/rpd/nch512

Chakraborty, D. P., & Berbaum, K. S. (2004a). Jackknife free-response ROC methodology. In D. P. Chakraborty & M. P. Eckstein (Eds.), *Proceedings of SPIE* (Vol. 4686, p. 144). http://doi.org/10.1117/12.533319

Chakraborty, D. P., & Berbaum, K. S. (2004b). Observer studies involving detection and localization: Modeling, analysis, and validation. *Medical Physics*, *31*(8), 2313–2330. http://doi.org/10.1118/1.1769352

Chakraborty, D. P., & Winter, L. H. (1990). Free-response methodology: alternate analysis and a new observer-performance experiment. *Radiology*, *174*(3), 873–881. http://doi.org/10.1148/radiology.174.3.2305073

Cohen, G., McDaniel, D. L., & Wagner, L. K. (1984). Analysis of variations in contrast-detail experiments. *Medical Physics*, *11*(4), 469–473. http://doi.org/10.1118/1.595539

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.Hilsdale." NJ: Lawrence Earlbaum Associates 2*.

Cohen, M. D., Markowitz, R., Hill, J., Huda, W., Babyn, P., & Apgar, B. (2012a). Quality assurance: A comparison study of radiographic exposure for neonatal chest radiographs at 4 academic hospitals. *Pediatric Radiology*, *42*(6), 668–673. http://doi.org/10.1007/s00247-011-2290-1

Cohen, M. D., Markowitz, R., Hill, J., Huda, W., Babyn, P., & Apgar, B. (2012b). Quality assurance: A comparison study of radiographic exposure for neonatal chest radiographs at 4 academic hospitals. *Pediatric Radiology*, *42*(6), 668–673. http://doi.org/10.1007/s00247-011-2290-1

Cole, J., Mackenzie, A., Clinch, J., Kazantzi, M., Clinch, P., & & Lawinski, C. (2009). *Surveying dose levels for computed radiography in the UK*. *KCARE report*.

http://doi.org/10.1007/978-3-642-03902-7-158

Commission of the European Communities (CEC). (1996a). *European Guidelines on Quality Criteria for Diagnostic Radiographic Images:(EUR 16260 EN).* Luxembourg: CEC. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/d59ccc60-97ed-4ce8-b396-3d2d42b284be/language-en

Commission of the European Communities (CEC). (1996b). *European Guidelines on quality criteria for diagnostic radiographic images in paediatrics:(UR 16261EN).* Luxembourg: CEC. Retrieved from ftp://ftp.cordis.lu/pub/fp5-euratom/docs/eur16261.pdf

Conway, B. J., Butler, P. F., Duff, J. E., Fewell, T. R., Gross, R. E., Jennings, R. J., … Showalter, C. K. (1984). Beam quality independent attenuation phantom for estimating patient exposure from x-ray automatic exposure controlled chest examinations. *Medical Physics*, *11*(6), 827–832. http://doi.org/10.1118/1.595611

Cowen, A. R., Davies, A. G., & Kengyelics, S. M. (2007). Advances in computed radiography systems and their physical imaging characteristics. *Clinical Radiology*, *62*(12), 1132–1141. http://doi.org/10.1016/j.crad.2007.07.009

Crop, A. De. (2015). *Image quality evaluation in X-ray medical imaging based on Thiel embalmed human cadavers*. Doctoral dissertation, Ghent University.

D'Ercole, L., Thyrion, F. Z., Bocchiola, M., Mantovani, L., & Klersy, C. (2012). Proposed local diagnostic reference levels in angiography and interventional neuroradiology and a preliminary analysis according to the complexity of the procedures. *Physica Medica*, *28*(1), 61–70. http://doi.org/10.1016/j.ejmp.2010.10.008

De Crop, A., Bacher, K., Van Hoof, T., Smeets, P. V, Smet, B. S., Vergauwen, M., … Thierens, H. (2012). Correlation of Contrast-Detail Analysis and Clinical Image Quality Assessment in Chest Radiography with a Human Cadaver Study. *Radiology*, *262*(1), 298–304. http://doi.org/10.1148/radiol.11110447

De Hauwere, A., Bacher, K., Smeets, P., Verstraete, K., & Thierens, H. (2005). Analysis of image quality in digital chest imaging. *Radiation Protection Dosimetry*, *117*(1–3), 174–177. http://doi.org/10.1093/rpd/nci748

Desai, N., Singh, A., & Valentino, D. J. (2010). Practical evaluation of image quality in computed radiographic (CR) imaging systems (Vol. 7622, p. 76224Q). http://doi.org/10.1117/12.844640

Dewerd L.A, K. M. (2014). *The Phantoms of Medical and Health Physics*. (L. A. DeWerd & M. Kissick, Eds.). New York.: Springer.

Dobbins III James T., Samei, E., Chotas, H. G., Warp, R. J., Baydush, A. H., Floyd, C. E., & Ravin, C. E. (2003). Chest radiography: optimization of X-ray spectrum for cesium iodide-amorphous silicon flat-panel detector. *Radiology*, *226*(1), 221–230. http://doi.org/10.1148/radiol.2261012023

Doyle, P., Martin, C. J., & Gentle, D. (2006). Application of contrast-to-noise ratio in optimizing beam quality for digital chest radiography: comparison of experimental measurements and theoretical simulations. *Physics in Medicine and Biology*, *51*(11), 2953–2970. http://doi.org/10.1088/0031-9155/51/11/018

Dragusin, O., Smans, K., Jacobs, J., Inal, T., & Bosmans, H. (2008). Evaluation of the contrast-detail response of a cardiovascular angiography system and the influence of equipment variables on image quality. In J. Hsieh & E. Samei (Eds.), *Proceedings of International Conference "Medical Physics in the Baltic States"* (Vol. 6913, p. 69134R). http://doi.org/10.1117/12.769780

Drost, W. T., Reese, D. J., & Hornof, W. J. (2008). Digital radiography artifacts. *Veterinary Radiology & Ultrasound*, *49*(SUPPL. 1), S48–S56. http://doi.org/10.1111/j.1740-8261.2007.00334.x

Egbe, N. O., Inyang, S. O., Eduwem, D. U., & Ama, I. (2009). Doses and image quality for chest radiographs in three Nigerian hospitals. *European Journal of Radiography*, *1*(1), 30–36. http://doi.org/10.1016/j.ejradi.2008.09.001

Ekpo, E. U., Hoban, A. C., & McEntee, M. F. (2014). Optimisation of direct digital chest radiography using Cu filtration. *Radiography*, *20*(4), 346–350. http://doi.org/10.1016/j.radi.2014.07.001

Elshiekh, E., Suliman, I. I., & Habbani, F. (2015). A comparative study of adult patient doses in film screen and computed radiography in some sudanese hospitals. *Radiation Protection Dosimetry*, *165*(1), 1–4. http://doi.org/10.1093/rpd/ncv186

Eurostat Statistics Explained. (2017). Overweight and obesity - BMI statistics. Retrieved June 18, 2018, from http://ec.europa.eu/eurostat/statistics-explained/index.php/Overweight_and_obesity_-_BMI_statistics

Fauber, T. L. (2016). *Radiographic Imaging and Exposure*. Retrieved from https://books.google.com/books?id=yVpPAQAAQBAJ&pgis=1

Fausto, A. M. F., Lopes, M. C., de Sousa, M. C., Furquim, T. A. C., Mol, A. W., & Velasco, F. G. (2017). Optimization of Image Quality and Dose in Digital Mammography. *Journal of Digital Imaging*, *30*(2), 185–196. http://doi.org/10.1007/s10278-016-9928-3

Fernandez, J. M., Ordiales, J. M., Guibelalde, E., Prieto, C., & Vano, E. (2008). Physical image quality comparison of four types of digital detector for chest radiology. *Radiation Protection Dosimetry*, *129*(1–3), 140–143. http://doi.org/10.1093/rpd/ncn026

Filed, A. (2009). *Discovering statistics using SPSS statistics* (Third edit). Londan: Sage publications Ltd.

Fischbach, F., Ricke, J., Freund, T., Werk, M., Spors, B., Baumann, C., … Felix, R. (2002). Flat panel digital radiography compared with storage phosphor computed radiography: assessment of dose versus image quality in phantom studies. *Investigative Radiology*, *37*(11), 609–14. http://doi.org/10.1097/01.RLI.0000031078.14395.E4

Fletcher, J. G., Hara, A. K., Fidler, J. L., Silva, A. C., Barlow, J. M., Carter, R. E., … McCollough, C. H. (2015). Observer performance for adaptive, image-based denoising and filtered back projection compared to scanner-based iterative reconstruction for lower dose CT enterography. *Abdominal Imaging*, *40*(5), 1050–1059. http://doi.org/10.1007/s00261-015-0384-1

Gallacher, D. J., MacKenzie, A., Batchelor, S., Lynch, J., & Saunders, J. E. (2003). Use of a quality index in threshold contrast detail detection measurements in television fluoroscopy. *British Journal of Radiology*, *76*(907), 464–472. http://doi.org/10.1259/bjr/82965103

Geijer, H., Beckman, K.-W., Andersson, T., & Persliden, J. (2001). Image quality vs radiation dose for a flat-panel amorphous silicon detector: a phantom study. *European*

Radiology, *11*(9), 1704–1709. http://doi.org/10.1007/s003300100851

Geijer, H., & Persliden, J. (2005). Varied tube potential with constant effective dose at lumbar spine radiography using a flat-panel digital detector. *Radiation Protection Dosimetry*, *114*(1–3), 240–245. http://doi.org/10.1093/rpd/nch509

Geleijns, J., Schultze Kool, L. J., Zoetelief, J., Zweers, D., & Broerse, J. J. (1993). Image Quality and Dosimetric Aspects of Chest X Ray Examinations: Measurements with Various Types of Phantoms. *Radiation Protection Dosimetry*, *49*(1–3), 83–88. http://doi.org/10.1093/oxfordjournals.rpd.a081907

Geuer, H., Norrman, E., & Persliden, J. (2009). Optimizing the tube potential for lumbar spine radiography with a flat-panel digital detector. *The British Journal of Radiology*, *82*(973), 62–68. http://doi.org/10.1259/bjr/56572915

Green, D. M., and Swets, J. A. (1988). *Signal Detection Theory and Psychophysics.*

Grewal, R. K., Young, N., Colins, L., Karunnaratne, N., & Sabharwal, N. (2012). Digital chest radiography image quality assessment with dose reduction. *Australasian Physical & Engineering Sciences in Medicine / Supported by the Australasian College of Physical Scientists in Medicine and the Australasian Association of Physical Sciences in Medicine*, *35*(1), 71–80. http://doi.org/10.1007/s13246-012-0125-5

Guerra, a Del. (2004). *Ionizing Radiation Detectors for Medical Imaging*. Retrieved from http://books.google.com/books?hl=en&lr=&id=UzOxXhBKpYEC&oi=fnd&pg=PA1&dq=Ionizing+Radiation+Detectors+for+Medical+Imaging&ots=5GV6emfaN3&sig=XsaBWkvLXJa3E8BXhwMXZ7AQvyg

Håkansson, M., Båth, M., Börjesson, S., Kheddache, S., Allansdotter Johnsson, Å., & Månsson, L. G. (2005). Nodule detection in digital chest radiography: Effect of system noise. *Radiation Protection Dosimetry*, *114*(1–3), 97–101. http://doi.org/10.1093/rpd/nch525

Håkansson, M., Båth, M., Börjesson, S., Kheddache, S., Grahn, A., Ruschin, M., … Månsson, L. G. (2005). Nodule detection in digital chest radiography: Summary of the radius chest trial. *Radiation Protection Dosimetry*, *114*(1–3), 114–120. http://doi.org/10.1093/rpd/nch574

Hamer, O. W., Sirlin, C. B., Strotzer, M., Borisch, I., Zorger, N., Feuerbach, S., & Völk, M. (2005). Chest radiography with a flat-panel detector: image quality with dose reduction after copper filtration. *Radiology*, *237*(2), 691–700. http://doi.org/10.1148/radiol.2372041738

Hamer, O. W., Völk, M., Zorger, N., Borisch, I., Büttner, R., Feuerbach, S., & Strotzer, M. (2004). Contrast-detail phantom study for x-ray spectrum optimization regarding chest radiography using a cesium iodide-amorphous silicon flat-panel detector. *Investigative Radiology*, *39*(10), 610–618. http://doi.org/00004424-200410000-00004 [pii]

Hamer, O. W., Völk, M., Zorger, Z., Feuerbach, S., & Strotzer, M. (2003). Amorphous silicon, flat-panel, x-ray detector versus storage phosphor-based computed radiography: contrast-detail phantom study at different tube voltages and detector entrance doses. *Investigative Radiology*, *38*(4), 212–20. http://doi.org/10.1097/01.RLI.0000057032.41715.15

Hampel, J. R., & Pascoal, A. (2018). Comparison and optimization of imaging techniques in suspected physical abuse paediatric radiography. *The British Journal of Radiology*, *91*(December 2017), 20170650. http://doi.org/10.1259/bjr.20170650

Hart, D. W. B. F., Wall, B., & Shrimpton, P. (2000). *Reference doses and patient size in paediatric radiology (No. NRPB-R--318).* (National Radiological Protection Board, Ed.).

Hart, D. (2003). The UK National Patient Dose Database: now and in the future. *British Journal of Radiology*, *76*(906), 361–365. http://doi.org/10.1259/bjr/18536168

Hart, D., Hillier, M. C., & Shrimpton, P. C. (2012). *Doses to Patients from Radiographic and Fluoroscopic X-ray Imaging Procedures in the UK – 2010 Review*. *Health Protection Agency (UK) Report*. Retrieved from http://www.hpa.org.uk/Publications/Radiation/CRCEScientificAndTechnicalReportSeries/HPACRCE034

Hart, D., & Shrimpton, P. C. (2012). Fourth review of the UK national patient dose database. *The British Journal of Radiology*, *85*(1018), e957-8. http://doi.org/10.1259/bjr/28778022

Hart, D., Wall, B., Hillier, M., & Shrimpton, P. (2010). *Frequency and collective dose for*

medical and dental X-ray examinations in the UK, 2008.

Hauge, A. I. H. R., Aandahl, J., Baranzelli, P., Coelho, M., Eriksen, K., Hadebe, S., … Sciences, A. (2017). Digital Radiography : Impact of lower tube voltages on image quality and radiation dose in chest phantom radiography . In *In (Eds.) A. Meijer, C. Buissink and P. Hogg .Radiation dose and image quality optimisation in medical imaging, Erasmus Intensive Programme OPTIMAX#2017*.

Hay, G. A., Clarke, O. F., & Eng, T. (1985). A set of X-ray test objects for quality control in television fluoroscopy. *British Journal of Radiology*, *58*(688), 335–344.

Hemdal, B. (2009). *Evaluation of absorbed dose and image quality in mammography*. *Physics*.

Hendee, W. R., & Ritenour, E. R. (2002). *Medical Imaging Physics*. (A. J. Roentgenol & J. W. & Sons, Eds.) (4th ed.). New York, USA.

Hintenlang, K., Williams, J., & Hintenlang, D. (2002). A survey of radiation dose associated with pediatric plain-film chest X-ray examinations. *Pediatric Radiology*, *32*(11), 771–777. http://doi.org/10.1007/s00247-002-0734-3

Hirose, M., Ikeda, M., Ito, K., Ishigaki, T., & Sakuma, S. (1993). Considerations for standard chest radiography: the long film-focus distance technique. *Nagoya Journal of Medical Science*, *55*(1–4), 33–39. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8247104

Hofmann, B. (2016). Obesity as a Socially Defined Disease: Philosophical Considerations and Implications for Policy and Care. *Health Care Analysis*, *24*(1), 86–100. http://doi.org/10.1007/s10728-015-0291-1

Hogg, P. and Blindell, P. (2012). Software for image quality evaluation using a forced choice method. In *In United Kingdom radiological conference (p. 139). British Institute of Radiology, Manchester/London, UK.*

Honey, I. D., & Mackenzie, A. (2009). Artifacts Found During Quality Assurance Testing of Computed Radiography and Digital Radiography Detectors. *Journal of Digital Imaging*, *22*(4), 383–392. http://doi.org/10.1007/s10278-008-9109-0

ICRP. (1996). *International Commission on Radiological Protection (ICRP). Publication 73: Radiological Protection and Safety in Medicine (Vol. 22). Elsevier Health Sciences*. *Annals of the ICRP*. http://doi.org/10.1016/S0146-6453

Im, J. G., Chung, M. J., Kim, H. Y., Lee, H. J., Kang, J. W., Kim, J. H., … Lee, Y. J. (2002). Detection of simulated chest lesions by using soft-copy reading: comparison of an amorphous silicon flat-panel-detector system and a storage-phosphor system. *Radiology*, *224*(1), 242―246. http://doi.org/10.1148/radiol.2241011441

Institute of Physics and Engineering in Medicine. (2005). *IPEM report 91: recommended standards for the routine performance testing of diagnostic x-ray imaging systems*. York: IPEM.

International Atomic Energy Agency. (2012). *Radiation Protection in Paediatric radiology*. Vienna, Austria: IAEA.

International Atomic Energy Agency. (2013). *Dosimetry in Diagnostic Radiology for Paediatric Patients*. *Internat. Atomic Energy Agency*.

Jäkel, F., & Wichmann, F. A. (2006). Spatial four-alternative forced-choice method is the preferred psychophysical method for naïve observers. *Journal of Vision*, *6*(11), 1307–22. http://doi.org/10.1167/6.11.13

Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *Journal of Experimental Psychology: General*, *138*(2), 291–306. http://doi.org/10.1037/a0015525

Jansson, M., Geijer, H., Persliden, J., & Andersson, T. (2006). Reducing dose in urography while maintaining image quality—a comparison of storage phosphor plates and a flat-panel detector. *European Radiology*, *16*(1), 221–226. http://doi.org/10.1007/s00330-005-2772-3

Jessen, K. A. (2004). Balancing image quality and dose in diagnostic radiology. *European Radiology, Supplement*, *14*(1), 9–18. http://doi.org/10.1007/s10406-004-0003-7

Jessop, M., Thompson, J. D., Coward, J., Sanderud, A., Jorge, J., de Groot, M., … Hogg, P. (2015). Lesion Detection Performance: Comparative Analysis of Low-Dose CT Data of the Chest on Two Hybrid Imaging Systems. *Journal of Nuclear Medicine*

Technology, *43*(1), 47–52. http://doi.org/10.2967/jnmt.114.147447

Jiang, X., Baad, M., Reiser, I., Feinstein, K. A., & Lu, Z. (2016). Effect of comfort pads and incubator design on neonatal radiography. *Pediatric Radiology*, *46*(1), 112–118. http://doi.org/10.1007/s00247-015-3450-5

Johnston, D. A., & Brennan, P. C. (2000). Reference dose levels for patients undergoing common diagnostic X-ray examinations in Irish hospitals. *British Journal of Radiology*, *73*(868), 396–402. http://doi.org/10.1259/bjr.73.868.10844865

Jones, A. K. (2006). *Dose Versus Image Quality in Pediatric Radiology: Studies Using a Tomographic Newborn Physical Phantom with an Incorporated Dosimetry System*. *Doctoral dissertation, University of Florida*. http://doi.org/10.1017/CBO9781107415324.004

Kepler, K., & Vladimirov, A. (2008). Optimisation strategies introduced for CR at health care centres in Estonia. *Radiation Protection Dosimetry*, *129*(1–3), 127–131. http://doi.org/10.1093/rpd/ncn004

Kiljunen, T., Tietäväinen, a, Parviainen, T., Viitala, a, & Kortesniemi, M. (2009). Organ doses and effective doses in pediatric radiography: patient-dose survey in Finland. *Acta Radiologica*, *50*(1), 114–124. http://doi.org/10.1080/02841850802570561

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. http://doi.org/10.1016/j.jcm.2016.02.012

Kostova-Lefterova, D., Taseva, D., Ingilizova, K., Hristova-Popova, J., & Vassileva, J. (2011). Potential for optimisation of paediatric chest X-ray examination. *Radiation Protection Dosimetry*, *147*(1–2), 168–170. http://doi.org/10.1093/rpd/ncr328

Kotter, E., & Langer, M. (2002). Digital radiography with large-area flat-panel detectors. *European Radiology*, *12*(10), 2562–2570. http://doi.org/10.1007/s00330-002-1350-1

Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception & Psychophysics*, *72*(5), 1205–1217. http://doi.org/10.3758/APP.72.5.1205

Kundel, H. L., & Revesz, G. (1976). Lesion conspicuity, structured noise, and film reader error. *American Journal of Roentgenology*, *6*(126), 1233–1238.

Kundel, H. L. (1979). Images, image quality and observer performance: new horizons in radiology lecture. *Radiology*, *132*(March), 265–271. http://doi.org/10.1148/132.2.265

Kundel, H. L., Nodine, C. F., & Carmody, D. (1978). Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology*, *13*(3), 175–81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/711391

L pez, M., Morant, J. J., Geleijns, K., & Calzado, A. (2000). A Regional Dose and Image Quality Survey for Chest, Abdomen and Pelvis Radiographs in Paediatrics. *Radiation Protection Dosimetry*, *90*(1), 275–278. http://doi.org/10.1093/oxfordjournals.rpd.a033137

Lança, L. J. O. C. (2011). *Radiological imaging in digital systems: the effect of exposure parameters in diagnostic quality and patient dose*.

Lanca, L., & Silva, A. (2012). *Digital Imaging Systems for Plain Radiography*. *Springer Science & Business Media*.

Lança, L., & Silva, A. (2009a). Digital radiography detectors - A technical overview: Part 1. *Radiography*, *15*(1), 58–62. http://doi.org/10.1016/j.radi.2008.02.004

Lança, L., & Silva, A. (2009b). Digital radiography detectors – A technical overview: Part 2. *Radiography*, *15*(2), 134–138. http://doi.org/10.1016/j.radi.2008.02.005

Lanconelli, N., Rivetti, S., Golinelli, P., Serafini, M., Bertolini, M., & Borasi, G. (2007). Comparison of Human observers and CDCOM software reading for CDMAM images. *Proc. of SPIE*, *6515*(March 2015), 65150E–1. http://doi.org/10.1117/12.710154

Lanhede, B., Båth, M., Kheddache, S., Sund, P., Björneld, L., Widell, M., … Månsson, L. G. (2002). The influence of different technique factors on image quality of chest radiographs as evaluated by modified CEC image quality criteria. *British Journal of Radiology*, *75*(889), 38–49. http://doi.org/10.1259/bjr.75.889.750038

Lazar, M. A., Plocher, E. K., & Egol, K. A. (2010). Obesity and its relationship with pelvic and lower-extremity orthopedic trauma. *American Journal of Orthopedics (Belle Mead, N.J.)*, *39*(4), 175–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20512170

Lee, S. C., Wang, J. N., Liu, S. C., & Jiang, S. H. (2007). Evaluation of dose-image-quality optimization in digital chest radiography. *Nuclear Instruments and Methods in Physics*

Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, *580*(1 SPEC. ISS.), 544–547. http://doi.org/10.1016/j.nima.2007.05.226

Lin, Y., Luo, H., Dobbins, J. T., Page McAdams, H., Wang, X., Sehnert, W. J., … Samei, E. (2012). An image-based technique to assess the perceptual quality of clinical chest radiographs. *Medical Physics*, *39*(11), 7019–7031. http://doi.org/10.1118/1.4760886

Lu, Z. F., Nickoloff, E. L., So, J. C., & Dutta, A. K. (2003). Comparison of computed radiography and film/screen combination using a contrast-detail phantom. *Journal of Applied Clinical Medical Physics*, *4*(1), 91–98. http://doi.org/10.1120/jacmp.v4i1.2548

Ludewig, E., Richter, A., & Frame, M. (2010). Diagnostic imaging - Evaluating image quality using visual grading characteristic (VGC) analysis. *Veterinary Research Communications*, *34*(5), 473–479. http://doi.org/10.1007/s11259-010-9413-2

Lyra, M. E., Kordolaimi, S. D., & Salvara, A.-L. N. (2010). Presentation of Digital Radiographic Systems and the Quality Control Procedures that Currently Followed by Various Organizations Worldwide. *Recent Patents on Medical Imaging*, *2*(1), 5–21. http://doi.org/10.2174/1877613201002010005

M nsson, L. G. (2000). Methods for the Evaluation of Image Quality: A Review. *Radiation Protection Dosimetry*, *90*(1), 89–99. http://doi.org/10.1093/oxfordjournals.rpd.a033149

Ma, W. K., Hogg, P., Tootell, A., Manning, D., Thomas, N., Kane, T., … Kitching, J. (2013). Anthropomorphic chest phantom imaging – The potential for dose creep in computed radiography. *Radiography*, *19*(3), 207–211. http://doi.org/10.1016/j.radi.2013.04.002

Mackenzie, A., Warren, L. M., Wallis, M. G., Given-Wilson, R. M., Cooke, J., Dance, D. R., … Young, K. C. (2016). The relationship between cancer detection in mammography and image quality measurements. *Physica Medica*, *32*(4), 568–574. http://doi.org/10.1016/j.ejmp.2016.03.004

Manning, D. J. (1998). *Applications of signal detection theory to the performance of imaging systems, human observers and artificial intelligence in radiography*.

Manning, D. J., Ethell, S. C., & Donovan, T. (2004). Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph. *British Journal of Radiology*,

77(915), 231–235. http://doi.org/10.1259/bjr/28883951

Månsson, L. G., Båth, M., & Mattsson, S. (2005). Priorities in optimisation of medical X-ray imaging—a contribution to the debate. *Radiation Protection Dosimetry*, *114*(1–3), 298–302. http://doi.org/10.1093/rpd/nch578

Marshall, N. W., Faulkner, K., Kotre, C. J., & Robson, K. (1992). Analysis of variations in contrast-detail measurements performed on image intensifier-television systems. *Physics in Medicine and Biology*, *37*(12), 2297–2302. http://doi.org/10.1088/0031-9155/37/12/011

Martin, C. J. (2007). Optimisation in general radiography. *Biomedical Imaging and Intervention Journal*, *3*(2). http://doi.org/10.2349/biij.3.2.e18

Martin, C. J., & Vano, E. (2018). Diagnostic reference levels and optimisation in radiology: Where do we go from here? *Journal of Radiological Protection*, *38*(1), E1–E4. http://doi.org/10.1088/1361-6498/aa9cfd

Mathews, J. D., Forsythe, A. V., Brady, Z., Butler, M. W., Goergen, S. K., Byrnes, G. B., … Darby, S. C. (2013). Cancer risk in 680 000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. *BMJ*, *346*(may21 1), f2360–f2360. http://doi.org/10.1136/bmj.f2360

Mc Laughlin, P., Neill, S. O., Fanning, N., Mc Garrigle, A. M., Connor, O. J. O., Wyse, G., & Maher, M. M. (2012). Emergency CT brain: Preliminary interpretation with a tablet device: Image quality and diagnostic performance of the Apple iPad. *Emergency Radiology*, *19*(2), 127–133. http://doi.org/10.1007/s10140-011-1011-2

McAdams, H. P., Samei, E., Dobbins, J., Tourassi, G. D., & Ravin, C. E. (2006). Recent Advances in Chest Radiography. *Radiology*, *241*(3), 663–683. http://doi.org/10.1148/radiol.2413051535

McEntee, M., Frawley, H., & Brennan, P. C. (2007). A comparison of low contrast performance for amorphous Silicon/caesium iodide direct radiography with a computed radiography: A contrast detail phantom study. *Radiography*, *13*(2), 89–94. http://doi.org/10.1016/j.radi.2006.01.004

Miracle, A. C., & Mukherji, S. K. (2009). Conebeam CT of the Head and Neck, Part 1:

Physical Principles. *American Journal of Neuroradiology*, *30*(6), 1088–1095. http://doi.org/10.3174/ajnr.A1653

Moore, C. S., Avery, G., Balcam, S., Needler, L., Swift, A., Beavis, A. W., & Saunderson, J. R. (2012). Use of a digitally reconstructed radiograph-based computer simulation for the optimisation of chest radiographic techniques for computed radiography imaging systems. *The British Journal of Radiology*, *85*(1017), e630–e639. http://doi.org/10.1259/bjr/47377285

Moore, C. S., Liney, G. P., Beavis, A. W., & Saunderson, J. R. (2011). A method to produce and validate a digitally reconstructed radiograph-based computer simulation for optimisation of chest radiographs acquired with a computed radiography imaging system. *The British Journal of Radiology*, *84*(1006), 890–902. http://doi.org/10.1259/bjr/30125639

Moore, C. S., Wood, T. J., Avery, G., Balcam, S., Needler, L., Smith, A., … Beavis, A. W. (2015). Investigating the use of an antiscatter grid in chest radiography for average adults with a computed radiography imaging system. *The British Journal of Radiology*, *88*(1047), 20140613. http://doi.org/10.1259/bjr.20140613

Moore, C. S., Wood, T. J., Beavis, A. W., & Saunderson, J. R. (2013). Correlation of the clinical and physical image quality in chest radiography for average adults with a computed radiography imaging system. *The British Journal of Radiology*, *86*(1027), 20130077. http://doi.org/10.1259/bjr.20130077

Morrell, R. E. (2006). *Dosimetry and optimisation in high dose fluoroscopic and fluorographic procedure(Doctoral dissertation, University of Nottingham).*

Mraity, H., England, A., Akhtar, I., Aslam, A., De Lange, R., Momoniat, H., … Hogg, P. (2014). Development and validation of a psychometric scale for assessing PA chest image quality: A pilot study. *Radiography*, *20*(4), 312–317. http://doi.org/10.1016/j.radi.2014.03.007

Mraity, H., Hogg, P., & England, A. (2015). *Optimisation of radiation dose and image quality for AP pelvis radiographic examination.*

Murata, K., & Nitta, N. (2016). Multipurpose Chest Phantom N1 “LUNGMAN.”

Neitzel, U., Pralow, T., Schaefer-prokopb, C., & Prokopb, M. (1998). Influence of scatter reduction on lesion signal-to-noise ratio and lesion detection in digital chest radiography. *International Society for Optics and Photonics*, *3336*(May), 337–348. http://doi.org/10.1117/12.317033

Norrman, E., Gårdestig, M., Persliden, J., & Geijer, H. (2005). A Clinical Evaluation of the Image Quality Computer Program, CoCIQ. *Journal of Digital Imaging*, *18*(2), 138–144. http://doi.org/10.1007/s10278-004-1036-0

Norweck, J. T., Seibert, J. A., Andriole, K. P., Clunie, D. A., Curran, B. H., Flynn, M. J., … Wyatt, M. (2013). ACR-AAPM-SIIM technical standard for electronic practice of medical imaging. *Journal of Digital Imaging*, *26*(1), 38–52. http://doi.org/10.1007/s10278-012-9522-2

Obuchowski, N. a. (2003). Receiver operating characteristic curves and their use in radiology. *Radiology*, *229*(1), 3–8. http://doi.org/10.1148/radiol.2291010898

Ohlmann-Knafo, S., Morlo, M., Tarnoki, D. L., Tarnoki, A. D., Grabowski, B., Kaspar, M., & Pickuth, D. (2016). Comparison of image quality characteristics on Silent MR versus conventional MR imaging of brain lesions at 3 Tesla. *The British Journal of Radiology*, *89*(1067), 20150801. http://doi.org/10.1259/bjr.20150801

Olgar, T., Onal, E., Bor, D., Okumus, N., Atalay, Y., Turkyilmaz, C., … Koc, E. (2008). Radiation exposure to premature infants in a neonatal intensive care unit in Turkey. *Korean Journal of Radiology*, *9*(5), 416–419. http://doi.org/10.3348/kjr.2008.9.5.416

Otto, D., Ludwig, K., Fessel, A., Bernhardt, T. M., Kästner, A., Reißberg, S., & Döhring, W. (2000). Digital selenium radiography: detection of subtle pulmonary lesions on images acquired with and without an additional antiscatter grid. *European Journal of Radiology*, *36*(2), 108–114. http://doi.org/10.1016/S0720-048X(00)00265-5

Pallant, J. (2010). *SPSS Survival Manual A Step by Step Guide to Data Analysis using SPSS for Windows, Open University Press*.

Park, S. H., Goo, J. M., & Jo, C.-H. (2004). Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean Journal of Radiology*, *5*(March), 11–8. http://doi.org/10.3348/kjr.2004.5.1.11

Pascoal, a, Lawinski, C. P., Honey, I., & Blake, P. (2005). Evaluation of a software package for automated quality assessment of contrast detail images—comparison with subjective visual assessment. *Physics in Medicine and Biology*, *50*(23), 5743–5757. http://doi.org/10.1088/0031-9155/50/23/023

Paulana, H. V, Wigati, K. T., & Soejoko, D. S. (2015). Mean Glandular Dose and Cdmam Phantom Image Quality for Siemens Mammomat Inspiration. *Journal of Medical Physics and Biophysics*, *30*(2).

Peer, S., Giacomuzzi, S. M., Peer, R., Gassner, E., Steingruber, I., & Jaschke, W. (2003). Resolution requirements for monitor viewing of digital flat-panel detector radiographs: a contrast detail analysis. *European Radiology*, *13*(2), 413–7. http://doi.org/10.1007/s00330-002-1479-y

Peer, S., Neitzel, U., Giacomuzzi, S. M., Peer, R., Gassner, E., Steingruber, I., & Jaschke, W. (2001). Comparison of low-contrast detail perception on storage phosphor radiographs and digital flat panel detector images. *IEEE Transactions on Medical Imaging*, *20*(3), 239–242. http://doi.org/10.1109/42.918474

Pelli, D. G., & Farell, B. (2010). *Psychophysical methods. Handbook of Optics, Third edition, Volume iii: Vision and Vision Optics* (Vol. III).

Petoussi-Henss, N., Zankl, M., Drexler, G., Panzer, W., & Regulla, D. (1998). Calculation of backscatter factors for diagnostic radiology using Monte Carlo methods. *Physics in Medicine and Biology*, *43*(8), 15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9725601

Pinto, A. (2010). Spectrum of diagnostic errors in radiology. *World Journal of Radiology*, *2*(10), 377. http://doi.org/10.4329/wjr.v2.i10.377

Portney, L., & Watkins, M. (2000). *Foundations of clinical research: applications to practice.New Jersey: Prentice Hall*.

Precht, H., Tingberg, A., Waaler, D., & Outzen, C. B. (2014). New Developed DR Detector Performs Radiographs of Hand, Pelvic and Premature Chest Anatomies at a Lower Radiation Dose and/or a Higher Image Quality. *Journal of Digital Imaging*, *27*(1), 68–76. http://doi.org/10.1007/s10278-013-9635-2

Raissaki, M. T. (2004). Pediatric radiation protection. *European Radiology Supplements*, *14*(S1), 74–83. http://doi.org/10.1007/s10406-004-0011-7

Raoof, S., Feigin, D., Sung, A., Raoof, S., Irugulpati, L., & Rosenow, E. C. (2012). Interpretation of Plain Chest Roentgenogram. *Chest*, *141*(2), 545–558. http://doi.org/10.1378/chest.10-1302

Rassow, J., Schmaltz, A. A., Hentrich, F., & Streffer, C. (2000). Effective doses to patients from paediatric cardiac catheterization. *The British Journal of Radiology*, *73*(866), 172–183. http://doi.org/10.1259/bjr.73.866.10884731

RaySafe X2. (2016). The unnfors solid state dosimeter RaySafe X2 user manual.

Redlich, U., Hoeschen, C., & Doehring, W. (2005). Assessment and optimisation of the image quality of chest-radiography systems. *Radiation Protection Dosimetry*, *114*(1–3), 264–268. http://doi.org/10.1093/rpd/nch559

Revesz, G., Kundel, H.L. and Graber, M. A. (1974). The influence of structured noise on the detection of radiologic abnormalities. *Investigative Radiology*, *9*(6), 479–486.

Riverti di Val Cervo, S. (2011). *Performance evaluation of detectors for digital radiography*. *Performance Evaluation*.

Rojas, L. J., Fausto, A. M. F., Mol, A. W., Velasco, F. G., Abreu, P. O. S., Henriques, G., & Furquim, T. A. C. (2017). Optimization of the exposure parameters in digital mammography using contrast-detail metrics. *Physica Medica*, *42*, 13–18. http://doi.org/10.1016/j.ejmp.2017.08.001

Rong, X. J., Shaw, C. C., Liu, X., Lemacks, M. R., & Thompson, S. K. (2001). Comparison of an amorphous silicon/cesium iodide flat-panel digital chest radiography system with screen/film and computed radiography systems - A contrast-detail phantom study. *Medical Physics*, *28*(11), 2328–2335. http://doi.org/10.1118/1.1408620

Rose, a. (1974). Vision: human and electronic, in Applied Solid State Physics. *Springer.*, 79–160. http://doi.org/10.1016/B978-1-4832-3089-4.50032-7

Rose, A. (1948). The sensitivity performance of the human eye on an absolute scale. *Journal of the Optical Society of America*, *38*(2), 196–208. http://doi.org/10.1364/JOSA.38.000196

Rzeszotarski, M. S. (1999). Counting statistics. *Radiographics : A Review Publication of the Radiological Society of North America, Inc*, *19*, 765–82. http://doi.org/10.1148/radiographics.19.3.g99ma33765

Saito, K., Hiramoto, S., Gomi, T., Muramoto, N., Seki, M., Tsukimura, K., & Suzuki, H. (2008). Evaluation of chest and abdominal exposure dose appropriate for a digital image reader system incorporating a columnar-crystal structured phosphor plate and a contrast-detail phantom. *Radiological Physics and Technology*, *1*(2), 238–243. http://doi.org/10.1007/s12194-008-0035-0

Samei, E., & Flynn, M. J. (2003). Performance of digital radiographic detectors: factors affecting sharpness and noise. *Advances in Digital Radiography*, 37–47.

Samei, E. (2003). Performance of digital radiography detectors: factors affecting sharpness and noise. *Advances in Digital Radiography. RSNA*, 49–61.

Samei, E., Dobbins, J. T., Lo, J. Y., & Tornai, M. P. (2005). A framework for optimising the radiographic technique in digital X-ray imaging. *Radiation Protection Dosimetry*, *114*(1–3), 220–229. http://doi.org/10.1093/rpd/nch562

Samei, E., Eyler, W., & Baron, L. (2000). Effects of anatomical structure on signal detection. In *Handbook of medical imaging. Volume 1. Physics and psychophysics* (pp. 655–682).

Samei, E., Flynn, M. J., & Eyler, W. R. (1999). Detection of Subtle Lung Nodules: Relative Influence of Quantum and Anatomic Noise on Chest Radiographs. *Radiology*, *213*(3), 727–734. http://doi.org/10.1148/radiology.213.3.r99dc19727

Samei, E., Flynn, M. J., Peterson, E., & Eyler, W. R. (2003). Subtle Lung Nodules: Influence of Local Anatomic Variations on Detection. *Radiology*, *228*(1), 76–84. http://doi.org/10.1148/radiol.2273020509

Samei, E., Seibert, J. A., Andriole, K., Badano, A., Crawford, J., Reiner, B., … Chang, P. (2004). AAPM/RSNA Tutorial on Equipment Selection: PACS Equipment Overview. *RadioGraphics*, *24*(1), 313–334. http://doi.org/10.1148/rg.241035137

Sandborg, M., McVey, G., Dance, D. R., & Carlsson, G. A. (2000). Comparison of Model Predictions of Image Quality with Results of Clinical Trials in Chest and Lumbar Spine Screen-film Imaging. *Radiation Protection Dosimetry*, *90*(1), 173–176.

http://doi.org/10.1093/oxfordjournals.rpd.a033112

Sandborg, M., Tingberg, A., Dance, D. R., Lanhede, B., Almén, A., McVey, G., … Alm Carlsson, G. (2001). Demonstration of correlations between clinical and physical image quality measures in chest and lumbar spine screen–film radiography. *The British Journal of Radiology*, *74*(882), 520–528. http://doi.org/10.1259/bjr.74.882.740520

Sandborg, M., Tingberg, A., Ullman, G., Dance, D. R., & Alm Carlsson, G. (2006). Comparison of clinical and physical measures of image quality in chest and pelvis computed radiography at different tube voltages. *Medical Physics*, *33*(11), 4169–4175. http://doi.org/10.1118/1.2362871

Schaefer-Prokop, C., Neitzel, U., Venema, H. W., Uffmann, M., & Prokop, M. (2008). Digital chest radiography: an update on modern technology, dose containment and control of image quality. *European Radiology*, *18*(9), 1818–1830. http://doi.org/10.1007/s00330-008-0948-3

Schultz, F. W., Geleijns, J., & Zoetelief, J. (1994). Calculation of dose conversion factors for posterior-anterior chest radiography of adults with a relatively high-energy X-ray spectrum. *British Journal of Radiology*, *67*(800), 775–785. http://doi.org/10.1259/0007-1285-67-800-775

Seeram, E., Davidson, R., Bushong, S., & Swan, H. (2016). Optimizing the Exposure Indicator as a Dose Management Strategy in Computed Radiography. *Radiologic Technology*, *87*(4), 380–391.

Seeram, Euclid, Davidson, R., Bushong, S., & Swan, H. (2014). Image Quality Assessment Tools for Radiation Dose Optimization in Digital Radiography: An Overview. *Radiologic Technology*, *85*(5), 555–562.

Seibert, J. A. (2009). Digital adiography: The bottom line comparison of CR and DR technology. *Applied Radiology*, *38*(5), 21–28.

Seibert, J. A., Boone, J. M., Cooper, V. N., & Lindfors, K. K. (2004). Cassette-based Digital Mammography. *Technology in Cancer Research & Treatment*, *3*(5), 413–427. http://doi.org/10.1177/153303460400300502

Seo, D., Jang, S., Kim, J., Kim, J., Sung, D., Kim, H., & Yoon, Y. (2014). Acomparative

assessment of entrance surface doses in analogue and digital radiography during common radiographic examinations. *Radiation Protection Dosimetry*, *158*(1), 22–27. http://doi.org/10.1093/rpd/nct189

Shet, N., Chen, J., & Siegel, E. L. (2011). Continuing challenges in defining image quality. *Pediatric Radiology*, *41*(5), 582–587. http://doi.org/10.1007/s00247-011-2028-0

Shetty, C. M., Barthur, A., Kambadakone, A., Narayanan, N., & Kv, R. (2011). Computed radiography image artifacts revisited. *American Journal of Roentgenology*, *196*(1), 37–47. http://doi.org/10.2214/AJR.10.5563

Shrimpton, P. C., Wall, B. F., Jones, D. G., Fisher, E. S., Hillier, M. C., Kendall, G. M., & Harrison, R. M. (1986). Doses to patients from routine diagnostic X-ray examinations in England. *British Journal of Radiology*, *59*(704), 749–758. http://doi.org/10.1259/0007-1285-59-704-749

Simpson, P. D., Martin, C. J., Darragh, C. L., & Abel, R. (1998). A study of chest radiography with mobile X-ray units. *British Journal of Radiology*, *71*(JUNE), 640–645. http://doi.org/10.1259/bjr.71.846.9849388

Singh, G., Martin, C. J., McCurrach, A., & Phanco, G. (2013). Optimisation of performance for computed radiography in the West of Scotland. *Radiation Protection Dosimetry*, *154*(3), 293–300. http://doi.org/10.1093/rpd/ncs238

Singh, V. H., & Pradhan, H. (2015). Neonatal chest radiography - Comparing image quality and dose for contact-techniques vs. under-tray techniques. *Radiography Open*, *2*(1), 65. http://doi.org/10.7577/radopen.1530

Smans, K., Struelens, L., Smet, M., Bosmans, H., & Vanhavere, F. (2010). Cu filtration for dose reduction in neonatal chest imaging. *Radiation Protection Dosimetry*, *139*(1–3), 281–286. http://doi.org/10.1093/rpd/ncq061

Smedby, Ö., & Fredrikson, M. (2010). Visual grading regression: Analysing data from visual grading experiments with regression models. *British Journal of Radiology*, *83*(993), 767–775. http://doi.org/10.1259/bjr/35254923

Smet, M. H., Breysem, L., Mussen, E., Bosmans, H., Marshall, N. W., & Cockmartin, L. (2018). Visual grading analysis of digital neonatal chest phantom X-ray images: Impact

of detector type, dose and image processing on image quality. *European Radiology*, *28*(7), 2951–2959. http://doi.org/10.1007/s00330-017-5301-2

Speets, A. M., van der Graaf, Y., Hoes, A. W., Kalmijn, S., Sachs, A. P. E., Rutten, M. J. C. M., … Mali, W. P. T. M. (2006). Chest radiography in general practice: Indications, diagnostic yield and consequences for patient management. *British Journal of General Practice*, *56*(529), 574–578. http://doi.org/10.1093/fampra/cml027

Struelens, L., Hambach, L., Buls, N., Smans, K., Malchair, F., Hoornaert, M. T., … Bosmans, H. (2008). Image quality assessment using the CD-DISC phantom for vascular radiology and vascular surgery. *European Journal of Radiology*, *67*(2), 348–356. http://doi.org/10.1016/j.ejrad.2007.07.019

Suliman, I. I., & Mohammedzein, T. S. (2014). Estimation of adult patient doses for common diagnostic X-ray examinations in Wad-madani, Sudan: Derivation of local diagnostic reference levels. *Australasian Physical and Engineering Sciences in Medicine*, *37*(2), 425–429. http://doi.org/10.1007/s13246-014-0255-z

Sun, Z., Lin, C., Tyan, Y., & Ng, K. H. (2012). Optimization of chest radiographic imaging parameters: A comparison of image quality and entrance skin dose for digital chest radiography systems. *Clinical Imaging*, *36*(4), 279–286. http://doi.org/10.1016/j.clinimag.2011.09.006

Sund, P., Båth, M., Kheddache, S., & Månsson, L. G. (2004). Comparison of visual grading analysis and determination of detective quantum efficiency for evaluating system performance in digital chest radiography. *European Radiology*, *14*(1), 48–58. http://doi.org/10.1007/s00330-003-1971-z

Sund, P., Herrmann, C., Tingberg, A., Kheddache, S., Månsson, L. G., Almén, A., & Mattsson, S. (2000). Comparison of two methods for evaluating image quality of chest radiographs. *Proceedings of SPIE*, *3981*(April), 251–257.

Swets, J. A., & Pickett, R. M. (1982). *Evaluation of diagnostic systems. Methods from signal detection theory.* http://doi.org/http://dx.doi.org/10.1016/B978-0-12-679080-1.50006-2

Swets, J. a, Green, D. M., Getty, D. J., & Swets, J. B. (1978). Signal detection and identification at successive stages of observation. *Perception & Psychophysics*, *23*(4),

275–289. http://doi.org/10.3758/BF03199711

Szczepura, K., Faqir, T., & Manning, D. (2017). Effects of increased compression with an ultrasound transducer on the conspicuity of breast lesions in a phantom. *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, *10136*, 1–11. http://doi.org/10.1117/12.2254263

Szczepura, K. R., & Manning, D. J. (2016). Validated novel software to measure the conspicuity index of lesions in DICOM images. In C. K. Abbey & M. A. Kupinski (Eds.), *SPIE 9787, Medical Imaging 2016: Image Perception, Observer Performance, and Technology Assessment* (p. 978703). California, United States: International Society for Optics and Photonics. http://doi.org/10.1117/12.2216211

Szczepura, K., Tomkinson, D., & Manning, D. (2017). Impact of tube current modulation on lesion conspicuity index in hi-resolution chest computed tomography. *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, *10136*, 1–12. http://doi.org/10.1117/12.2254256

Tang, K., Wang, L., Li, R., Lin, J., Zheng, X., & Cao, G. (2012). Effect of Low Tube Voltage on Image Quality, Radiation Dose, and Low-Contrast Detectability at Abdominal Multidetector CT: Phantom Study. *Journal of Biomedicine and Biotechnology*, *2012*(Lcd), 1–6. http://doi.org/10.1155/2012/130169

Tapiovaara, M., & , Siiskonen, T. (2008). *A Monte Carlo program for calculating patient doses in medical x-ray examinations* (2nd ed.). Helsinki,Finland: Radiation and Nuclear Safety Authority.

Tapiovaara, M. J. (2008). Review of relationships between physical measurements and user evaluation of image quality. *Radiat Prot Dosimetry*, *129*(1–3), 244–248. http://doi.org/10.1093/rpd/ncn009

Tapiovaara, M. J., & Sandborg, M. (2004). How should low-contrast detail detectability be measured in fluoroscopy? *Medical Physics*, *31*(9), 2564–2576. http://doi.org/10.1118/1.1779357

The Royal College of Radiologists. (2012). *Picture archiving and communication systems (PACS) and guidelines on diagnostic display devices*.

Thompson, J. D., Manning, D. J., & Hogg, P. (2013). The Value of Observer Performance Studies in Dose Optimization: A Focus on Free-Response Receiver Operating Characteristic Methods. *Journal of Nuclear Medicine Technology*, *41*(2), 57–64. http://doi.org/10.2967/jnmt.112.116566

Tingberg, A. (2000). *Quantifying the quality of medical X-ray images. An evaluation based on normal anatomy for lumbar spine and chest radiography*. *Thesis, Lund University, Malmo¨ , Sweden*. http://doi.org/10.1179/174328108X301688

Tingberg, A., Båth, M., Håkansson, M., Medin, J., Besjakov, J., Sandborg, M., … Månsson, L. G. (2005). Evaluation of image quality of lumbar spine images: A comparison between FFE and VGA. *Radiation Protection Dosimetry*, *114*(1–3), 53–61. http://doi.org/10.1093/rpd/nch566

Tingberg, A., Båth, M., Håkansson, M., Medin, J., Sandborg, M., Alm-carlsson, G., … Månsson, L. G. (2000). Comparison of two methods for evaluation of the image quality of lumbar spine radiographs. *Radiation Protection Dosimetry*, *90*(1), 165–168. http://doi.org/10.1093/oxfordjournals.rpd.a033110

Tingberg, A., Herrmann, C., Besjakov, J., Rodenacker, K., Almén, A., Sund, P., … Månsson, L. G. (2000). Evaluation of lumbar spine images with added pathology. *Proc. SPIE*, *3981*, 34–42.

Tingberg, A., Herrmann, C., Lanhede, B., Almén, A., Sandborg, M., McVey, G., … Zankl, M. (2004). Influence of the characteristic curve on the clinical image quality of lumbar spine and chest radiographs. *British Journal of Radiology*, *77*(915), 204–215. http://doi.org/10.1259/bjr/22642890

Tingberg, A., & Sjöström, D. (2005). Optimisation of image plate radiography with respect to tube voltage. *Radiation Protection Dosimetry*, *114*(1–3), 286–293. http://doi.org/10.1093/rpd/nch536

Tsai, D.-Y., Lee, Y., & Matsuyama, E. (2008). Information Entropy Measure for Evaluation of Image Quality. *Journal of Digital Imaging*, *21*(3), 338–347. http://doi.org/10.1007/s10278-007-9044-5

Tugwell, J., Everton, C., Kingma, A., Oomkens, D. M., Pereira, G. A., Pimentinha, D. B., … Hogg, P. (2014). Increasing source to image distance for AP pelvis imaging – Impact

on radiation dose and image quality. *Radiography*, *20*(4), 351–355. http://doi.org/10.1016/j.radi.2014.05.012

Tung, C. J., Tsai, H. Y., Shi, M. Y., Huang, T. T., Yang, C. H., & Chen, I. J. (2007). A phantom study of image quality versus radiation dose for digital radiography. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *580*(1), 602–605. http://doi.org/10.1016/j.nima.2007.05.232

Uffmann, M., Schaefer-Prokop, C., Neitzel, U., Weber, M., Herold, C. J., & Prokop, M. (2004). Skeletal Applications for Flat-Panel versus Storage-Phosphor Radiography: Effect of Exposure on Detection of Low-Contrast Details. *Radiology*, *231*(2), 506–514. http://doi.org/10.1148/radiol.2312021662

Ujah, F. O., Agba, N. B. A. E. H., & Iortile, T. J. (2012). A comparative study of patients radiation levels with standard diagnostic reference levels in federal medical centre and bishop murray hospitals in Makurdi, *4*(2), 800–804.

Uppot, R. N., Sahani, D. V., Hahn, P. F., Gervais, D., & Mueller, P. R. (2007). Impact of obesity on medical imaging and image-guided intervention. *American Journal of Roentgenology*, *188*(2), 433–440. http://doi.org/10.2214/AJR.06.0409

Vaishnav, J. Y., Jung, W. C., Popescu, L. M., Zeng, R., & Myers, K. J. (2014). Objective assessment of image quality and dose reduction in CT iterative reconstruction. *Medical Physics*, *41*(7), 071904. http://doi.org/10.1118/1.4881148

Van Soldt, R. T. M., Zweers, D., Van Den Berg, L., Geleijns, J., Jansen, T. M., & Zoetelief, J. (2003). Survey of posteroanterior chest radiography in The Netherlands: Patient dose and image quality. *British Journal of Radiology*, *76*(906), 398–405. http://doi.org/10.1259/bjr/76222078

Vano, E., Geiger, B., Schreiner, A., Back, C., & Beissel, J. (2005). Dynamic flat panel detector versus image intensifier in cardiac imaging: dose and image quality. *Physics in Medicine and Biology*, *50*(23), 5731–5742. http://doi.org/10.1088/0031-9155/50/23/022

Varchena, V. (2002). Pediatric phantoms. *Pediatric Radiology*, *32*(4), 280–284. http://doi.org/10.1007/s00247-002-0681-z

Veldkamp, W. J. H., Kroft, L. J. M., Boot, M. V., Mertens, B. J. A., & Geleijns, J. (2006). Contrast-detail evaluation and dose assessment of eight digital chest radiography systems in clinical practice. *European Radiology*, *16*(2), 333–341. http://doi.org/10.1007/s00330-005-2887-6

Veldkamp, W. J. H., Kroft, L. J. M., & Geleijns, J. (2009). Dose and perceived image quality in chest radiography. *European Journal of Radiology*, *72*(2), 209–217. http://doi.org/10.1016/j.ejrad.2009.05.039

Vennart, W. (1997). *ICRU Report 54: Medical imaging—the assessment of image quality: ISBN 0-913394-53-X. April 1996, Maryland, USA*.

Wall, B. F., Harrison, R. M., & Spiers, F. W. (1988). Patient Dosimetry Techniques in Diagnostic Radiology. *Nstitute of Physical Sciences in Medicine.*, *53*.

Wall, B. F., & Hart, D. (1997). Revised radiation doses for typical X-ray examinations: Report on a recent review of doses to patients from medical X-ray examinations in the UK by NRPB. *British Journal of Radiology*, *70*(MAY), 437–439. http://doi.org/doi:10.1259/bjr.70.833.9227222

Warren-Forward, H. M., Haddaway, M. J., McCall, I. W., & Temperton, D. H. (1996). Influence of dose reduction recommendations on changes in chest radiography techniques. *British Journal of Radiology*, *69*(824), 755–761. http://doi.org/10.1259/0007-1285-69-824-755

Warren-Forward, H. M., & Millar, J. S. (1995). Optimization of radiographic technique for chest radiography. *British Journal of Radiology*, *68*(815), 1221–1229. http://doi.org/10.1259/0007-1285-68-815-1221

Warren, L. M., Mackenzie, A., Cooke, J., Given-Wilson, R. M., Wallis, M. G., Chakraborty, D. P., … Young, K. C. (2012). Effect of image quality on calcification detection in digital mammography. *Medical Physics*, *39*(6Part1), 3202–3213. http://doi.org/10.1118/1.4718571

Ween, B., Olstad, M., Jakobsen, J. Å., & Olsen, D. R. (2009). Pediatric digital chest radiography, comparison of grid versus non-grid techniques. *European Journal of Radiography*, *1*(4), 201–206. http://doi.org/10.1016/j.ejradi.2010.09.002

Williams, M. B., Krupinski, E. A., Strauss, K. J., Breeden, W. K., Rzeszotarski, M. S., Applegate, K., … Seibert, J. A. (2007). Digital Radiography Image Quality: Image Acquisition. *Journal of the American College of Radiology*, *4*(6), 371–388. http://doi.org/10.1016/j.jacr.2007.02.002

Willis, C. E. (2009). Optimizing digital radiography of children. *European Journal of Radiology*, *72*(2), 266–273. http://doi.org/10.1016/j.ejrad.2009.03.003

World Health Organization. (2000). *Obesity : preventing and managing the global epidemic : report of a WHO consultation.* World Health Organization.

Yanch, J. C., Behrman, R. H., Hendricks, M. J., & McCall, J. H. (2009). Increased Radiation Dose to Overweight and Obese Patients from Radiographic Examinations. *Radiology*, *252*(1), 128–139. http://doi.org/10.1148/radiol.2521080141

Yano, Y., Yabuuchi, H., Tanaka, N., Morishita, J., Akasaka, T., Matsuo, Y., … Sasaki, M. (2013). Detectability of simulated pulmonary nodules on chest radiographs: Comparison between irradiation side sampling indirect flat-panel detector and computed radiography. *European Journal of Radiology*, *82*(11), 2050–2054. http://doi.org/10.1016/j.ejrad.2013.05.036

Zarb, F., Rainford, L., & McEntee, M. F. (2010). Image quality assessment tools for optimization of CT images. *Radiography*, *16*(2), 147–153. http://doi.org/10.1016/j.radi.2009.10.002

Zoetelief, J. (2013). *Radiological protection in paediatric diagnostic and interventional radiology*. Retrieved from https://academic.oup.com/rpd/article-abstract/155/3/380/1612791