

**Task-specific observer performance and image quality:
direct and indirect relationships
in low-dose CT images.**

PhD Thesis

Maryam Jessop

(Linda Maree Maryam Jessop)

2021

Student ID: 00346914

School of Health and Society

University of SALFORD

Supervisors (Current):

Dr J.D. Thompson

Dr R.N. Higgins

Contents:

I.	Acknowledgements	Page 7
II.	List of Tables	Page 8
III.	List of Figures (including graphs)	Page 13
IV.	List of Appendices	Page 25
V.	Image quality metrics preview	Page 26
VI.	Abstract	Page 27
1.0	Chapter One: Introduction	Page 28
1.1	Context	Page 28
1.2	Aim	Page 29
1.3	Method	Page 29
1.4	Results and discussion	Page 30
1.5	Summary	Page 31
2.0	Chapter Two: Literature review	Page 32
2.1	Introduction	Page 32
2.2	Computed tomography (CT) and low-dose CT (LDCT)	Page 33
2.2.1	Computed tomography (CT)	Page 33
2.2.2	Low dose CT (LDCT)	Page 38
2.2.2.1	LDCT thorax	Page 40
2.2.2.2	LDCT abdomen and pelvis	Page 42
2.3	Image quality and image interpretation	Page 45
2.3.1	Physical image quality assessment	Page 46

2.3.1.1	Noise	Page 46
2.3.1.2	Mathematical image quality measures	Page 56
2.3.1.3	Perceptual characteristics – image quality measures	Page 57
2.3.2	Visual perception	Page 60
2.3.3	Medical imaging	Page 64
2.3.4	Psychophysical measures	Page 67
2.3.4.1	Visual grading analysis (VGA)	Page 68
2.3.4.2	Receiver operating characteristic (ROC) analysis	Page 70
2.3.4.3	Visual grading characteristics (VGC)	Page 72
2.3.4.4	Equivalency	Page 72
2.4	Statistical relationships	Page 73
3.0	Chapter Three: Hypothesis and Null hypothesis	Page 76
3.1	Hypothesis	Page 76
3.2	Null hypothesis	Page 76
4.0	Chapter Four: Method	Page 77
4.1	Introduction	Page 77
4.2	Methodology	Page 78
4.3	Equipment	Page 79
4.3.1	CT equipment	Page 79
4.3.2	Imaging phantoms	Page 80
4.3.2.1	Catphan®	Page 80

4.3.2.2	'LUNGMAN' [®] anthropomorphic chest phantom	Page 82
4.4	Data acquisition	Page 84
4.4.1	Preparatory images	Page 84
4.4.2	Main study	Page 85
4.5	Image analysis	Page 88
4.5.1	Image analysis using MATLAB [®]	Page 89
4.5.2	Image analysis using LIFEx	Page 92
4.5.3	Image analysis using ImageJ	Page 94
4.6	Observer performance: Visual Grading Characteristics (VGC) and Receiver Operating Curve (ROC) analysis	Page 97
4.6.1	Recruitment	Page 98
4.6.2	Viewing and environmental conditions	Page 99
4.6.3	VGC and ROC data collection	Page 100
4.6.4	VGC analysis	Page 103
4.6.5	JAFROC	Page 103
5.0	Chapter Five: Results - Image Quality Metrics (IQM)	Page 106
5.1	Introduction	Page 106
5.2	Radiation exposure information	Page 107
5.3	Mean squared error (MSE)	Page 107
5.4	Peak signal-to-noise ratio (PSNR)	Page 114
5.5	Signal-to-noise ratio (SNR)	Page 120

5.6	Contrast-to-noise ratio (CNR)	Page 126
5.7	Structural similarity index metric (SSIM)	Page 132
5.8	Non-shift edge ratio (NSER)	Page 139
5.9	Noise power spectrum (NPS)	Page 146
5.10	Texture analysis	Page 155
	5.10.1 Contrast	Page 155
	5.10.2 Correlation	Page 163
	5.10.3 Energy	Page 171
	5.10.4 Homogeneity	Page 179
	5.10.5 Entropy	Page 187
5.11	Best performing image quality metrics (IQM)	Page 195
6.0	Chapter Six: Results - Observer Performance	Page 196
6.1	Participant demographics	Page 196
6.2	VGC analysis	Page 197
6.3	Fulfilment of Image Quality Criteria	Page 200
6.4	ROC analysis	Page 204
6.5	Missed lesion detection	Page 205
6.6	Time-case-lesion analysis	Page 207
7.0	Chapter Seven: Correlations and Associations	Page 210
7.1	Introduction	Page 210
7.2	Multivariate correlations	Page 211
	7.2.1 Pearson product-correlation (Pearson's r coefficient)	Page 211

7.2.2	Spearman's ρ correlation coefficient	Page 215
7.2.3	Summary IQM multivariate correlations and associations	Page 217
7.2.4	Correlation between IQMs and Observer Performance	Page 218
8.0	Chapter Eight: Discussion	Page 223
9.0	Chapter Nine: Conclusion	Page 232
	References	Page 235
	Glossary	Page 249
	Appendices	Page 251

I. Acknowledgements:

I wish to thank my supervisors, past and present, and acknowledge all their kind efforts in bringing my research to fruition. Previous supervisors were Professor Peter Hogg who has been on this journey from the very beginning, and Professor David Manning whose expertise in the field was gratefully appreciated. I also wish to thank the post-graduate support team for all their kind assistance throughout this time.

As this project has been undertaken part-time while I have been working in a busy Imaging department, I could not have managed without the support of the team at Brighton and Sussex University Hospitals NHS Trust. I give a special thanks to the Clinical Physics team for their kind assistance and Emma for adapting some of her coding to assist in multi-slice analysis. Last, but not least, I thank Professor Sabina Dizdarevic for her continued support and encouragement throughout.

My inspiration has always been my daughters in New Zealand. Our dreams are there for us to grasp and turn into reality, by the grace of God.

II. List of tables:

Table 1	Classification of artefacts on the basis of appearance. (<i>table 9-2 in Seeram, 2009f</i>)	Page 50
Table 2	Gestalt principles. (<i>table 1 in Sabih et al., 2011</i>)	page 63
Table 3	Examples for visual grading rating scales. (<i>Table 1 in Ludewig, Richter, & Frame, 2010</i>)	Page 69
Table 4	Summary of ROC methods. (Chakraborty & Berbaum 2004; Chakraborty 2004, 2005, 2006; <i>table 1 in Zarb et al., 2010</i>)	Page 71
Table 5	Further ROC developments. (Chakraborty, 2013; Chakraborty, 2015)	Page 71
Table 6	Simulated lesion size placement (Legend).	Page 86
Table 7	Image acquisition and reconstruction parameters.	Page 87
Table 8	Image quality metrics (IQMs) and related image characteristics they assess.	Page 88
Table 9	Texture descriptors in a GLCM (Mathworks, 1994-2020).	Page 94
Table 10	ROI size reduction to geometric unsharpness.	Page 95
Table 11	CT dose information.	Page 107
Table 12	Correlation of MSE graph values with averaged data series (Pearson correlation coefficient r).	Page 109
Table 13	T-test for average MSE data series with and without attenuation jackets.	Page 110
Table 14	MSE upper zone where improved values at lower amperage translates into reduced radiation exposure (CTDIvol).	Page 111
Table 15	Correlation of PSNR graph values with averaged data series. (Pearson correlation coefficient r).	Page 115

Table 16	T-test for average PSNR data series with and without attenuation jackets.	Page116
Table 17	PSNR mid zone where improved values at lower amperage translates into reduced radiation exposure (CTDI _{vol}).	Page 117
Table 18	Correlation of SNR graph values with averaged data series (Pearson correlation coefficient r).	Page 121
Table 19	Correlation of CNR graph values with averaged data series (Pearson correlation coefficient r).	Page 128
Table 20	Correlation of SSIM graph values with averaged data series (Pearson correlation coefficient values r).	Page 133
Table 21	T-test for average SSIM data series with and without attenuation jackets.	Page 134
Table 22	SSIM mid zone where improved values at lower amperage translates into reduced radiation exposure (CTDI _{vol}).	Page 135
Table 23	Correlation of NSER graph values with averaged data series (Pearson correlation coefficient r).	Page 140
Table 24	T-test for average NSER data series with and without attenuation jackets.	Page 141
Table 25	NSER mid zone where improved values at lower amperage translates into reduced radiation exposure (CTDI _{vol}).	Page 142
Table 26	Peak groups: normalized frequency ranges.	Page 149
Table 27	Contrast: Lower zone left lung for example in <i>figure 93</i> , where improved values at lower amperage potentially translate into reduced radiation exposure (CTDI _{vol}).	Page 157
Table 28	Contrast: Difference in values between 10mm and 8mm simulated lesions at varying density, expressed as a percentage.	Page 160
Table 29	Correlation mid zone left lung for example in <i>figure 101</i> , where improved values at lower amperage potentially translate into reduced radiation exposure (CTDI _{vol}).	Page 166

Table 30	Correlation: Difference in values between 10mm and 8mm simulated lesions at varying density, expressed as a percentage.	Page 168
Table 31	Correlation of Energy graph values with averaged data series (Pearson correlation coefficient values r).	Page 172
Table 32	T-test for averaged Energy data series with and without attenuation jackets.	Page 173
Table 33	Energy: for averaged data series where improved values at lower amperage translate into potential reduction in radiation exposure (CTDI _{vol}).	Page 174
Table 34	Difference in Energy values between 10mm and 8mm simulated lesions at varying density, expressed as a percentage.	Page 176
Table 35	Correlation of Homogeneity values with averaged data series (Pearson correlation coefficient values r).	Page 180
Table 36	T-test for averaged homogeneity data series with and without attenuation jackets.	Page 181
Table 37	Homogeneity: for averaged data series where improved values at lower amperage translate into potential reduction in radiation exposure (CTDI _{vol}).	Page 182
Table 38	Difference in Homogeneity values between 10mm and 8mm simulated lesions at varying density, expressed as a percentage.	Page 184
Table 39	Correlation of Entropy values with averaged data series (Pearson correlation coefficient values r).	Page 188
Table 40	T-test for averaged Entropy data series with and without attenuation jackets.	Page 189
Table 41	Entropy: for averaged data series where improved values at lower amperage translate into potential reduction in radiation exposure (CTDI _{vol}).	Page 190
Table 42	Difference in Entropy values between 10mm and 8mm simulated lesions at varying density, expressed as a	Page 192

percentage.

Table 43	Effect of using attenuation jackets seen in graphed data with changes in gradient and y-intercept.	Page 195
Table 44	Participant backgrounds and CT viewing experience.	Page 196
Table 45	Legend (<i>Table 41</i>).	Page 196
Table 46	Image quality criteria: p-values for Mann-Whitney significance testing within regions.	Page 200
Table 47	Image quality criteria: p-values for Mann-Whitney significance testing between regions (with and without attenuation jackets).	Page 201
Table 48	Incidence of missed lesion detection relating to lesion density, use of attenuation jackets and size.	Page 205
Table 49	Pearson's r values with strongest positive correlations (without attenuation jackets).	Page 212
Table 50	Pearson's r values with strongest negative correlations (without attenuation jackets).	Page 212
Table 51	Pearson's r values with strongest positive correlations (with attenuation jackets).	Page 214
Table 52	Pearson's r values with strongest negative correlations (with attenuation jackets).	Page 214
Table 53	Pearson's r values with correlations for IQMs with and without attenuation jackets.	Page 214
Table 54	Spearman's ρ values with strongest positive correlations (without attenuation jackets).	Page 215
Table 55	Spearman's ρ values with strongest negative correlations (without attenuation jackets).	Page 215
Table 56	Spearman's ρ values with strongest positive correlations (with attenuation jackets).	Page 216

Table 57	Spearman's ρ values with strongest negative correlations (with attenuation jackets).	Page 216
Table 58	Spearman's ρ values with correlations for IQMs with and without attenuation jackets.	Page 216
Table 59	Table 55. Spearman's ρ values with strongest positive and negative correlations (with and without attenuation jackets), for proximal structures (upper zone). Weakest correlations highlighted in grey.	Page 219
Table 60	Spearman's ρ values with strongest positive and negative correlations (with and without attenuation jackets), for proximal structures (lower zone). Weakest correlations highlighted in grey.	Page 219
Table 61	Spearman's ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (upper zone). Weakest correlations highlighted in grey.	Page 220
Table 62	Spearman's ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (middle zone). Weakest correlations highlighted in grey.	Page 221
Table 63	Spearman's ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (lower zone). Weakest correlations highlighted in grey.	Page 221

III. List of Figures (including graphs):

Figure 1	Computed Tomography Process. (<i>Fig 16-1 in Computed Tomography [RadiologyKey], 2016</i>)	page 34
Figure 2	Filtered back-projection. (<i>based on information from Seeram, 2009d</i>)	Page 34
Figure 3	Data reconstruction in Filtered-Back-Projection. (<i>Fig 1 in Geyer et al., 2015</i>).	Page 34
Figure 4	Logic behind iterative reconstruction. (<i>Fig 10-55 in Bushberg et al., 2012b</i>)	Page 36
Figure 5	Schematic representation of the principal steps of iterative image algorithms. (<i>Fig 2 in Geyer et al., 2015</i>)	page 36
Figure 6	Diagrammatic representation of Helical pitch.	Page 38
Figure 7	Dose modulation (CARE Dose4D) (<i>in Siemens, 2009</i>)	Page 39
Figure 8	Image noise in CT (<i>Fig 10-12 in Bushberg et al., 2012b</i>)	Page 47
Figure 9	Electronic noise in CT detectors (<i>Fig 1 in Duan et al., 2013</i>)	Page 48
Figure 10	Electronic noise (<i>Fig 4-21 B. in Bushberg et al., 2012a</i>)	Page 49
Figure 11	Streak artefact at low mA using anthropomorphic chest phantom.	Page 51
Figure 12	Noise power spectrum NPS(f) is shown for uncorrelated and correlated noise. (<i>fig 4-28 in Bushberg et al., 2012a</i>)	Page 53
Figure 13	NPS 3D: Same object slice acquired at (A)10mA and (B)200mA, from images acquired using an anthropomorphic chest phantom and analysed in MATLAB®, producing a pseudospectrum estimate via the Eigenvector method.	Page 55
Figure 14	Diagrammatic representation of visual perception. (<i>information from Gonzalez & Woods, 2008a</i>)	Page 60
Figure 15	Interpretation errors. (<i>information from Samei & Krupinski, 2010</i>)	Page 65

Figure 16	Inter-relation of thesis major sub-components.	Page 77
Figure 17	CATPHAN®. (in "CATPHAN 600 [Brochure 'CATPHAN 500/600']," n.d.)	Page 81
Figure 18	CATPHAN® inserts for module CTC515. (in "Catphan 500 and 600 Manual," 2015)	Page 81
Figure 19	Image of CATPHAN® phantom using CTP515 low-contrast module.	Page 81
Figure 20	CATPHAN® annuli. (in "Catphan accessories," 2018)	Page 81
Figure 21	Image of CATPHAN® using a CTP515 low-contrast module with annulus model CT579.	Page 81
Figure 22	'LUNGMAN'® (in Kagaku, n.d.)	Page 82
Figure 23	Figure 23. Mediastinum and pulmonary vessel insert for 'LUNGMAN'® (in Kagaku, n.d.)	Page 82
Figure 24	Simulated lesions 5mm, 8mm, 10mm. (partial image from Kagaku, n.d.)	Page 82
Figure 25	Simulated lesion insertion. (in Kagaku, n.d.)	Page 83
Figure 26	Image of 'LUNGMAN'® anthropomorphic chest phantom.	Page 83
Figure 27	Acquisition workflow.	Page 85
Figure 28	Simulated lesion placement.	Page 86
Figure 29	Graphic of Non-shift-edge ratio: (C) reference image matrix; (D) distorted image matrix: similarities to (C) in green and shift-edges indicated in red; (E) frequency diagram indicating zero-crossings with arrows.	Page 90
Figure 30	Power Spectral Density (at 10mA) obtained from the Yule Walker AR (autocorrelation) method.	Page 92
Figure 31	Variance from image analysis in figure 30, plotted via the Eigenvector method.	Page 92
Figure 32	Regions of interest for texture analysis in LIFEx.	Page 93

Figure 33	Background ROI ImageJ.	Page 95
Figure 34	ROI ImageJ, over simulated lesion at full size.	Page 95
Figure 35	ROI ImageJ, over simulated lesion – smaller to reduce geometric unsharpness.	Page 95
Figure 36	Effect of ROI size reduction for simulated lesions to reduce geometric unsharpness relating to designated HU (100HU).	Page 95
Figure 37	Effect of ROI size reduction for simulated lesions to reduce geometric sharpness relating to designated HU (-630HU and -800HU).	Page 96
Figure 38	Observer performance sequence of events: recruitment to participation.	Page 99
Figure 39	ViewDEX display screen.	Page 101
Figure 40	Area requiring participant attention when assessing peripheral vessels indicated by the purple colour.	Page 102
Figure 41	Sample size calculation. (graph) (<i>information from Obuchowski, 2000</i>)	Page 105
Figure 42	CT dose index $CTDI_{vol}$ (mA vs mGy)	Page 107
Figure 43	MSE graphs (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets - expanded to match graph A, (C) with attenuation jackets.	Page 108
Figure 44	MSE: Graphs of averaged data series, with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).	Page 109
Figure 45	(A) MSE upper zone with improvement (from red line to green line): 90mA to 80mA, 50mA to 45mA, 35mA to 30mA; (B) MSE upper zone when attenuation jackets used, with improvement (from red line to green line): 90mA to 80mA, 60mA to 50mA.	Page 110
Figure 46	MSE: Frequency meta-analysis, of occurrence at mA where	Page 111

improvement was possible, and at mA where improvement occurred.

Figure 47	MSE: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 112
Figure 48	MSE: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 113
Figure 49	PSNR graphs (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets.	Page 114
Figure 50	PSNR: Graphs of averaged data series, with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).	Page 115
Figure 51	PSNR mid zone with improvement (from red line to green line): 100mA to 80mA, 70mA to 45mA, 40mA to 30mA; (B) PSNR mid zone when attenuation jackets used, with improvement (from red line to green line): 90mA to 80mA, 70mA to 40mA.	Page 117
Figure 52	PSNR: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.	Page 118
Figure 53	PSNR: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 118
Figure 54	PSNR: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 119
Figure 55	SNR graphs (relating to the simulated lesions): (A) without attenuation jackets , (B) with attenuation jackets.	Page 120
Figure 56	SNR for -630HU 8mm lesion located in lower zone (A) without attenuation jacket, with improvement (from red line to green line): 90mA to 80mA, 70mA to 60mA, 50mA to 30mA, 25mA to 15mA; (B) with attenuation jackets, with improvement (from red line to green line): 80mA to 70mA, 60mA to 35mA, 30mA to 20mA.	Page 122

Figure 57	Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.	Page 122
Figure 58	SNR: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 123
Figure 59	SNR: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 123
Figure 60	SNR demonstrates change with size and density of simulated lesion (A) 100HU 8mm and 10mm with attenuation jackets (B) -800HU 8mm and 10mm with attenuation jackets.	Page 124
Figure 61	SNR demonstrates little change in comparison of simulated Lesions with and without attenuation jackets.	Page 125
Figure 62	CNR graphs (relating to the simulated lesions): (A) 100HU without attenuation jackets, (B) 100HU with attenuation jackets.	Page 126
Figure 63	CNR graphs (relating to the simulated lesions): (A) -630HU and -800HU without attenuation jackets, (B) -630HU and -800HU with attenuation jackets.	Page 127
Figure 64	CNR (A) for 100HU 10mm lesion located in mid zone with improvement (from red line to green line) from : 100mA to 80mA, 60mA to 50mA, 50mA to 45mA, 40mA to 35mA, 20mA to 15mA; (B) for 100HU 10mm lesion located in mid zone when attenuation jackets used, with improvement (from red line to green line): from 100mA to 90mA, 70mA to 60mA, 40mA to 35mA, 25mA to 20mA.	Page 129
Figure 65	CNR: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.	Page 130
Figure 66	CNR: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 130
Figure 67	CNR: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 131

Figure 68	SSIM graphs (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets.	Page 132
Figure 69	SSIM: Graphs of averaged data series, with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).	Page 133
Figure 70	SSIM mid zone (A) with improvement (from red line to green line): 100mA to 90mA, 70mA to 60mA, 45mA to 40mA, 35mA to 25mA, 20mA to 15mA; (B) when attenuation jackets used, with improvement (from red line to green line): 100mA to 90mA, 80mA to 70mA, 60mA to 50mA, 40mA to 35mA.	Page 135
Figure 71	SSIM: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.	Page 136
Figure 72	SSIM: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 137
Figure 73	SSIM: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 137
Figure 74	NSER graphs (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets.	Page 139
Figure 75	NSER: Graphs of averaged data series, with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).	Page 140
Figure 76	A) NSER lower zone with improvement (red line to green line) from: 60A to 45mA, 35mA to 30mA; (B) NSER lower zone when attenuation jackets used, with improvement (red line to green line) from 80mA to 70mA, 45mA to 35mA.	Page 141
Figure 77	NSER: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.	Page 142
Figure 78	NSER: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 143

Figure 79	NSER: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 144
Figure 80	NPS: lower zone.	Page 146
Figure 81	NPS: lower zone magnified (normalized frequency range 0.15 - 0.4), with examples of peak locations marked by arrows.	Page 147
Figure 82	NPS: lower zone (with attenuation jackets).	Page 148
Figure 83	NPS: lower zone (with attenuation jackets) magnified (normalised frequency range 0.35 – 1), with examples of peak locations marked by arrows.	Page 149
Figure 84	NPS peak grouping (lower zone).	Page 150
Figure 85	NPS peak grouping (lower zone with attenuation jackets).	Page 150
Figure 86	Simplified NPS lower zone.	Page 151
Figure 87	Simplified NPS lower zone with attenuation jackets.	Page 151
Figure 88	NPS peak values / amperage, with linear trend line.	Page 151
Figure 89	NPS zone comparison with datasets obtained at 50mA. Image display from top to bottom for upper zone, middle zone and lower zone respectively.	Page 152
Figure 90	NPS Peak distributions for middle zone (A) Kurtosis and (B) Skewness.	Page 153
Figure 91	Contrast: Graphs for mid zone (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets.	Page 155
Figure 92	Contrast: Graphs for upper and lower zones (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets.	Page 156
Figure 93	Contrast: Lower zone zones (where HU refers to the simulated lesion density related to that image series) (A) without attenuation jackets, with improvement (red line to green line)	Page 157

from: 80mA to 70mA, 60mA to 50mA, 25mA to 10mA;
 (B) with attenuation jackets, with improvement (red line to green line) from 100mA to 80mA, 45mA to 35mA, 30mA to 20mA.

Figure 94	Contrast: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.	Page 158
Figure 95	Contrast: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 159
Figure 96	Contrast: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 159
Figure 97	Contrast: Lower zone left lung (where HU refers to the simulated lesion density relating to each image series, and Average aj refers to the average of these data series only). Consistent peaks at 45mA and 90mA.	Page 160
Figure 98	Contrast: Graphs for 10mm and 8mm simulated lesions showing differences in values relating to density (HU): (A) 10mm without attenuation jackets, (B) 10mm with attenuation jackets, (C) 8mm without attenuation jackets, (D) 8mm with attenuation jackets; NB. % differences are expressed in comparison to 100HU data series.	Page 161
Figure 99	Correlation: Graphs for mid zone (where HU refers to the simulated lesion density relating to that image series): (A) without attenuation jackets, (B) with attenuation jackets.	Page 163
Figure 100	Correlation: Graphs for upper and lower zones (where HU refers to the simulated lesion density related to that image series): (A) without attenuation jackets, (B) with attenuation jackets.	Page 164
Figure 101	Correlation: Mid zone (where HU refers to the simulated lesion density related to that image series) (A) without attenuation jackets, with improvement (red line to green line) from: (A) 100mA to 80mA, 70mA to 60mA, 45mA to 40mA; (B) with attenuation jackets, with improvement (red line to green line) from 90mA to 80mA, 70mA to 60mA, 45mA to 40mA, 35mA to 30mA.	Page 165

Figure 102	Correlation: Frequency meta-analysis of occurrence at mA where improvement was possible, and where improvement occurred.	Page 166
Figure 103	Correlation: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 167
Figure 104	Correlation: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 167
Figure 105	Correlation: Lower zone right lung (where HU refers to the simulated lesion density relating to each image series, and Average aj refers to the average of these data series only). Consistent peaks at 45mA and 90mA.	Page 168
Figure 106	Correlation: Graphs for 10mm and 8mm simulated lesions showing differences in values relating to density (HU): (A) 10mm without attenuation jackets, (B) 10mm with attenuation jackets, (C) 8mm without attenuation jackets, (D) 8mm with attenuation jackets; NB. percentage differences are expressed in comparison to 100HU data series.	Page 169
Figure 107	Energy: Averaged data graphed (A) without attenuation jackets, (B) with attenuation jackets.	Page 171
Figure 108	Energy: Averaged data graphs with and without attenuation jackets showing logarithmic trend lines and standard error bars (SE).	Page 172
Figure 109	Energy: Averaged data series, with improvement (red line to green line) from: 90mA to 80mA, 45mA to 40mA (without attenuation jackets).	Page 173
Figure 110	Energy: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.	Page 174
Figure 111	Energy: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 175

Figure 112	Energy: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 176
Figure 113	Energy: Graphs for 10mm and 8mm simulated lesions showing differences in values relating to density (HU): (A) 10mm without attenuation jackets, (B) 10mm with attenuation jackets, (C) 8mm without attenuation jackets, (D) 8mm with attenuation jackets; NB. percentage differences are expressed in comparison to 100HU data series.	Page 177
Figure 114	Homogeneity: Averaged data series graphed (A) without attenuation jackets, (B) with attenuation jackets.	Page 179
Figure 115	Homogeneity: Graphs of averaged data series, with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).	Page 180
Figure 116	Homogeneity mid zone (where HU refers to the simulated lesion density related to that image series) (A) without attenuation jackets, with improvement (red line to green line) from: 100mA to 90mA, 70mA to 60mA, 35mA to 30mA; (B) with attenuation jackets, with improvement (red line to green line) from 100mA to 90mA, 50mA to 45mA.	Page 181
Figure 117	Homogeneity: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.	Page 182
Figure 118	Homogeneity: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 183
Figure 119	Homogeneity: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 184
Figure 120	Homogeneity: Graphs for 10mm and 8mm simulated lesions showing differences in values relating to density (HU): (A) 10mm without attenuation jackets, (B) 10mm with attenuation jackets, (C) 8mm without attenuation jackets, (D) 8mm with attenuation jackets; NB. percentage differences are expressed in comparison to 100HU data series.	Page 185

Figure 121	Entropy: Averaged data series graphed (A) without attenuation jackets, (B) with attenuation jackets.	Page 187
Figure 122	Entropy: Averaged data series graphed with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).	Page 188
Figure 123	Entropy lower zone (where HU refers to the simulated lesion density related to that image series) (A) without attenuation jackets, with improvement (red line to green line) from: 100mA to 90mA, 60mA to 50mA, 20mA to 15mA; (B) with attenuation jackets, with improvement (red line to green line) from 90mA to 80mA, 60mA to 50mA, 45mA to 40mA.	Page 189
Figure 124	Entropy: Frequency meta-analysis, of occurrence at mA where improvement possible, and at mA where improvement occurred.	Page 191
Figure 125	Entropy: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.	Page 191
Figure 126	Entropy: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.	Page 192
Figure 127	Entropy: Graphs for 10mm and 8mm simulated lesions showing differences in values relating to density (HU): (A) 10mm without attenuation jackets, (B) 10mm with attenuation jackets, (C) 8mm without attenuation jackets, (D) 8mm with attenuation jackets; NB. percentage differences are expressed in comparison to 100HU data series.	Page 193
Figure 128	AUC _{VGC} with asymmetric 95% confidence intervals.	Page 198
Figure 129	Binormal VGC curve for first question set (proximal bronchi). [Reference condition = fulfilment of image quality criterion; Test condition = how much noise interfered].	Page 199
Figure 130	Binormal VGC curve for second question set (peripheral lung vessels). [Reference condition = fulfilment of image quality criterion; Test condition = how much noise interfered].	Page 199

Figure 131	Fulfilment of image quality criterion: Proximal structures in upper, middle and lower zones. [1 = criterion fulfilled; 5 = criterion not fulfilled].	Page 202
Figure 132	Fulfilment of image quality criterion: Peripheral structures in upper, middle and lower zones. [1 = criterion fulfilled; 5 = criterion not fulfilled].	Page 203
Figure 133	Incidence of missed lesion detection at varying amperage.	Page 207
Figure 134	Example of case versus time spent.	Page 208
Figure 135	Peaks in time taken related to images containing simulated lesions of different density.	Page 208
Figure 136	Peaks in time taken related to images acquires at varying amperage.	Page 209
Figure 137	Scatter plot matrix of relationships between IQMs (no attenuation jackets). Yellow squares indicate those relationships with Pearson's r values > 0.9 or < -0.9 . (Values for individual IQMs on both axes).	Page 211
Figure 138	Scatter plot matrix of relationships between IQMs (with attenuation jackets). Yellow squares indicate those relationships with Pearson's r values > 0.9 or < -0.9 . (Values for individual IQMs on both axes).	Page 213
Figure 139	Spearman's ρ values with strongest positive and negative correlations (with and without attenuation jackets), for proximal structures (upper zone). Weakest correlations highlighted in grey.	Page 219
Figure 140	Spearman's ρ values with strongest positive and negative correlations (with and without attenuation jackets), for proximal structures (lower zone). Weakest correlations highlighted in grey.	Page 219
Figure 141	Spearman's ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (upper zone). Weakest correlations highlighted in grey.	Page 220

Figure 142	Spearman's ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (middle zone). Weakest correlations highlighted in grey.	Page 221
Figure 143	Spearman's ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (lower zone). Weakest correlations highlighted in grey.	Page 221
Figure 144	Frequency meta-analysis, for Spearman's ρ correlation coefficients for IQM performance when correlated with observer performance data for fulfilment of image quality criteria.	Page 222

IV. List of Appendices:

Appendix 1	IQM formulas	Page 251
Appendix 2	University of Salford ethics approval	Page 253
Appendix 3	NHS Health Research Authority ethics approval	Page 254
Appendix 4	Participant information sheet	Page 255
Appendix 5	Research participant consent form	Page 258
Appendix 6	ROC input file legend	Page 260

V. Image quality metrics preview:

- Signal-to-noise ratio (SNR) assessed the detectability of an object by using mean pixel values over an object in relation to the standard deviation in background (sections 2.3.1.2, 4.5.3, and 5.5).
- Mean squared error (MSE) assessed the degree of distortion or error by comparing two images and producing a score (sections 2.3.1.2, 4.5.1, and 5.3).
- Peak signal-to-noise (PSNR) compared the peak signal in two images (sections 2.3.1.2, 4.5.1, and 5.4).
- Contrast-to-noise ratio (CNR) assessed the signal level of an object in relation to ambient noise (sections 2.3.1.2, 4.5.3, 5.6).
- Structural similarity index (SSIM) looked at patterns in local pixel intensities, using a multiplicative combination of structure, contrast, and luminance (sections 2.3.1.3, 4.5.1, and 5.7).
- Noise power spectrum (NPS), as a function occurring in the frequency domain, examined the texture of the noise in an image (sections 2.3.1.1, 4.5.1, and 5.9).
- Non-shift-edge ratio (NSER) examined change in edge definition when comparing two images (sections 2.3.1.3, 4.5.1, and 5.8).
- Texture analysis looked at Entropy, Energy (uniformity), Contrast, Correlation, and Homogeneity, where aspects such as smoothness/coarseness and randomness/regularity were obtained using spatial variation of pixel intensities (sections 2.3.1.3, 4.5.2, and 5.10).

VI. Abstract:

Task-specific observer performance and image quality: direct and indirect relationships in low-dose CT images.

Aim: The aim of this research was to examine image quality in low-dose CT images, as determined by a range of image quality measures (IQM), in addition to psychophysical assessment, and look at direct or indirect relationships.

Method: CT images of an anthropomorphic chest phantom were obtained using increasing tube current (mA) to vary image quality. Tube current was increased incrementally (15mA- 100mA) and 200mA (reference image). Three sets of simulated lesions (sizes 5mm, 8mm and 10mm) of density 100HU, -630HU and -800HU were imaged one density at a time, using a 16 slice CT. A normal series was also acquired. These image sets were repeated using attenuation jackets, to achieve further image degradation. Images (5mm slice thickness) were reconstructed using filtered back-projection. Image analyses were carried out on 235 images of which 39 were normal. IQMs used were signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), mean square error (MSE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), non-shift-edge ratio (NSER), texture analysis and 3D noise power spectrum (NPS). Visual grading characteristic (VGC) observer performance studies were performed with seven observers using a localisation task, and observer tasks involving visualisation of structures with how well images fulfilled international CT quality criteria, and the impact of noise on those decisions.

Results and conclusion: The relationship of mA to IQM values was mainly logarithmic, with patterns of undulations in graphed data indicating potential for reduction in radiation dose. Additionally, attenuation jackets significantly affected IQM values, as did regional variation. NSER, Energy(uniformity) and Homogeneity showed the strongest inter-correlation and an inverse relationship with Entropy. Observer performance studies demonstrated no significant difference in lesion localisations at different amperage, however there was strong correlation between the impact of image noise and the fulfilment of visualisation criteria in CT images. NSER, Energy (uniformity), Homogeneity and Entropy had moderate to strong correlation with human observers. This research demonstrated the importance of simulating body habitus, and using the appropriate IQM, when assessing image quality for the task intended.

1.0 Chapter One: Introduction

1.1 Context

With the advent of new imaging technologies, there comes the opportunity to examine more closely the information we wish to obtain from images that are produced, and the radiation dose needed for this to be achieved. Hybrid imaging, where technologies such as single photon emission computed tomography (SPECT) and computed tomography (CT) are combined, offers a synergy of functional and anatomical information. The low-dose CT (LDCT) component of this technology is not only used for attenuation correction (CTAC), but makes lesion localization and characterization possible. LDCT is also used in conventional CT for things such as lung cancer screening programmes (MacRedmond et al., 2004), and examining the abdomen for example in CT colonography screening programmes (Yee et al., 2013). As these more complex technologies are increasingly used, so is the need to lower radiation dose to the patient by means of software, hardware, and decisions by clinicians (Schindera, 2017). How much information is needed in the images will be determined by the requirements of the diagnostic task or referral question being asked.

The challenge of how to assess the images produced and to determine their suitability for the diagnostic task being undertaken, is an international dilemma. The problem of assessing image quality in medical images has been contemplated for nearly half a century (Rossman & Wiley, 1970). Research has been presented on both sides of the Atlantic at international conferences (Dirks et al., 2017; Schindera, 2017), with research also coming from and in collaboration with countries such as China, Japan and Sweden (Fisichella et al., 2010; Kamath et al., 2011; Li et al., 2013). While quality assessment is not a static process but rather continuous (Yee, 2013), there is still no easy-to-use and meaningful measure of CT image quality (Schindera, 2017).

How we perceive an image in terms of quality may not necessarily correspond to diagnostic findings, and increased radiation dose may even hinder rather than help. Manning et al. (2004) suggest that complex images with many confounding factors may cause even the most experienced reader to miss lesions that are 'visually obvious'. In a study looking at ultra-low dose

CT of the abdomen, Pickhardt et al. (2012) concluded that image quality and dose did not necessarily correlate with diagnostic findings. Using a variety of image evaluation methods, other studies also show that detection of abnormalities and image quality do not always correspond (Barrett & Myers, 2003; ESR, 2011; Jessop et al., 2015; Manning et al., 2005). In addition, the human visual system (HVS) has a remarkable ability to adapt to the level of noise in an image (Abbey, 2013). Ultimately a diagnosis is based on what is seen or recognized in an image, where a mixture of sharpness of outline, size and complexity of surrounding structures affect visual impact of image features (Manning et al., 2004; Ulzheimer, 2013).

1.2 Aim

The aim of this research is to determine if there are any quantifiable relationships, whether they be direct or indirect, between image characteristics and how well we see image features. During the process, image quality metrics (IQM) are assessed to determine their effectiveness. The conundrum that is faced, is how to match IQM outcomes to visual perception, and the determination of causal relationships between image quality and perception. Should this situation be rectified, it would provide the potential to distinguish between image dependent errors and non-image dependent errors. This would impact on decision making of clinicians when determining CT acquisition parameters for the required diagnostic task, and on their confidence levels when reporting, as well as ongoing quality assessment of images. This could also impact on training regimens which could be established to recognise the significance of lesion detection at lower radiation doses.

1.3 Method

The main body of research utilises an anthropomorphic chest phantom (Kagaku, n.d.), with images acquired over a wide range of milliamperes (mA) concentrating on the lower mA range. Image characteristics being examined look at image noise, image contrast, edge definition and complexity of surrounding structures. Observer performance studies have been carried out for specific tasks relating to clear visualization of structures, the effects of image noise on the decision-making process and lesion localization.

Using a variety of image quality measures, image characteristics are examined using IQM ranging from those more signal based eg. peak signal-to-noise ratio (PSNR), to those closer approximating the HVS eg. structural similarity index (SSIM) and image texture. Most authors present research that currently uses a limited number of these (only two or three in many cases), the nature of which are influenced by the emphasis of their work. To date, no single IQM is able to capture all aspects of image quality. By using a range of IQM in this research and examining the relationship between them, quantification of this relationship (or non-relationship) may then be correlated with observer performance.

Using an interactive computing environment in the form of MATLAB[®] allowed comprehensive image analysis to take place, enabling image characteristics to be measured (Hunt et al., 2006; Roberson, 2015). 'ImageJ' was used to perform some image analysis tasks. ImageJ is an image processing platform that is public domain and developed by the US National Institutes of Health (Dougherty, 2009a; NIH, 2020). In addition, 'LIFEx' (radiomics feature extraction freeware) enabled texture analysis to be performed (Nioche et al., 2018).

Observer performance studies were performed using visual grading characteristics (VGC) and lesion localization tasks. Images were displayed in 'ViewDEX', a dedicated software program for image display in observer performance (Håkansson et al., 2010). VGC uses visual grading and ratings to provide ordinal data, which can then be analysed using dedicated software for VGC analysis to produce a VGC curve and quantification of the area under the curve (AUC). Statistical analysis of image grading against specified tasks, provides one aspect of the information to be gleaned from the ViewDEX data. Further statistical examination on the impact of 'noise' provides another aspect. Receiver operating characteristic (ROC) analysis was performed on lesion localization data.

1.4 Results and discussion

Results and discussion surround the specifics of the image quality measures used, the rationale for their use and their effectiveness individually and in combination. Observer performance is correlated with these. The observer performance methods used are discussed, as is their

relevance. Correlation of these results facilitates examination of any direct or indirect relationships pertaining to image quality and observer performance.

1.5 Summary

The importance of this research lies in its potential to increase our understanding of visual perception in the context of image quality (Jessop, 2015). With a deeper understanding of the images produced for reporting, visual perception and how this impacts on the diagnostic task at hand, clinicians would have greater confidence in their imaging reports (Jessop, 2015). An international drive towards dose reduction and dose optimization continues. Introduction of newer technology drives the development of image quality measures that are increasingly complex and precise. There is an urgent need for more precise answers that relate to diagnostic accuracy. By using the specific range of IQM and observer performance combined in this research, and in attempting to define these relationships, understanding the interaction of image quality and visual perception is one step closer.

2.0 Chapter Two: Literature Review

2.1 Introduction

Literature reviewed in the process of this research ranged from the tangible of what lies in an image to the less tangible of how the mind perceives that image. Knowledge and expertise were required for image analysis platforms intertwined with observer performance methodology. An understanding of physics, signal detection and the frequency domain was combined with visual perception.

It is sensible to review the types of images that have been examined, and their use in clinical practice. The emphasis of this research is on CT images, in particular those acquired with a lower dose of radiation. Acquisition and processing aspects of CT imaging are considered, to further understand the types of analysis carried out, and resulting limitations.

'An image is always a *sampled* version of the input scene or object' to quote Robert Wagner (Wagner, 2010a). The CT data, obtained from projected data sets, undergoes algorithmic transformation to produce a representation of what has been imaged (Bushberg et al., 2012b). Concepts, in general terms, of what happens to data in the frequency domain, need to be understood to analyse and interpret the IQMs used. Principles of image reconstruction and image compression are also useful to understand to examine other authors work in quality metrics.

The image quality metrics used in this research were specifically chosen as a cross section of both traditional and modern metrics, to encompass signal and structural content. Each of these are dealt with in depth: contrast-to-noise ratio (CNR), Noise power spectrum (NPS), signal-to-noise ratio (SNR), peak signal-to-noise ratio (PSNR), mean square error (MSE), non-shift edge ratio (NSER), structural similarity index (SSIM) and image texture (correlation, contrast, uniformity or energy, homogeneity, entropy). The different types of noise in an image are also reviewed, and the concept of conspicuity is examined.

The definition of image quality is dependent on the application (Watson, 1993). Depending on the task for which the image is intended, and according to the perspective of the end user, performance of the observer in this task defines the image quality, where they cannot perform

the task intended if image quality is not suitable (Park et al., 2003). The performance of the observer is affected by their visual perception and how this impacts on image interpretation. Vision is an information processing task, where the information is 'represented' in the brain (Marr, 1982). How the observer perceives an image is impacted by the spatial arrangements within an image (lesion or structure shape, size, surroundings) and the way they discriminate image texture (Marr, 1982).

Visual perception plays an important role in the accurate interpretation of medical images and yet easily ignored (Krupinski, 2011; Zuckier, 2011). In the late 1940's human observer studies began in radiology and their development over time has given rise to observer performance methodologies that include ROC analysis and its permutations, visual grading and visual grading characteristics (Burgess, 2011; Båth & Månsson, 2007).

2.2 Computed tomography (CT) and low-dose CT (LDCT)

2.2.1 Computed tomography (CT)

Image acquisition occurs as the x-ray tube and detectors rotate around the patient, after which exit radiation that has been measured by the detectors multiple times, is digitized through an analogue to digital converter (ADC) and the data sent to a computer (*figure 1*) (*Computed Tomography [RadiologyKey], 2016; Seeram, 2009a*). Modern, fourth generation CT scanners may have a detector array that is stationary completing the entire circle (Seeram, 2009b). The acquired data is converted to images through the use of reconstruction algorithms (Seeram, 2009c). According to Knuth (1977, cited in Seeram, 2009), an algorithm is "a set of rules or directions for getting a specific output from a specific input" (Seeram, 2009c). These algorithms are unique to each manufacturer, and specific to the CT equipment used (Ulzheimer & Flohr, 2009). The shape and geometry of the x-ray beam is taken into account, as is the noise specific to the imaging system, along with any blur introduced as part of the reconstruction process (Geyer et al., 2015; Ulzheimer & Flohr, 2009). Analytic reconstruction algorithms such as filtered-

back-projection (FBP) are used (figure 3), or in more recent years iterative reconstruction has been available (Geyer et al., 2015; Seeram, 2009d).

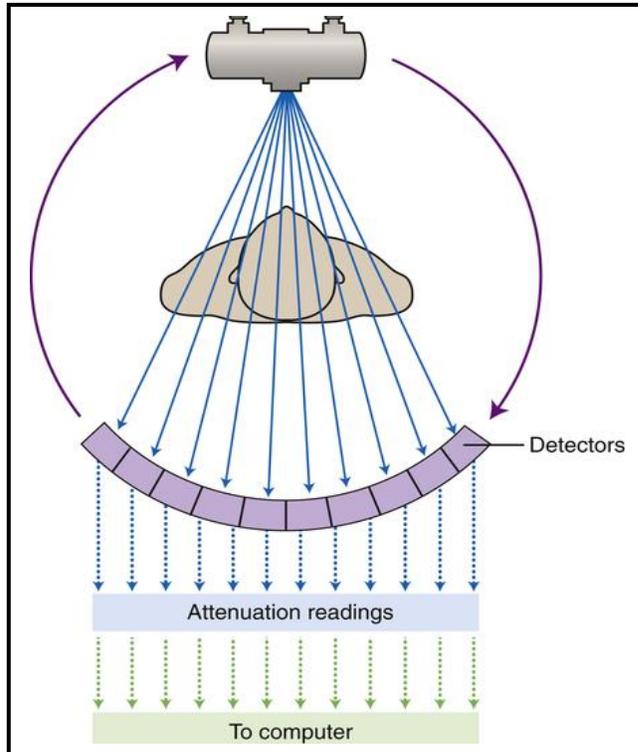


Figure 1. Computed Tomography Process (Fig 16-1 in *Computed Tomography [RadiologyKey], 2016*)

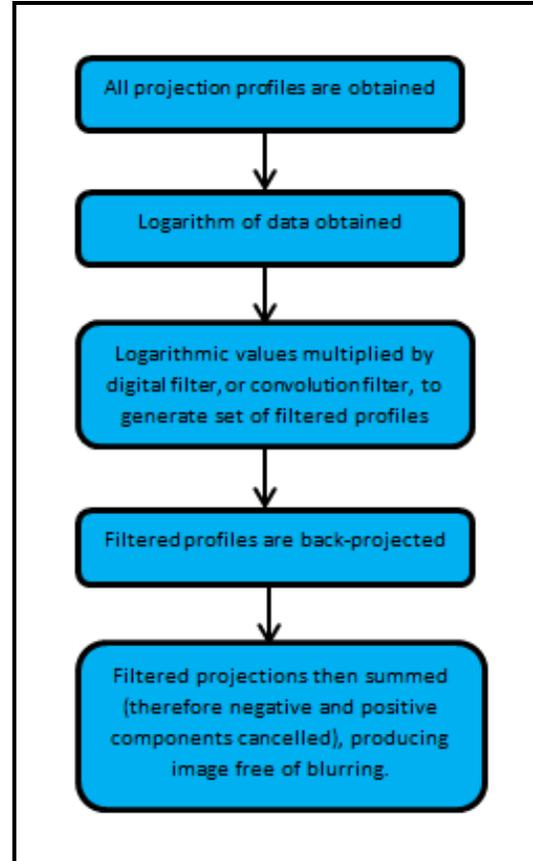


Figure 2. Filtered back-projection (based on information from Seeram, 2009d)

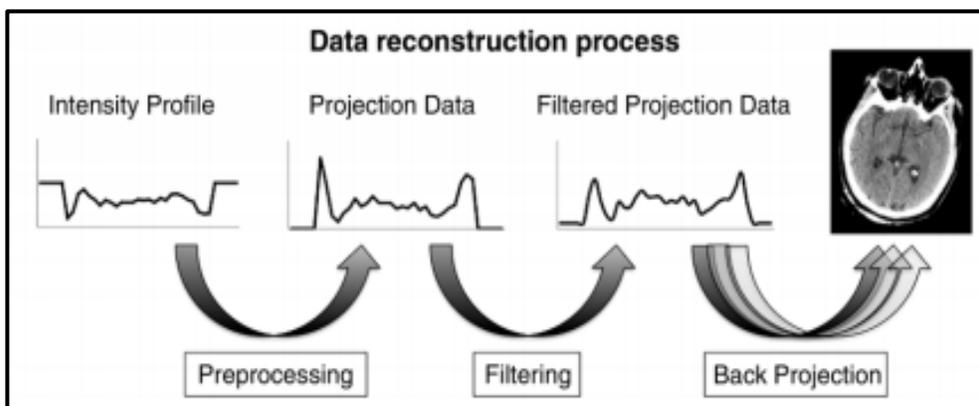


Figure 3. Data reconstruction in Filtered-Back-Projection (Fig 1 in Geyer et al., 2015).

Transformation of intensity values into attenuation values (projection data) occurs during pre-processing (*figure 3*) (Geyer et al., 2015). Different reconstruction algorithms or kernels are then used on the projection data depending on the purpose for which the images are required (whether a specific filter for attenuation correction, a softer filter for soft tissues or a sharp filter for high-contrast visualization in bone) (*figure 3*) (Geyer et al., 2015). Using the process of back-projection, the resultant projection data is then reproduced back into in the spatial domain to produce images (*figures 2 and 3*) (Geyer et al., 2015; Seeram, 2009d).

Iterative reconstruction variably weights image data using statistical and geometric models, while preserving resolution and reducing noise and producing images of improved quality when compared to FBP, which Seibert (2014) claims could lower the radiation dose for most exams in the clinical setting by 20-40% (Seibert, 2014). Bushberg, Seibert, Leidholt and Boone (2012) explain the process (*figure 4*) beginning with an initial estimate of the image, followed by a forward projection, which is then compared against a measured projection (Bushberg et al., 2012b). An error matrix is created of the difference and applied by the iterative algorithm to the next iteration. The iterative algorithm is applied to subsets within the data and the data as a whole. The process then continues until the error matrix is so small that the image is close to what was actually scanned (Bushberg et al., 2012b). By making better use of the image data than FBP, resultant images have a higher signal-to-noise ratio (SNR) when using the same radiation dose (Bushberg et al., 2012b). Subsequently using lower radiation doses can still produce similar SNR values in resultant images (Bushberg et al., 2012b). Geyer et al. (2015) describe six main steps carried out by iterative image algorithms in *figure 5*, where the final image is generated once a predefined condition set by the algorithm is satisfied.

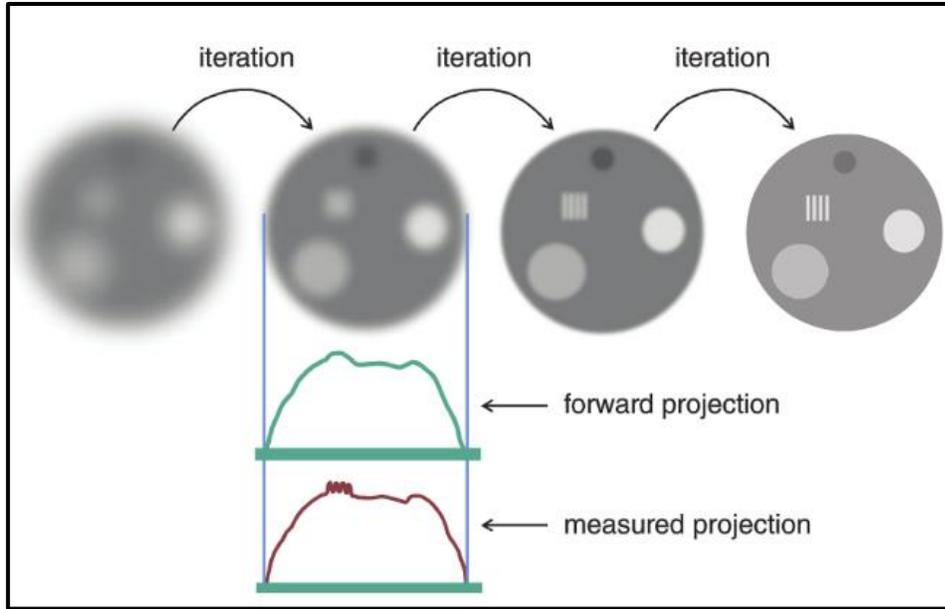


Figure 4. Logic behind iterative reconstruction. (Fig 10-55 in Bushberg et al., 2012b)

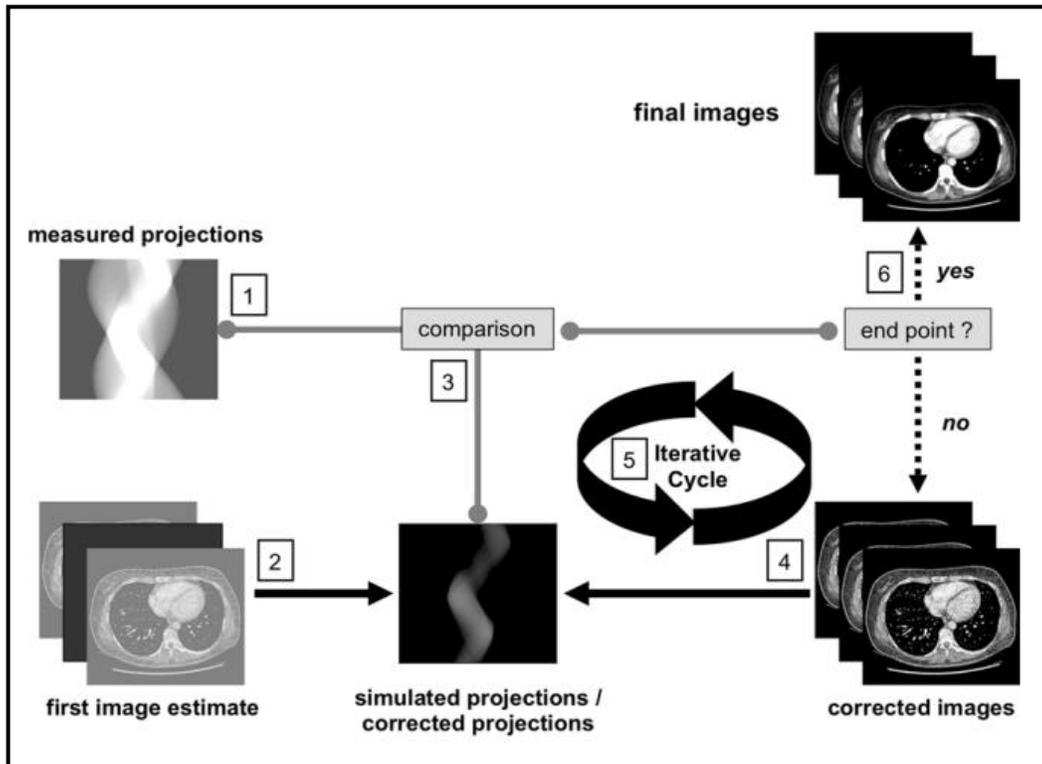


Figure 5. Schematic representation of the principal steps of iterative image algorithms. (Fig 2 in Geyer et al., 2015)

Algorithms required for CT reconstruction convert acquisition data from the spatial signal domain into the frequency domain (Bushberg et al., 2012b). By using a 'Fourier' transform the signal data is decomposed into series of sine waves of different phase, frequency and amplitude, which replicate that signal data when summed (Bushberg et al., 2012a). Subsequently, in the spatial frequency domain, higher frequencies correspond to smaller objects and lower frequencies correspond to larger objects (Bushberg et al., 2012a). This is a consideration in understanding the amount of detail available in an image, in relation to smoothness or sharpness produced by a selected filter or kernel, and the amount of noise in the resultant the image (Bushberg et al., 2012b; Seeram, 2009d).

Slice thickness of images also impacts of image quality. While image data can be acquired in finer slices, this is reconstructed according to image quality and diagnostic requirements. Thinner slices provide improved spatial resolution, however this is accompanied by a decreased SNR (Seeram & Lapointe, 2009e). Thicker slices, while providing an increase in image quality from noisier images acquired at lower radiation dose, can produce partial volume artefacts (Romans, 2013). These artefacts occur when several types of tissues are seen within a CT voxel (Bushberg et al., 2012b). The resultant CT number (usually expressed in Hounsfield units) is an average for all the tissues seen in that voxel (Seeram & Hsieh, 2009f).

Whatever the diagnostic task of the images being acquired, there is an obligation to use only the radiation dose that is necessary to produce these images (Seeram, 2009g). This process is called 'dose optimisation' which is described by Seeram (2009) as reducing radiation dose 'without compromising the required image quality needed for making a diagnosis' (Seeram, 2009h). The object of this radiation protection principle is to keep patient doses to as low as reasonably practicable (ALARP). Being mindful of the ALARP principle has gained increasing importance especially when, as we are reminded by comments in a report published in 2010 from the United Nations Scientific Committee on the Effects of Atomic Radiation when they assessed global medical exposures for a decade (1997 – 2007), that digital technologies and their increasing use are resulting in a higher radiation burden to the global population (UNSCEAR, 2010). Lai and Frush (2006) emphasize how important it is to justify examinations, with consideration given to the net benefit (hazards and clinical benefit) associated with every exposure as described in government

legislation 'Ionising Radiation (Medical Exposure) Regulations 2017' (IRMER) (Lai & Frush, 2006; Seeram, 2009g, UK statutory instruments, 2017).

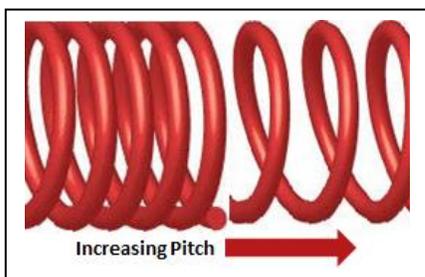
2.2.2 Low dose CT (LDCT)

There are a number of ways in which to affect the radiation dose that the patient receives during image acquisition. These include consideration of the energy of the x-rays, the number of x-rays, the volume covered by the x-ray tube with each rotation (affected by 'pitch') and the use of dose modulation (Seeram, 2009g). They do however impact on image quality therefore playing a role in diagnostic outcomes (Seeram, 2009g).

The energy of the x-ray beam can be produced at varying voltages (kV), where the penetrating power of the photons is determined by the peak kilovoltage (kVp) (Seeram, 2009g). Lower kVp means that photons have lower energy and penetrate thick objects less effectively (Seeram, 2009g). It is worth noting the exponential (rather than linear) relationship of kVp with patient dose when considering what level to use (Lai & Frush, 2006; Murphy & Goel, 2020).

The quantity of photons produced is determined by the milliamperage (mA), where the dose to the patient over the time of the scan is determined by milliamperage-seconds (mAs) (Seeram, 2009g). Less photons, however, means that because there are not as many resultant signals, the images are noisier with poorer contrast resolution (Sil et al., 2012).

In helical or spiral CT, the distance that the table travels per x-ray tube rotation is called 'pitch' (Seeram, 2009g). The radiation dose that the patient receives is inversely proportional to the



pitch, where doubling the pitch can halve the patient dose (Seeram, 2009g). As illustrated by the springs in *figure 6*, the paths followed by the X-ray beam are farther apart during rotations in helical scanning with a higher pitch setting.

Figure 6. Diagrammatic representation of Helical pitch.

Dose modulation is an automatic process that takes place during data acquisition so dose to the patient is optimised while image quality is maintained (Seeram, 2009g). The tube current (mA) is automatically controlled in the x-y and z axes, and takes into consideration attenuation differences of the body tissues as well as the patient size, where original mA settings based on a topogram are then adjusted on a rotational basis (*figure 7*) (Lee et al., 2008; Seeram, 2009g; Siemens, 2009).

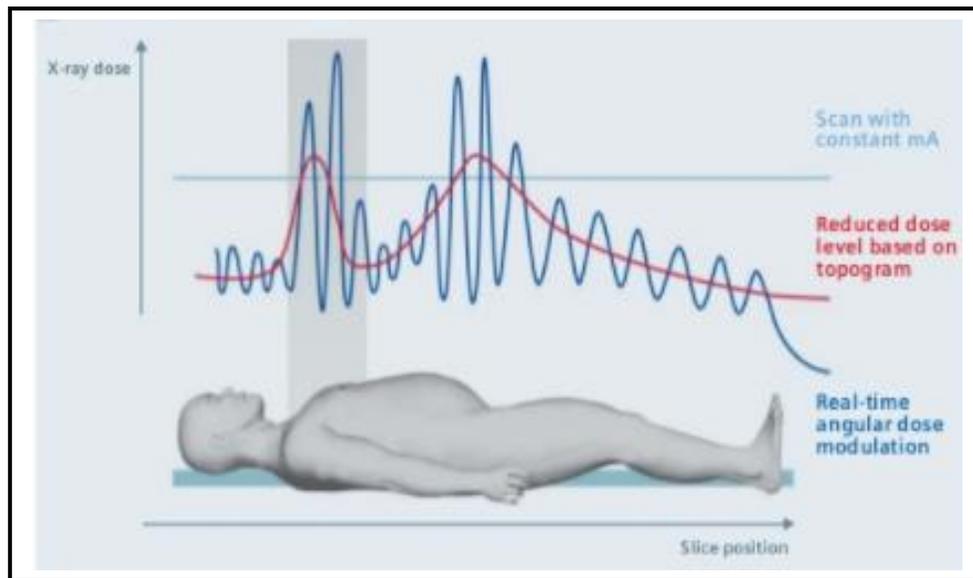


Figure 7. Dose modulation (CARE Dose4D) (*in Siemens, 2009*)

Once the required image quality has been decided for the procedure by the Imaging reporting consultants, the Image Quality Reference mAs is set, and if the automatic exposure control (AEC) is set up so smaller or larger patients can be dealt with before the minimum or maximum tube current is reached, then there is no need to adjust for individual patients (even those very large or very small) to maintain acceptable noise levels (Lee et al., 2008).

2.2.2.1 LDCT thorax

MacRedmond et al. (2004) published results of a study using low dose spiral chest computed tomography (LDCCT) on 449 volunteers. At that time they quoted lung cancer as the most common cancer in the world, which accounted for 20.3% of cancer related deaths in Ireland (MacRedmond et al., 2004). They quoted a resurgence of interest for lung cancer screening with the development of LDCCT (MacRedmond et al., 2004). Their baseline screening tests used the following settings: spiral 10mm, pitch 2 and low dose (50mA or less). High resolution CT (HRCT) was then used to image and stage patients in whom lung nodules were detected, ground glass opacities, or previously undetected benign calcification (MacRedmond et al., 2004). A total of 145 nodules <10mm were detected in 19.4% of patients. Initial screening produced incidental findings in 61.5% of patients, the most common of which were emphysema and coronary artery calcification (MacRedmond et al., 2004).

Aziz, Padley and Hansell (2004) from the Royal Brompton Hospital and the Chelsea and Westminster hospitals in London, published a paper on CT techniques for imaging the lung. They commented on a number of investigators who had recently used LDCT for screening high-risk individuals for lung cancer who were asymptomatic (Aziz et al., 2004). These studies reported detecting smaller size (<2cm diameter) lung cancers and at an earlier stage than chest radiography (Aziz et al., 2004). Protocols from the larger studies and consensus arrived at by the Society of Thoracic Radiology show the use of 120-140 kV, mAs ranging from 20-80, and Pitch ranging from 1.5- 3.75 (Aziz et al., 2004). They refer to a study published by Itoh et al. in 2000 who showed equal diagnostic utility of images obtained at 10mAs and 50mAs, for the task of detecting 6mm simulated lesions (Itoh et al., 2000). However they also referred to concerns about missing some lung cancers raised by Li et al. in a study published in 2002, where 39% of lung cancers were missed on LDCT screening, of which a significant proportion was due to detection errors, and the others missed due to interpretation (Li et al., 2002).

In December 2013, the United States Preventive Services Task Force (USPSTF) issued a recommendation for “annual screening for lung cancer with LDCT in persons who are at high risk based on age and cumulative tobacco smoke exposure”(USPSTF, 2013). The importance of the

programme was justified in terms of the prevalence of lung cancer in the US, with lung cancer their leading cause of cancer death (USPSTF, 2013). With approximately 37% of US adults who were current or former smokers at that time, and the increasing incidence of lung cancer with age, the benefit of a LDCT screening programme was evidenced in a large national, randomized, controlled trial (USPSTF, 2013). While lung cancer has a poor prognosis, early detection and treatment, especially of operable early-stage non-small cell lung cancer (NSCLC), gives the possibility of improved prognosis (USPSTF, 2013). Harms listed included both false negative and false positive results, over-diagnosis and radiation exposure

At the conference for the Radiological Society of North America, December 2013, results of a large Japanese lung cancer screening study were presented (Barnes, 2013a). This study found that there are some solid ground-glass nodules (SGGN), 5mm and smaller, which can grow to become invasive adenocarcinomas even over a longer period than 5 years (Barnes, 2013a). From their study, where subjects in this category were followed for a mean of 5.7 years, their recommendation was that “ patients with solid ground-glass nodules smaller than 5mm should be rescanned within no less than 6 years” (Barnes, 2013a). The significance of this is the growing understanding of SGGNs, and that some may stay unchanged for years (Yoshihisa, 2013).

For high-risk individuals, LDCT is now accepted as an effective lung cancer screening method (Pedersen et al., 2016). It has been implemented in both the United States and China, and being considered in others (Pedersen et al., 2016). Zhao and Wu (2015) published a review on the current status of LDCT screening in China. As in many western countries, lung cancer is the leading cause of cancer-related death, however mortality from lung cancer is far more prevalent in China when compared to the world-wide average and has increased 465% in the last three decades (Zhao & Wu, 2015). Zhao and Wu (2015) concluded that for now, LDCT is the most promising tool for lung cancer screening on a large-scale population basis. In July 2017, Chinese researchers Wang and Han et al. published in the Chinese Journal of Cancer, results using a statistical model, stating that LDCT screening could reduce deaths by 24% from lung cancer in China (Casey, 2017). They recommended an optimal age range of 55 to 75, focusing on old smokers in urban China (Casey, 2017). In October 2017, results of the Pan-Canadian Early Detection of Lung Cancer

[PanCan] study were published in the Lancet Oncology journal where they assessed a model using low-dose CT in patients who were current and former smokers between the ages of 50 and 75 years (Tammemagi et al., 2017). They described the efficacy of resultant early detection, in terms of treatment outcomes and monetary savings to the healthcare system (Tammemagi et al., 2017). In 2018, the findings of a large-scale trial for CT lung cancer screening in Japan, led by Dr Takeshi Nawa, were published in the Japanese Journal of Oncology, which reported a 51% reduction in lung cancer mortality (Kim, 2018).

From a technical perspective it is interesting to note that Cody et al. (2010) found a variation in normalized dose by a factor of almost two, as part of the quality assurance programme for the National Lung Screening Trial in the United States. They inferred that 'mAs alone cannot be used as a universal indicator of image quality', because of the variation of output radiation per mAs between CT scanners (Cody et al., 2010).

In a slightly different direction, researchers in Boston have tried a 'sparse sampling' technique to reduce radiation dose to the patient (Barnes, 2013b). Using iterative reconstruction for best results, they maintained a high tube current but reduced the projection fields, and performed non-continuous reconstructions 'with angular gaps between subsequent projections' (Barnes, 2013b). Dr Khawaja of Massachusetts General Hospital stated that with this technique a >90% reduction in dose is achievable, when compared to a full chest CT (Barnes, 2013b).

2.2.2.2 LDCT abdomen and pelvis

Low-contrast detectability (LCD) refers to how well two objects of similar properties can be differentiated, such as those with similar Hounsfield units in CT (Radiology-tip-database, 2018). LCD is a key consideration in CT of this part of the body, where increases in image noise can severely affect the diagnostic outcome (Funama et al., 2005; Kalra et al., 2004). While radiation can be reduced by lowering tube current and subsequent mAs settings, the contrast-to-noise ratio (CNR) is reduced as a result (Funama et al., 2005). Funama et al. (2005) presented results of a phantom study using a low-tube voltage (90kV), where they determined a direct correlation

between CNR and radiation dose and assessed low-contrast detectability by using receiver operating characteristic (ROC) observer performance. They concluded in practical terms, that radiation dose can be reduced by using low tube voltage and still maintain acceptable levels of low-contrast detectability, with 71% radiation dose at 90kV compared with 120kV (at identical CNR) (Funama et al., 2005). They made a specific recommendations that for 90KV, the volume CT dose index (CTDI) be set at 70% of that used at 120kV, and the technique should only be used for those patients who weigh <80kg due to level of noise present in images of those above this weight (Funama et al., 2005).

The Centers for Medicare and Medicaid Services (CMS) (2009) published a decision memo for computed tomography colonography (CTC) screening. In this memo they name colorectal cancer (CRC) as one of the top three cancers in the United States. However, they concluded that the evidence to date was inadequate to grant coverage, when compared to other screening techniques at the time (CMS, 2009). It is worth noting that they acknowledged the recent use of “multidetector or multislice CT technology” (MDCT) which “shortens scan time and reduces radiation dose while preserving high spatial resolution” (CMS, 2009).

Despite funding concerns, the American College of Radiology (ACR) white paper on managing incidental findings on abdominal CT (2010) comments on the increasing use of CT screening examinations of the abdomen, mentioning CT colonography as an example (Berland et al., 2010). Subsequent concerns were also noted about radiation exposure (Berland et al., 2010).

Marin et al. (2010) from the Duke University Medical Centre presented results of a study looking at low-tube-voltage and adaptive statistical iterative reconstruction (ASIR). While the research maintained a high-tube-current, its application had implications for the use of lower tube current and dose modulation in the clinical setting (Marin et al., 2010). The study demonstrated lower image noise and improved image quality, with ASIR producing noise reduction that while more apparent at the higher spatial frequencies was also quantifiable at lower spatial frequencies (Marin et al., 2010). This meant that there was less interference from noise in the regions of spatial frequencies corresponding to fine detailed texture features (Marin et al., 2010).

Pickhardt et al. (2012) reported preliminary results of an ongoing prospective trial of ultralow-dose abdominal MDCT using model-based iterative reconstruction (MBIR). They emphasised the need to prioritise dose reduction, however they warned that an acceptable image quality and adequate maintenance of diagnostic accuracy were also priorities (Pickhardt et al., 2012). They described MBIR as an advance beyond ASIR, with potential to increase resolution while reducing dose in a substantial manner (Pickhardt et al., 2012). Based on the diagnostic CTDI, images were acquired targeting a radiation dose reduction of 70-90% (Pickhardt et al., 2012). Acquired at 120kV, parameters adjusted to achieve this targeted reduction were tube current range, noise and slice thickness (Pickhardt et al., 2012). Qualitative differences were readily seen between MBIR and other reconstruction techniques at very low doses, but the compromise of lesion detection by ultralow-dose MBIR in comparison with standard FBP was a matter to be carefully considered when balancing diagnostic performance with dose reduction (Pickhardt et al., 2012). They cautioned that no firm conclusions could be made at the time of publication, given that it was an interim analysis and the small sample size to date, and highlighted an important finding that required further investigation which was lack of correlation between objective and subjective image quality measurements (Pickhardt et al., 2012).

Gervaise et al. (2014) from France, published research comparing standard and low-dose abdominal and pelvic CT using FBP and AIDR 3D. While this study was limited to only a small number of patients, and diagnostic performance was not assessed, it had definite findings that concluded a halving of radiation dose with ASIR 3D iterative reconstruction to achieve an equivalent image quality to that of standard FBP (Gervaise et al., 2014).

Yee, Weinstein, Morgan, Alore and Aslam (2013) from the USA, published a review on the advances in CTC for colorectal cancer which acknowledged a 50% reduction in radiation dose that could be employed when using iterative reconstruction. The use of AEC and low tube current were listed as techniques to achieve this dose reduction (Yee et al., 2013). CTC screening is considered valid for average-risk individuals according to a joint guideline from the American College of Radiology, the United State Multi-Society Task Force on Colorectal Cancer and the American Cancer Society, and starting at 50, it is recommended that screening CTC happens

every 5 years (Yee et al., 2013). With modern MDCT scanners that allow thinner slices, lower radiation dose and use improving reconstruction techniques which give better quality images, the risks are outweighed by the benefits and CTC screening programmes are increasingly being used (Yee et al., 2013).

Pickhardt, Correale, Delsanto, Regge and Hassan (2018) reported a higher detection rate for cancer in adults over 65 years of age (in comparison with younger adults) when using CTC. In addition, Ricci, Koby and Yee (2020) highlighted the advantage of performing CTC in patients who, due to comorbidities, have an increased risk for performing optical colonoscopy.

NHS England - National Cancer Programme (2019) published the protocol for targeted lung cancer screening in the UK, enabling nation-wide participation for anyone between 55-75 years of age who had smoked at any point in their life. Assessment of individual risk determines suitability for LDCT (NHSE- National Cancer Programme, 2019).

2.3 Image quality and image interpretation

Difficult to define, image quality should be looked at in relation to the intended task of the image, and how effective its use is for this task (ICRU; Zarb et al., 2010). Keeping in mind that a medical image is a representation of some structure or function of the body, then image quality can be defined by the accuracy with which this representation occurs (ICRU, 1996; Zarb et al., 2010). However, a balance is required when optimizing radiation dose to be as low as reasonably practicable (ALARP), while diagnostic confidence is maintained from using images that are of acceptable quality (Zarb et al., 2010). Reproducing the pathology in question is an essential requirement when considering the adequacy of image quality (Manning, 2005). Dr Siegal (2013) from the University of Maryland, brought to attention of attendees at an International Symposium on MDCT, that it is difficult to optimize dose without a mechanism to determine image quality and subsequently assess any improvement (Yee, 2013).

The complex task of determining optimal image quality involves using physical measures that are objective, in conjunction with observer performance measures that are subjective (Zarb et al.,

2010). The challenge lies in establishing applicable image quality measures of suitable accuracy and precision, a challenge which has been taken up by the European Society of Radiology (ESR) 'Eurosafes Imaging Workgroup: Appropriate Image Quality' established in 2016 (Schindera, 2017). Given the visual nature of medical image information, defining the quality of information at the 'point of presentation to the observer' would make sense (Manning, 2005).

2.3.1 Physical image quality assessment

Physical image quality is measured in image space, for example IQMs examining image noise, and contrast to noise, or in frequency space examining aspects such as noise power spectra (NPS) (Berta et al., 2014). Some objective measures require the existence of a reference image (one that is free of distortion) that is used as a comparison for the image being measured. According to Eckert and Bradley, due to the dependence of visibility of artefacts on local image content, the characterisation of spatial variations across an image is required of an IQM (Eckert & Bradley, 1998). The ideal objective measure would be one that approximates the HVS, and takes into account structural content and variations in image texture (Gebejes & Huertas, 2013; Lundström, 2013).

2.3.1.1 Noise

The concept of 'noise' in an image is not a simple matter. Noise is defined in common terms as 'unwanted sound or noise pollution', or in terms of communication, whether it be human communication or electronic in nature, noise can block or distort what is being communicated (Burgess, 2010a). There are three types of noise in CT images: (1) random noise which includes statistical noise, electronic noise and noise caused by reconstruction parameters; (2) artefactual noise; and (3) structural noise (Hanson, 1981).

Quantum noise:

The simplest way of defining image noise in CT is by the standard deviation (σ) in a region of interest (ROI) over an homogenous area (Bushberg et al., 2012b). The root mean square equation is seen in *figure 8* (Bushberg et al., 2012b). Quantum limited imaging such as LDCT, demonstrates well this relationship where 'noise is proportional to the square root of the dose' so in practical terms doubling the radiation dose causes a noise reduction of 40% (Bushberg et al., 2012b).

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (HU_i - \overline{HU})^2}{N-1}}$$

Figure 8. Image noise in CT
(Fig 10-12 in Bushberg et al., 2012b)

Random noise arises from a random (or stochastic) process which according to Burgess 'can be described only in terms of probability distributions' (Burgess, 2010b). Statistical noise, or quantum noise is a result of limitations in the number of x-rays transmitted during CT image acquisition that produce the attenuation data detected, from which the resultant image is produced (Hanson, 1981). As mA or kV reduce, there is less attenuation data collected.

Electronic noise:

Electronic noise is inherent in the electronics of the imaging system, with analogue circuits the most susceptible (Hanson, 1981). Digital circuits are relatively impervious to this problem, because of the signal shape which allows the same information to be carried digitally even with noise contamination (*figure 9*) (Duan et al., 2013; Hanson, 1981). In work published by Duan et al. (2013) they demonstrated that the use of integrated electronics in a CT detector showed significantly improved results than with conventional distributed electronics. They commented that in efforts to reduce patient dose by lowering technical parameters affecting CT output, when less radiation dose is used there is an increase in image noise (Duan et al., 2013). Duan et al. (2013) also commented that while this image noise is mainly from quantum noise, that electronic noise still makes a contribution, it is completely unrelated to numbers of photons detected and has no diagnostic usefulness (as demonstrated by Bushberg (2012) in *figure 10*) (Bushberg et al.,

2012a; Duan et al., 2013). Their work showed a potential avenue to reduce patient dose even further (Duan et al., 2013).

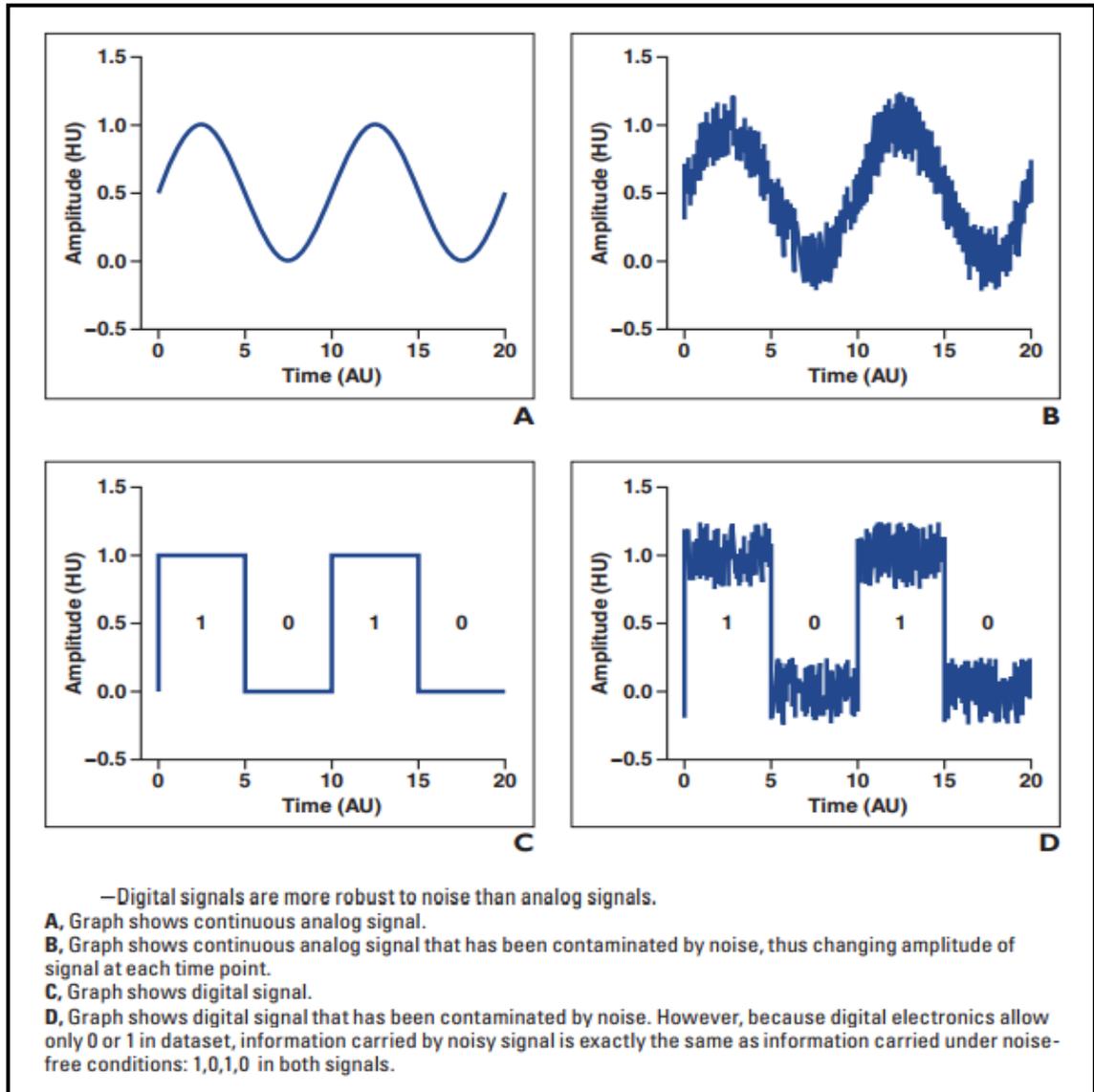


Figure 9. Electronic noise in CT detectors (Fig 1 in Duan et al., 2013)

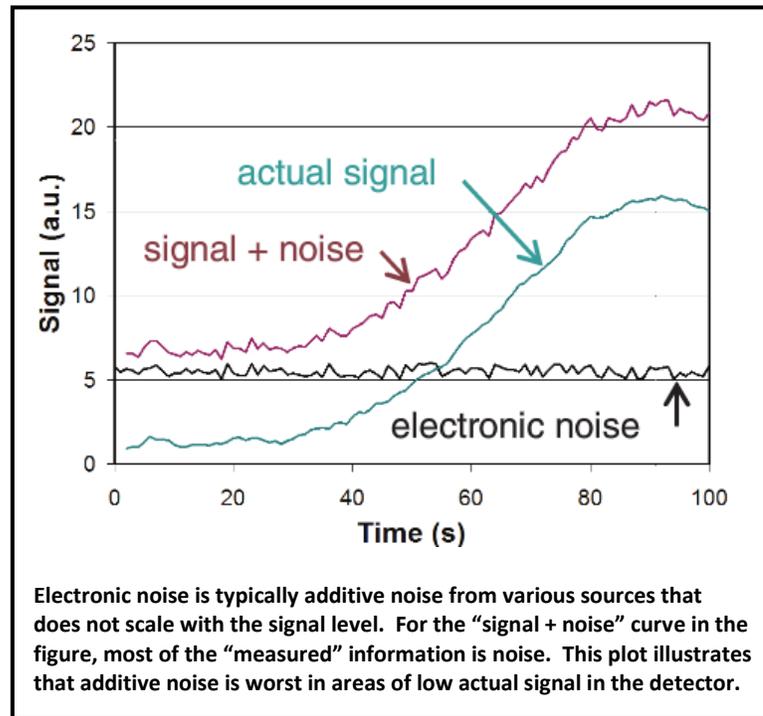


Figure 10. Electronic noise (Fig 4-21 B. in Bushberg et al., 2012a)

Reconstruction noise:

During the reconstruction process noise can be introduced due to ‘roundoff’ errors (Hanson, 1981). These errors occur due to the representation of numbers in the computer and limitations of bits used in these representations, where the results of computations are rounded to the least significant bit used (Hanson, 1981). These types of errors can be kept to a minimum by a suitably powerful computer and suitable programming (Hanson, 1981).

The selection of image reconstruction kernel will influence the amount of image noise present (Seeram & Hsieh, 2009f). Sharper (or high-resolution) kernels which provide greater edge enhancement, produce increased noise (Gierada et al., 2010; Seeram & Hsieh, 2009f). These kernels preserve the image contents that present at higher frequencies, and noise usually presents in this range also (Seeram & Hsieh, 2009f).

Artefactual noise:

Morgan & Miller (1983, cited in Seeram and Heish, 2009) defined artefact as ‘a distortion or error in an image that is unrelated to the subject being studied’ (Seeram & Hsieh, 2009f). Specifically for CT, Hsieh (1995) defined an image artefact as “any discrepancy between the reconstructed CT numbers in the image and the true attenuation coefficients of the object” (Seeram & Hsieh, 2009f). Due to their interference with image interpretation, artefacts can be thought of as image noise (Hanson, 1981). This type of degradation to image quality may not only affect how details in the image are perceived, but may even cause misdiagnosis (Bushberg et al., 2012b). These artefacts can be divided into four main categories which include shading, bands, streaks and miscellaneous as seen in *table 1*, where these artefacts classified on the basis of appearance have been tabulated against probable causes (Seeram & Hsieh, 2009f).

Appearance	Cause
Streaks	Improper sampling of data; partial volume averaging; patient motion; metal; beam hardening; noise; spiral/helical scanning; mechanical failure.
Shading	Partial volume averaging; beam hardening; spiral/helical scanning; scatter radiation; off-focal radiation; incomplete projections.
Rings and bands	Bad detector channels in third generation CT scanners.

Table 1. Classification of artefacts on the basis of appearance (*table 9-2 in Seeram, 2009f*)

Due to the nature of an image being a ‘sampled’ version of an object, it follows that under-sampling could cause image patterns not corresponding to the original object (Wagner, 2010b). Less incident photons on the detectors causes more noise and less signal detected, and this ‘photon starvation’ combined with electronic noise can cause severe streaking (*figure 11*) (Seeram & Hsieh, 2009f). Streaking is a predominant artefact seen in this research project.

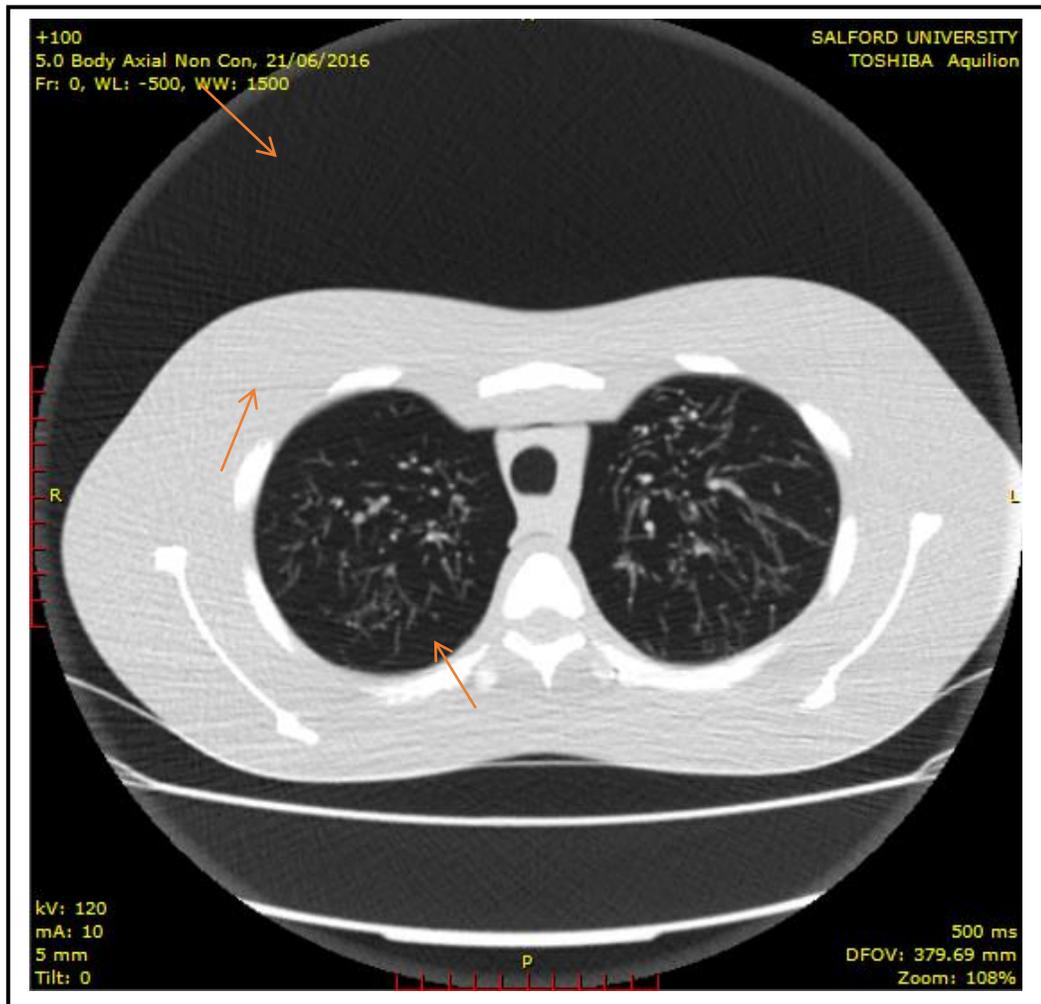


Figure 11. Streak artefact at low mA using anthropomorphic chest phantom.

Anatomical / structural noise:

As a pattern in the image, patient anatomy may or may not be of diagnostic importance (Bushberg et al., 2012a). Varying densities in an object being imaged, if interfering with diagnosis, may be referred to as structural noise, and while superimposition of structures is not so much an issue in CT with the spatial isolation it achieves, awareness of contribution from structures in adjacent slices and structural variation itself is important (Bushberg et al., 2012a; Hanson, 1981).

Some authors have made a distinction between anatomical noise and anatomical background which is referring to anatomical structures that are distinguishable (Hoeschen et al., 2005; Tapiovaara, 2008; Tischenko et al., 2003). However, other studies have concluded that it was

anatomical background (that is the anatomical structures) that limited lesion detection (Båth et al., 2005; Sund et al., 2004; Tapiovaara, 2008). In work published by Gurney (1996) which looked at assessing detection of lung cancer, it was found that to be distinguished from surrounding structures that are normal, the lung cancer needs to grow large enough. Detection errors were also found by Aziz et al. (2004) to have related to difficulty in distinguishing lesions from normal structures.

Noise power spectrum (NPS):

There is no information obtained on noise spatial characteristics when using just the standard deviation of noise, from the simple equation at the beginning of this section on noise (Boedeker & McNitt-Gray, 2007). To characterise (more fully) the noise in an image, the 'noise power spectrum' (NPS) is used (Seeram & Hsieh, 2009f). Describing the texture of noise contained in an image, the NPS correlates the noise at one point in an image with its surrounds, and in so doing describes the noise frequency content in reconstructed data (Bushberg et al., 2012a; Miéville et al., 2012). By enabling analysis of noise correlation induced by the various processing steps (including reconstruction and filtering steps) and characterising spatial frequency distribution and magnitude, this provides a more complete description of noise in an image (Miéville et al., 2012).

The NPS measurement looks at the noise variance (σ^2) of the spectral data characterised by frequency (Bushberg et al., 2012a). 'White noise' for a two-dimensional (2D) image, where in each of the pixels the noise levels are independent of their surroundings, the NPS function plot against spatial frequency shows a straight line (*figure 12*) (Bushberg et al., 2012a). This is due to the absence of noise correlation .

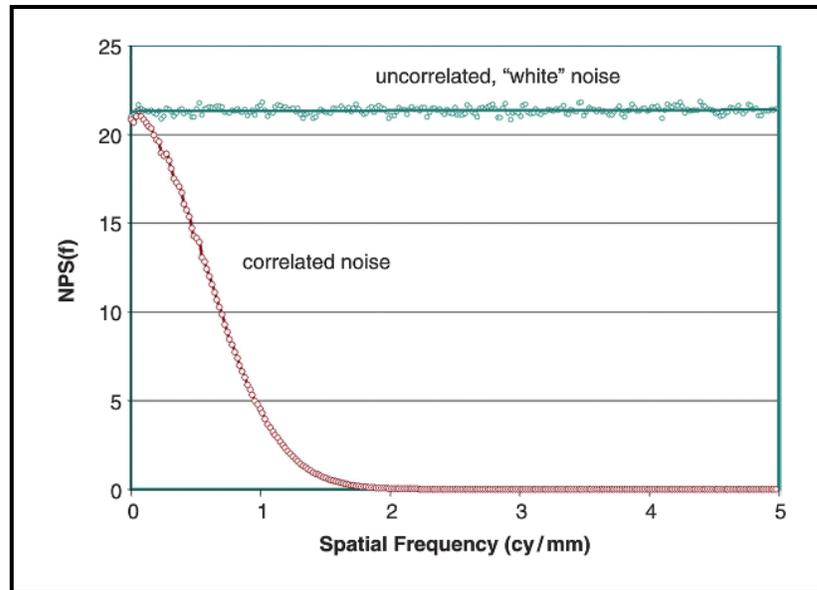


Figure 12. Noise power spectrum $NPS(f)$ is shown for uncorrelated and correlated noise. (fig 4-28 in Bushberg et al., 2012a)

Algorithms for iterative reconstruction are non-linear, therefore the imaged object will affect how the algorithm performs (Funama et al., 2014). Marin et al. (2010) published work that demonstrated significant reduction of image noise with iterative reconstruction, however the effect was non-uniform across the whole spatial frequency spectrum. They demonstrated that how the noise appeared (spatial frequency distribution of the noise) along with the magnitude, could affect image quality and subsequent lesion detectability (Marin et al., 2010).

In 2012, Pineda et al. described characterization of image noise in three-dimensional (3D) x-ray CT as critical. The reasons they stated for this characterization of image noise being critical, were performance of multiple diagnostic tasks governed by patient radiation dose and the relationship to image noise (Pineda et al., 2012). Proper dose levels should be maintained, appropriate to image quality required for the images obtained for the clinical task (Pineda et al., 2012). They stated that simple measures of noise levels fail to account for noise correlation in this context (Pineda et al., 2012). Ultimately, noise characterization and task-based assessment can be linked to further identify factors determining image quality (Pineda et al., 2012).

The power spectrum (or power spectral density (PSD)) demonstrates the average power of a signal when distributed according to frequency (Mirchandani, 2012). A common method

for examining image noise in CT is by computing the location averaged NPS in the Fourier domain, where the NPS as an integral over all frequencies in the CT image, yields the noise variance (σ^2) from the Fourier coefficients of that noise (Bushberg et al., 2012a; Pineda et al., 2012).

In 2D NPS created from a single simulated projection, aliasing effects were demonstrated by Kijewski and Judy (1987), as was apodisation (change in the shape of the mathematical function) due to linear interpolation. By using contributions from all projections a complete image NPS is obtained, and by employing a 3D Fourier transform of the 3D data, a 3D NPS is produced (Kijewski & Judy, 1987; Pineda et al., 2012). Siewerdsen et al. (2002) emphasize that meaningful NPS results can only be obtained when the 'full dimensionality' of image data is appreciated.

In CT, noise has 'non-stationary' properties which means that across the spatial domain there are variations in correlation and magnitude, in addition to which the reconstruction algorithm affects variability of the NPS in characterisation of noise in reconstructed images (Bushberg et al., 2012a; Miéville et al., 2012; Pineda et al., 2012). CT images are non-stationary due to factors such as quantum noise, finite number of projection views (view aliasing effects), 'spatiotemporal blur', discrete nature of the images and the variation in x-ray exposure and their detection (Pineda et al., 2012). NPS asymmetry can also impact on structure detectability across reconstructed planes (Siewerdsen et al., 2002). Stochastic variations in image signal can significantly impact on low-contrast and fine structure detectability, depending on the magnitude and frequency content of these stochastic variations (Siewerdsen et al., 2002). Siewerdsen et al. (2002) describe the importance of quantitative measurement of noise characteristics when optimizing imaging technologies. An example of changes in NPS caused by variation in acquisition parameters, in this instance the quantity of photons produced to obtain the image (mA), is demonstrated in *figure 13*.

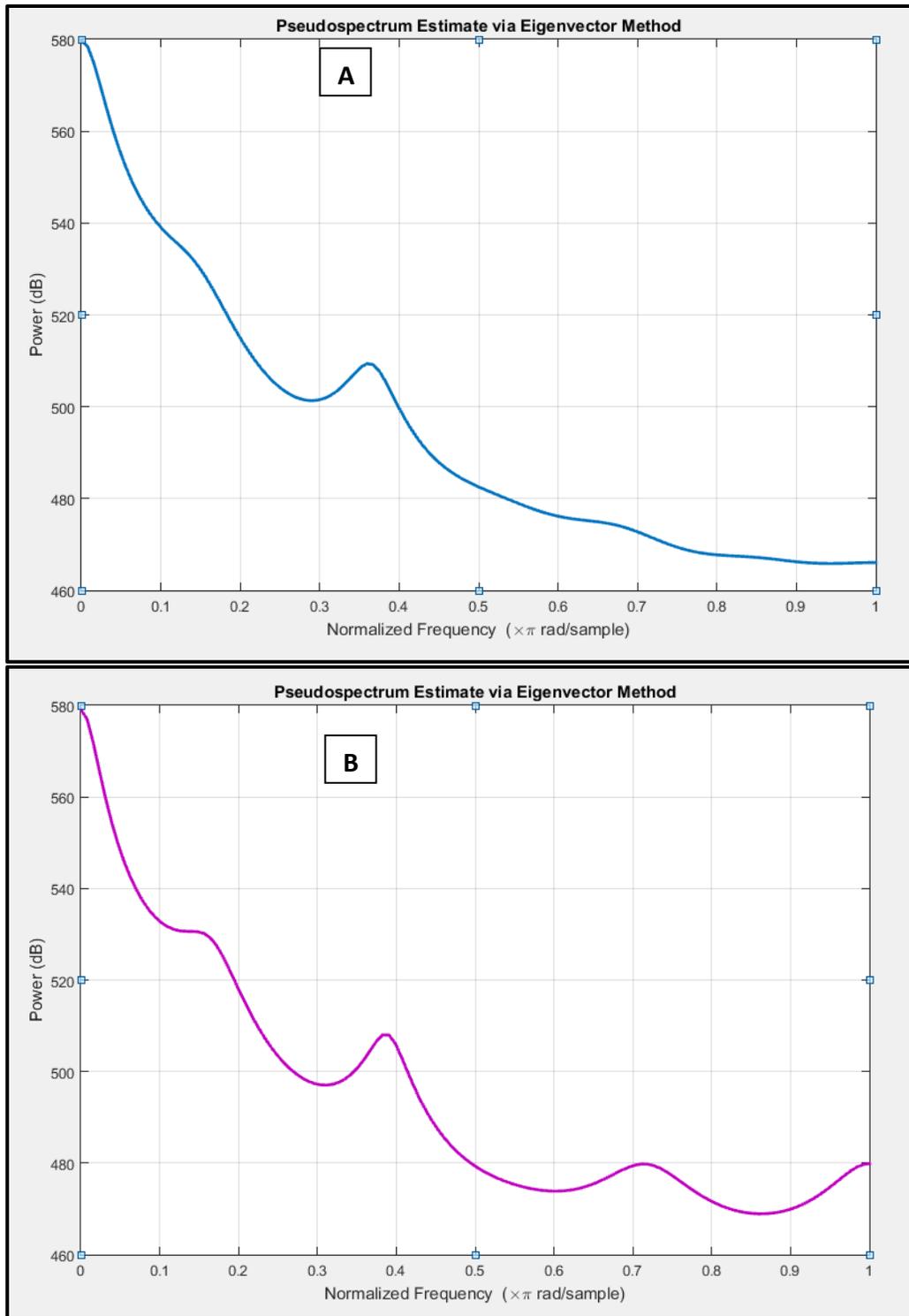


Figure 13. NPS 3D: Same object slice acquired at (A)10mA and (B)200mA, from images acquired using an anthropomorphic chest phantom and analysed in MATLAB®, producing a pseudospectrum estimate via the Eigenvector method.

NPS can be used to assess image characteristics in a number of ways. In France, Pasquier et al. (2017) used a NPS measurement tool on clinical CT images to investigate the impact of iterative reconstructions. In a joint project between Sweden and the United States, Engstrom et al. (2009) showed correlation with a specific power-law exponent and amount of structure in mammographic breast background, and its effect on lesion detectability. In the United States, McKenney et al. (2014) used a power-law function to translate protocols and convert automatic exposure control (AEC) settings between scanner platforms, concluding that preserving dose levels between platforms will not necessarily produce image quality that is identical.

2.3.1.2 Mathematical image quality measures

Mathematical metrics, such as 'mean square error' (MSE), 'signal-to-noise ratio' (SNR) and 'peak signal-to-noise ratio' (PSNR), while not performing well where image content is significantly different, have ease of use as an advantage (Eckert & Bradley, 1998). These do not however take into account spatial frequency content, and could in fact produce the same result for images of marked differences in image quality (Boedeker & McNitt-Gray, 2007). Boedeker and McNitt-Gray (2007) found this to be the case for SNR when comparing two images of different appearance and spatial characteristics but the same standard deviation for noise, in which their example had a result differing by a factor of 3.6. They went on to describe how SNR values increased with mAs, but not always as predicted by photon statistics (Boedeker & McNitt-Gray, 2007).

'Contrast-to-noise ratio' (CNR) describes the signal amplitude in relation to the ambient noise in an image (Bushberg et al., 2012a). Marin et al. (2010) found it a useful metric when evaluating the use of an 'adaptive statistical iterative reconstruction' algorithm (ASIR) on the quality of CT images. In a phantom study, Gayou (2012) used CNR as one of the metrics to assess the influence of acquisition parameters on CT image quality for pre-treatment images required for radiation therapy planning.

CNR and SNR have often been used in combination to assess image quality in various modalities over a range of applications. Manotta et al. (2006) compared different magnetic resonance (MR) imaging scanners in a multicentre imaging study using these metrics. The importance of this

endeavour was to be able to merge data across sites, and to do so they needed to assess the ability to produce images of the same quality at all sites (Magnotta et al., 2006). They felt that the CNR metric in particular, was clinically relevant (Magnotta et al., 2006). Fanous, Kashani, Jiminez and Murphy (2012) used these two metrics as part of an evaluation of image quality and radiation dose when using different tube potentials (100 and 120kVp) when performing pulmonary CT. Li et al. (2013) compared low tube voltage CT acquisitions using iterative and FBP reconstruction for contrast-enhanced chest imaging, concluding that the iterative reconstruction produce better image quality with higher SNR and CNR values. Welvaert and Rosseel (2013), while using SNR and CNR for functional MR data, concluded that to achieve consensus on a common definition for these metrics in this context would be difficult. Bhosale, Wagner-Bartak, Wei, Kundra and Tamm (2015) used CNR and SNR when examining image quality using different image reconstruction techniques.

Correlation based metrics such as PSNR and MSE can be used where the image in question is compared to a reference image (Sasi varnan et al., 2011). Sasi varnan, Jagan, Kaur, Jyoti, and Rao (2011) described PSNR and MSE as the most widely used full-reference IQMs in the spatial domain. These two metrics are interrelated where the MSE is inversely proportional to the PSNR and is a component of the PSNR calculation (Sasi varnan et al., 2011). As similarity measures, they are acceptable when examining a simple change in distortion of a particular type, however these are signal based metrics and do not approximate the HVS perception of image quality (Keimel & Diepold; Sasi varnan et al., 2011). In a review paper on 'quantitative image quality assessment of medical ultrasound images' Rangaraju et al. (2012) described the digital pixel values, on which the computation of MSE is based, as not necessarily representative of the light stimulus that enters the eye.

2.3.1.3 Perceptual characteristics - image quality measures

Perceptual IQMs involve an awareness of the human visual system (HVS) and estimate differences perceived by the HVS (Eckert & Bradley, 1998). According to Sasi varnan et al. (2011),

the premise on which this is based is that humans look at image attributes including brightness, contrast, image texture and shape rather than perceiving the image as actual signals.

Included in this category are IQMs that look at contrast sensitivity function (CSF) and luminance adaptation (Eckert & Bradley, 1998). The threshold at which an image feature becomes visible is defined by contrast sensitivity, and therefore has significant clinical relevance (Pelli & Bex, 2013). The luminance level has a marked effect on CSF which may also have a noise related component, where the effect of noise is less marked at higher contrasts (Chesters, 1992). In addition, background luminance levels may change visibility of image features, as does local luminance levels on a regional basis (Choi & Lee, 2011; Eckert & Bradley, 1998). Visual noise caused by variations in luminance causes an increase in the minimum contrast that is needed to detect a signal, however a uniform background in medical images would not necessarily occur therefore quantifying absolute values for minimum contrast would be difficult (Chesters, 1992). Assessing the impact of three characteristics of an image (luminance, contrast and structure), the structural similarity index metric (SSIM) was found by Liu et al. (2011) to predict results of CT image quality more effectively than other popular objective image quality measures (Liu et al., 2011). Working on the premise that the HVS is good at extracting structurally based information, the SSIM evaluates structural changes between two complex structured signals (Mathworks, 1994-2020; Veeramallu et al., 2013).

Edge related information, which is part of crucial structural information in an image that is extracted by the early HVS (ie. before any event interpretation or recognition occurs in the visual process), can be useful for assessing image quality (Zhang et al., 2011). Zhang, Mou and Zhang (2011) proposed a new IQM called the 'non-shift edge ratio' (NSER). This metric investigates whether the edges contained in an image remain in the same location as the original (Zhang et al., 2011). Variation in the number of edge points can measure image quality, as increasing distortion produces increasing change in the edge positions (Veeramallu et al., 2013). 'Zero crossings' computations are used to create the non-shift edge (NSE) maps, where the information carried by the edges as they cross zero in the frequency domain (Veeramallu et al., 2013).

'Radiomics' is an emerging field where quantitative metrics are extracted from medical images (Mayerhoefer et al., 2020). Within this field, image texture, while having no formal description,

provides measures of certain aspects of an image such as smoothness or coarseness and randomness or regularity (Gonzalez & Woods, 2008b). Description of texture in a region can use statistical, structural and spectral approaches (Gonzalez & Woods, 2008b). Texture analysis, using a gray-level co-occurrence matrix (GLCM), characterises regions of an image, attempting to quantify qualities such as smoothness which are intuitively perceived by the HVS, by examining spatial variation of pixel intensities (Gebejes & Huertas, 2013; Mathworks, 1994-2020). Al-Hinnawi and Daear (2012) found the use of five co-occurrence matrix descriptors reliable when assessing the use of a bilateral non-linear filter on MSCT images as a means to reduce image noise. Gebejes and Huertas (2013) examined the use of texture characterization using GLCM, on images of food items with varying degrees of degradation.

2.3.2 Visual perception

Vision is an information processing task, where the information is then ‘represented’ by the brain (Marr, 1982). The brain spends a large part of our waking lives receiving and analysing images, and to quote Confucius, “a picture is worth a thousand words” (Dougherty, 2009b). The eye contains structures that make it analogous to a sensor array and gives it the ability to resolve detail (Gonzalez & Woods, 2008a). In addition, humans then have the ability to integrate our experience and apply our intelligence to what is seen with the eye in the perceptual process (figure 14) (Gonzalez & Woods, 2008a).

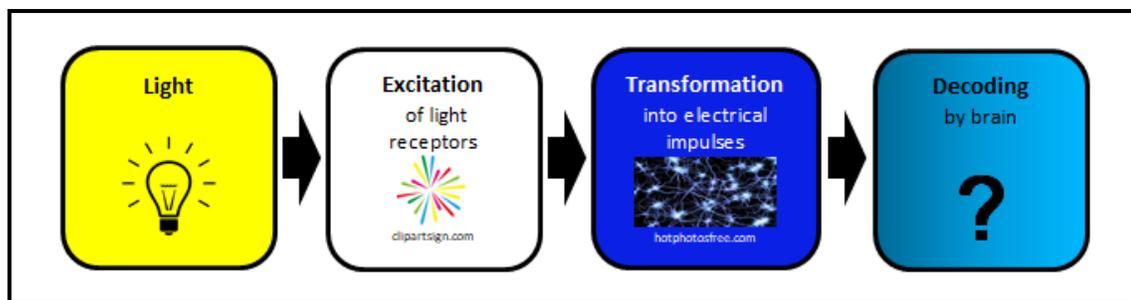


Figure 14. Diagrammatic representation of visual perception. (information from Gonzalez & Woods, 2008a)

The selective mechanism with which the human visual system (HVS) deals with the vast amount of data travelling into our eyes is called ‘visual attention’ (Zhang, 2017). Two types of mechanisms guide this: (1) the bottom-up mechanism and (2) the top-down mechanism (Zhang, 2017). The bottom-up mechanism is stimulus driven by attributes of what is being seen such as motion, colour, contrast and orientation (Koch & Ullman, 1985; Zhang, 2017). The top-down mechanism is linked to cognitive processes such as memory, cultural background and experience (Koch & Ullman, 1985; Zhang, 2017). This top-down mechanism is also called visual salience or saliency, which is the state or quality by which an item stands out relevant to its neighbours (Itti, 2007; Zhang, 2017).

When interpreting a visual representation, the physical assumptions underlying this interpretation may include the existence of surfaces, organisational aspects of differing emphasis (structure), similarity (for example size, contrast, spatial organisation), continuity (curves, lines, more complex patterns) and discontinuities (Marr, 1982). As well as taking into account spatial arrangement, texture discrimination plays a role (Marr, 1982). While some textures are

distinguished straight away in what is described as 'preattentive perception', some require greater 'scrutiny' involving more detailed and lengthier study (Marr, 1982). In addition, disjunctive eye movements may occur, where two lines of sight are converging or diverging (Marr, 1982). Subsequently, a perceptual boundary can also result from variations in visual texture (Marr, 1982).

Perceptual grouping, a visual phenomenon closely associated with perceptual organisation, is when some elements of the visual field are perceived as belonging together more than others (Wagemans et al., 2012). This may be based on values such as luminance or position of these elements, rather than the features themselves (Wagemans et al., 2012). In practical terms, a task of the visual system is to group together portions of an image that projects from an object, to 'exploit contours for object segmentation' effectively, using a form of perceptual contour that is relaxed in order to work reliably (Wagemans et al., 2012). An example of this would be seeing an animal in the woods that is partially obscured by trees (Wagemans et al., 2012). Sabih et al. (2011), in an article which had as part of the title 'Image perception and interpretation of abnormalities; can we believe our eyes?', included a useful table of Gestalt principles (*table 2*) which includes the phenomena described. According to Gestalt theory, perception cannot be reduced to the physical attributes of an image but is rather a global phenomenon (Sabih et al., 2011). Of particular interest in image interpretation, are where image elements may be perceived as a group (whether proximity or similarity), symmetry, and continuity or discontinuity of lines, although all of the phenomena described in *table 2* are relevant in this context.

The HVS can adapt to a wide range of light intensities, with experimental evidence indicating that how the HVS perceives light intensity is logarithmic in nature (Gonzalez & Woods, 2008a). The HVS cannot operate over this wide range simultaneously however, and in a phenomenon known as 'brightness adaptation' it changes its sensitivity to light as the situation dictates (Gonzalez & Woods, 2008a). In psychophysics, a 'just-noticeable difference' or JND is described as the smallest detectable difference between two stimuli (Weber's Law) (Stern & Johnson, 2010). At any specific adaptation level the HVS can discriminate changes in light intensity, and how it detects these changes is affected by the degree of change, where a large change in intensity

requires only poor brightness discrimination but a small change in intensity requires good brightness discrimination (Gonzalez & Woods, 2008a).

Weber's law can also be applied to 'spatial interval discrimination', where the JND is thresholded by line separation (Levi & Klein, 1990). However a critical factor limiting spatial vision is blur, where the 'spread' of an image impacts on the ability of the HVS to resolve the lines (Levi & Klein, 1990).

According to Krupinski, spatial resolution and contrast resolution are the two main aspects of vision that are important for most medical images to be interpreted (Krupinski, 2010b). Visual acuity, while dependent on physiological factors, is also dependent on the physical separation of what is being looked at, the light's wavelength, background illumination and the viewer's dark/light adaptation (Krupinski, 2010b). Of particular importance is the interdependence of contrast and visual acuity (Krupinski, 2010b). Winkler and Mohandas comment that the HVS has difficulty distinguishing distortion at high frequency and difficulty distinguishing content when there is lots of 'image activity' (Winkler & Mohandas, 2008).

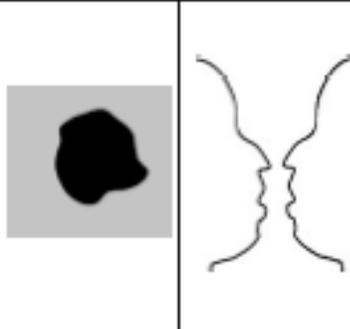
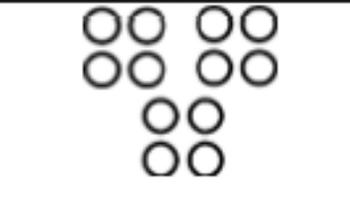
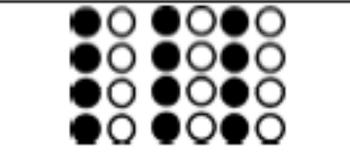
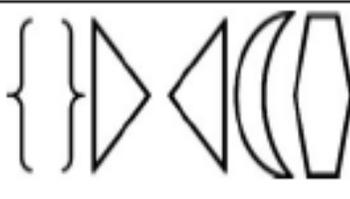
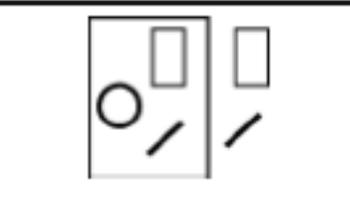
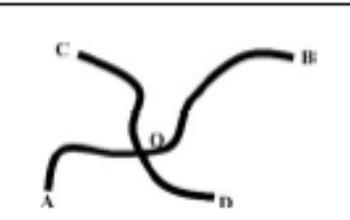
Gestalt Principles		
Figure-Ground/ Multistability	Enables attention to be focused on one part of the image. Usually this is smaller, brighter or otherwise more prominent; the rest of the scene becomes background (image on left). When figure-ground cues are absent, perception alternates between possibilities; perception in the accompanying figure alternates between a vase or two faces (right).	
Proximity	Elements in an image that are close together are perceived as a group. The accompanying picture is perceived to have 3 groups of four circles.	
Similarity	Objects that are alike in form, colour, brightness or size tend to be grouped in the mind.	
Closure	Tendency to fill in spaces; this figure is perceived as a triangle instead of an accurate description in the mind.	
Symmetry	Similar shapes are grouped together, regardless of their proximity; in this case the crescent and hexagon are the odd ones out, despite their proximity.	
Common fate	Elements in a common region are perceived as a group, regardless of similarity or proximity. In this case the figures inside the box look like a group.	
Continuity	The mind prefers to perceive continuous rather than discontinuous lines, and the figure is perceived as the intersection of two curved lines (AOB and COD) rather than the confluence of four lines (AO, OB, CO, OD)	

Table 2. Gestalt principles. (table 1 in Sabih et al., 2011)

2.3.3 Medical imaging

Image perception underlies image interpretation, where in order for the images to be interpreted they have to be perceived first of all (Manning et al., 2005). Perception is by nature imperceptible, an element of diagnosis that can be ignored, and deficiency and subsequent fallibility sometimes hard to admit from a cultural perspective (Zuckier, 2011).

Medical image perception research escalated in the late 1940's when a significant variation was noticed in viewing tuberculosis in x-rays (Krupinski, 2000). The formation of the Medical Image Perception Society (MIPS) in 1997 provided a forum for 'discussion of perceptual, psychophysical and cognitive issues (Krupinski, 2000). This eventually led to an outline of basic goals for perception research to further understand the detection of abnormalities, in unknown locations, and how experience and knowledge impacted on detection (Krupinski, 2000). They also aimed to develop image quality standards that were perception based, computer-aided perception tools, quantitative methods for describing images, and performance measures for human performance of detection and recognition (Krupinski, 2000). Samei and Krupinski (2010) commented that research in image perception is motivated by two main things: (1) human factors which are a main limitation on imaging system performance and (2) human error and error in image interpretation (Samei & Krupinski, 2010).

Krupinski commented that 'medical images are not self-explanatory'. Even with good physical measures of image quality, significant inter- and intraobserver variation occurs (Krupinski, 2011). Ultimately the human eye and brain perform the interpretation, even with the most sophisticated of technologies (Manning et al., 2005). 'Diagnostic accuracy' refers to how well a system predicts presence and extent of disease (or absence thereof), but perception and cognition are at the heart of the interpretation process and diagnostic performance (Krupinski, 2010a).

Samei and Krupinski (2010) pointed out that one cannot define diagnostic accuracy independent of interpretation. How images are used in terms of technical limitations or whether they are suboptimal, has a significant impact on clinical and therapeutic decisions that are possible as a result of their use (Manning, 2010). In addition, it is important that the observer understands

what the image indicates (Manning, 2010). Cognition and perception are intertwined, where new perceived data is continuously interacting with stored knowledge, thus facilitating recognition of image features (Manning, 2010). A number of factors can contribute to inaccurate diagnosis as seen in *figure 15* (Samei & Krupinski, 2010). Cognitive interpretation errors involve a number of psychophysical factors: anatomical noise, 'visual search', peripheral vision during interpretation, dwell times and 'satisfaction of search' (Samei & Krupinski, 2010).

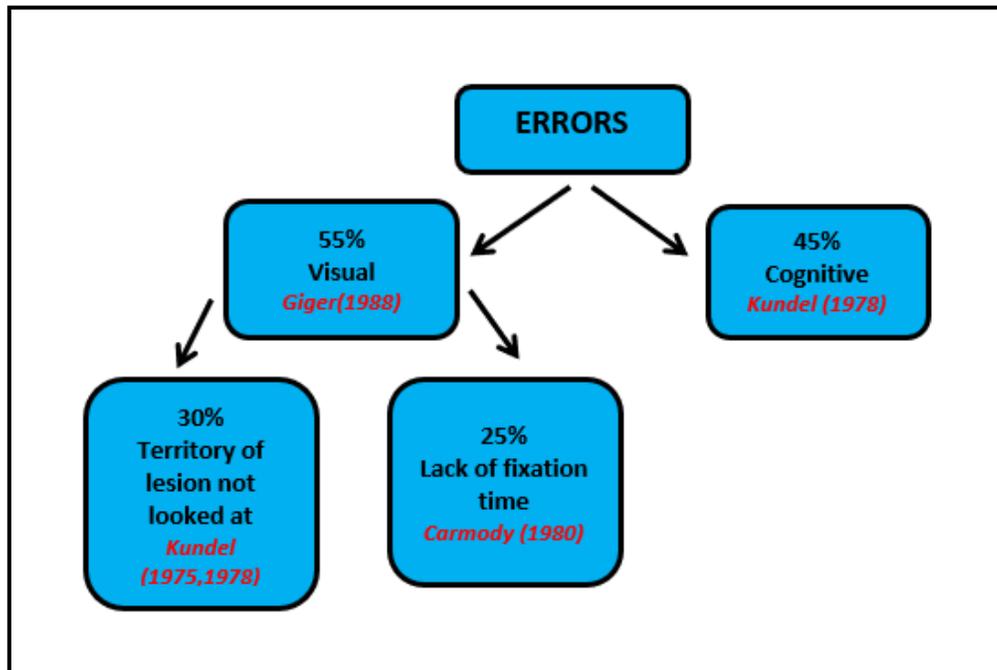


Figure 15. Interpretation errors (information from Samei & Krupinski, 2010)

Chesters (1992) described perception as the challenge of analysing a complex visual scene (with no prior information) and then deciding the importance of features within that scene (Chesters, 1992). Recognition, on the other hand, is described as the ability to memorise and recognise, which interacts with a permanent storage of visual information (Chesters, 1992). Medical images are extensive and complex, the pathologies they contain may be numerous, of varying size and complexity (Chesters, 1992). Conspicuity of any given feature, meaning how well it stands out from the surrounding structures, will also impact on detectability (Revesz, 1985). However, Dendy and Heaton (2013) argued that dwell times were more pertinent than conspicuity, and more likely a greater influence on radiologists (Dendy & Heaton, 2013). Dwell time, refers to the

amount of time image features are fixated on, and contributes to false negative interpretation errors where the dwell time spent is insufficient (Krupinski, 2000).

In a process called 'visual search', in most cases an observer will search for abnormalities in the clinical image (Chesters, 1992). It is the limited angular extent of the 'high-fidelity' foveal vision of the human eye, that makes this necessary (Samei & Krupinski, 2010). Eye tracking studies have shown that distortion in an image can alter gaze patterns, the degree of which is influenced by the extent of the distortion (Zhang, 2017). Whatever the circumstance, search strategies may differ and may not be as equally effective as one another (Chesters, 1992). An example of search strategy was described by Nodine (1987) where a global impression is obtained which is then followed by a detailed examination moving the eye around the image (Manning, 2010). Krupinski on the other hand suggests there is alternation between global and focal patterns of search (Krupinski, 2011). Compounded by unsuspected pathologies, errors in diagnosis can occur some of which can be attributed to an inadequate search strategy (Chesters, 1992). Tuddenham (1962) commented that sources of errors in perception may result from 'incomplete coverage, unpatterned search and premature abandonment'. Berbaum et al. (1986) commented that clinical information, while improving the perception of the image, this is not independent of improved detection because the observer is prompted to look further and adopt a certain search strategy to try and find specific image feature. Conversely, 'satisfaction of search' (SOS) may occur when the existing radiographic findings (which may relate to the clinical information) interfere with further radiographic findings, where clinical importance of existing findings cause termination of the visual search (Ashman et al., 2000; Medical_Image_Perception_Laboratory, 2017). However, Manning, Ethel and Donovan (2004) concluded from eye-tracking studies for chest images, that failures in decision rather than detection were the cause of many missed lesions. Therefore, the phenomenon of SOS is suggested to be one of suppression of recognition rather than a problem completely related to scan or search (Manning, 2010).

Samei and Krupinski (2010) defined medical imaging expertise as "the ability to efficiently use contextual medical knowledge to make accurate and consistent diagnoses". The perceptual and

cognitive analysis that they say is required for this involves a 'rich and structured knowledge' of normal, 'perturbations' from normal and an efficient search strategy (Samei & Krupinski, 2010). Knowledge of perception issues contributes to increased care in image creation and appreciation for image quality and its effect on accuracy and precision (Saunders & Samei, 2010). Understanding is essential that the 'ultimate quality' of an image is not one of how pleasing the image is to look at, but whether the information it provides enables an accurate and confident interpretation (Krupinski & Samei, 2010). However, tissue characteristics and anatomical location of a clinical condition are on what optimal detection or assessment depends (Bushberg et al., 2012a).

2.3.4 Psychophysical measures

Manning (2005) ascribed the success of medical imaging to a dependence on a subjective notion of image quality and on influencing factors on the observer being able to interpret the image information. He divided these image dependent factors into two categories: (1) visual conspicuity of clinically relevant features and (2) primarily cognitive factors relating to the observer's knowledge about the visual information being presented (Manning et al., 2005).

Wagner and Brown (1985) pointed out that all the relevant statistical information in an image is unable to be used by human observers. Surrounding ribs and blood vessels influence the detection of solitary pulmonary nodules according to Revesz et al. (1974), where the effect of 'structured noise' and 'conspicuity' is demonstrated (Chesters, 1992). The same would be if vascular detail overlapped a nodule, this would make detection more difficult according to Kundel et al. (1985), and Ohara et al. (1989) pointed out that detail adjacent to a lesion may also provide some difficulty in detection (Chesters, 1992).

Eckert and Bradley on examining psychophysical validation of IQMs, referred to the human observer as the 'gold standard' to which an IQM is validated against (Eckert & Bradley, 1998). Therefore, the importance of observer performance studies is highlighted as they are closely interlinked with validation of any perceptual IQM (Eckert & Bradley, 1998).

It has been known for a long time that an image of high physical quality is not always of high diagnostic quality (Leitz et al., 1993). Therefore, the complex assessment of clinical image quality is a process that is task-dependent (Berta et al., 2014). Psychophysical studies have their challenges including their time-consuming nature, difficulty to control, the number of images and observers required for robust statistical reliability (Barrett, n.d.). The use of human observers in the assessment of image quality also brings intra- and inter-observer variability (Lundin, 2012). For this very reason, the method used for assessment of observer performance is important (Carrino, 2002).

Psychophysical measurements look at an observer's response to visual stimuli (Zarb et al., 2010). These measures can be obtained using clinical images or phantoms, and while there are a number of ways in which to perform the assessments, the image quality can be measured against fulfilment of a set of criteria (Zarb et al., 2010). Examples of such criteria from a clinical perspective, are contained in the European Guidelines on Quality Criteria for CT (EUR16262) (EC., 2000).

2.3.4.1 Visual grading analysis (VGA)

VGA provides a simple way of quantifying subjective opinions and scientifically evaluating fulfilment of image quality criteria (Sund et al., 2004; Zarb et al., 2010). VGA can be applied using relative grading or absolute grading (Zarb et al., 2010). Relative VGA employs the use of a reference image against which test images are compared for grading, however great care must be taken in the selection of the reference image (Tingberg et al., 2000; Zarb et al., 2010). Absolute grading (also called comparative grading) uses a method where there is comparison and grading of images against each other (Zarb et al., 2010).

The use of image quality criteria such as the European guidelines for quality of CT images (EUR16262), helps identify problem areas in performance of equipment so that corrective action can be taken in the clinical setting (EC., 2000; Zarb et al., 2010). Examples of scales for rating fulfilment of criteria are listed in *table 3 (Ludewig et al., 2010)*. If selection of anatomical

structures is clinically relevant then validity of the study is considered greater (Båth, 2010). There are the added advantages of a low learning threshold and only moderate time requirements (Båth, 2010).

(A) Image Criteria (IC) - study	
1	Clear confident that the criterion is fulfilled
2	Somewhat confident that the criterion is fulfilled
3	Indecisive whether the criterion is fulfilled or not
4	Somewhat confident that the criterion is not fulfilled
5	Clear confident that the criterion is not fulfilled
(B) relative Visual Grading Analysis (VGA) - study	
1	Test image clear superior to reference image
2	Test image somewhat superior to reference image
3	Test image equal to reference image
4	Test image somewhat inferior to reference image
5	Test image clear inferior to reference image
(C) absolute Visual Grading Analysis (VGA) - study	
1	Excellent image quality: no limitations for clinical use
2	Good image quality: minimal limitations for clinical use
3	Sufficient image quality: moderate limitations for clinical use
4	Restricted image quality: relevant limitations for clinical use, clear loss of information
5	Poor image quality: image not usable, loss of information, image must be repeated

Table 3. Examples for visual grading rating scales (Table 1 in Ludewig, Richter, & Frame, 2010)

A disadvantage of this method is that normalization of differences between readers requires high numbers of readers (Zarb et al., 2010). In addition, VGA may also prove problematic where there is pathological distortion of anatomical structures (Zarb et al., 2010). A common error occurs in statistical analysis of the data, where the ordinal ratings are handled as if belonging to an interval

scale, as ordinal data should not be handled as numbers (Bath & Hansson, 2016; Hansson et al., 2016).

2.3.4.2 Receiver operating characteristic (ROC) analysis

Båth (2010) describes the fundamental task of someone reading medical images is to decide if the image demonstrates whether a patient is healthy or diseased. Metz (2006) stated that ROC analysis 'provided the most comprehensive description of diagnostic accuracy to date' due to combinations of sensitivity and specificity it is able to report. ROC analysis, which is based on statistical decision theory, uses the binary ROC paradigm enabling classification of observer responses regarding the presence of a lesion in an image (Börjesson et al., 2005; D. Chakraborty, 2011a). This process incorporates the whole imaging chain (unlike physical measures) and is a widely used method for the subjective evaluation of medical images (Chakraborty, 2010; Zarb et al., 2010).

By comparing responses to the actual 'true' status, a Figure of Merit (FOM) is computed which is an indicator of performance (D. P. Chakraborty, 2011b). This can be used for comparison of image acquisition or processing parameters and their accuracy, or comparison of different imaging systems, by examining the statistical significance of differences in FOMs (D. P. Chakraborty, 2011b; Zarb et al., 2010). Sensitivity is defined by the proportion of correct positive identifications made, and the specificity is defined by the proportion of responses correctly identifying no lesion present (Bewick et al., 2004). From this perspective, ROC analysis is sufficiently complex and becomes particularly useful when examining diagnostic efficacy of a procedure or piece of equipment (Metz, 2006).

In a review article, Zarb et al. described several developments that ROC methodology has undergone (*table 4*), in order to overcome various limitations such as the number of abnormalities able to be categorised in an image, numbers of images needed and numbers of readers needed to obtain greater statistical power, and also developments in how false positive and false negative data is used (Zarb et al., 2010). Chakraborty (2013, 2015) has described further

developments, describing EFROC in an article (2013) on the history of the FROC paradigm, and released RJAFROC software in 2015 (*table 5*).

Abbreviation	Full name	Description	Comments
ROC	Receiver Operator Characteristics	Discriminates between normal and abnormal images.	Unable to discriminate between more than one abnormality. Does not consider the location of the lesions. Low statistical power
LROC	Localization ROC	Identifies localization of one abnormality in the images.	Incorrect location of lesion considered as false positive.
FROC	Free-response ROC	Identifies and localizes several abnormalities within the images	Issues with search satisfaction and how to treat false positives and false negatives within same image. FROCFIT software available.
JAFROC	Jack-knife Analysis of FROC	A re-sampling method that does not assume independence of responses within the same study applied to FROC.	Very stable method but requires multiple readers (approx.5) and substantial number of images (approx.100) for high statistical power.
AFROC	Alternative free-response ROC	An alternative way to analyse FROC. Images contain noise looking like a signal (False Positive Image)	Similar to FROC. Assumed independence between ratings observed on the same image. Criticized with limited acceptance of use. ROCFIT software available.
FFE	Free-response forced error	Does not assume independence	Has not been validated. Statistical power may be poor.
DROC	Differential ROC	Determines differences between modalities.	High statistical power.

Table 4. Summary of ROC methods (Chakraborty & Berbaum 2004; Chakraborty 2004, 2005, 2006; table 1 in Zarb et al., 2010)

Abbreviation	Full name	Description/comment
EFROC	Exponential free-response ROC	○ Uses multiple non-lesion localizations on normal images rather than just the highest rated one (as in AFROC)
RJAFROC	R jack-knife free-response ROC	○ allows lesion weighting according to their importance ○ allows choice between DBM or OR significance testing ○ choices of several FOM ○ sample size estimation tools

Table 5. Further ROC developments (Chakraborty, 2013; Chakraborty, 2015)

2.3.4.3 Visual grading characteristics (VGC)

A development using visual grading, was published by Båth and Månsson in the (2007) British Journal of Radiology presenting their work on VGC analysis. Developing VGA a step further, in VGC the observer is asked to rate his confidence about how well image quality criteria are fulfilled (Båth & Månsson, 2007). Rating data is then analysed in a similar manner to ROC analysis (Båth, 2010; Båth & Månsson, 2007). VGC analysis of visual grading data is defined as a 'non-parametric rank-invariant' method (Bath & Hansson, 2016).

To facilitate implementation of VGC studies, Håkansson et al. released their ViewDEX 2.0 software that superseded the 2007 version 1.0 (Håkansson et al., 2009). This software allows medical images to be displayed while registering the observer's response simultaneously (Håkansson et al., 2009).

The basic idea of visual grading relies on the assumption of correlation between the reproduction of anatomy and the possibility of detection of pathology (Båth & Månsson, 2007). Kundel (1979) argues that images of the highest diagnostic quality enable the most accurate report of diagnostically relevant structures and features (Båth & Månsson, 2007). In today's environment of dose optimization and ALARA, to balance radiation dose against clinical indication, it is necessary to assess images for their fitness for purpose against international standards, therefore using subjective methodology such as VGC in a quantitative manner proves useful (Ledelius et al., 2010).

2.3.4.4 Equivalency

When comparing diagnostic tests or technologies, not finding statistical evidence of difference between them is not evidence of equivalency (Tourassi, 2010). Observer performance may be equivalent at a certain level but not the images, or the numbers may be too small to adequately detect a difference (with the lack of statistical power causing a probability type 2 error) (Tourassi, 2010). Obuchowski (1997) stated that definition of equivalency is often poor, and proper

statistical testing rarely carried out. A false claim of equivalency could result in the subsequent use of inferior technology potentially putting the public at risk (Obuchowski, 1997).

What is relevant is the clinical equivalence of the technologies or tests being compared due to the potential clinical implications (Obuchowski, 1997). Different statistical testing needs to be carried out to prove equivalency, and these are not new (Tourassi, 2010). Bioavailability of drug formulations are often checked in Pharmacology using these methods (Obuchowski, 1997). As an example, Chen et al. used multireader multicase ROC (MRMC ROC) to assess equivalency when comparing different imaging modalities (Chen et al., 2012).

An equivalence test that can be used is based on the two one-sided tests (TOST) method (Tourassi, 2010). This method defines equivalence as 'the presence of difference that is bounded between a lower and upper bound' where clinical criteria set the bounds (Tourassi, 2010). Obuchowski notes that the interval does not necessarily have to be symmetrical around zero (Obuchowski, 1997).

2.4 Statistical relationships

While assessing the effectiveness of various IQMs, the object of this research is to identify if there are any relationships between observer performance studies and image characteristics, and then if possible, determine the nature of any relationships found. However, Tapiovaara (2008) commented that although work on establishing this kind of relationship has been pursued for a number of years, it is still not fully understood.

The relevance of a single statistical result tells only part of the story (Altman, 1991). Altman (1991) states that p-values as informative as they are, only tell part of the story and need to be looked at in context of what actually was observed. Describing something as 'statistically significant' is not a meaningful interpretation on its own, where confidence intervals are needed as it is only significant if the 95% confidence intervals do not include zero, and variations within a patient group or therapy should be examined for example (Altman, 1991). While a correlation coefficient indicates the strength of an association as a single number, to describe the

relationship between two variables then regression analysis would be needed (Altman, 1991). 'Goodness-of-fit' provides an assessment of how well a model fits the data or how well the dependent variable is predicted, and 'analysis of residuals' provides an assessment of how good a model's prediction is for values of the dependent variable for individuals (Altman, 1991). As demonstrated, all aspects of statistical analysis should not happen in isolation but operate together with perfect precision giving attention to detail to produce the correct result. Of particular note is the importance of distinguishing between interaction, correlation and causation, all of which are quite different concepts (Aldrich, 1995; Altman, 1991).

Interaction in a statistical context refers to when the simultaneous influence of variables on another variable is not additive (Cox, 1984; Oxford_Dictionary_of Statistical_Terms, 2006). Occurring most commonly in the context of regression analysis, detecting interaction can be achieved by using F-tests and graphical representations of data (Cox, 1984). To interpret interactions in greater depth, calculation of the F-ratio for the various individual transformations could be useful (Cox, 1984). P-values, may provide a useful guide but are not always necessary for interpretation (Cox, 1984). Graphical methods may be advantageous when there a large number of component interactions or to highlight areas of large contrast needing more detailed examination, to assist in deciding which contrasts are reasonable for use in determining estimations of error or detection of error structure previously undetected, and obviously graphical representation is useful when presenting results (Cox, 1984).

Correlation refers to the relationship (or dependency) between two variables in both strength and direction (2020 a). This gives some understanding of how the variables co-vary, the extent to which common causes govern them, their dependence on each other and their predictive relationship (Aldrich, 1995). A correlation coefficient quantifies the strength of a relationship and the linearity of that relationship (Bewick et al., 2003). Common errors of its use can happen if a non-linear relationship or one that is not 'monotonic' is inadequately described by the various correlation coefficient methodologies available, the presence of clusters resulting from distinctive subgroups or the presence of another variable's interaction (spurious relationship) (Bewick et al., 2003; Shaughnessy, Zechmeister & Zechmeister, 2012; Viera & Garrett, 2005). In

addition, high correlation does not necessarily imply agreement between two methods therefore in some instances Bland-Altman plots, which allow for outliers and systemic difference between the measurements, might be more appropriate (Bewick et al., 2003; Giavarina, 2015; Shaughnessy et al., 2012; Viera & Garrett, 2005).

Causation or causality is examined when determining the cause of a relationship, but cannot necessarily be inferred from knowing a correlation (Shaughnessy et al., 2012). A combination of a multimethod approach along with correlational evidence can help identify causal relationships (Shaughnessy et al., 2012). Predictions for correlated variables can also be facilitated by correlation evidence (Shaughnessy et al., 2012). Sophisticated statistical techniques can be employed to identify causality such as 'path analysis' (Shaughnessy et al., 2012). This technique involves identifying mediator and moderator variables, where the mediator variable looks at correlation between two variables and the moderator variable affects the direction or strength (Shaughnessy et al., 2012).

3.0 Chapter Three: Hypothesis and Null hypothesis

The aim of this research was to examine image quality in low-dose CT images, as determined by a range of image quality measures (IQM), in addition to psychophysical assessment involving specific tasks. This was followed by correlation of IQM data and observer performance to determine any quantifiable relationships. The hypothesis and null hypothesis have been devised with this intent.

3.1 Hypothesis

There is a quantifiable relationship between image quality and observer performance in specified tasks.

3.2 Null hypothesis

There is no quantifiable relationship between image characteristics and observer performance in specified tasks.

4.0 Chapter Four: Method

4.1 Introduction

This thesis has a number of inter-related components, of which the major sub-components are outlined in *figure 16*. This chapter explains the methodological approach of the thesis sub-components relating to image data and image analysis. Equipment used, including imaging phantoms, and details of image acquisition are covered. In addition, image analysis in combination with the software platforms used, and observer performance studies and their analysis are detailed. How the method related to the aim of the thesis is addressed in the Methodology section which follows, in addition to why this approach was used.

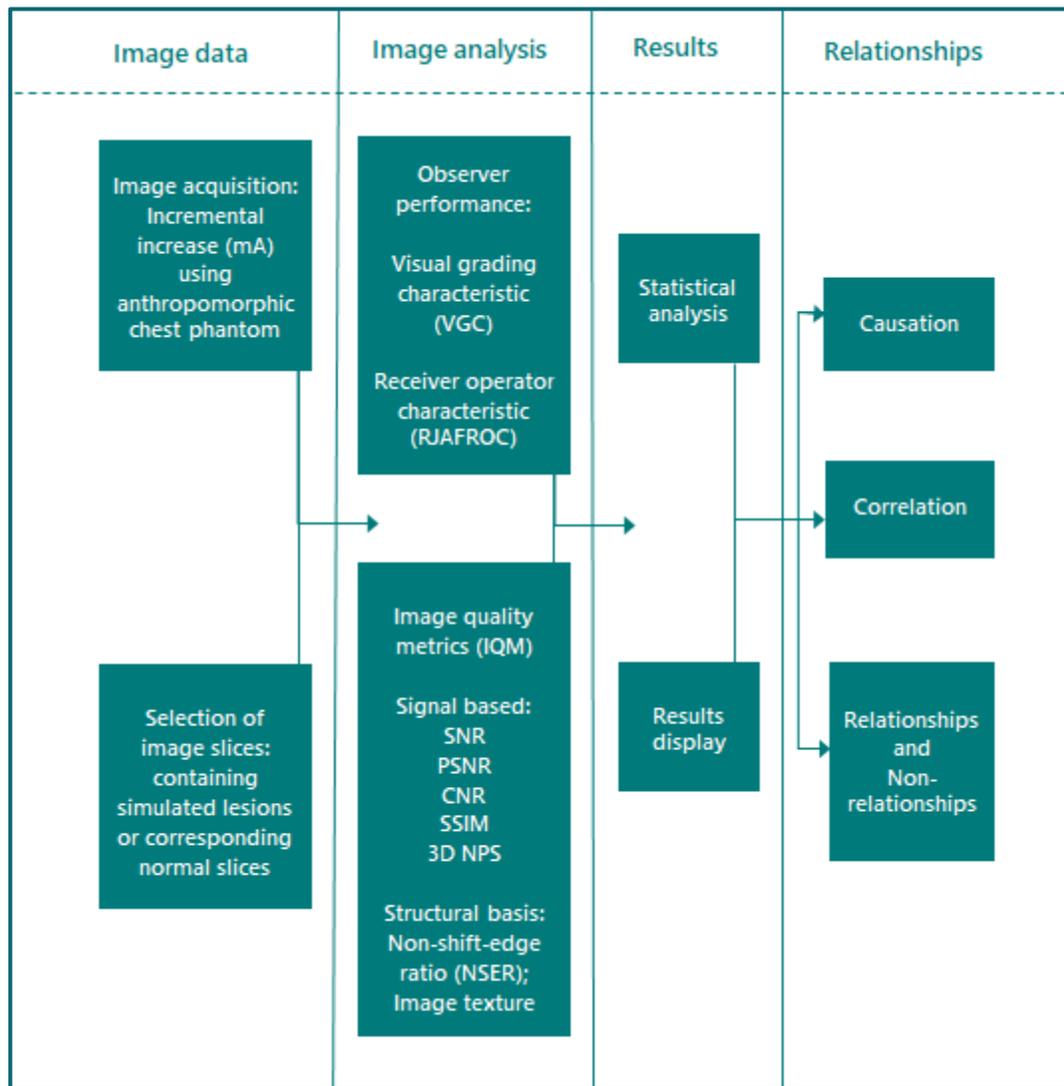


Figure 16. Inter-relation of thesis major sub-components.

4.2 Methodology

The aim of the research is to relate both the physical characteristics of an image and visual perception of that image, with varied radiation exposures. To facilitate this, phantom images were obtained at varying amperage (with particular emphasis on lower amperage values), which were then analysed to assess physical characteristics of the images. Statistical analyses of these results provided information not only about the effectiveness of the image quality metrics (IQMs) themselves, but also how physical image quality related to the lower doses of radiation that were used to acquire the images. The same images were also used in psychophysical testing, where observer performance provided information on how the images were perceived in relation to the quality of images acquired at varying radiation exposures. Specific tasks were carried out relating to clear visualization of structures according to international image quality criteria (EC., 2000), the effects of image noise on the decision-making process and lesion localization.

A variety of image quality measures (IQM) were used to assess image noise, image contrast, edge definition and complexity of surrounding structures. Image characteristics were examined using IQM ranging from those more signal based such as peak signal-to-noise ratio (PSNR), to those closer approximating the human visual system (HVS) such as structural similarity index (SSIM), Non-shift-edge ratio (NSER) and image texture. Most authors have published research in journals that used a limited number of IQM (only two or three in many cases), the nature of which are influenced by the emphasis of their work. Examples of this are articles by Pahn et al. (2016) in their work looking at standardisation of image quality assessment quality using SNR and CNR, or Sasi varnan et al. (2011) in their work looking at image quality assessment using MSE, PSNR and SSIM (Pahn et al., 2016; Sasi varnan et al., 2011). To date, no single IQM is able to capture all aspects of image quality. Using a range of IQM (from signal based to those more perceptually based), this thesis examined each IQM individually and looked at the relationship between them. Quantification of these relationships (or non-relationships) was looked at in addition to correlation with observer performance.

Preparatory images were obtained on a hybrid PET/CT system and a stand-alone CT unit, using a physics phantom and an anthropomorphic chest phantom respectively, to assist in developing

analysis software and establishing IQMs suitable for the project. Key considerations at the outset were: (1) IQMs and what they measure, what they demonstrate and how to measure them, and why they were appropriate, (2) image parameters and how to alter them and effects this has on the images produced, (3) observer performance studies including their types, what tasks would be asked of the observers and their suitability for specific aspects of the research, (4) establishing the method and (5) preparatory image acquisition.

Estimation of observer performance using VGC analysis of visual grading data, was chosen. This was performed using 'ViewDEX' software (Håkansson et al., 2009). Data was provided linking visually assessed image quality and international standards for CT image criteria. In addition, statistical examination on the impact of 'noise' was facilitated. This software also provided an opportunity for an image localization task, which enabled ROC analysis to be performed.

The intended primary outcome of the thesis was to produce correlative statistics defining correlation or non-correlation between visual performance and image characteristics. Secondary outcome measures were also anticipated such as validation of the various IQMs used, as well as defining any further statistical relationships.

4.3 Equipment

4.3.1 CT equipment

There were two CT units used in the acquisition of the phantom-based images. The first, used only in the acquisition of preparatory physics phantom data, was the Biograph 64 which is a PET/CT unit incorporating a 64 slice CT. The second CT unit, used to create images of an anthropomorphic phantom for the main body of the thesis, was the Aquilion 16 which is a stand-alone 16 slice CT unit. These units complied with equipment standards as defined by the Quality Assurance Reference Centre for the European Commission, which contained parameters with very specific suspension levels above which the equipment should not be in operation (EC., 2012).

4.3.2 Imaging phantoms

Two imaging phantoms were used for this thesis. Preparatory images were created using a CATPHAN[®] 600 physics phantom manufactured by The Phantom Laboratory (The Phantom Laboratory, Incorporated, PO Box 511, Salem, NY 12865-0511) ("Catphan 500 and 600 Manual," 2015). Images used for the main body of the thesis were created using an anthropomorphic chest phantom, N1 'Lungman', manufactured by Kyoto Kagaku (Head office: 15 Kitanechoya-cho, Fushimi-ku, Kyoto, 612-8388, JAPAN) (Kagaku).

4.3.2.1 Catphan[®]

The 'CATPHAN[®] 500/600' series are physics phantoms of a modular design (*figure 17*), designed for multi-slice CT scanners to measure performance of the equipment for geometric and low-contrast sensitivity, sensitometry and uniformity ("CATPHAN 600 [Brochure 'CATPHAN 500/600']," n.d.). In particular, the CTP515 module is a low contrast module containing a series of supra-slice (2-15mm diameter) and sub-slice (3-9mm diameter) cylindrical rods with three contrast levels (*figures 18 and 19*) ("Catphan 500 and 600 Manual," 2015). This module was used to readily observe the effect of attenuation and scatter on image quality when an attenuation ring was used (*figure 21*). The attenuation ring called an 'annulus' cast from uniformity material, slipped around the phantom, and could be used to mimic body scatter. The type used in image acquisition was an oval annulus, model CT579 (*figure 20*), that had dimensions 25cm (height) x 35cm (width)("Catphan accessories," 2018).

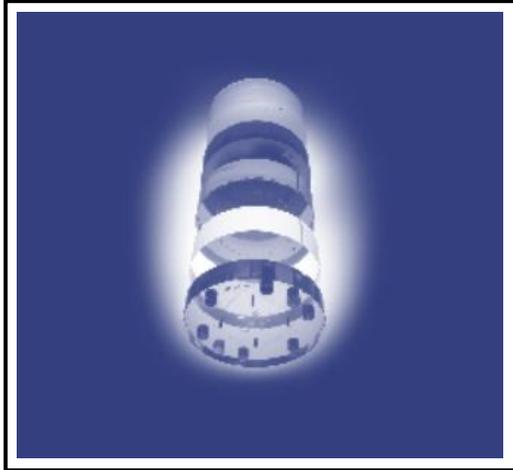


Figure 17. CATPHAN® (in "CATPHAN 600 [Brochure 'CATPHAN 500/600']," n.d.)

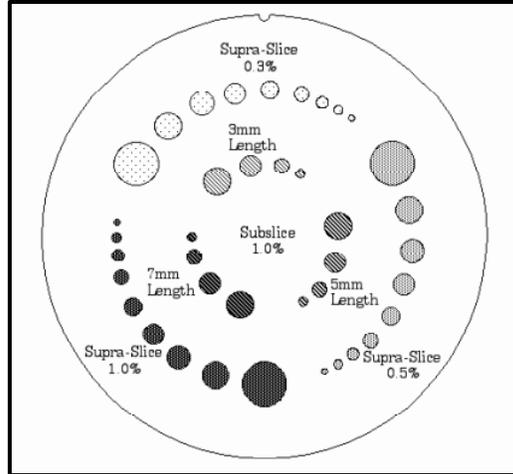


Figure 18. CATPHAN® inserts for module CTC515 (in "Catphan 500 and 600 Manual," 2015)

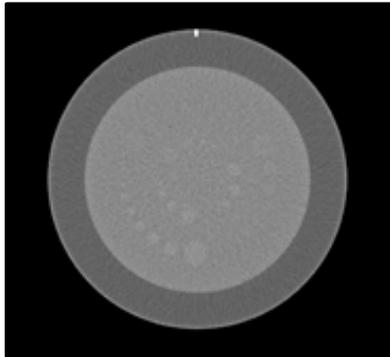


Figure 19. CATPHAN® phantom image using CTP515 low-contrast module.

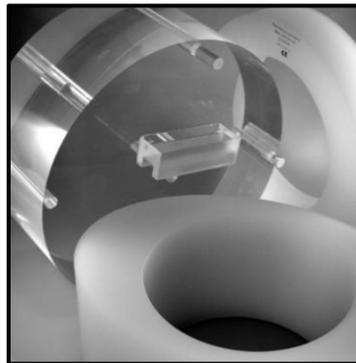


Figure 20. CATPHAN® annuli ("Catphan accessories," 2018)

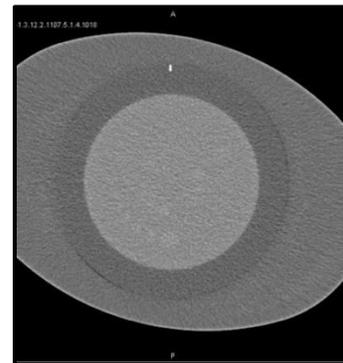


Figure 21. Image of CATPHAN® using a CTP515 low-contrast module with annulus model CT579.

4.3.2.2 'LUNGMAN'[®] anthropomorphic chest phantom



Figure 22. 'LUNGMAN'[®] (Kagaku, n.d.)



Figure 23. Mediastinum and pulmonary vessel insert for 'LUNGMAN'[®] (Kagaku, n.d.)

This is a multipurpose anthropomorphic chest Phantom (*figure 22*) which is an accurate life size, torso made from materials that approximate x-ray absorption rates of human tissue (Kagaku, n.d.). The soft tissue and vessels are made from Polyurethane (gravity 1.06, referring to its density), and the synthetic bones from epoxy resin and calcium carbonate (*figures 22 and 23*) (Kagaku, n.d.; Zimmerman Jones, 2018). A CT image of this phantom is displayed in *figure 26*. Chest plates are also provided that mimic additional soft tissue attenuation (*figure 22*) and are commonly known as an 'attenuation jacket' or a 'fat jacket' referring to the way additional body fat is mimicked in this region (Kagaku, n.d.). This phantom was particularly suitable as it approximated the human body while allowing for repeated radiation exposures.



Figure 24. Simulated lesions 5mm, 8mm, 10mm (partial image from Kagaku, n.d.)

There are varying sizes of simulated lesions (3mm,5mm,8mm,10mm,12mm) supplied with three different sets of Hounsfield units (approximating +100HU,-630HU, -800HU) (*figure 24*) (Kagaku, n.d.). The +100HU simulated lesions are made from Polyurethane, SZ50 and Hydroxyapatite (Kagaku, n.d.). The -630HU and -800HU simulated lesions are made from Urethane foam. These are placed carefully in the pulmonary vessels and removed carefully upon completion (*figure 25*).



Figure 25. Simulated lesion insertion (*Kagaku, n.d.*)

The importance of the HU values is that they approximate structures and pathologies found in the lungs. For example, ground glass opacities are found at approximately -600HU (Kitami et al., 2012) and lung tumour densities typically lie in the same range as blood vessels in the lung between +100HU and +200HU (Peskin & Dima, 2010).



Figure 26. 'LUNGMAN'® anthropomorphic chest phantom image

4.4 Data acquisition

4.4.1 Preparatory images

Preparatory images were obtained using the CATPHAN[®] 600 physics phantom, with the CTC515 insert which was designed to assess parameters found in low CT contrast areas, for example the abdomen or soft tissues of the neck. Images of these areas are prone to be more 'noisy' and sensitive to improved photon flux with increased x-ray exposure as a result of increasing amperage ("Catphan 500 and 600 Manual," 2015). This precise knowledge of the components in this physics phantom was useful, as image features were able to be clearly defined.

The PET/CT scanner used was a 'Biograph 64' which has a 64-slice CT integrated into the equipment. Once the correct positioning of the phantom was established, the voltage was selected to be 120kV as this was a Siemens[®] standard setting for this purpose. Both acquisition slice thickness and reconstruction slice thickness were set at 5mm, to mimic reconstruction parameters often found in LDCT images. Filtered back-projection was used in reconstructing the images and the B31s kernel was selected as a filter for this process because, while the B30s or B40s are standard kernels recommended for soft tissue parts of the body, the low contrast detectability of the B30s setting is improved by using the B31s kernel (Siemens, 2007). Optimised image reconstruction parameters, therefore, enabled a focus on image quality while minimising their contribution to any changes between the images.

Acquisitions were initially obtained over 6 slices through the CTC515 insert at incremental amperage settings. The CT579 oval annulus was then used to mimic body scatter. The images obtained showed incremental changes in quality with a known variable, and also demonstrated that image quality was markedly affected by use of the annulus. As the purpose of this thesis was to examine changes in image quality using a number of image quality metrics (IQM), these images were suitable to be used for setting up image analysis software and methods. Example of images obtained are seen in *figure 19* and *figure 21*.

4.4.2 Main study

Images for the main study were acquired at the University of Salford imaging laboratory, on the 'Aquilion 16' CT system. The anthropomorphic chest phantom ('LUNGMAN'[®]) was used for the image acquisitions.

Simulated lesions of sizes 5mm, 8mm and 10mm were placed in the phantom for each of the three densities available (+100HU, -630HU and -800HU). The very smallest size simulated lesion (3mm) was excluded as identifying this lesion would not be possible, especially at lower amperage levels, due to its similarity in size to the smaller lung vessels and image viewing would not contain any scroll facility. The largest size simulated lesion available (12mm) was also excluded because it would add little to the visual assessments and image analysis due its obvious nature.

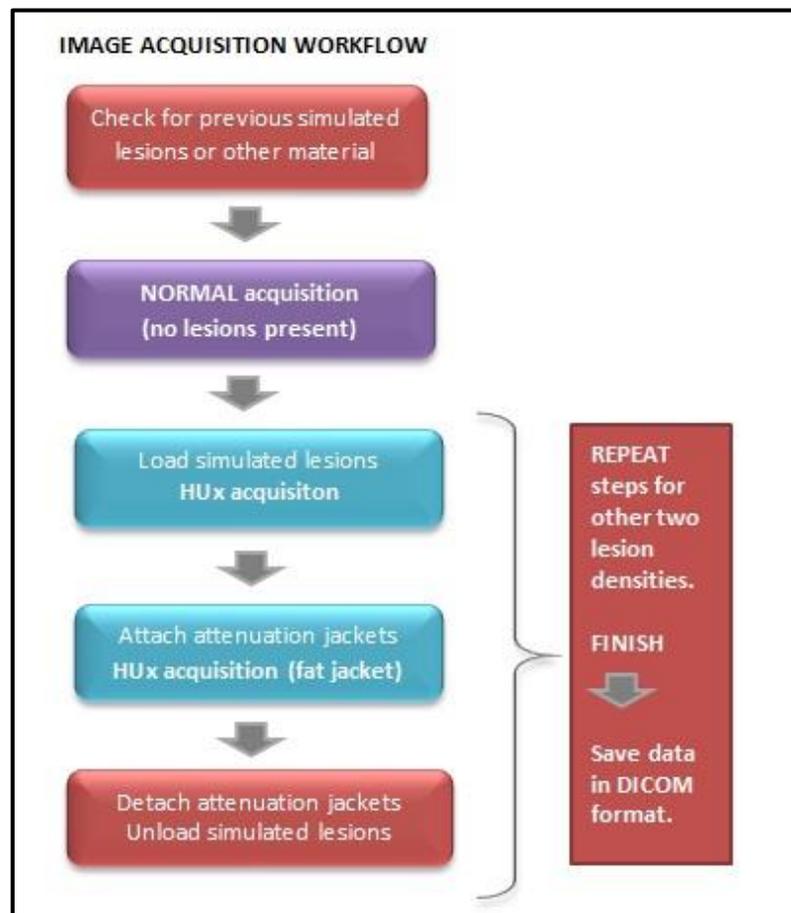


Figure 27. Acquisition workflow.

A concise acquisition workflow is seen in *figure 27*. In the first instance, it was necessary to obtain normal image data with no added simulated lesions. While providing images for observer performance studies, this also functioned as a quality control measure to assess the integrity and quality of the phantom thus ensuring that no simulated lesions or other materials remained in the phantom from previous use. Simulated lesions were then placed in the phantom just three at a time, of the same lesion density. These were placed in a specific pattern covering the upper, middle and lower regions of the lung (*figure 28 and table 6*), according to zones assigned for assessment of Chest X-rays (Radiology-Masterclass, 2020). The 10mm lesion was placed centrally on the left side, close to major vessels to assess the effect of these structures, on image lesion detectability. The 5mm lesion was placed near the right apex to assess peripheral effects, and the 8mm lesion was placed in lower part of the right lung (*figure 28*). This process was repeated for each of the lesion densities. Before removing each set of lesions, a further image series using attenuation jackets was acquired, to obtain further images with increased x-ray scatter and lower image quality. Each of these sets of acquisitions were acquired only once, from which a large amount of data were produced from the analyses. This was also limited by time and distance from the scanner and imaging phantom.

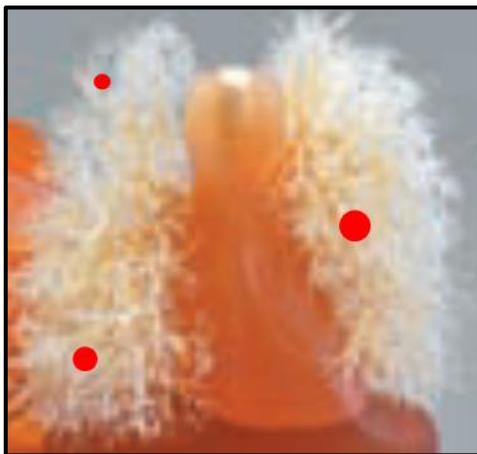


Figure 28. Simulated lesion placement

Lesion size	Indicates Lesion placement
5mm	
8 mm	
10mm	

Table 6. Simulated lesion size placement (Legend)

Helical acquisitions were obtained, without dose modulation. Care was taken to start the acquisition in the same position each time, so that direct comparison could be made between all images in the same series and avoid any anomalies as much as practicable.

Acquisition and reconstruction settings are listed in *table 7*. The acquisition protocol used was a default thorax protocol which was then modified for voltage and slice thickness. The voltage (120kV) is a typical voltage used for low-dose CT (LDCT), and the slice thickness was to enable appropriate reconstruction for the purposes of analysis. Two sets of reconstructions were performed. The first set of reconstructions, using small slice thickness, were used to assist with any partial-volume issues in determining correct image slice selection for analysis.

Acquisition	
voltage	120kV
helical pitch	23
pitch factor	1.438
slice thickness	1.0mm x 16 slices
Reconstruction(1)	
slice thickness	1.0mm
reconstruction interval	1.0mm
Reconstruction(2)	
slice thickness	5.0mm
reconstruction interval	5.0mm

Table 7. Image acquisition and reconstruction parameters.

Images were obtained using amperage settings in 5mA increments at lower amperage (10mA to 50mA), and 10mA increments thereafter. An additional acquisition was obtained at 200mA as this was required for reference images during image analysis. The smaller 5mA increments used in the lower end of the amperage range were needed, as graphed data with suitable detail was then able to be obtained from image analysis in that range.

The relevant image slices containing simulated lesion data were identified and stored in files ready for image analysis. This was achieved by using the reference images acquired at 200mA and checking through each of the image series to confirm the slice best displaying the simulated lesion. This was of particular importance where partial-volume effects were apparent for 5mm lesions due to the reconstructed 5mm slice thickness, in addition to facilitating consistent selection of image slices for all lesion sizes.

The windowing of the image display was set at a window of 1500HU, and centred at -500HU. In a special report, Bankler et al. (2017) described most studies investigating lung nodule measurement, for both accuracy and variability, had used wide lung window settings centred

between -700HU and -500HU with the window width ranged between 1500HU and 2000HU. While in the clinical setting the lung window may typically be centred at -600HU, for the purposes of this research, centering the window at -500HU rather than -600HU enabled the effect of amperage on image quality to be viewed more readily with the streaking artifact more apparent. In addition the image data was more centred at -500HU when the HU spectrum was viewed in MATLAB®.

4.5 Image analysis

There were eight IQMs specifically chosen for image analysis of the anthropomorphic chest phantom, which are listed in *table 8* with the image characteristics they assess. Included are a mixture of both signal-based IQMs and structure-based IQMs (which approximate human visual perception more closely). A variety of software platforms were used for these analyses which included MATLAB®, LIFEx and ImageJ. Details pertaining to the IQMs and software used, are seen further on in this section. After checking for reproducibility, final image analyses were carried out only once.

<i>IQM</i>	<i>Acronym</i>	<i>Image characteristic</i>	
Mean squared error	MSE	Distortion / error	
Peak signal-to-noise ratio	PSNR	Comparison of peak signal in two images	
Signal-to-noise ratio	SNR	Detectability of object	
Contrast-to-noise ratio	CNR	Signal level of object in relation to ambient noise	
Structural similarity index	SSIM	Patterns in local pixel intensities	
Noise power spectrum	NPS	Noise characterisation/ noise texture	
Non-shift-edge ratio	NSER	Edge definition	
Texture analysis: Entropy, energy (uniformity), contrast, correlation, homogeneity		Spatial variation of pixel intensities. Examines aspects such as smoothness or coarseness and randomness or regularity in relation to image structure.	

Table 8. Image quality metrics (IQMs) and related image characteristics they assess.

4.5.1 Image analysis using MATLAB®

Global image analysis (looking at each image in its entirety) was carried out in MATLAB®, which provides an environment in which to compute, analyse and visualize data (Hunt et al., 2006). It contains comprehensive image and signal analysis tools making it possible to analyse image characteristics (Mathworks, 1994-2020). MATLAB® is a product of Mathworks Inc. and provides online assistance in some instances.

Designed to work on arrays and matrices, MATLAB® is abbreviated from 'matrix laboratory' (Mathworks, 1994-2020). The programming language uses syntax that combines simple mathematical operations with 'call functions', where the functions are incorporated into the code being written to carry out computational tasks. (Mathworks, 1994-2020).

CT images were uploaded into MATLAB® in DICOM format, where they were then selected for processing and converted to a multi-dimensional array (Mathworks, 1994-2020). Coding was created to facilitate data analysis. Some of the IQMs required a reference image, therefore following the selection of the image requiring analysis, a second image was selected for this purpose.

All images were converted in MATLAB® to a double precision 'floating point' format. The 'double' function (standing for double precision), converts the image file to a 'floating point' format (Roberson, 2015). In floating-point arithmetic terms, this refers to the decimal point which can be moved or floated relative to the significant digits in a number, or in computer terms this refers to the binary point. This is of particular importance when representing numbers of different orders of magnitude, and to ensure compatibility of image file formats when being operated on within the MATLAB® environment (Roberson, 2015).

Image quality metrics analysed using MATLAB®

a) The **mean square error (MSE)** aims to compare two images giving a score describing distortion or error (Rangaraju et al., 2012). The reference image for comparison was acquired at 200mA. The calculation (*Appendix 1*) produces values that are lower in the case of greater similarity (Sasi varnan et al., 2011). A lower value, therefore, implies an image of greater image quality.

b) The **peak signal-to-noise ratio (PSNR)** allows comparison of the peak signal in two images and is inversely proportional to MSE. The reference image for comparison was acquired at 200mA. The calculation (*Appendix 1*) produces values that are higher in cases of greater similarity (Sasi varnan et al., 2011). A higher value, therefore, implies a higher quality image.

c) The **non-shift edge ratio (NSER)** allows edge definition to be examined, by using 'zero-crossings' in the frequency domain (Zhang et al., 2011). Marr defines a zero-crossing as 'a place where the value of a function passes from positive to negative' (Marr, 1982). The zero-crossings, which represent edges corresponding to spatial locations in an image, change with blurring or distortion as the edges become less clearly defined (Zhang et al., 2011). Where two image matrices are analysed and comparison is made looking for edges that haven't shifted (the non-shift edges (NSE)), then a ratio can be calculated (*Appendix 1*) (Xue & Mou, 2011), and the amount of degradation expressed as an inverse of this ratio (Zhang et al., 2011). The reference image for comparison was acquired at 200mA.

The graphic in *figure 29* illustrates the principle of this IQM, where similarities to the reference image matrix (C) are still seen in the distorted image matrix (D) remaining green. Where the zero-crossings have changed (demonstrated in the frequency diagram *figure 29* (E)), these are then indicated in the distorted image matrix (D) as red. The similarities (green) in the distorted image matrix (D) are the edges that have not shifted. These are divided by the edges represented as green in the reference image matrix (C), to give the non-shift-edge ratio. The closer this ratio is to 1, the better the quality of the image.

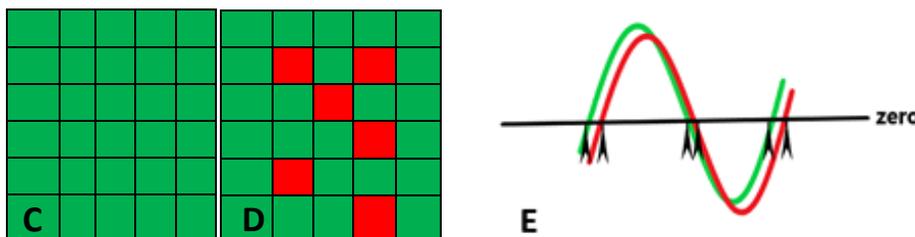


Figure 29. Graphic of Non-shift-edge ratio:
 (C) reference image matrix;
 (D) distorted image matrix: similarities to (C) in green and shift-edges indicated in red;
 (E) Frequency diagram indicating zero-crossings with arrows.

d) The **structural similarity index metric (SSIM)** uses the three characteristics of structure, contrast and luminance to perform a comparison of an image to a reference image, looking at local patterns of pixel intensities which have been normalized for contrast and luminance (Choi & Lee, 2011; Mathworks, 1994-2020). The multiplicative combination of all three terms provides the overall index (*Appendix 1*). The closer to an index of 1, the greater the similarity (National_Instruments, 2015). To achieve this metric in MATLAB[®], the call function 'ssim' was used.

e) The **noise power spectrum (NPS)** enables a fuller characterisation of noise in an image (Seeram & Hsieh, 2009f), describing the texture of noise contained in an image by correlating the noise frequency content at one point with its surrounds (Bushberg et al., 2012a). The analysis of spatial frequency distribution and magnitude provides a more complete description of noise in an image (Miéville et al., 2012).

NPS analysis was performed using a separate MATLAB[®] code. This method provided a 3D NPS as recommended by the International Commission on Radiation Units and Measurements (ICRU, 2012). Due to potential for variation in local measurements of NPS in the axial plane, and for consistency throughout all the image analyses, NPS was calculated on a global basis rather than on a regional basis (Pineda et al., 2012).

To create the power spectral density (PSD) component a parametric method was chosen, because higher resolution was possible with short signal lengths, which corresponds to small features and noise in an image (MATLAB, 1994-2020). In addition, the PSD obtained would be less prone to issues relating to spectral leakage, which causes distortion when a frequency component is spread over adjacent frequency 'bins'. The Yule-Walker AR method (autocorrelation method) was used specifically, because autocorrelation methods are more suitable where the PSD data has a wide variation and is large at certain frequencies, and it produces a stable model (MATLAB, 1994-2020). The call function 'pyulear' was used to obtain the PSD estimate (*figure 30*), where the number of discrete Fourier transform (DTF) points was altered to 512 to match the matrix size and avoid overlap (MATLAB, 1994-2020). The variance (σ^2) was then calculated to provide the variables required to display the results.

To display these results in a meaningful manner, a ‘pseudospectrum estimate via the Eigenvector method’ was plotted (*figure 31*) and the resultant display file of the graph saved (MATLAB, 1994-2020). The ‘Eigenvector method’ (also known as a ‘high-resolution method’) which is particularly suited to spectra of sinusoidal signals, can detect sinusoids buried in the noise especially when signal to noise ratios are low (Mirchandani, 2012). This was of particular relevance when analysing images obtained at lower amperage.

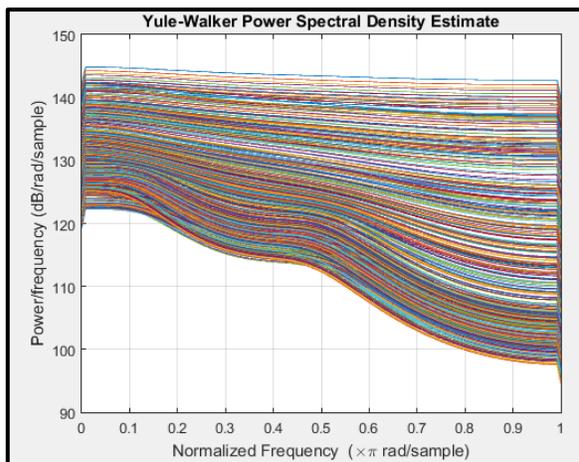


Figure 30. Power Spectral Density (at 10mA) obtained from the Yule Walker AR (autocorrelation) method.

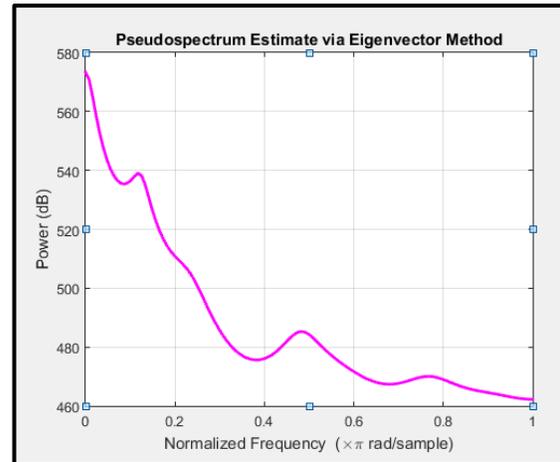


Figure 31. Variance from image analysis in figure 30, plotted via the Eigenvector method.

4.5.2 Image analysis using LIFEx

Texture analysis enables intuitive qualities such as the smoothness or regularity in an image to be examined, in relation to image structure, using spatial variation in pixel intensities (Gonzalez & Woods, 2008b; Mathworks, 2019). This was performed using Local Image Features Extraction (LIFEx) software, which is medical software licensed by the French Atomic Energy Agency (CEA) (lifex.org, n.d.; Nioche et al., 2018). LIFEx enabled regional image texture analysis and was particularly useful for analysis of areas local to the simulated lesions. Image perception is a global phenomenon rather than being reduced to physical attributes, therefore the area surrounding an image feature is also important (Sabih et al., 2011).

DICOM image data sets were uploaded into LIFEx, and texture analysis selected. Images were then windowed using the automatic lung window setting. Regions of interest were defined over the left lung using an irregular descriptor. A circular descriptor, producing volumes that approximated 2 ml, was used to define regions of interest around the simulated lesion and the proximal area. The right lung was also defined, using an irregular descriptor, if the simulated lesion was located on that side. These regions are demonstrated in *figure 32*. A further reason for the left lung to be defined, was that the observer tasks in the observer performance studies were related to the left side.

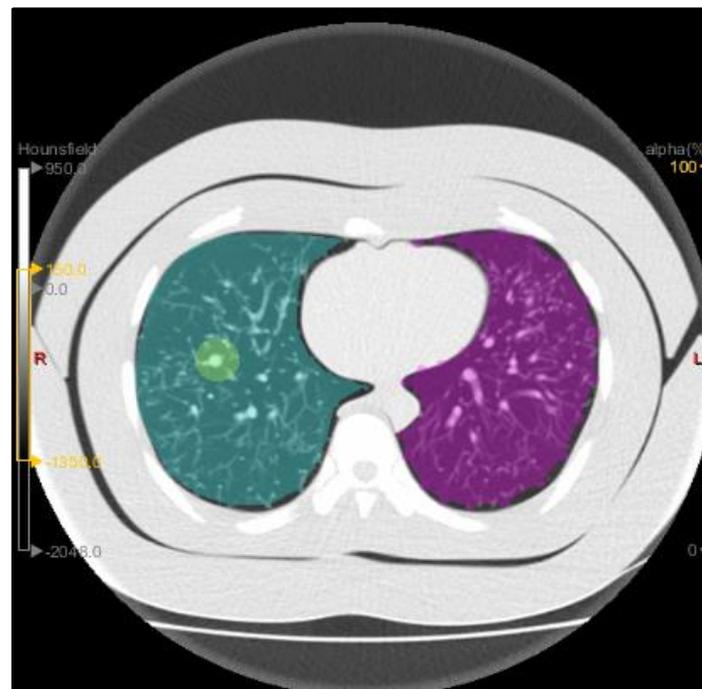


Figure 32. Regions of interest for texture analysis in LIFEx

Gray-Level-Co-occurrence-Matrix (GLCM) data were recorded from the texture analysis. Radiomic features providing information on distribution of intensities and relative pixel positions are provided using this method therefore relating to structural content, rather than texture analysis using only histograms from calculated statistical moments (Gonzalez & Woods, 2008b; Nioche et al., 2018). IQMs noted for the purposes of this thesis were contrast, correlation, energy, homogeneity and entropy, as described in *table 9*.

IQM	Description (over the whole region of interest)	Interpretation
Contrast (variance or inertia)	A measure of the intensity contrast between a pixel and its neighbour.	0 for constant image Range = $[0 (\text{size}(\text{GLCM},1) - 1)^2]$
Correlation	A measure of how correlated a pixel is to its neighbour.	1 or -1 for a perfectly positively or negatively correlated image. Range = $[-1 1]$
Energy (uniformity)	Returns the sum of squared elements in the GLCM.	1 for constant image Range = $[0 1]$
Homogeneity (relating to structures)	Returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.	1 for a diagonal GLCM Range = $[0 1]$
Entropy	Measure of variability or randomness	0 for constant image

Table 9. Texture descriptors in a GLCM (Mathworks, 1994-2020)

4.5.3 Image analysis using ImageJ

ImageJ imaging analysis software is open source therefore available in the public domain. It contains many basic image operations, and developers can create plugins to extend the software capabilities (Mateos-Pérez & Pascau, 2013). This software enabled small regions of interest (ROI) to be drawn for image analysis over an image series,

The **contrast-to-noise ratio (CNR)** describes the signal level in relation to the ambient noise present (Bushberg et al., 2012a). This IQM is performed applying the same regions of interest within each image series. The CNR value is obtained by using mean pixel values in regions of interest drawn over an object and the background surrounding that object (*Appendix 1*) (Kamath et al., 2011; Xu et al., 2013).

The background was measured by averaging measurements from four equally sized circular ROI placed at positions of 0°, 90°, 180° and 270°, at a distance of 2 pixels from the simulated lesion (*figure 33*). The size of the regions approximated the size of the simulated lesion as seen on the image. The background ROIs were of similar size to the simulated lesion to approximate visual assessment of a lesion, which is relative to surrounding structures. In addition, a small background ROI could be placed by chance over a single but not necessarily dominant structure in the vicinity, producing an erroneous result.

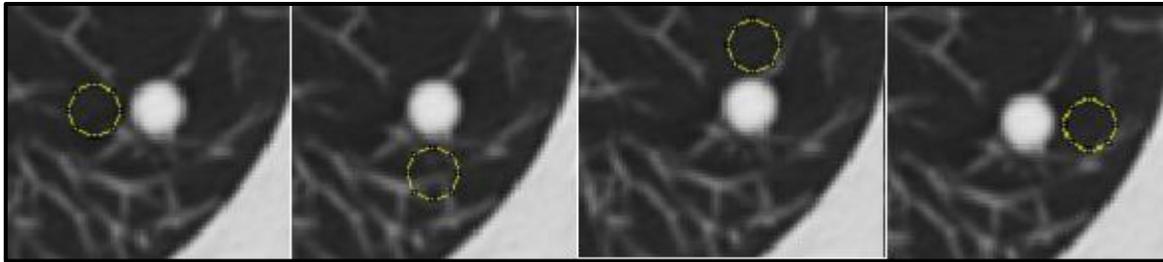


Figure 33. Background ROI ImageJ

The circular ROI placed over the simulated lesion was reduced in size to reduce variation due to geometric unsharpness in measurements obtained (figures 34 and 35), as the calculations are based on mean values rather than absolute values. The decision on the amount of ROI size reduction for each lesion was based on both visual assessment and the resulting measurements obtained from analysis. Variation due to geometric unsharpness was particularly noticeable due to the spherical nature of the simulated lesions, which was potentially compounded by partial volume effects. As this is a signal based IQM, this improved accuracy and consistency over all images analysed. While the size of the area (in pixels) was reduced (table 10), the maximum measurement remained the same. In addition, less error was introduced due to potential variation in ROI placement.

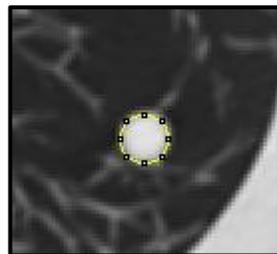


Figure 34. ROI ImageJ, over simulated lesion at full size.

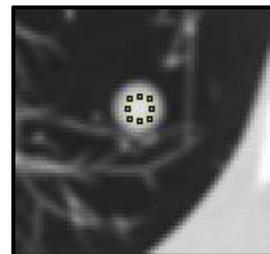


Figure 35. ROI ImageJ, over simulated lesion – smaller to reduce geometric unsharpness.

Lesion size (mm)	ROI full size (pixels)	ROI reduced (pixels)	% reduction
10	57.104	17.571	69
8	28.552	6.589	77
5	4.92	0.549	89

Table 10. ROI size reduction to reduce geometric unsharpness

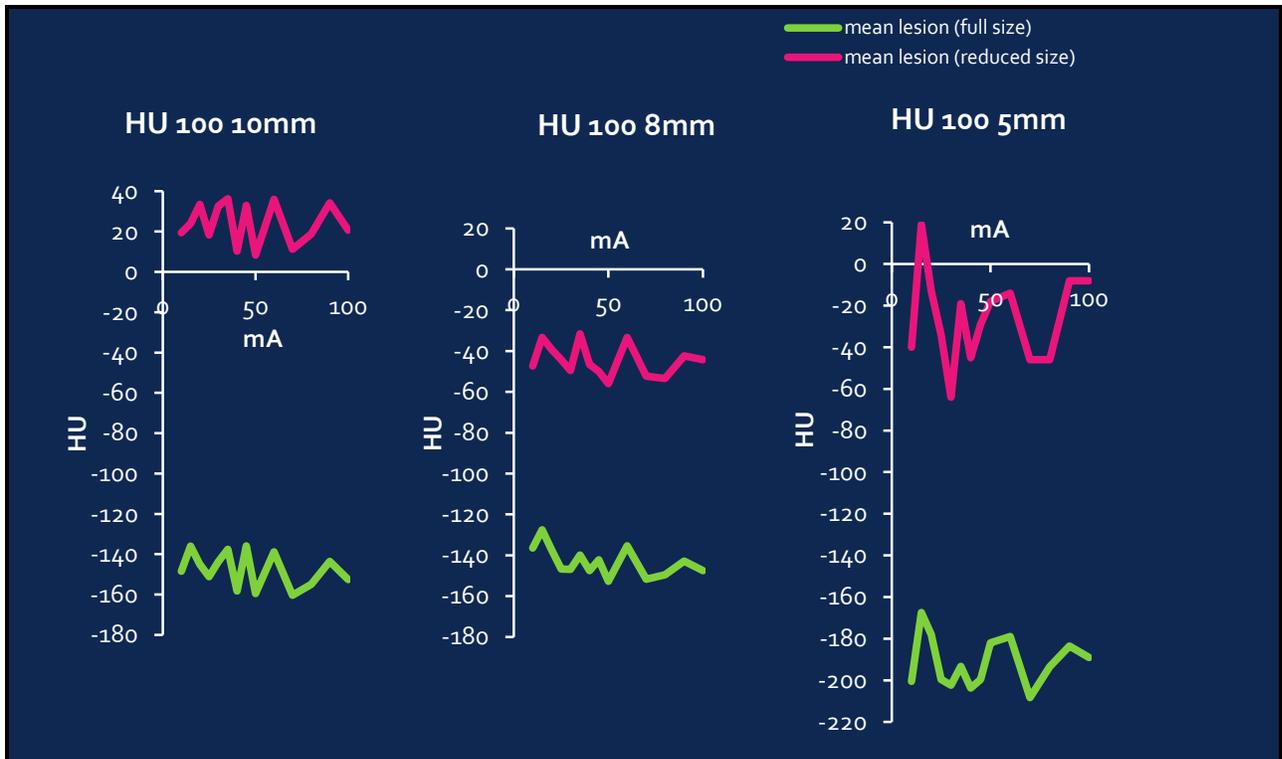


Figure 36. Effect of ROI size reduction for simulated lesions to reduce geometric unsharpness relating to designated HU (100HU).

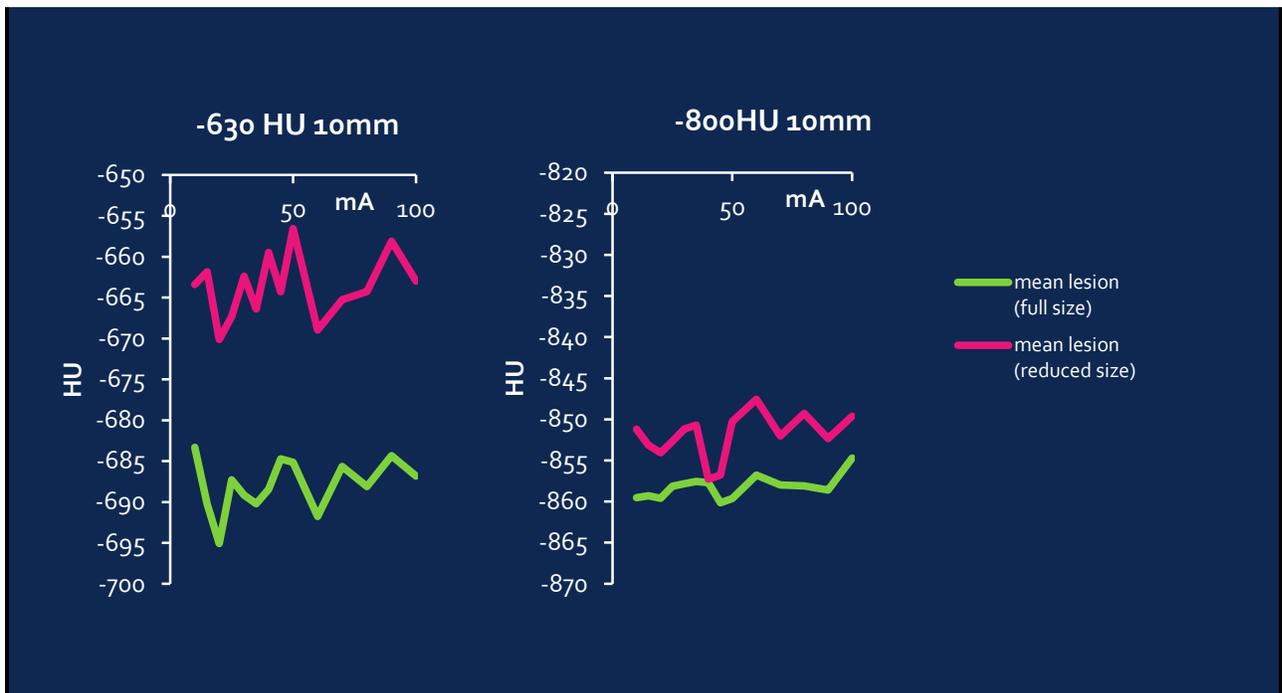


Figure 37. Effect of ROI size reduction for simulated lesions to reduce geometric unsharpness relating to designated HU (-630HU and -800HU).

Effects of this size reduction in ROI are demonstrated in curves on graphs (HU vs mA) (*figure 36*). These curves contained many similar characteristics but with some differences, and variations within the graphs were of greater magnitude. In addition, results on the vertical axis became closer to the designated densities of the simulated lesions (*figures 36 and 37*), however the difference from designated densities could potentially relate to a problem with calibration. Repeated measures and varying simulated lesion placement may have produced more accurate results. The effect on results of the ROI reduction for the different sized simulated lesions was more apparent for the smallest (5mm) simulated lesion (*figure 36*).

The **signal-to-noise ratio (SNR)** is related to the detectability of object, where according to the 'Rose Criterion' if the SNR is greater than 5 then the object will always be recognised or detected, with degrading detection performance as the value approached zero (Bushberg et al., 2012a). The SNR value is obtained by using the mean of pixel values in a ROI over the object, divided by the standard deviation in a background ROI (*Appendix 1*) (*Gaillard et al., n.d.; Magnotta et al., 2006*). The SNR was calculated using the averaged measurements from the four background regions, as used for the CNR, reflecting results within the image itself. The circular ROI over the simulated lesion was reduced in size to avoid penumbral variation, as previously described in the CNR analysis.

4.6 Observer performance: Visual Grading Characteristics (VGC) and Receiver Operating Curve (ROC) analysis

Observer Performance studies involve psychophysical measurements looking at an observer's response to visual stimuli (Zarb et al., 2010). The human observer performs the assessments, where image quality can be measured against fulfilment of a set of criteria (Zarb et al., 2010). In addition, the importance of observer performance studies is closely interlinked with validation of any perceptual IQM (Eckert & Bradley, 1998).

Observer performance studies were performed using ViewDEX which had VGC and ROC analysis components (Svalkvist et al., 2016). Selection of observer tasks was of particular importance because not only did they need to be based on valid criteria, but the psychophysical results needed to be able to relate to the physical IQM results in some way. The nature of the task also needed to relate to the adequacy of the information obtained and the nature of the classification or estimation involved (Barrett & Myers, 2003). As a basis for this, image criteria in the European guidelines for CT quality were used (EC., 2000). In these guidelines are categories for visualisation, critical reproduction and visual sharpness which contain a number of criteria, divided into sections for different parts of the body (EC., 2000). Participants were asked to undertake observer tasks that were divided into two sets (one relating to visualisation of central vessels and the other relating to visualisation of peripheral vessels), and a localization task.

4.6.1 Recruitment

Recruitment, explanation and informed consent occur at the beginning of any observer performance study. Ethical approvals were granted (*Appendices 2 and 3*), and the protocol as prescribed, was complied with. Recruitment was achieved by emails and printouts (posted on imaging and radiotherapy staff notice boards) comprising an invitation to participate accompanied by a participant information sheet (PIS) (*Appendix 4*). The invitation was directed at participants who had familiarity in viewing CT images from groups including Radiologists, Radiographers, Nuclear Medicine Technologists and Physicists. The professional range of these groups was intended to limit bias and enable focus on image quality. After going through the information sheet with the participant and any queries answered, they were given the consent form to sign (*Appendix 5*). The participant then completed the registration form which obtained, in particular, information regarding their imaging experience and CT image viewing experience. They were then assigned a study ID for anonymity. All personal information provided was stored in the site file located in a secured office, as a responsibility stated in the ethics application process. A time was then arranged for the observer performance study to take place. The sequence of events up to this point can be seen in *figure 38*.

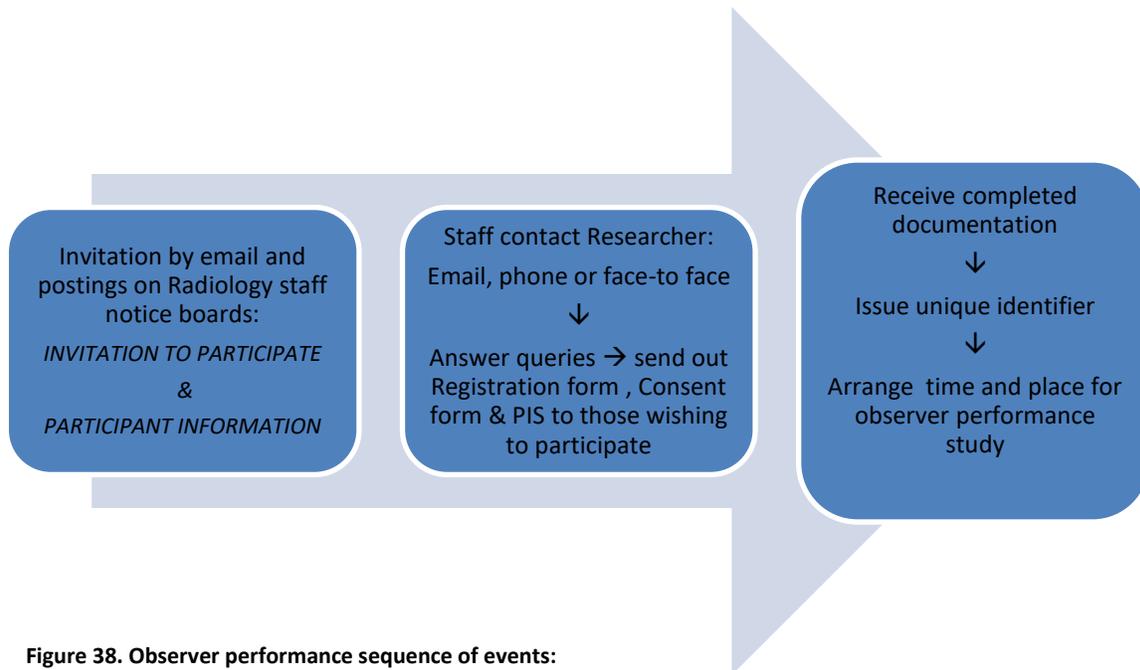


Figure 38. Observer performance sequence of events: recruitment to participation.

4.6.2 Viewing and environmental conditions

The observer studies were performed in an environment without distraction. The computer monitor (HP EliteDisplay E232) on which images were displayed, complied with standards as defined by the Royal College of Radiologists (Royal_College_of_Radiologists, 2019). When reviewing images, these standards require: (1) at least 2 MP resolution and a luminance range 0.8–250cd/m² and (2) to be within 20% of the ‘DICOM GSDF’ which examines the grayscale standard display function (Royal_College_of_Radiologists, 2019).

Muted ambient room lighting was used (59-80 Lux at right angles to the viewing screen), as high levels of ambient lighting can result in reduced visualisation of lesion (ALA, 2017; McEntee et al., 2007). Alternatively, shifting between the high luminance of a monitor in a darkened room can cause an increase in visual fatigue while potentially degrading diagnostic performance (Pollard et al., 2012). When assessing ‘object detectability at increased ambient lighting conditions’, Pollard, Chawla, Delong, Hashimoto and Samei (2008) recommended an optimum of 50 to 80 Lux to increase Radiologist comfort. According to the American Lighting Association, ambient lighting provides a comfortable overall illumination without glare (ALA, 2017).

Explanation and training were given (detailed in section 4.5.3) just prior to the observer performance study being undertaken, to ensure it would be carried out in accordance with the study protocol. This was important in keeping inter-operator variability to a minimum for any other reason that the actual task they were being asked to perform. Also, the participant needed to be comfortable with what they were doing, so that the participant's lack of ability to utilise the facilities and technology would not be a limitation, as is sometimes the case in the clinical realm (ICRU, 1996).

The image viewing sessions were carried out in one hour segments (Barrett & Myers, 2003). This was done to avoid eyestrain and fatigue. It has been demonstrated that Radiologists experience increasing symptoms of eyestrain and blurred vision and problems focussing as they read more images (Krupinski, 2010b). They are also fatigued after sustained image viewing, and this is likely to affect their ability to process the image information as well as their decision making ability (Krupinski, 2010b, 2011). There were no constraints on individual image viewing time (Kim et al., 2008; Park et al., 2003). Those participants with corrected vision wore glasses or contact lenses as they would normally, when viewing images.

4.6.3 VGC and ROC data collection

The aim of this observer performance study was for quality of the images to be assessed by human observers in a quantifiable manner, in addition to lesion detection. This made VGC a more suitable manner in which to perform the study rather than focus on a ROC study involving lesion detection only. However, there was a lesion detection component that facilitated a location sensitive ROC analysis.

ViewDEX version 2.48 was used to enable the VGC study to be performed, and data for the ROC study to be obtained (Håkansson et al., 2010). Input into the database, were 235 images, of which 39 images were 'normal' (without simulated lesions inserted). All images were in DICOM format and fixed at 'lung window' settings (-500HU to 1500 HU) as described in the method section (at the end of 4.4.2). The images were randomised for each participant.

The ViewDEX display is demonstrated in *figure 39*. Listed at the bottom right of the screen are confidence levels relating to the first and third tasks. The confidence levels are: (A) confident that the criterion is fulfilled, (B) somewhat confident that the criterion is fulfilled, (C) indecisive about whether the criterion are fulfilled, (D) somewhat confident that the criterion is not fulfilled and (E) confident that the criterion is not fulfilled. The second and fourth tasks use numeric responses, and the final task is a lesion localization task which uses a numeric response to indicate confidence of simulated lesion localization.

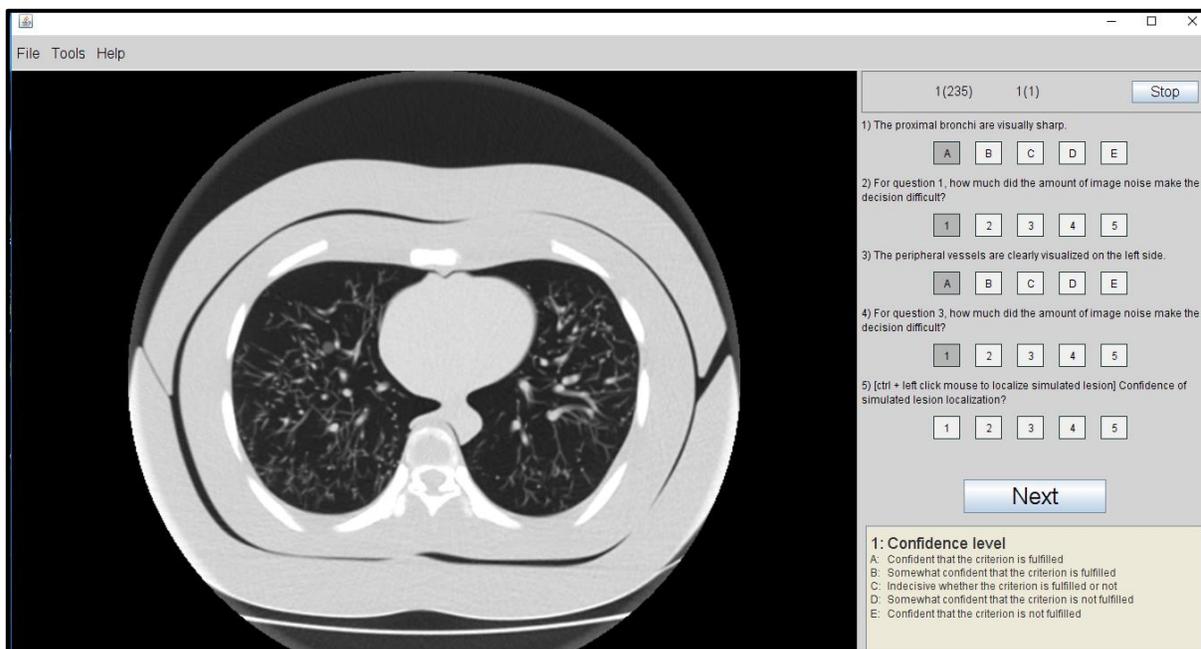


Figure39. ViewDEX display screen.

The tasks that were used are as follows: (1) 'The proximal bronchi are visually sharp' is the criterion against which the confidence levels are applied, (2) 'For question 1, how much did the amount of image noise make the decision difficult' to which a numeric response was applied (1 –not interfered to 5-significantly interfered), (3) 'The peripheral vessels are clearly visualised on the left side' is the criterion against which the confidence levels are applied, (4) 'For question 3, how much did the amount of image noise make the decision difficult' to which a numeric response was applied (1–not interfered to 5-significantly interfered) and (5) ' [ctrl + left click

mouse to localize simulated lesion] Confidence of simulated lesion localization? ' to which a numeric response was applied (1-not confident to 5-very confident). Questions one and three are based on European CT image quality guidelines for critical reproduction: 'standard 1.2.10 Visually sharp reproduction of the peripheral lung vessels' (EC., 2000; Håkansson et al., 2010). The tasks related to noise would be an additional help in relating the observer responses to the image analysis.

The participant was given specific training, including explanation of what was required, demonstration and practice, just prior to commencing the study. As the images had been acquired with a phantom, the vessels were not distributed as much or as far to the edges of the lung fields as in a clinical situation, so question three relating to the peripheral vessels was explained to the observer. They understood that the area requiring their attention was not as narrow as in the clinical environment and a diagram was shown to the participant for clarity (*figure 40*).



Figure 40. Area requiring participant attention when assessing peripheral vessels indicated by the purple colour.

4.6.4 VGC analysis

Output data from the ViewDEX software contained participant responses, localisation task coordinates and time taken. To achieve results that provided outcomes for the VGC study, output data were then formatted for results analysis. VGC analysis was performed using dedicated software, VGC Analyzer 1.0 (release 2), provided by the authors of ViewDEX (Bath & Hansson, 2016). VGC analysis is a 'non-parametric rank-invariant' method for analysing data from visual grading, and determined the area under the curve (AUC_{VGC}) using 'non-parametric re-sampling techniques' (Bath & Hansson, 2016).

Input data sheets for the VGC analyser, containing participant responses for questions 1-2, and 3-4, were created in Excel spreadsheets and saved as a text file. The text file was formatted for paired data, low ratings that corresponded to high image quality, the number of observers (seven in this instance), and the number of cases or images assessed (235 in this instance) which had to be the same for all observers and match the number of paired data input. In addition, a default value of 2000 was used for determining uncertainty of the AUC_{VGC} (Bath & Hansson, 2016). The reference condition was defined as the response for the confidence level regarding fulfilment of image criteria, and the test condition defined as the response to how much image noise interfered with that decision.

4.6.5 JAFROC

The location sensitive ROC analysis was performed using lesion localisation coordinates provided in the ViewDEX output data. ROC analyses were performed with RJAFROC (release date 2015) analysis using Dorfman- Berbaum-Metz analysis with Hillis improvements (DBMH) (Chakraborty, 2015). Output data was produced for random readers and fixed cases, as the images from which the input data were created were acquired using an imaging phantom. The type I error rate (α) was set at .05 therefore only p-values less than 0.05 would be statistically significant. While RJAFROC has the facility to weight lesions, this was not required in this instance as there was no distinction made between the simulated lesions in terms of any priority in detection.

To prepare data for the RJAFROC analysis, localisation data for all 235 cases were collated, relating to modality (the mA used for image acquisition, the case number (which related to the contents of the image in terms of lesion size and density or whether it was 'normal', and whether it was acquired with a 'fat jacket'), the lesion ID (which related to whether the image was normal or abnormal containing a simulated lesion), lesion weighting (which was zero for all cases as there was no distinction made between the simulated lesions), the reader ID (which was assigned to the individual participant ID) and the occurrence of false positives. False positives were defined. After observing lesion location information in the ViewDEX output data, it was found that lesion location was fairly precise when it was correct. Therefore, false positives were defined as any lesion localisations that occurred on normal images, and when any lesion localisations occurred more than ten pixel values from the 'truth' coordinates in any direction.

'Truth' data, or data relating to the true position of the simulated lesions, were established in the first instance by placement of simulated lesions at image acquisition. Subsequently, by viewing the higher quality images, the localisation coordinates for each of the images containing lesions ($235_{\text{total}} - 39_{\text{normal}} = 196_{\text{simulated lesions}}$) were obtained by going through the images using the VGC observer performance program on ViewDEX.

Input data were collected for those responses that corresponded to true positives and false positives. This was achieved by comparing each of the responses and any localisation coordinates obtained for all of the observers, with the 'truth' coordinates that had been established. The truth data input for the RJAFROC analysis were only required in terms of assigning whether a lesion was present (value =1) or not present (value =0).

Input files were created for the RJAFROC analysis which had to take place in more than session. This was due to limitations in the number of modalities that can be analysed simultaneously on this software, and to increase sensitivity of the data at lower and higher amperage ranges. As the modalities represented the mA at which the acquisitions took place, the inputs were divided into two, with the first five modalities ranging from 10mA to 35mA, and the second set of five modalities ranging from 35mA to 60mA as seen in *Appendix 6*. These covered the more useful

data required for comparison with image analysis, which becomes apparent when examining the results. If more modalities were to be analysed then the results from the seventeen images obtained at 70mA would have been excluded, as normal images were not available at 70mA, and this would only leave data from the remaining 80 mA, 90mA and 100mA acquisitions.

Sample size required for the RJAFROC analysis, was based on 85 ($17_{\text{cases}} \times 5_{\text{modalities}}$) as the resultant number of images evaluated per participant being used for each analysis session. According to Obuchowski's sample size calculation represented in *figure 41*, for moderate accuracy, 7 observers were required (Obuchowski, 2000). This would result in 595 images being considered for each analysis session. As a guide, published sample size tables were referred to for numbers required to detect a 'suspected difference in accuracy with sensitivity at a false-positive rate of less than or equal to 0.10' in ROC studies (Obuchowski, 2000). At moderate accuracy (high accuracy is usually applied to imaging techniques such with high structural definition such as MRI neurology) with expected small intra-operator variability (due to skill levels of observers), the number of observations required is indicated in the graph in *figure 41*.

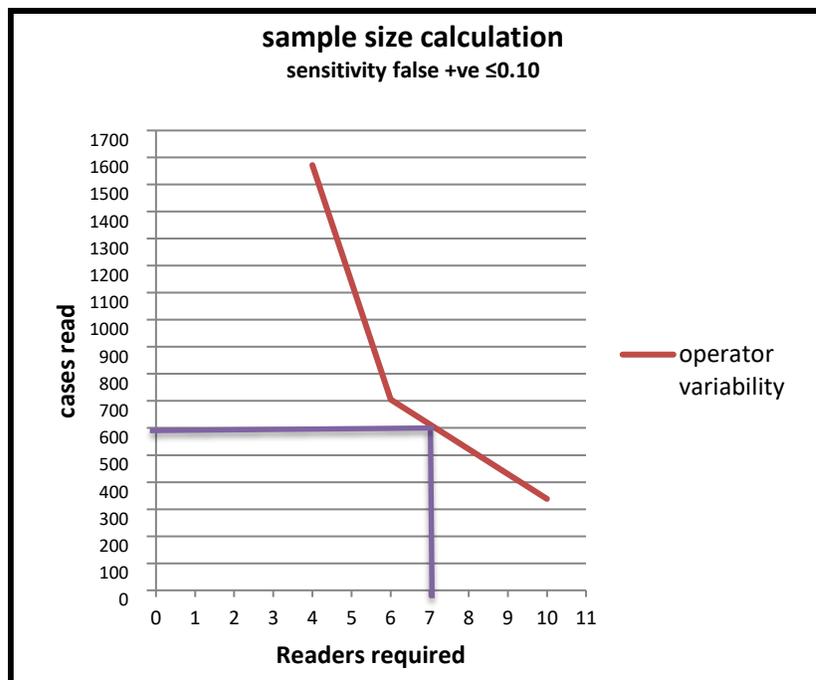


Figure 41. Sample size calculation. (Obuchowski, 2000)

5.0 Chapter Five: Results - Image Quality Metrics (IQM)

5.1 Introduction

The results presented in this chapter focus on physical measures of image quality. Eight IQMs specifically chosen for analysis included a mixture of both signal-based and perceptually based IQMs: mean squared error (MSE), peak signal-to-noise ratio (PSNR), signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), structural similarity index (SSIM), non-shift-edge ratio (NSER), noise power spectrum (NPS) and texture analysis (looking at Contrast, Correlation, Energy, Homogeneity and Entropy). Formulae for calculation of the relevant IQMs are listed in *Appendix 1*.

There were 14 series of images analysed, (ranging over the upper zone, mid zone and lower zone of the thoracic region) with an amperage range of 10mA to 100mA. Data were graphed, and trend lines applied for all but NPS. This facilitated viewing of undulations (peaks, troughs and plateaus) in the graphs and identification of any prominent features, which were then examined for any patterns or consistencies. This search was carried out with a view to potential radiation dose reductions. Frequency analyses were performed on how often patterns or features occurred at a particular frequency.

In addition, the data were averaged in two subsets, for data acquired with and without attenuation jackets. Correlation analysis was then performed producing Pearson correlation coefficient values (r) to examine the efficacy of using the averaged data series. Regression analysis produced R squared (R^2) values as a coefficient of determination, used here to determine the relation between the IQM values and logarithmic trending that had been applied. The equations, that described the shape of the trend line, enabled comparison between data acquired with and without attenuation jackets, where differences in gradient and y-intercept were calculated. T-tests were performed to determine the statistical significance between the two subsets. Initially, an F-test was performed to determine whether the T-test analysis should be one assuming equal variances or unequal variances. Features of NPS output data were also examined in terms of response with increasing amperage, and the effect of attenuation jackets.

At the beginning of each section, a brief introduction to the IQM is made, and resultant graphs are displayed to give an overview of the results. At the end of each section, the performance of the IQM is summarised. This includes consistency of the IQM, and whether a pattern was found that could provide potential to reduction in radiation exposure.

5.2 Radiation exposure information

When examining results of the IQM, it is useful to have in mind the radiation exposures that have produced that information. The CT_{volume} dose index ($CTDI_{vol}$), seen in *table 11* and *figure 42*, was identical for all series of acquisitions. The radiation exposures presented here are from an amperage range of 10mA to 200mA, where the $CTDI_{vol}$ dose displayed in *figure 42* (mA versus mGy) shows a linear response to amperage ($y = 0.0462x + 0.0087$).

mA	$CTDI_{vol}$ (mGy)
10	0.5
20	0.9
30	1.4
40	1.9
50	2.3
60	2.8
80	3.7
100	4.6
150	6.9
200	9.3

Table 11. CT dose information

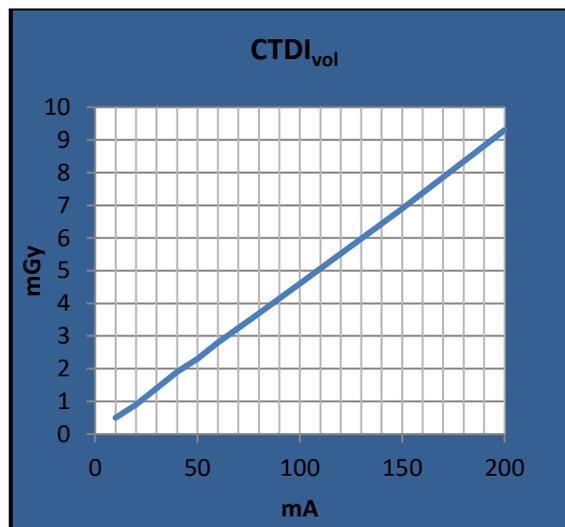


Figure 42. CT dose index $CTDI_{vol}$ (mA vs mGy)

5.3 Mean squared error (MSE)

Image comparison with a reference image (acquired at 200mA) produced values describing distortion, where values were lower in the case of greater similarity (Rangaraju et al., 2012; Sasi varnan et al., 2011). The analysis was performed using MATLAB®. Resultant graphs for MSE are shown in *figure 43*.

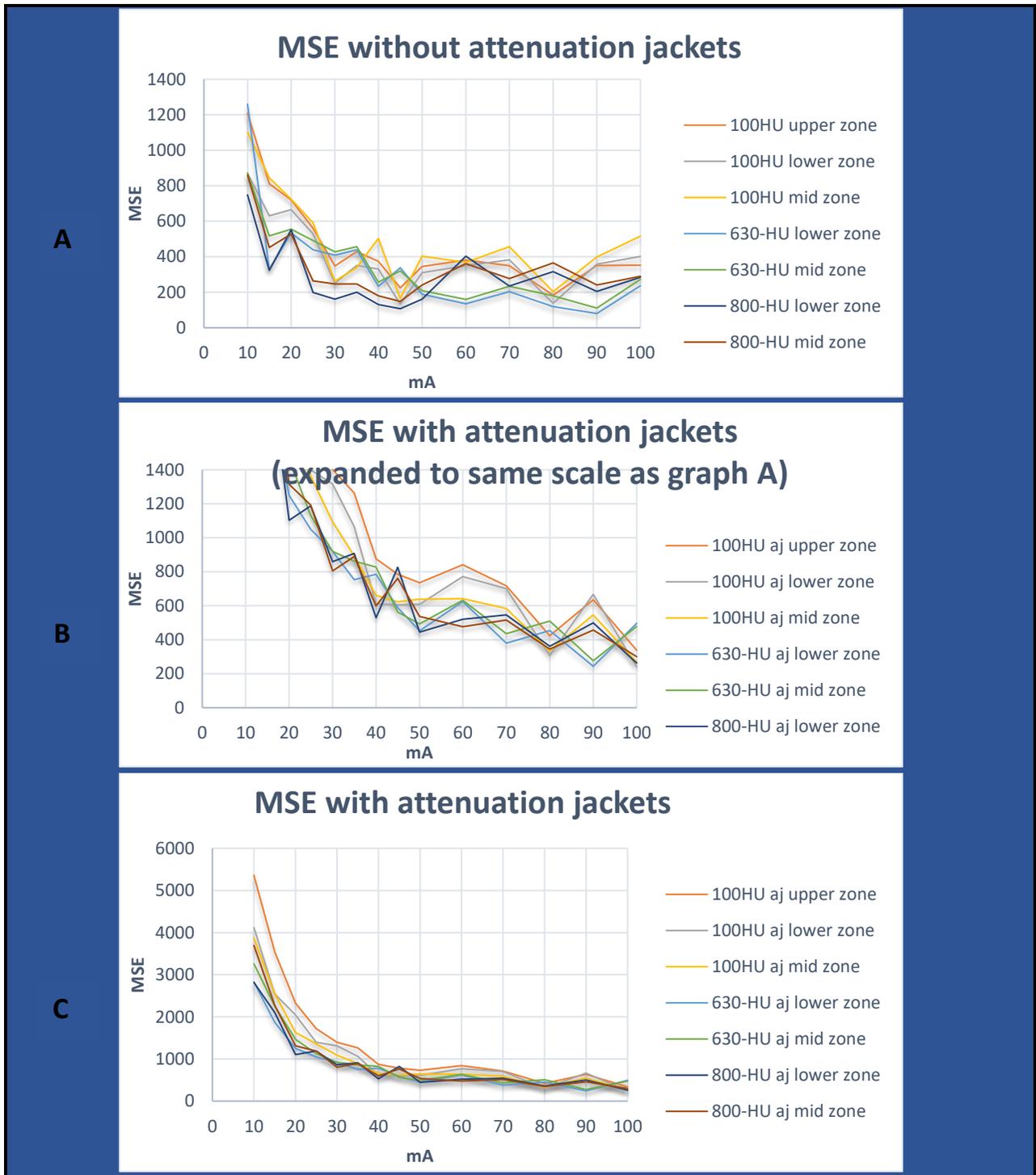


Figure 43. MSE graphs (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets - expanded to match graph A, (C) with attenuation jackets

Good correlation occurred both with and without attenuation jackets when compared to MSE values in the averaged data series for each. The results, as seen in *table 12*, show Pearson correlation values (r) for these data series, that ranged from 0.86 to 0.98 without attenuation jackets, and consistent values (r) of 0.99 when attenuation jackets were used. Consequently, the efficacy of using averaged data series was apparent when examining these MSE results.

MSE: (r)	Without attenuation jackets	With attenuation jackets
Averaged data series	1	1
100HU upper zone	0.984017	0.999369
100HU lower zone	0.941073	0.991518
100HU mid zone	0.928443	0.998753
630-HU lower zone	0.902325	0.98996
630-HU mid zone	0.904037	0.993059
800-HU lower zone	0.858682	0.987305
800-HU mid zone	0.923687	0.99467

Table 12. Correlation of MSE graph values with averaged data series (Pearson correlation coefficient r).

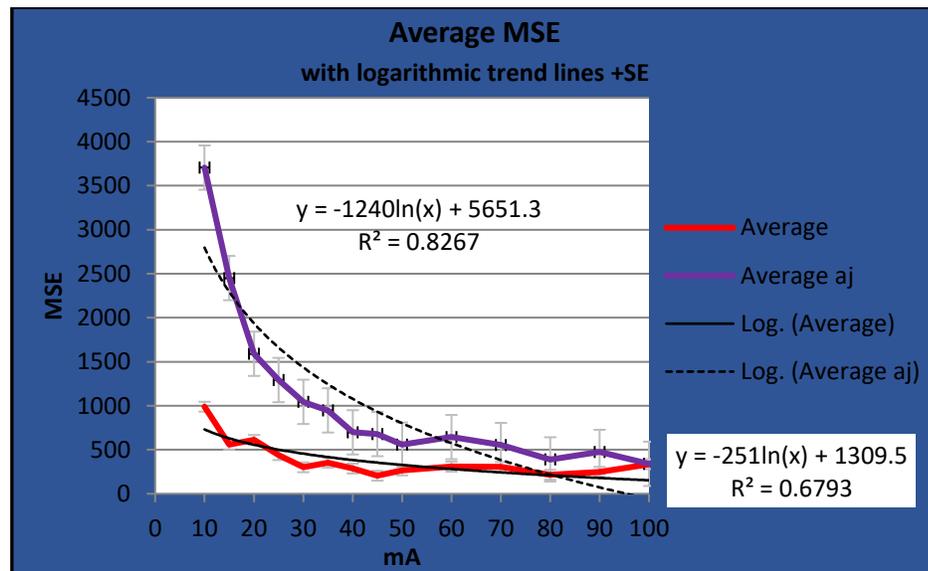


Figure 44. MSE: Graphs of averaged data series, with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).

MSE values of averaged data series, both with and without attenuation jackets, are graphed in *figure 44*, where logarithmic trend lines have been applied. Regression analysis produced an R^2 value that increased when attenuation jackets were used ($R^2=0.6793$ increased to $R^2=0.8267$) and

demonstrated some correlation of MSE values with a logarithmic trend. When attenuation jackets were used, the slope of the graph increased by 80%, and the y-intercept increased by 77%. When comparing results with and without attenuation jackets, the difference in MSE results was statistically significant ($p=0.01$). Results of the T-test, assuming equal variances, are detailed in *table 13* with the T-stat ($T=-2.76$) larger than the two-tail critical value ($T=2.06$).

t-Test: Two-Sample Assuming Equal Variances	
alpha	0.05
df	26
t Stat	-2.75546
P(T<=t) two-tail	0.010564
t Critical two-tail	2.055529

Table 13. T-test for average MSE data series with and without attenuation jackets.

When examining graphs of individual data series acquired over upper zone, mid zone and lower zones, peaks and plateaus were identified, where MSE values were similar or lower at lower amperage, in each case. Examples of this process are seen in *figure 45*, where a red line (mA1) indicates a value that was then improved upon at a lower amperage indicated by a green line (mA2).

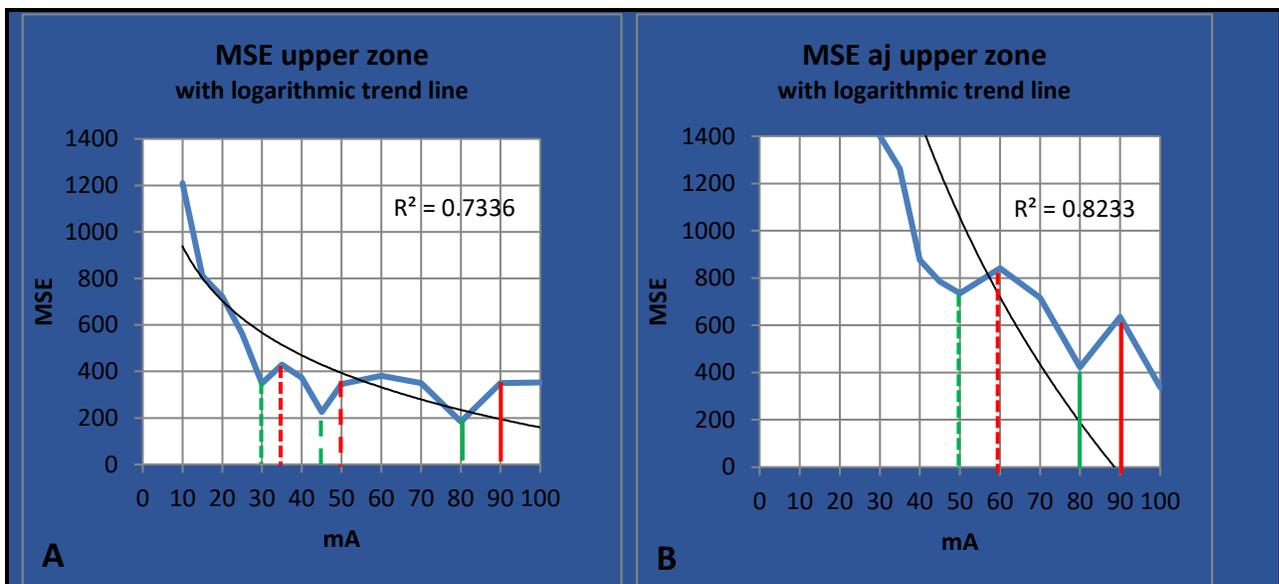


Figure 45. (A) MSE upper zone with improvement (from red line to green line): 90mA to 80mA, 50mA to 45mA, 35mA to 30mA; (B) MSE upper zone when attenuation jackets used, with improvement (from red line to green line): 90mA to 80mA, 60mA to 50mA.

The potential usefulness of this information is seen in *table 14*, where MSE values are similar or lower at reduced amperage (mA2). This is then translated into potential for reduced radiation exposure. In these examples, the potential for reduction in radiation exposure ranged from 8.7% to 17.9%. When looking at MSE data from all acquisitions in this study, potential for reduction in radiation exposure ranged from 8.7% to a maximum of 35.4%.

mA1	mA2 (lower)	MSE (mA1)	MSE (mA2)	MSE % difference	CTDI _{vol} mA1	CTDI _{vol} mA2	CTDI _{vol} % difference
Without attenuation jackets							
90	80	350.53	183.59	48	4.15	3.7	10.8
50	45	345.79	225.10	35	2.3	2.1	8.7
35	30	429.58	348.68	19	1.65	1.4	15.1
With attenuation jackets							
90	80	635.96	422.76	36	4.15	3.7	10.8
60	50	841.75	735.18	13	2.8	2.3	17.9

Table 14. MSE upper zone where improved values at lower amperage translates into reduced radiation exposure (CTDI_{vol}).

To identify any pattern and how often this occurred at any particular amperage, frequency analyses were carried out. Overall, the most frequent amperages where improvement was possible, were: 70mA (15%), 60mA (15%), 100mA (13%), 90mA (13%) and 35mA (13%), as seen in *figure 46*. Overall, the most frequent amperages where improvement occurred were: 80mA (15%), 30mA (15%), 50mA (13%), 90mA (11%) and 45mA (11%), also seen in *figure 46*.

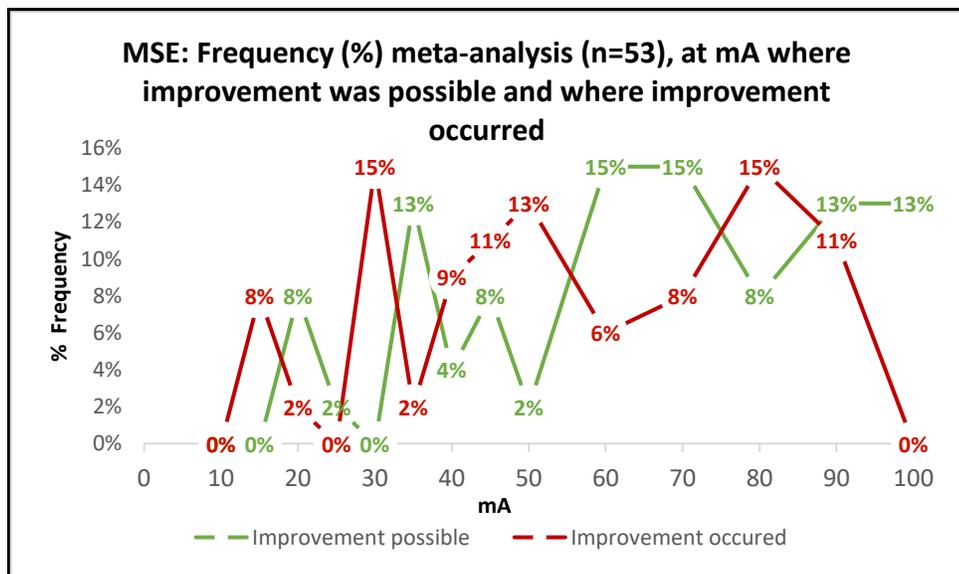


Figure 46. MSE: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperages where improvement was possible, without attenuation jackets, were: 70mA (20%), 100mA (17%), 35mA (17%), 20mA (13%), as seen in *figure 47*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 60mA (26%) and 90mA (22%), also seen in *figure 47*.

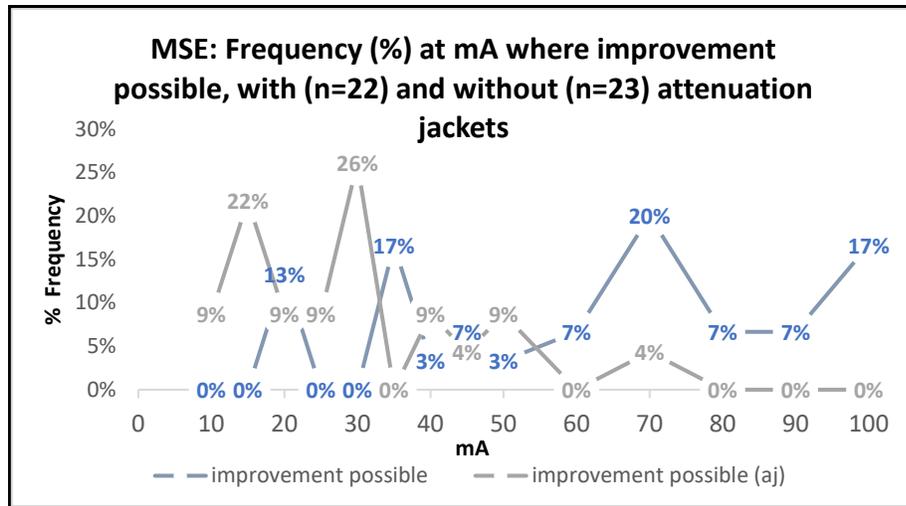


Figure 47. MSE: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets

The most common amperages where improvement occurred, without attenuation jackets, were: 30mA (20%), 45mA (17%), 90mA (13%), 15mA (13%) and 80mA (10%), as seen in *figure 48*. The most frequent amperages where improvement occurred, with attenuation jackets, were: 80mA (22%), 50mA (22%) and 40mA (13%), also seen in *figure 48*.

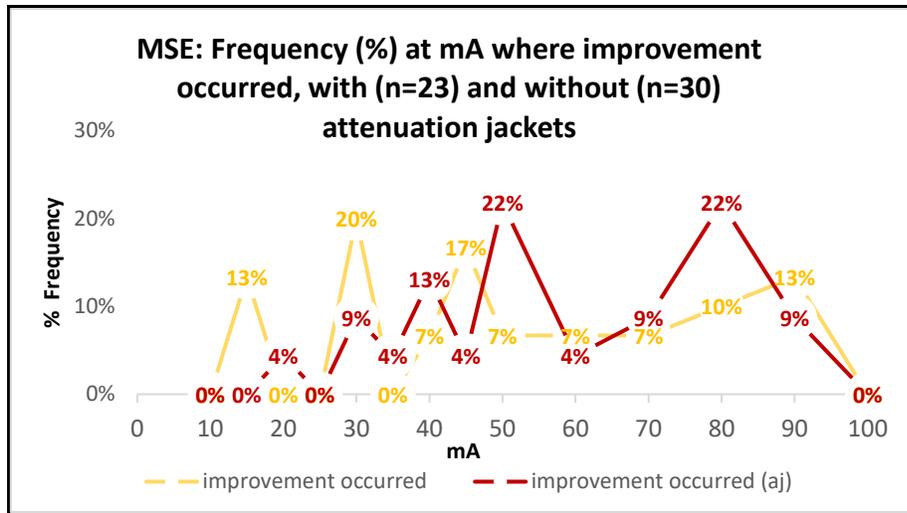


Figure 48. MSE: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.

In summary, MSE is a signal-based image quality metric that demonstrated some consistency. Good correlation with values in averaged data series facilitated examination of the results. Logarithmic trending was applied to the graphed data, where both the gradient and the y-intercept increased with the use of attenuation jackets. This would be expected with this IQM, as decreased spatial resolution and increased distortion occurred at lower mA, with attenuation jackets causing increased effect therefore higher values. In addition, regression analysis produced values (R^2) that increased when attenuation jackets had been used. MSE values related to the greatest improvement in quality between 10mA and 50mA. There was less improvement with increasing mA between 50mA to 100mA. With a view to finding a pattern that could provide potential to reduce radiation exposure, when examining the frequency of peaks and plateaus at any particular amperage, no consistent pattern was identified.

5.4 Peak signal-to-noise ratio (PSNR)

The peak signal in an image was compared with a reference image (acquired at 200mA), where values were higher in the case of greater similarity (Sasi varnan et al., 2011). The analysis was performed using MATLAB®. Resultant graphs for PSNR are shown in *figure 49*.

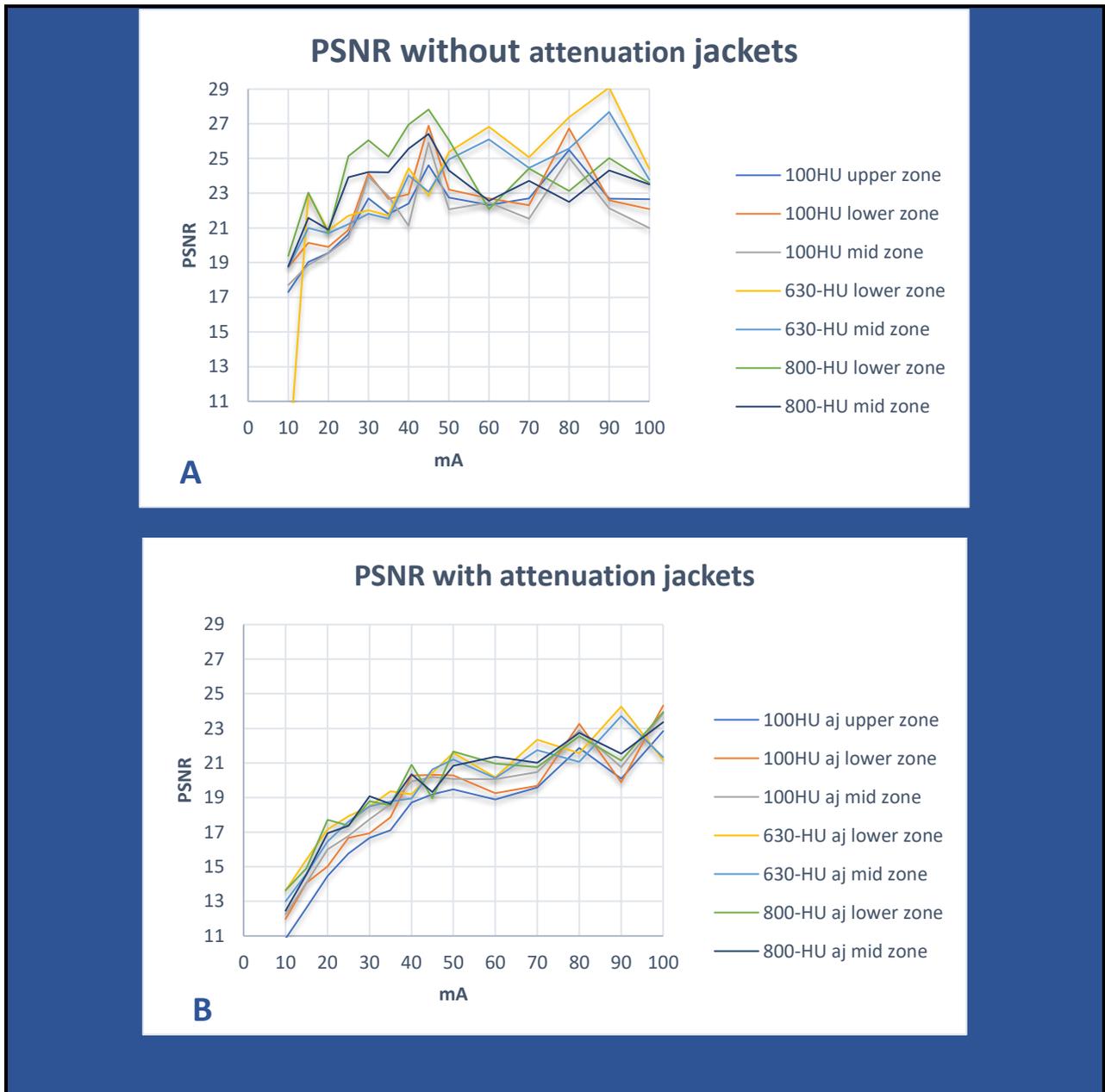


Figure 49. PSNR graphs (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets.

The outlier seen at the beginning of the graph labelled -630HU lower zone, is puzzling. Despite repeated analysis it remained apparent. This may be related to an equipment glitch or an issue with machine calibration.

While good correlation occurred both with and without attenuation jackets, when compared to PSNR values in the averaged data series for each, correlation was less without attenuation jackets. The results, as seen in *table 15*, show Pearson correlation values (*r*) for these data series, that ranged from 0.76 to 0.94 without attenuation jackets, and values (*r*) that ranged from 0.93 to 0.99 when attenuation jackets were used. Consequently, using averaged data series was still reasonable when examining these PSNR results.

PSNR: (<i>r</i>)	Without attenuation jackets	With attenuation jackets
Averaged data series	1	1
100HU upper zone	0.943765	0.995222
100HU lower zone	0.85793	0.968733
100HU mid zone	0.850048	0.990221
630-HU lower zone	0.868451	0.934444
630-HU mid zone	0.798347	0.951681
800-HU lower zone	0.756918	0.975916
800-HU mid zone	0.838627	0.988798

Table 15. Correlation of PSNR graph values with averaged data series. (Pearson correlation coefficient *r*).

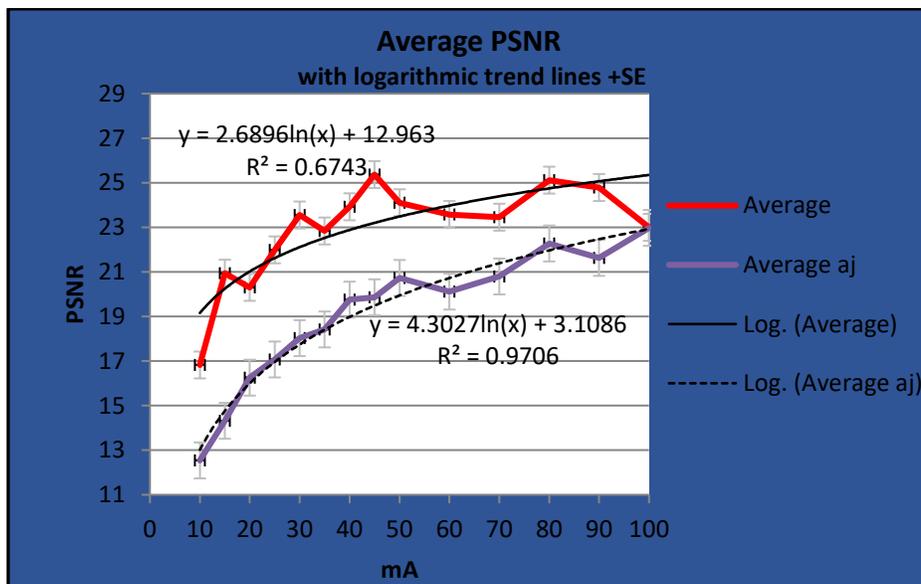


Figure 50. PSNR: Graphs of averaged data series, with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).

PSNR values of averaged data series, both with and without attenuation jackets, are graphed in *figure 50*, where logarithmic trend lines have been applied. Regression analysis produced an R^2 value that increased when attenuation jackets were used ($R^2=0.6743$ increased to $R^2=0.9706$) and demonstrated some correlation of PSNR values with a logarithmic trend. When attenuation jackets were used, the slope of the graph increased by 37%, and the y-intercept decreased by 76%. When comparing results with and without attenuation jackets, the difference in PSNR results was statistically significant ($p=0.0007$). Results of the T-test, assuming unequal variances, are detailed in *table 16* with the T-stat ($T=3.90$) higher than the two-tail critical value ($T=2.06$).

t-Test: Two-Sample Assuming Unequal Variances	
alpha	0.05
df	24
t Stat	3.899544571
P(T<=t) two-tail	0.000678912
t Critical two-tail	2.063898562

Table 16. T-test for average PSNR data series with and without attenuation jackets.

In graphs of individual data series acquired over upper zone, mid zone and lower zones, peaks and plateaus were identified, where PSNR values were similar or increased at lower amperage, in each case. Examples of this process are seen in *figure 51*, where a red line (mA1) indicates a value that was then improved upon at a lower amperage indicated by a green line (mA2).

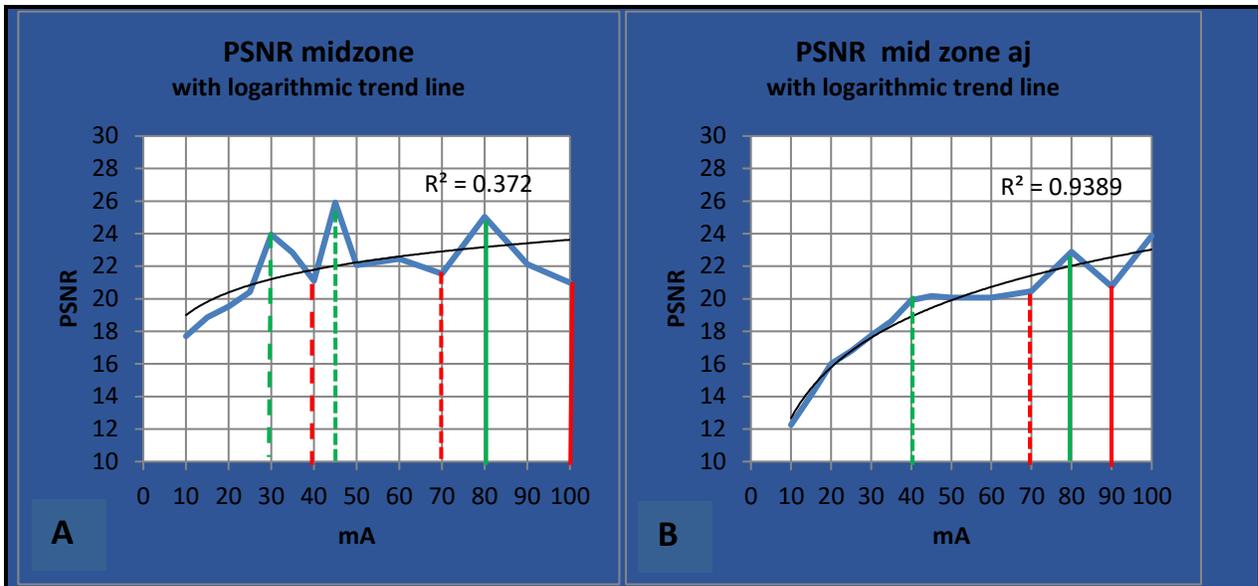


Figure 51. (A) PSNR mid zone with improvement (from red line to green line): 100mA to 80mA, 70mA to 45mA, 40mA to 30mA; (B) PSNR mid zone when attenuation jackets used, with improvement (from red line to green line): 90mA to 80mA, 70mA to 40mA.

The potential usefulness of this information is seen in *table 17*, where PSNR values are similar or increased at reduced amperage (mA²). This could translate into potential for reduced radiation exposure. In these examples, the potential for reduction in radiation exposure ranged from 11% to 42%. When looking at PSNR data from all acquisitions in this study, potential for reduction in radiation exposure is similar.

mA1	mA2 (lower)	PSNR (mA1)	PSNR (mA2)	PSNR % difference	CTDI _{vol} mA1	CTDI _{vol} mA2	CTDI _{vol} % difference
Without attenuation jackets							
100	80	21.00	25.04	16	4.6	3.7	20
70	45	21.54	25.91	17	3.25	2.1	35
40	30	21.12	23.94	12	1.9	1.4	26
With attenuation jackets							
90	80	20.76	22.88	9	4.15	3.7	11
70	40	20.47	20.18	-3	3.25	1.9	42

Table 17. PSNR mid zone where improved values at lower amperage translates into reduced radiation exposure (CTDI_{vol}).

To identify any pattern and how often this occurred at any particular amperage, frequency analyses were carried out. Overall, the most frequent amperages, where improvement was possible, were: 70mA (19%), 100mA (13%), 90mA (13%) and 35mA (13%), as seen in *figure 52*.

Overall, the most frequent amperages where improvement occurred were: 80mA (15%), 30mA (15%), 90mA (12%) and 40mA (12%), also seen in *figure 52*.

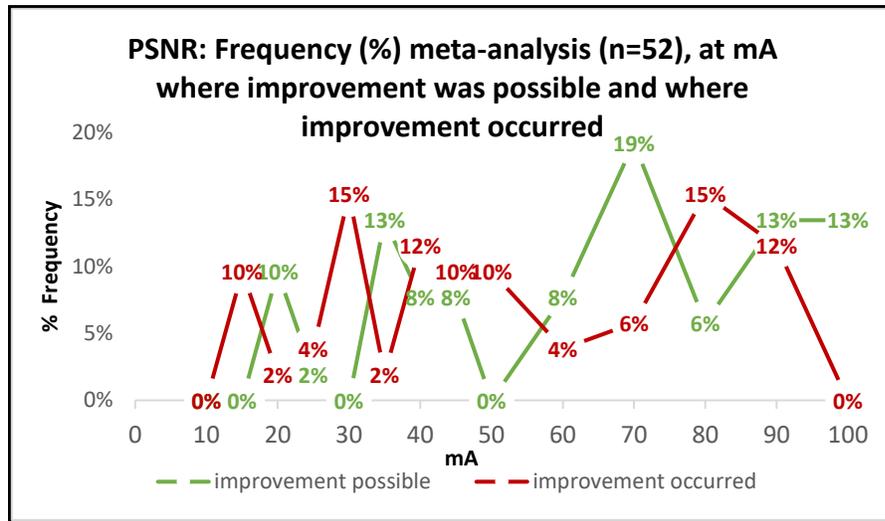


Figure 52. PSNR: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperages where improvement was possible, without attenuation jackets, were: 100mA (17%), 70mA (17%), 35mA (17%), 20mA (17%), as seen in *figure 53*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 90mA (23%) and 70mA (23%), also seen in *figure 53*.

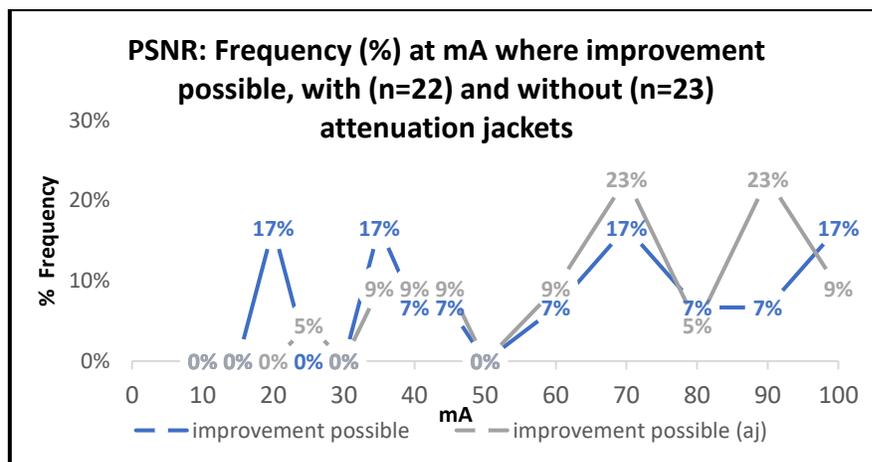


Figure 53. PSNR: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.

The most common amperages where improvement occurred, without attenuation jackets, were: 45mA (17%), 30mA (17%), 15mA (17%) and 90mA (13%), as seen in *figure 54*. The most frequent amperages where improvement occurred, with attenuation jackets, were: 80mA (23%), 50mA (23%), 40mA (18%) and 30mA (14%), also seen in *figure 54*.

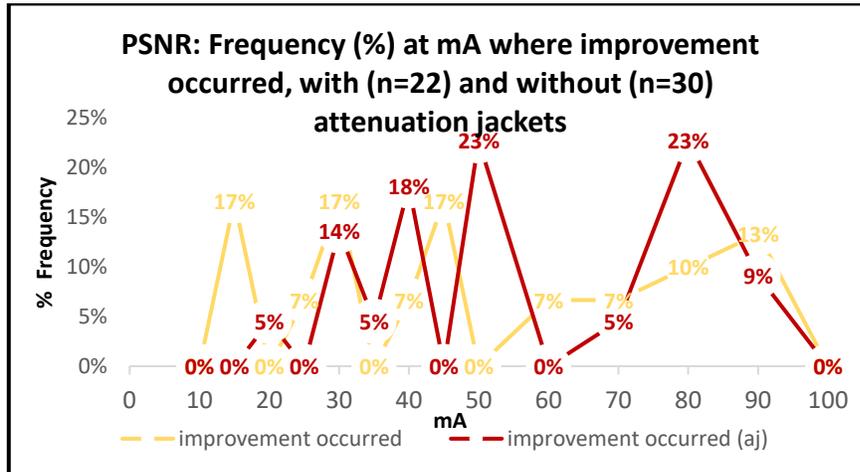


Figure 54. PSNR: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.

In summary, PSNR is a signal-based image quality metric that demonstrated some consistency when attenuation jackets were used with Pearson correlation (r) values that ranged from 0.93 to 0.99, and very good regression coefficients (R^2) in relation to logarithmic trending. However, in data series acquired without attenuation jackets there were lower regression coefficients (R^2) in relation to the logarithmic trending, and less correlation (r) with averaged data series, compared to when attenuation jackets were used. This IQM may therefore be more consistent in the clinical setting where body habitus has an impact on image quality. The graphed data demonstrated an increase in gradient and a decrease of the y-intercept with the use of attenuation jackets. PSNR values related to the greatest improvement in quality between 10mA and 50mA. There was less improvement with increasing mA between 50mA to 100mA.

Again, frequency of peaks and plateaus at any particular amperage was examined, with a view to finding a pattern that could provide potential to reduce radiation exposure. In the overall data there was some indication at 70mA that there may be potential for dose reduction, however no consistent pattern was identified.

5.5 Signal-to-noise ratio (SNR)

This IQM relates to the detectability of an object. According to the ‘Rose Criterion’, if the SNR is greater than 5 then the object will always be recognised or detected (Bushberg et al., 2012a). SNR values decreased with degrading detection performance (Bushberg et al., 2012a). The analysis was performed using ImageJ.

SNR was calculated with the average of four background regions placed within the phantom, proximal to the simulated lesion at the main compass points. Resultant graphs for SNR are shown in *figures 55* and *figures 56*.

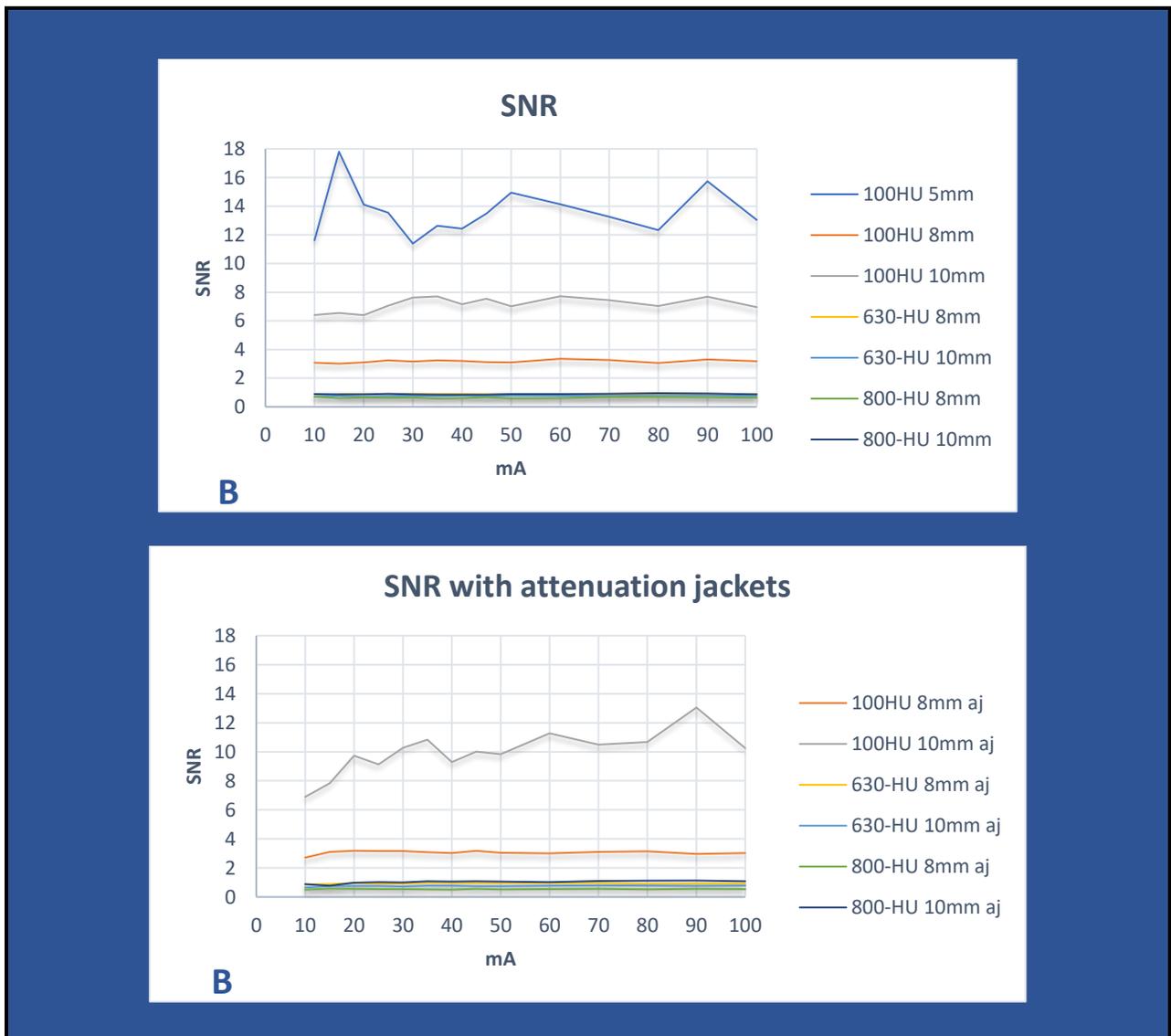


Figure 55. SNR graphs (relating to the simulated lesions): (A) without attenuation jackets , (B) with attenuation jackets.

Simulated lesions were placed with the 5mm lesions in the upper zone, the 10mm lesions in the mid-zone and the 8mm lesions in the lower zone. When trying to analyse the 5mm lesions, the effect of lesion size and density, in combination with an increase in attenuation and scatter with the use of attenuation jackets, became apparent. The 5mm 100HU simulated lesion was difficult to identify and define when attenuation jackets were used, therefore the SNR was not calculated. In addition, it was difficult to identify and define the 5mm -630HU and 5mm -800HU simulated lesions both with and without attenuation jackets, therefore the SNR was not calculated for these.

Very little correlation occurred both with and without attenuation jackets, when compared to SNR values in the averaged data series for each, and correlation was less without attenuation jackets. The results, as seen in *table 18*, show Pearson correlation values (r) for these data series, that ranged from -0.37 to 0.94 without attenuation jackets, and values (r) that ranged from 0.29 to 0.99 when attenuation jackets were used. Consequently, using averaged data series for regression analysis was not reasonable when examining these SNR results. In addition, useful trend lines were unable to be applied to the individual data series.

SNR: (r)	Without attenuation jackets	With attenuation jackets
Averaged data series	1	1
100HU 5mm	0.938797	
100HU 8mm	0.183531	0.353566
100HU 10mm	0.190359	0.994243
-630HU 8mm	0.456807	0.534078
-630HU 10mm	0.350177	0.737911
-800HU 8mm	-0.36833	0.289863
-800HU 10mm	0.033853	0.811878

Table 18. Correlation of SNR graph values with averaged data series (Pearson correlation coefficient r).

In graphs of individual data series acquired, peaks and plateaus were identified, where SNR values were similar or improved at lower amperage, in each case. Examples of this process are seen in *figure 56*, where a red line (mA1) indicates a value that was then improved upon at a lower amperage, indicated by a green line (mA2).

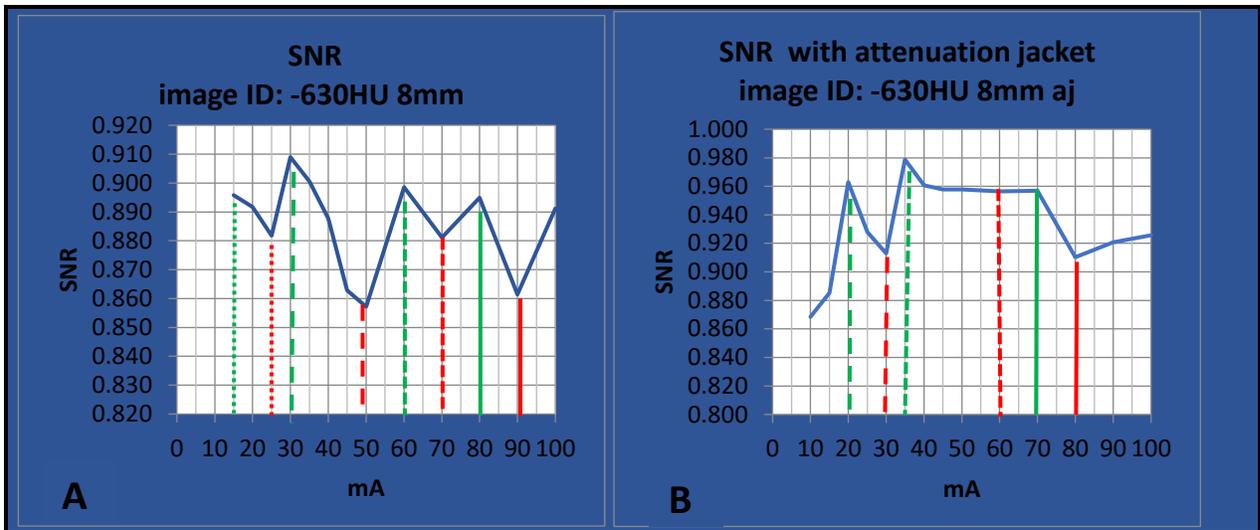


Figure 56. SNR for -630HU 8mm lesion located in lower zone (A) without attenuation jacket, with improvement (from red line to green line): 90mA to 80mA, 70mA to 60mA, 50mA to 30mA, 25mA to 15mA; (B) with attenuation jackets, with improvement (from red line to green line): 80mA to 70mA, 60mA to 35mA, 30mA to 20mA.

Given the inconsistent trending of SNR values obtained, there is little predictive value for radiation dose reduction. However, it is still worth noting how often peaks, plateaus and troughs occur at any particular amperage, for comparison with other IQMs. To identify any pattern, frequency analyses were carried out.

Overall, the most frequent amperages, where improvement was possible, were: 100mA (17%), 50mA (13%) and 80mA (13%), as seen in figure 57. Overall, the most frequent improvement occurred were: 90mA (12%) and 35mA (12%), also seen in figure 57.

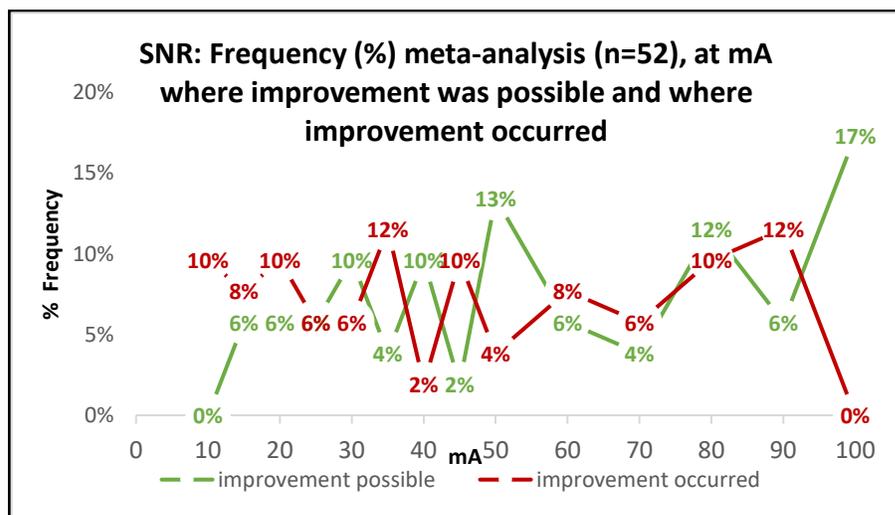


Figure 57. SNR: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperages, where improvement was possible, without attenuation jackets, were: 100mA (20%) and 50mA (13%), as seen in *figure 58*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 100mA (14%), 80mA (14%), 60mA (14%), 50mA (14%), 40mA (14%) and 30mA (14%) also seen in *figure 58*.

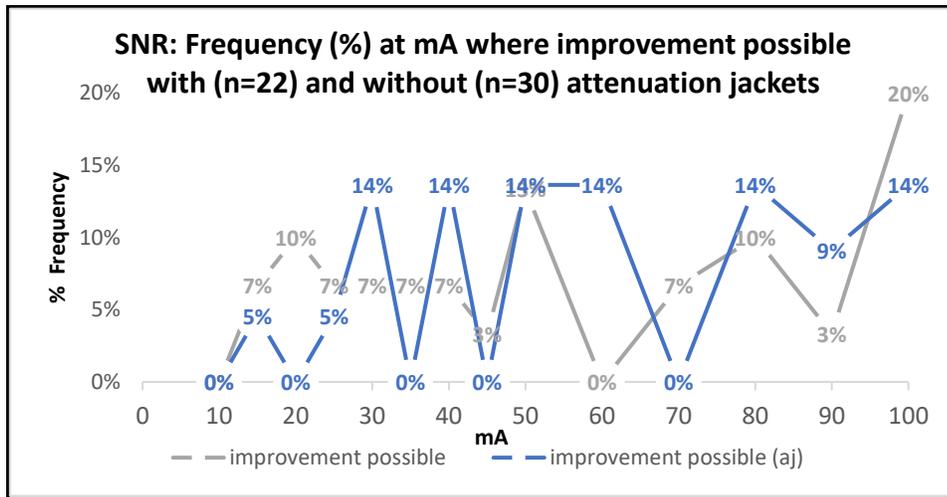


Figure 58. SNR: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.

The most frequent amperages where improvement occurred without attenuation jackets were: 80mA (13%) and 80mA (13%), as seen in *figure 59*. The most frequent amperages where improvement occurred, with attenuation jackets, were: 35mA (18%), 20mA (18%), 90mA (14%), 70mA (14%) and 45mA (14%), also seen in *figure 59*.

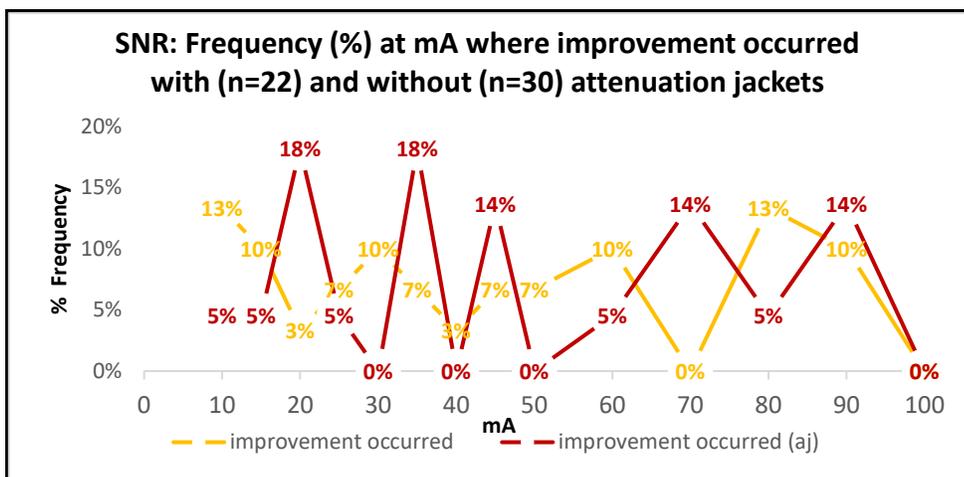


Figure 59. SNR: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.

SNR values increased in relation to size and density. This is demonstrated in *figure 60* where examples are given with simulated lesions that are 8mm and 10mm in size and 100HU and -800HU in density. However, there is little change seen when attenuation jackets are used, as demonstrated with the example in *figure 61*.

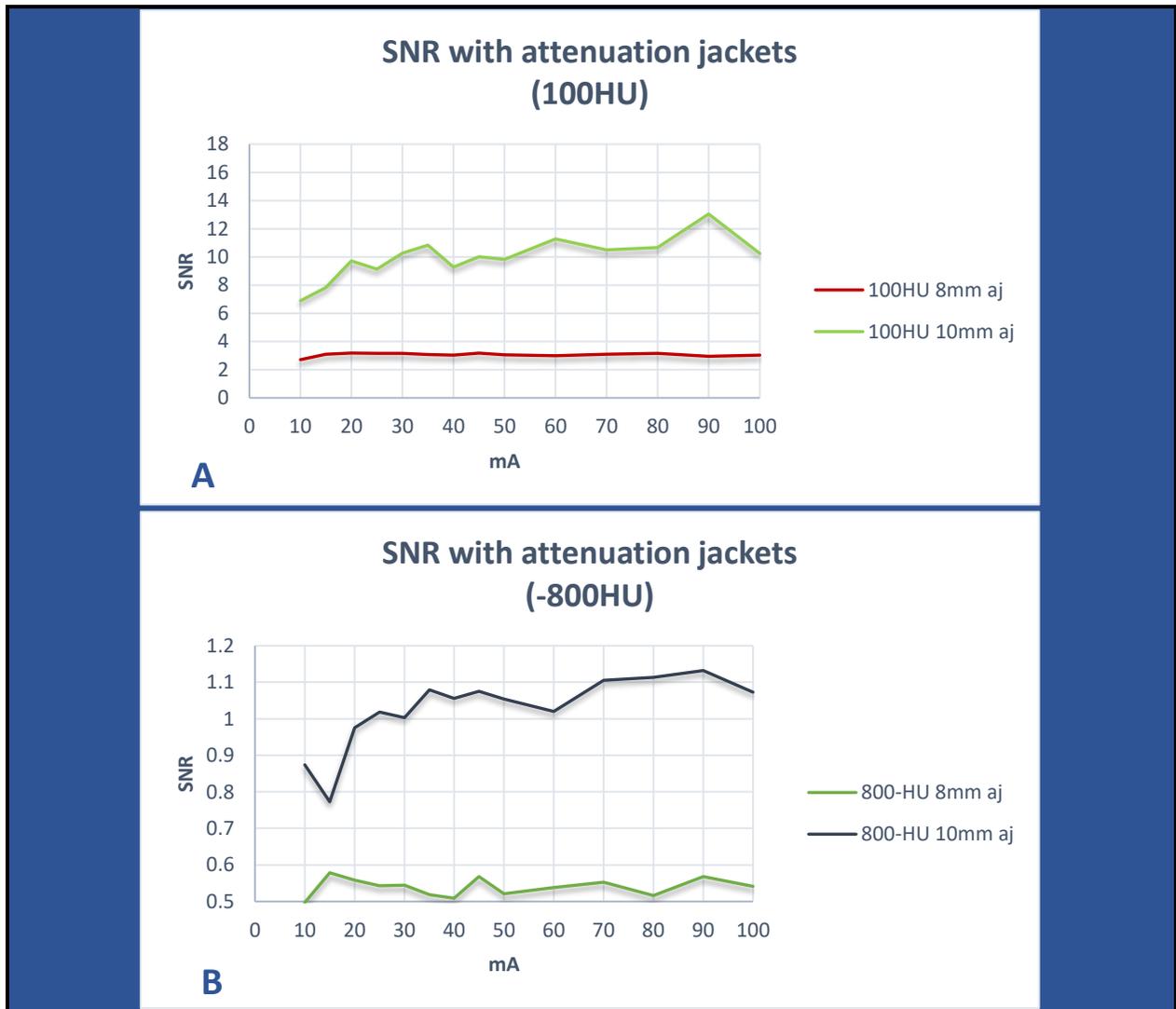


Figure 60. SNR demonstrates change with size and density of simulated lesion (A) 100HU 8mm and 10mm with attenuation jackets (B) -800HU 8mm and 10mm with attenuation jackets.

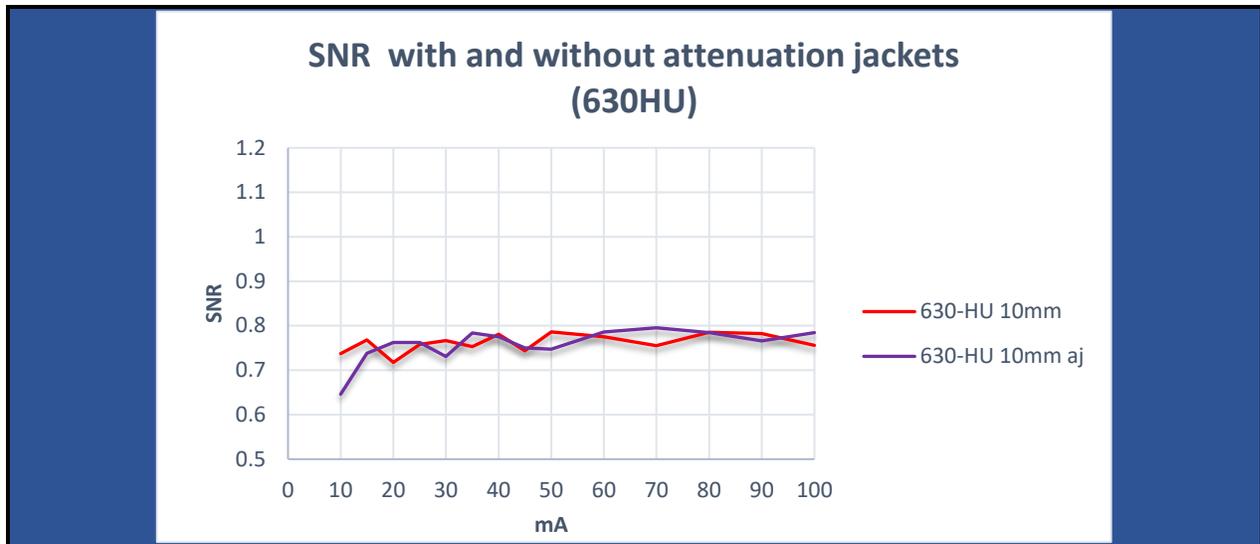


Figure 61. SNR demonstrates little change in comparison of simulated lesions with and without attenuation jackets.

In summary, SNR is a signal-based image quality metric that demonstrated limited information in the context of this research. There was inconsistent trending with little predictive value for radiation dose reduction. However, it was noted that both with and without attenuation jackets, the SNR values indicated a drop in image quality at 100mA.

SNR values did change in relation to size which is to be expected, as the size is explicit in the computation (Bushberg et al., 2012a). Density of the lesion also impacted on the SNR with increased values as density increased. This IQM is noticeably affected by the use of attenuation jackets, and translated into the clinical context would be subject to the effects of body habitus, relating to contrast resolution with subsequent effect on lesion detectability.

A problem with the use of this quality metric in the context of this research is that it required an homogenous background (Bushberg et al., 2012a). This was not the case within the thoracic region of the anthropomorphic chest phantom. In addition, streaking caused by image noise especially at the lower amperages, caused the background regions to be less homogenous. However, the smaller regions of interest relating the 100HU 5mm lesion (without attenuation jackets) may have contributed to the background appearing more homogenous with subsequently increased SNR values.

5.6 Contrast-to-noise ratio (CNR)

This is a signal based IQM that describes the signal level in relation to ambient noise (Bushberg et al., 2012a). CNR values increase with increased contrast. According to Bushberg et al. (2012), CNR is independent of size, however, the results of this research are not in agreement. Analysis was performed using ImageJ.

The simulated lesions were placed with the 5mm lesions in the upper zone, the 10mm lesions in the mid-zone and the 8mm lesions in the lower zone. As with SNR, the 5mm 100HU simulated lesion was difficult to identify and define when attenuation jackets were used, therefore is not included in the related graphs, as were the 5mm -630HU and 5mm -800HU simulated lesions both with and without attenuation jackets. Graphs of CNR calculated from the image data series are seen in *figures 62 and 63*.

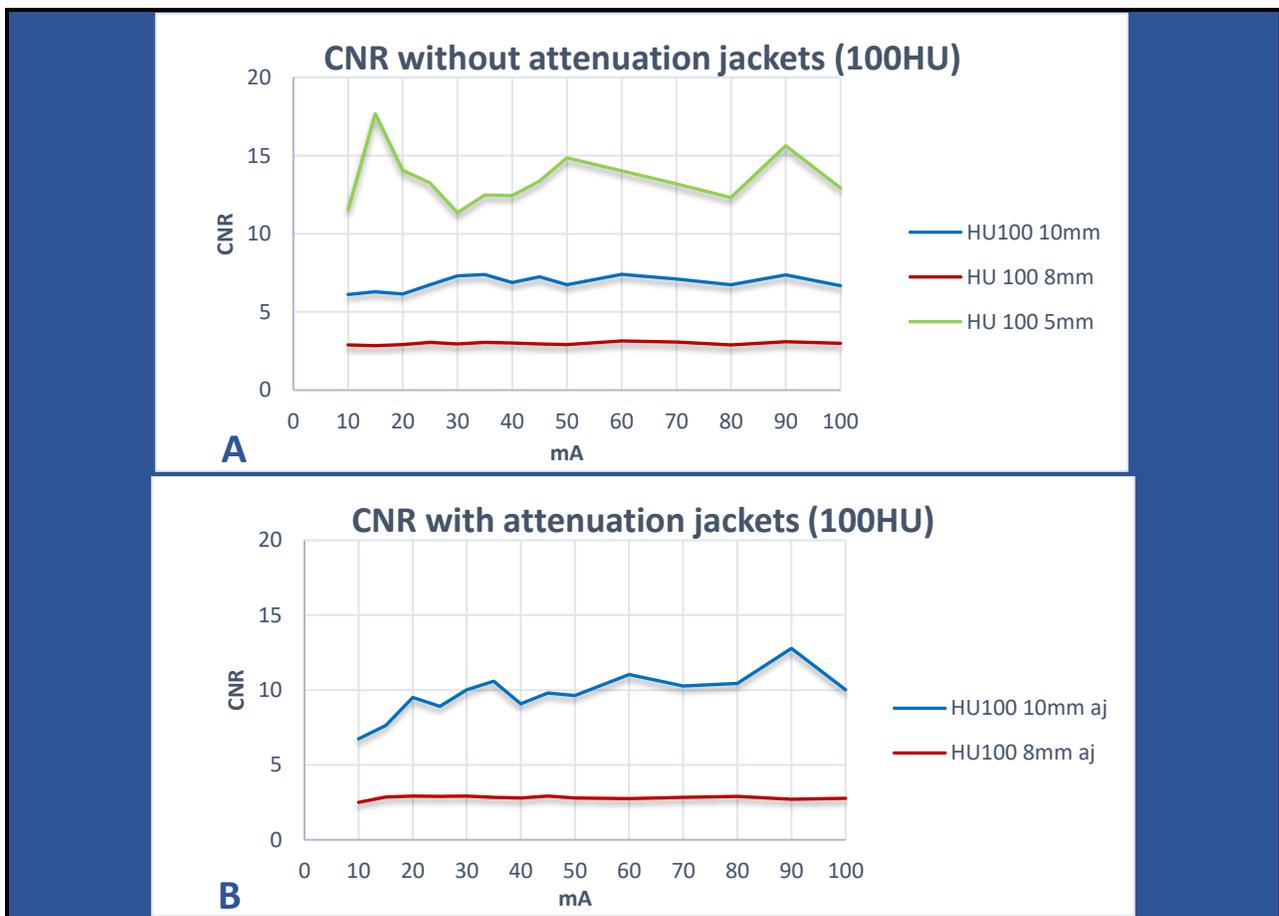


Figure 62. CNR graphs (relating to the simulated lesions): (A) 100HU without attenuation jackets, (B) 100HU with attenuation jackets.

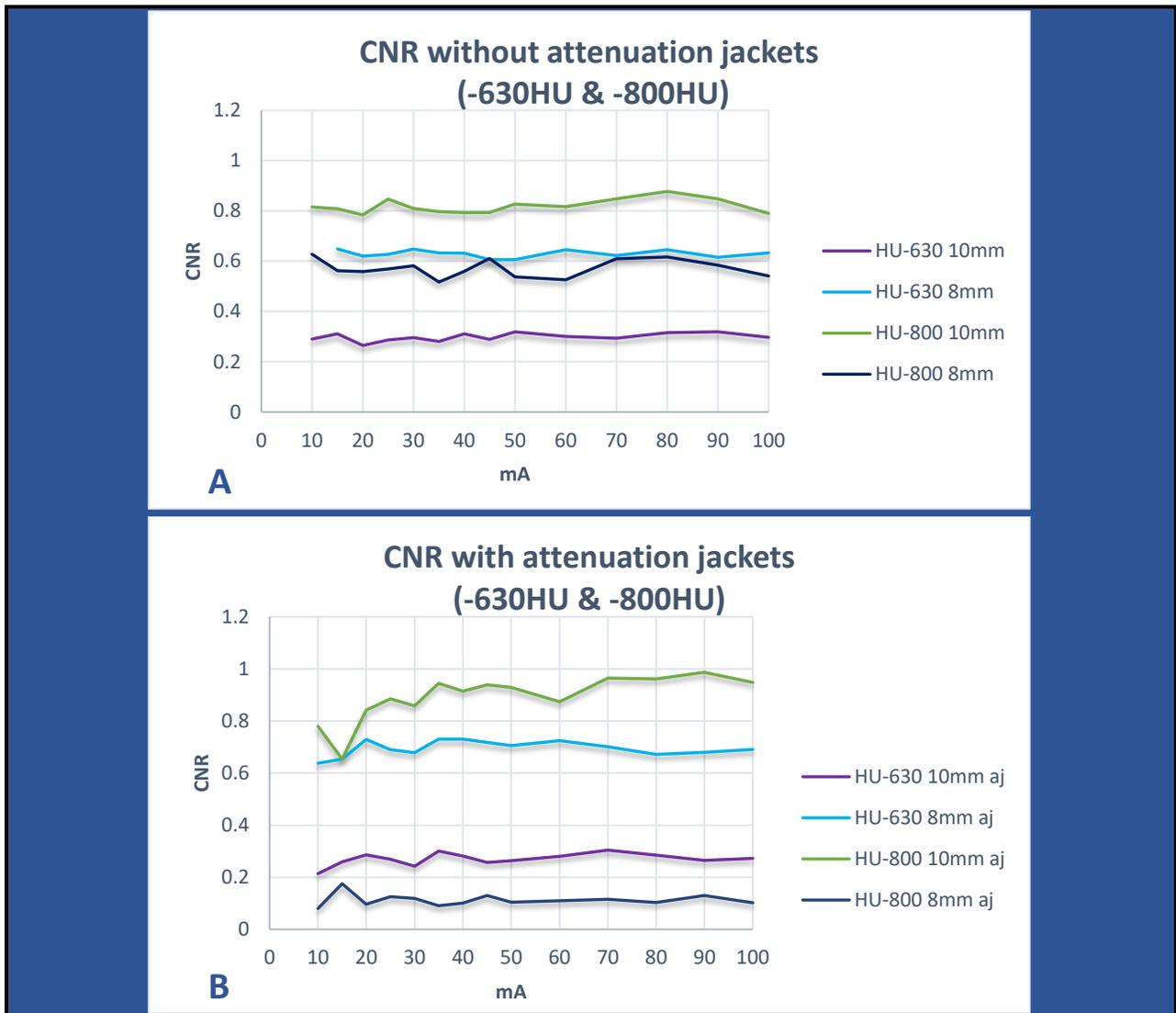


Figure 63. CNR graphs (relating to the simulated lesions): (A) -630HU and -800HU without attenuation jackets, (B) -630HU and -800HU with attenuation jackets.

The graphs in figures 62 and 63 demonstrate CNR values that were consistently higher with simulated lesions of all sizes with a density of 100HU. This was variable with densities of -630HU and -800HU, which may be more prone to influence by location of the simulated lesion than some other IQMs. The -800HU 8mm simulated lesion, located in the lower zone, was particularly affected by the use of attenuation jackets where CNR values decreased by 80%. Conversely, the 100HU 10mm simulated lesion, located in the mid zone, was affected in the

opposite direction, where CNR values increased with the use of attenuation jackets by 30%. Of interest is the -630HU 10mm simulated lesion located in the mid zone, where the CNR values are noticeably lower than those of the -800HU simulated lesion of the same size. This would indicate that lesion density is not always a predictor of improved CNR values, where one might have expected CNR to consistently improve with lesion density in similar locations.

Very little correlation occurred both with and without attenuation jackets, when compared to CNR values in the average data series for each, both with and without attenuation jackets. The results, as seen in *table 19*, show Pearson correlation values (r) for these data series that ranged from -0.098 to 0.892 without attenuation jackets, and values (r) that ranged from -0.001 to 0.995 when attenuation jackets were used. Therefore, as with SNR, using the averaged data series for regression analysis was not reasonable when examining these CNR results. In addition, as previously with SNR, useful trend lines were unable to be applied to the individual data series.

CNR: (r)	Without attenuation jackets	With attenuation jackets
Averaged data series	1	1
HU100 10mm	-0.08121	0.995485
HU 100 8mm	-0.00275	0.302924
HU 100 5mm	0.892227	
HU-630 10mm	0.381823	0.564268
HU-630 8mm	-0.13847	0.46676
HU-800 10mm	0.102877	0.761208
HU-800 8mm	-0.09873	-0.00129

Table 19. Correlation of CNR graph values with averaged data series (Pearson correlation coefficient r).

In graphs of individual data series acquired, peaks and plateaus were identified, where CNR values were similar or lower at lower amperage. Examples of this process are seen in *figure 64* where a value indicated by a red line (mA1) then improved at a lower amperage indicated by a green line (mA2).

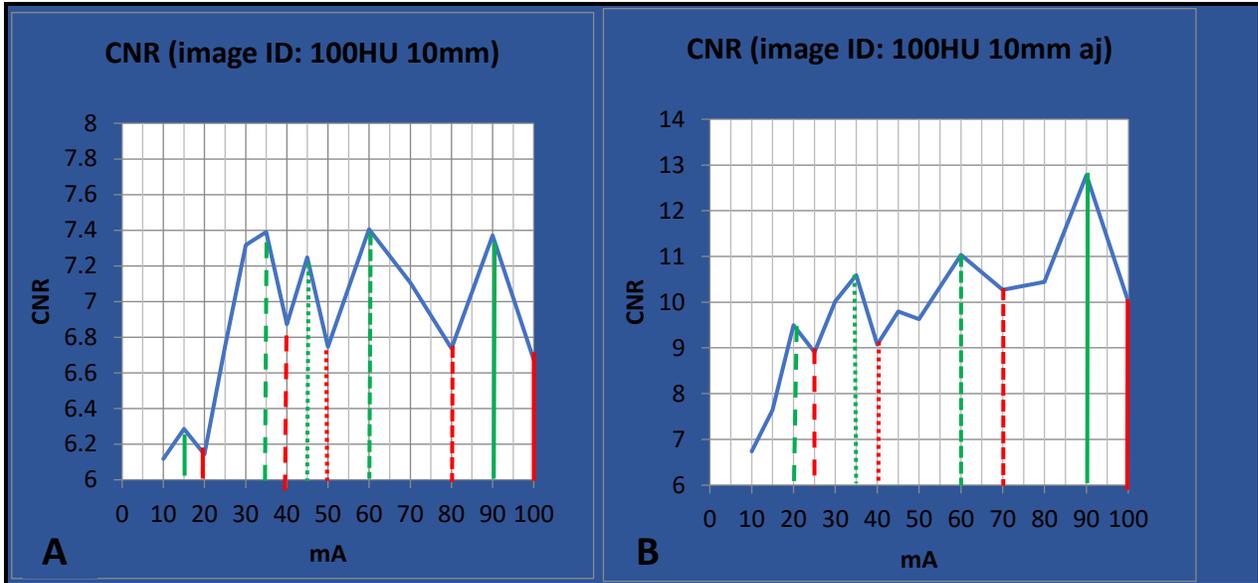


Figure 64. CNR (A) for 100HU 10mm lesion located in mid zone with improvement (from red line to green line) from : 100mA to 80mA, 60mA to 50mA, 50mA to 45mA, 40mA to 35mA, 20mA to 15mA; (B) for 100HU 10mm lesion located in mid zone when attenuation jackets used, with improvement (from red line to green line): from 100mA to 90mA, 70mA to 60mA, 40mA to 35mA, 25mA to 20mA.

As already mentioned, there is little predictive value for radiation dose reduction. However, it is still worth noting the frequency of occurrence at the amperages of where peaks, plateaus and troughs occur, for comparison with other IQMs. To identify any pattern and if this occurred at any particular amperage, frequency analyses were carried out.

Overall, the most frequent amperage where improvement was possible, was 100mA (17%), as seen in *figure 65*. Overall, the most frequent amperages where improvement occurred were: 90mA (13%) and 35mA (12%), also seen in *figure 65*.

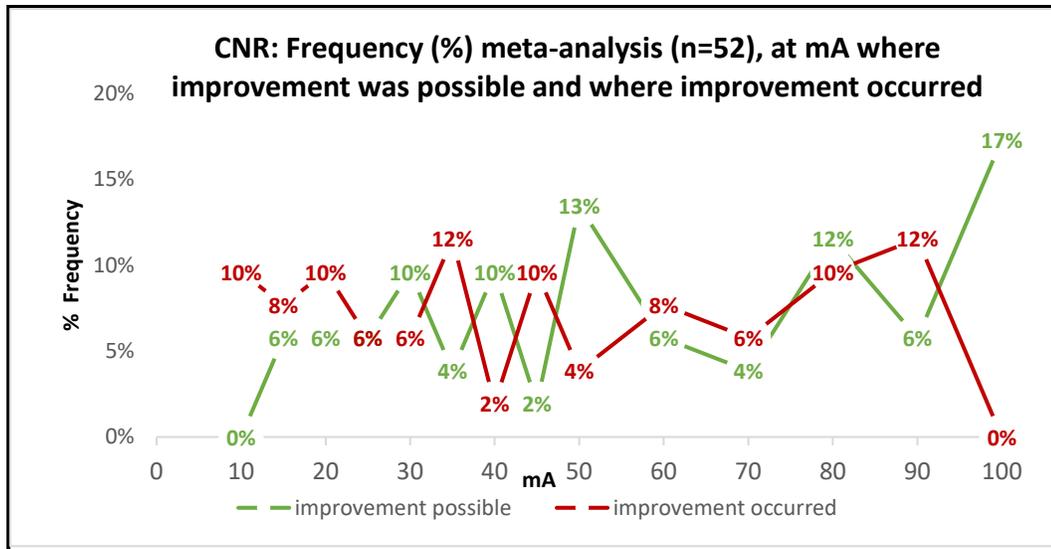


Figure 65. CNR: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperage, where improvement was possible, without attenuation jackets, was 100mA (21%), as seen in *figure 66*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 100mA (13%), 40mA (13%) and 30mA (13%), also seen in *figure 66*.

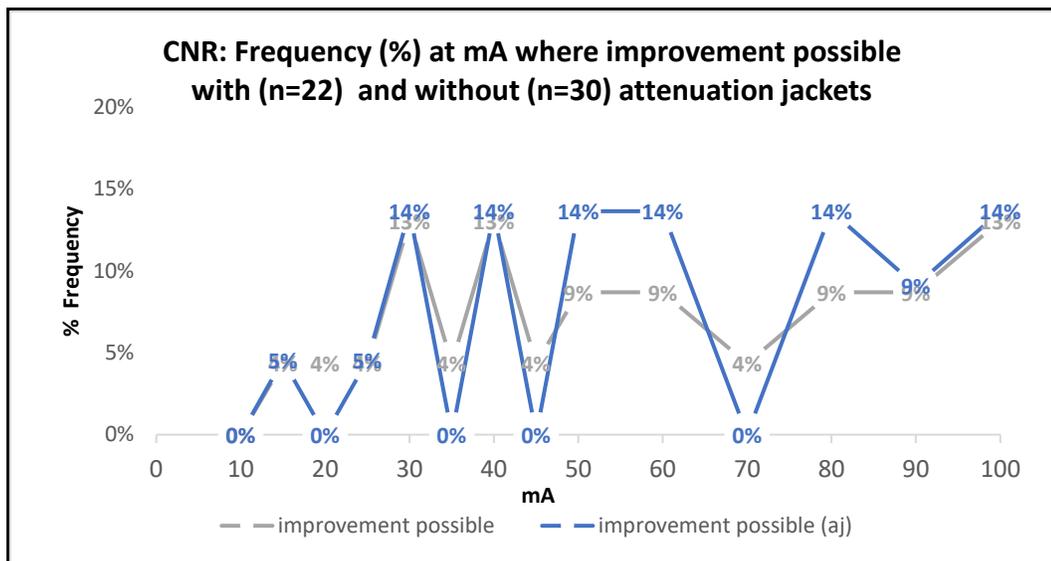


Figure 66. CNR: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.

The most frequent amperage where improvement occurred, without attenuation jackets, was: 90mA (14%), as seen in *figure 67*. The most frequent amperages where improvement occurred,

with attenuation jackets, were: 90mA (13%), 45mA (13%), 35mA (13%) and 20mA (13%), also seen in *figure 67*.

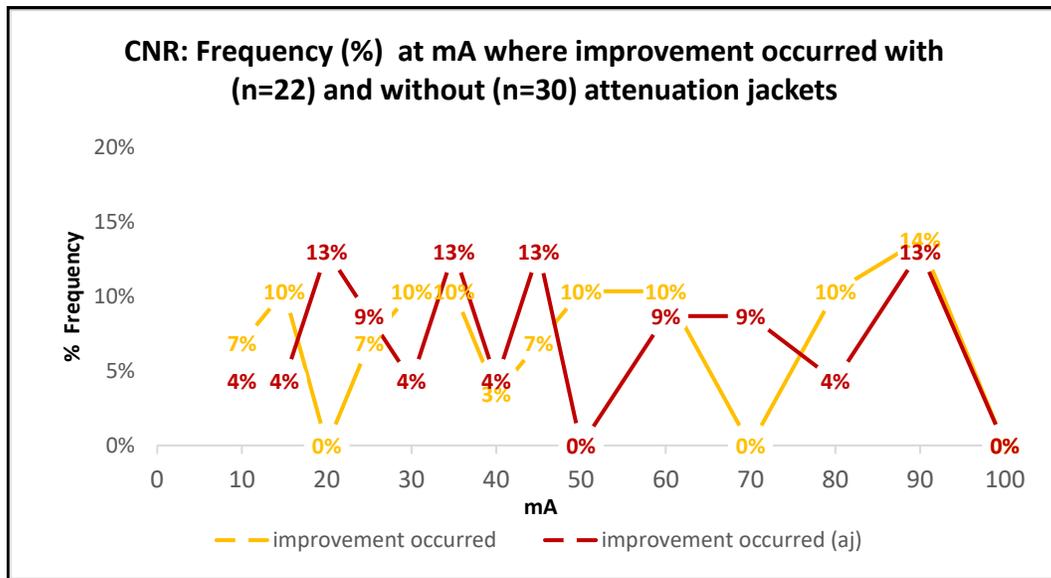


Figure 67. CNR: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.

In summary, CNR is a signal-based image quality metric that demonstrated limited information in the context of this research. The inconsistent trending provided little predictive value for radiation dose reduction. Of note, however, the most frequent amperage where improvement was possible, was 100mA. The most frequent amperage where this improvement occurred, was 90mA. The variability in the data may be in part due to the inhomogeneity of the background regions.

CNR values were consistently higher at a density of 100HU with simulated lesions of all sizes, with the influence of location more variable with -630HU and -800HU densities. The CNR values of the -800HU 8mm simulated lesion, located in the lower zone of the thoracic region, were noticeably influenced by the use of attenuation jackets where the CNR was reduced 80%. The CNR values of the 100HU 10mm simulated lesion, located in the mid zone, were also affected by the use of attenuation jackets, but in this instance the CNR values increased by 30%. In addition, the -800HU 10mm simulated lesion, located in the mid-zone, produced CNR values higher than those for the -630HU 10mm lesion located in the same region, indicating that lesion density is not always a predictor of improved CNR values.

5.7 Structural similarity index (SSIM)

SSIM uses a multiplicative combination of structure, contrast and luminance performing a comparison with a reference image (acquired at 200mA), to provide an overall index (Mathworks, 1994-2020; National_Instruments, 2015). The closer the index values are to 1 the greater the similarity with the reference image, therefore the value increases with improving image quality (National_Instruments, 2015). The analysis was performed using MATLAB®. Resultant graphs for SSIM are shown in *figure 68*.



Figure 68. SSIM graphs (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets.

Good correlation occurred both with and without attenuation jackets, when compared to SSIM values in the averaged data series for each, although there was less correlation without attenuation jackets. The results, as seen in *table 20*, show Pearson correlation values (*r*) for these data series, that ranged from 0.901 to 0.957 without attenuation jackets, and values (*r*) that ranged from 0.965 to 0.994 when attenuation jackets were used. Consequently, using averaged data series was reasonable when examining these SSIM results.

SSIM: (<i>r</i>)	Without attenuation jackets	With attenuation jackets
Averaged data series	1	1
100HU upper zone	0.929479	0.993583
100HU lower zone	0.950873	0.982106
100HU mid zone	0.904685	0.991134
630-HU lower zone	0.931138	0.964826
630-HU mid zone	0.905064	0.968318
800-HU lower zone	0.957383	0.98344
800-HU mid zone	0.901048	0.990834

Table 20. Correlation of SSIM graph values with averaged data series (Pearson correlation coefficient values *r*).

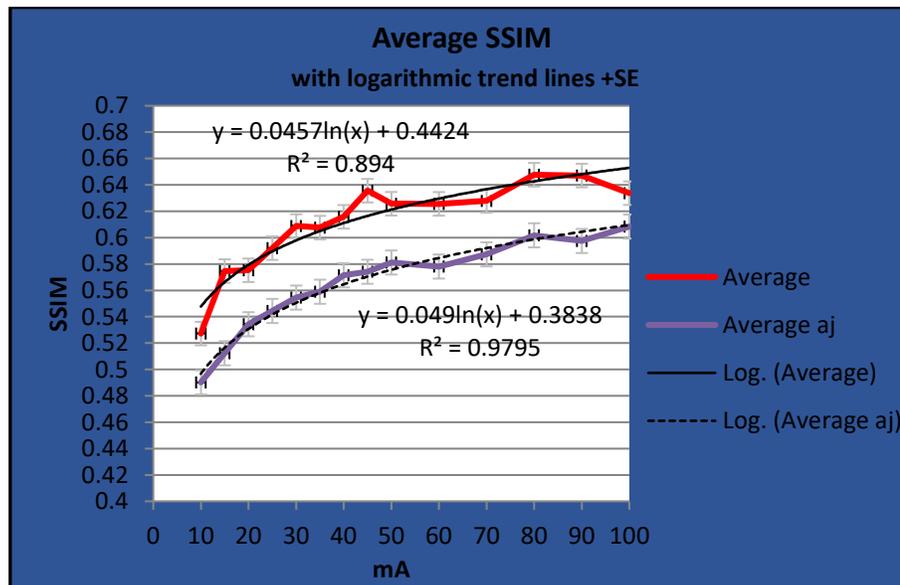


Figure 69. SSIM: Graphs of averaged data series, with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).

SSIM values of averaged data sets, both with and without attenuation jackets, are graphed in *figure 69*, where logarithmic trend lines have been applied. Regression analysis produced an R^2 value that increased when attenuation jackets were used ($R^2=0.894$ increased to $R^2=0.9795$) and demonstrated good correlation of SSIM values with a logarithmic trend. When attenuation jackets were used, the slope of the graph increased by 7%, and the y-intercept decreased by 13%. When comparing results with and without attenuation jackets, the difference in SSIM results was statistically significant ($p=0.0012$). Results of the T-test, assuming unequal variances, are detailed in *table 21* with the T-stat ($T=3.63$) higher than the two-tail critical value ($T=2.06$).

t-Test: Two-Sample Assuming Unequal Variances	
alpha	0.05
df	26
t Stat	3.634913375
P(T<=t) two-tail	0.001202165
t Critical two-tail	2.055529439

Table 21. T-test for average SSIM data series with and without attenuation jackets.

In graphs of individual data series acquired over upper zone, mid zone and lower zones, peaks and plateaus were identified, where SSIM values were similar or lower at lower amperage, in each case. Examples of this process for SSIM are seen in *figure 70*, where a red line (mA1) indicates a value that was then improved upon at a lower amperage indicated by a green line (mA2).

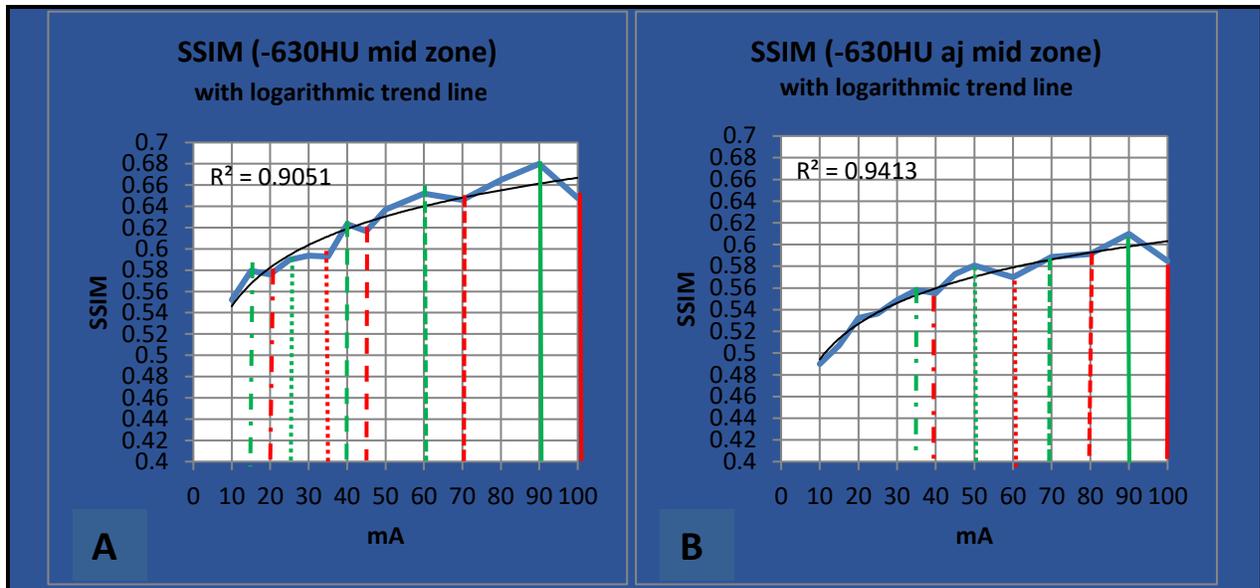


Figure 70. SSIM mid zone (A) with improvement (from red line to green line): 100mA to 90mA, 70mA to 60mA, 45mA to 40mA, 35mA to 25mA, 20mA to 15mA; (B) when attenuation jackets used, with improvement (from red line to green line): 100mA to 90mA, 80mA to 70mA, 60mA to 50mA, 40mA to 35mA.

The potential usefulness of this information is seen in *table 22*, where SSIM values are similar or lower at reduced amperage (mA²). This is then translated into potential for reduced radiation exposure. In these examples, the potential for reduction in radiation exposure ranged from 9.5% to 30%. When looking at SSIM data from all acquisitions in this study, potential for reduction in radiation exposure ranged from 8.7% to 35%.

mA1	mA2 (lower)	SSIM (mA1)	SSIM (mA2)	SSIM % difference	CTDI _{vol} mA1	CTDI _{vol} mA2	CTDI _{vol} % difference
Without attenuation jackets							
100	90	0.64771315	0.679893984	5	4.6	4.15	9.78
70	60	0.64596112	0.652056158	0.9	3.25	2.8	13.85
45	40	0.616627881	0.623252683	1	2.1	1.9	9.52
35	25	0.592822363	0.589900501	-0.5	1.65	1.15	30.30
20	15	0.576014897	0.579382015	0.5	0.9	0.7	22.22
With attenuation jackets							
100	90	0.584738614	0.609873064	4	4.6	4.15	9.78
80	70	0.591372011	0.588715961	-0.3	3.7	3.25	12.16
60	50	0.569756792	0.580748462	2	2.8	2.3	17.86
40	35	0.55513424	0.557998956	0.5	1.9	1.65	13.16

Table 22. SSIM mid zone where improved values at lower amperage translates into reduced radiation exposure (CTDI_{vol}).

To identify any pattern and if this occurred at any particular amperage, frequency analyses were carried out. Overall, the most frequent amperages where improvement was possible, were: 100mA (15%), 70mA (15%), 90mA (13%) and 35mA (11%), as seen in *figure 71*. Overall, the most frequent amperages where improvement occurred were: 80mA (17%), 45mA (13%), 90mA (11%), 50mA (11%) and 45mA (11%), also seen in *figure 71*.

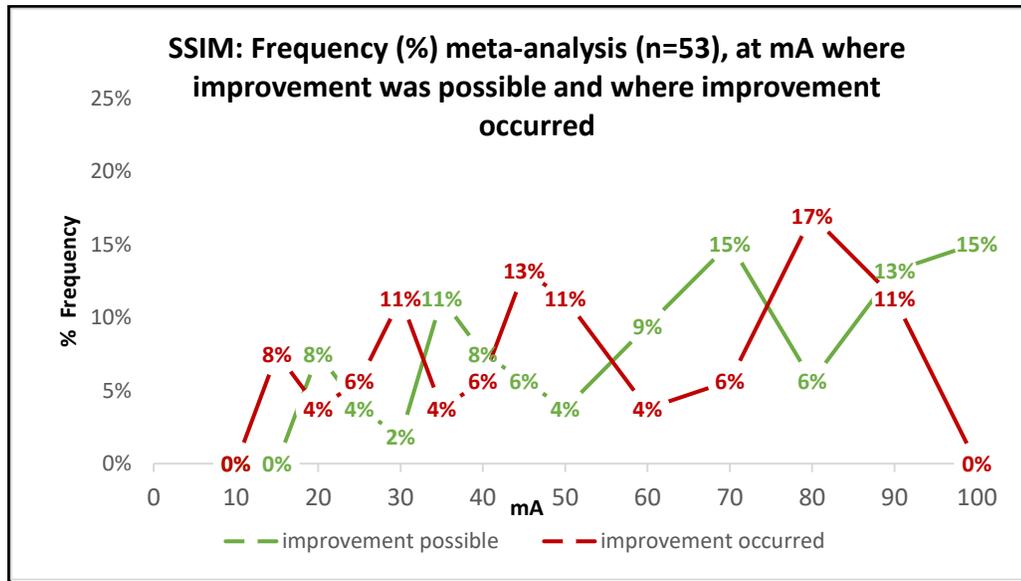


Figure 71. SSIM: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperages, where improvement was possible, without attenuation jackets, were: 100mA (22%), 20mA (15%), 70mA (11%), 60mA (11%) and 35mA (11%), as seen in *figure 72*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 90mA (23%), 70mA (19%) and 35mA (12%), also seen in *figure 72*.

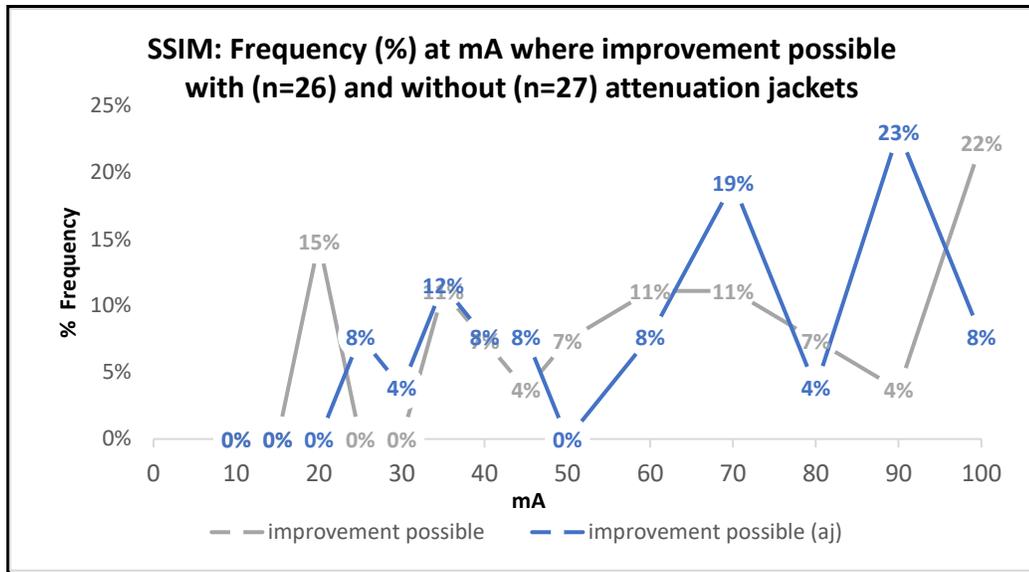


Figure 72. SSIM: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.

The most common amperages where improvement occurred, without attenuation jackets, were: 90mA (15%), 45mA (15%), 15mA (15%), 80mA (11%) and 30mA (11%), as seen in *figure 73*. The most frequent amperages where improvement occurred, with attenuation jackets, were: 80mA (23%), 50mA (15%), 45mA (12%) and 30mA (12%), also seen in *figure 73*.

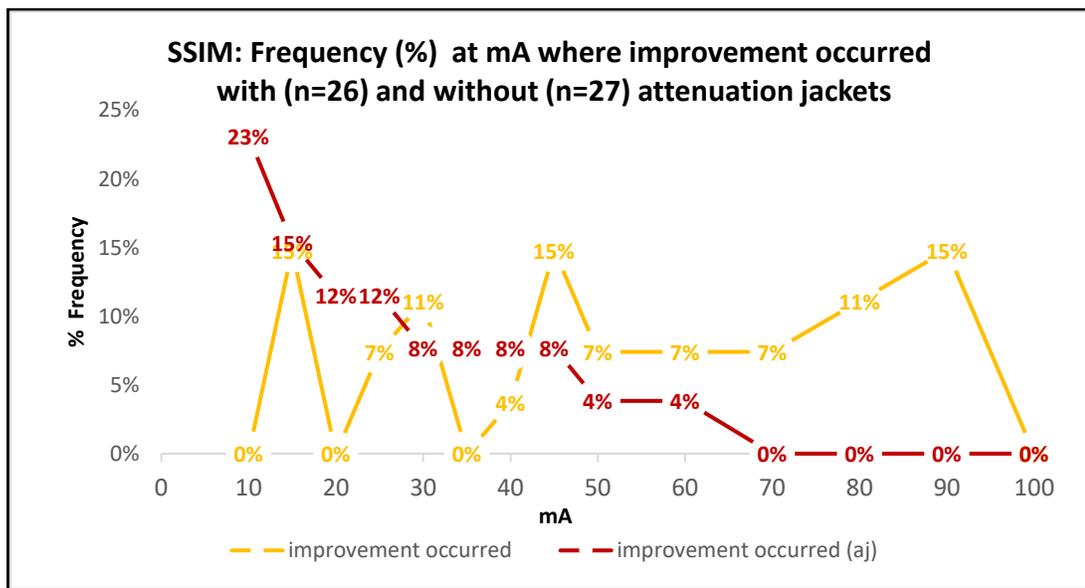


Figure 73. SSIM: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.

In summary, SSIM uses a multiplicative combination of structure, contrast and luminance. This IQM demonstrated some consistency. In data series acquired without attenuation jackets there were lower R^2 coefficients in relation to the logarithmic trending, and less correlation with SSIM values in averaged data series. However, there was still good correlation with SSIM values in averaged data series, which facilitated examination of the results. Logarithmic trending was applied to the graphed data, which demonstrated an increase in gradient and a decrease of the y-intercept with the use of attenuation jackets, with lower SSIM values. The greatest improvement in quality demonstrated by SSIM values occurred over a wide range of amperages, however there was less improvement with increasing mA between 80mA to 100mA. SSIM showed improved values with the use of attenuation jackets, this is most likely due to the decrease in image quality which made image features less distinctive thus appearing more similar. Frequency in occurrence of peaks and plateaus in the graphed data were examined with a view to finding a pattern that could provide potential to reduction in radiation exposure. In the overall data there was some indication (particularly in the upper range of amperages) that there may be potential for dose reduction, however no consistent pattern was identified.

5.8 Non-shift edge ratio (NSER)

NSER looks at edge definition by comparing an image matrix with that of a reference image in the frequency domain (Zhang et al., 2011). By looking at the change in blurring or distortion as edges become less defined, therefore shifting in the frequency domain, a comparison is made with the reference image and a ratio produced with values representing the proportion of edges that have not shifted (Xue & Mou, 2011). The analysis was performed using MATLAB®. Resultant graphs for NSER are shown in *figure 74*, demonstrating improved NSER values with increasing amperage.

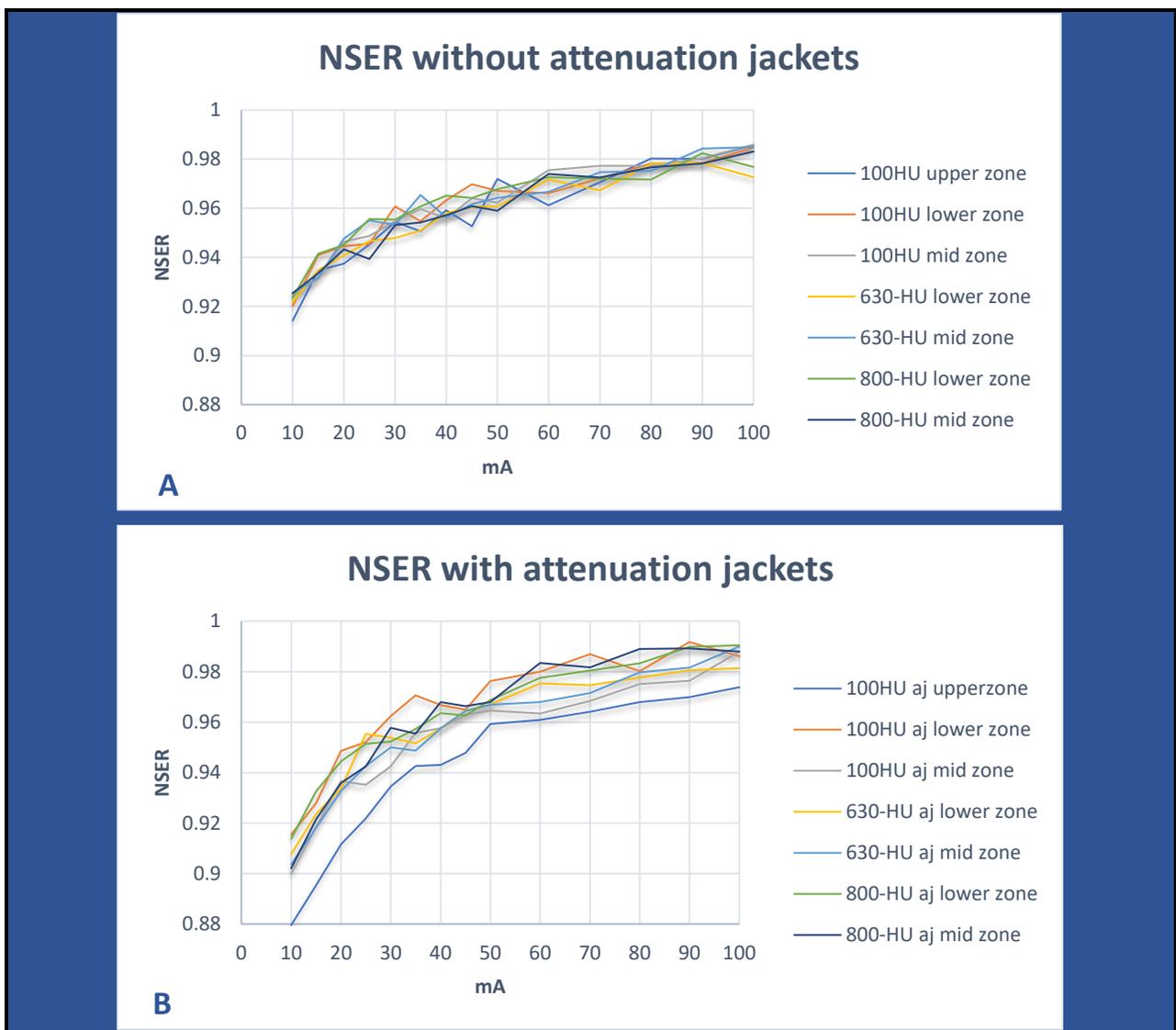


Figure 74. NSER graphs (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets.

Very high correlation occurred both with and without attenuation jackets, when compared to NSER values in the averaged data series for each, with correlation only slightly less without attenuation jackets. The results, as seen in *table 23*, show Pearson correlation values (*r*) for these data series, that ranged from 0.974 to 0.987 without attenuation jackets, and values (*r*) that ranged from 0.984 to 0.996 when attenuation jackets were used. Consequently, using averaged data series was reasonable when examining these NSER results.

Correlation: (<i>r</i>)	Without attenuation jackets	With attenuation jackets
Averaged data series	1	1
100HU upper zone	0.974569	0.996149
100HU lower zone	0.978957	0.984533
100HU mid zone	0.987293	0.98652
630-HU lower zone	0.983663	0.98707
630-HU mid zone	0.975612	0.993831
800-HU lower zone	0.979671	0.991991
800-HU mid zone	0.983049	0.99277

Table 23. Correlation of NSER graph values with averaged data series (Pearson correlation coefficient *r*).

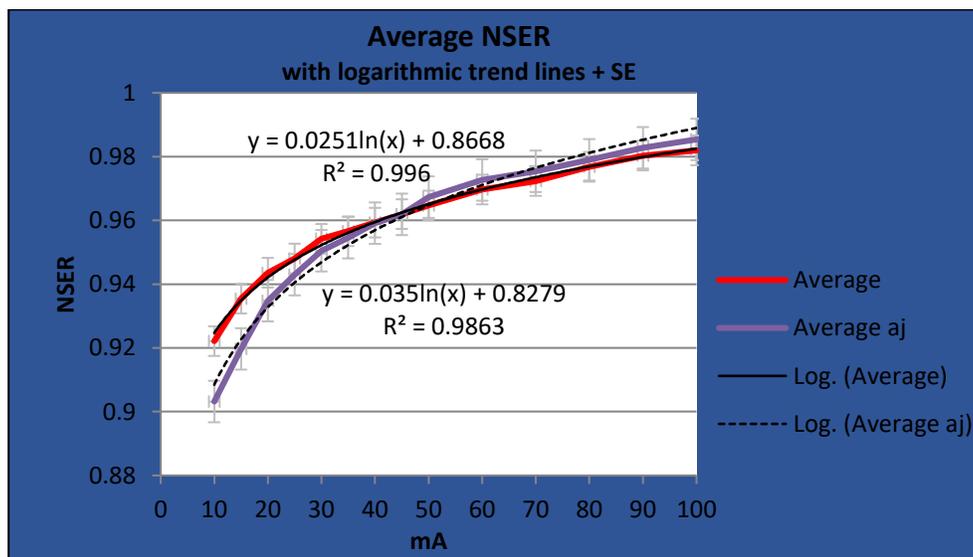


Figure 75. NSER: Graphs of averaged data series, with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).

NSER values of averaged data sets, both with and without attenuation jackets, are graphed in *figure 75*, where logarithmic trend lines have been applied. Regression analysis produced an R^2 value that slightly decreased when attenuation jackets were used ($R^2=0.996$ decreased to

$R^2=0.9863$) and demonstrated very good correlation of NSER values with a logarithmic trend. When attenuation jackets were used, the slope of the graph increased by 28%, and the y-intercept decreased by 4%. When comparing the results with and without attenuation jackets, the NSER results were not statistically significant ($p=0.74$). Results of the T-test, assuming unequal variances, are detailed in *table 24* with the T-stat ($T=0.338$) lower than the two-tail critical value ($T=2.06$).

t-Test: Two-Sample Assuming Unequal Variances	
alpha	0.05
df	24
t Stat	0.33801421
P(T<=t) two-tail	0.738291408
t Critical two-tail	2.063898562

Table 24. T-test for average NSER data series with and without attenuation jackets.

In graphs of individual data series acquired over upper zone, mid zone and lower zones, peaks and plateaus were identified, where NSER values were similar or higher at lower amperage. Examples of this process for NSER are seen in *figure 76*, where a red line (mA1) indicates a value that was then improved upon at a lower amperage indicated by a green line (mA2).

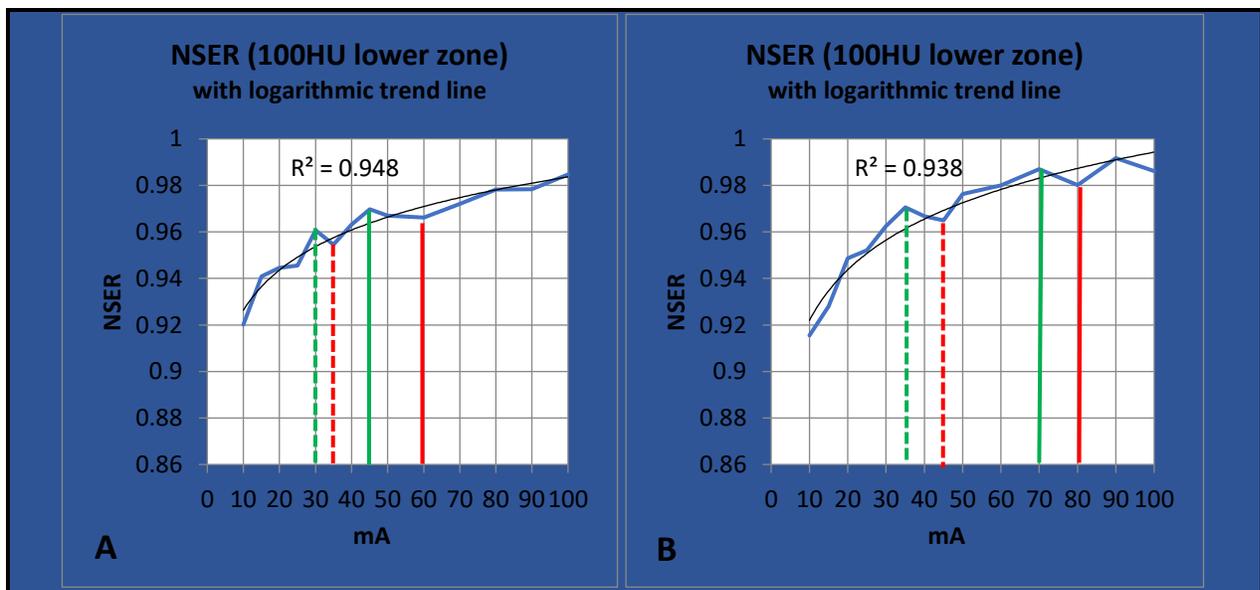


Figure 76. (A) NSER lower zone with improvement (red line to green line) from: 60A to 45mA, 35mA to 30mA; (B) NSER lower zone when attenuation jackets used, with improvement (red line to green line) from 80mA to 70mA, 45mA to 35mA.

The potential usefulness of this information is seen in *table 25*, where NSER values are similar or higher at reduced amperage (mA2). This is then translated into potential for reduced radiation exposure. In these examples, the potential for reduction in radiation exposure ranged from 12.2% to 25.0%. When looking at NSER data from all acquisitions in this study, potential for reduction in radiation exposure ranged from 8.7% to 30.3%.

mA1	mA2 (lower)	NSER (mA1)	NSER (mA2)	NSER % difference	CTDI _{vol} mA1	CTDI _{vol} mA2	CTDI _{vol} % difference
Without attenuation jackets							
60	45	0.966142861	0.969793749	0.4	2.8	2.1	25.0
35	30	0.95461963	0.960752584	0.6	1.65	1.4	15.2
With attenuation jackets							
80	70	0.980200602	0.986940432	0.7	3.7	3.25	12.2
45	35	0.964939309	0.970626218	0.6	2.1	1.65	21.4

Table 25. NSER mid zone where improved values at lower amperage translates into reduced radiation exposure (CTDI_{vol}).

To identify any pattern and how often this occurred at any particular amperage, frequency analyses were carried out. Overall, the most frequent amperages, where improvement was possible, were: 80mA (21%) and 60mA (17%), as seen *figure 77*. Overall, the most frequent amperages where improvement occurred were: 60mA (21%), 50mA (14%) and 40mA (14%), also seen *figure 77*.

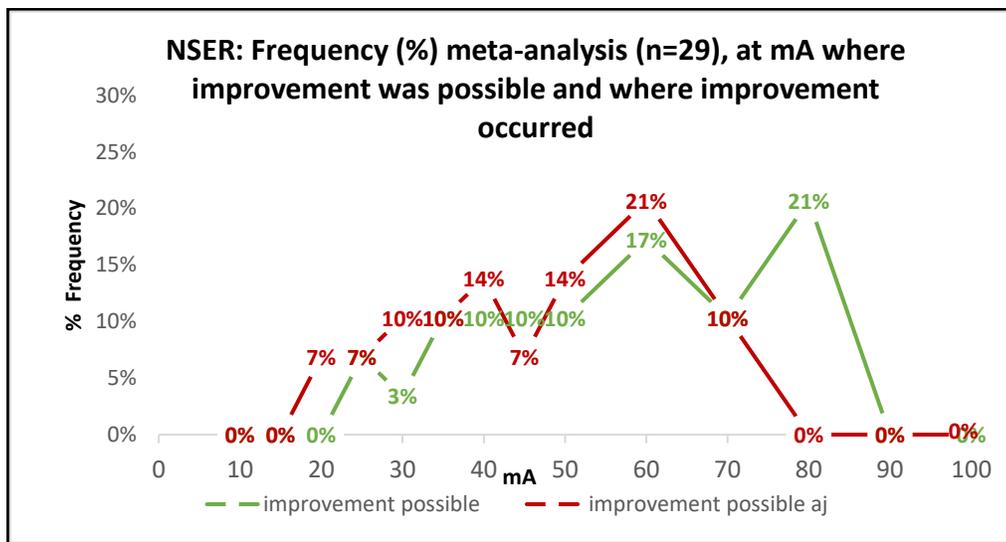


Figure 77. NSER: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperages where improvement was possible, without attenuation jackets, were: 80mA (20%), with 13% for 70mA, 60mA, 50mA and 40mA, as seen in *figure 78*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 80mA (21%), 60mA (21%), 45mA (14%) and 35mA (14%), also seen in *figure 78*.

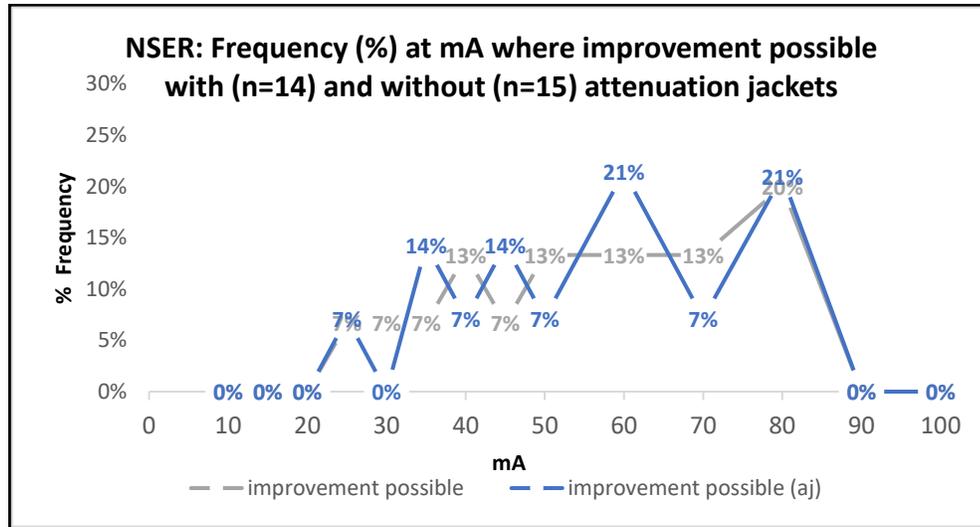


Figure 78. NSER: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.

The most common amperages where improvement occurred, without attenuation jackets, were: 60mA (27%), 45mA (13%), 40mA (13%), and 30mA (13%), as seen in *figure 79*. The most frequent amperages where improvement occurred, with attenuation jackets, were: 50mA (21%), with 14% for 70mA, 60mA, 40mA and 35mA, also seen in *figure 79*.

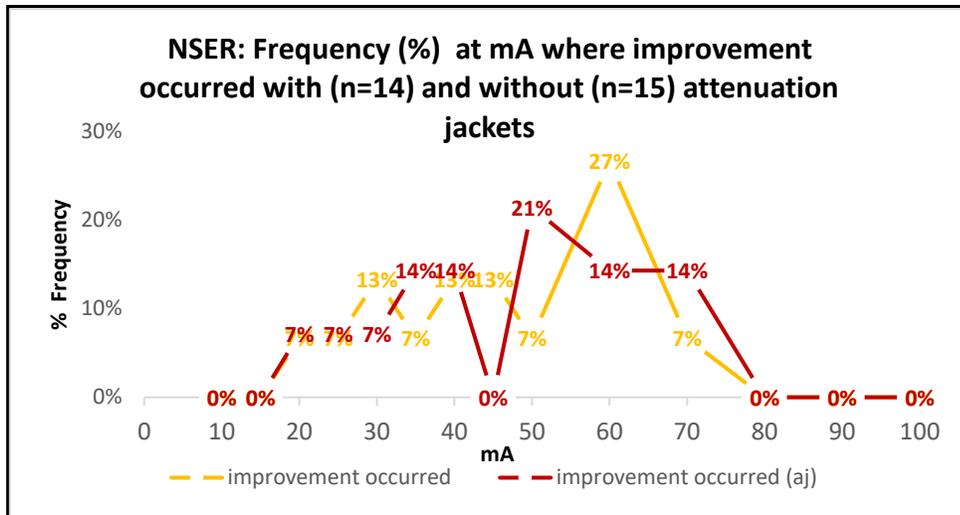


Figure 79. NSER: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets

In summary, NSER looked at edge definition by comparing an image matrix with that of a reference image in the frequency domain (Zhang et al., 2011). This IQM has produced the most useful results thus far.

There was a very good relation between data series, both with and without attenuation jackets, when compared to NSER values in the averaged data series, with correlation only slightly less without attenuation jackets. The averaged data demonstrated no significant difference whether attenuation jackets were used or not, however the slope of the graph increased by 28% indicating an increased influence of amperage in the lower range when attenuation jackets were used. The fact that NSER is less affected by the use of attenuation jackets than other IQMs may be caused by the relation to edge definition which takes spatial resolution into account more than contrast resolution.

Analysis of peaks and plateaus of individual data series highlighted the amperage at 80mA (both with and without attenuation jackets) where the NSER then improved at lower amperage. These improvements were highlighted at 60mA without attenuation jackets and in the meta-analysis, and improvements were also highlighted at 50mA when attenuation jackets were used. This raises the potential for radiation dose reduction with a pattern starting to emerge.

NSER values demonstrated a very good relation with logarithmic trending, with regression analysis producing high R^2 coefficients, and demonstrated a consistent response to increasing amperage regardless of region imaged (upper, middle or lower zone). This is useful in the context of predicting image quality and the potential for lowering amperage to produce a reduction in radiation dose, with an emerging pattern where there is potential for lower amperages while maintaining image quality.

5.9 Noise power spectrum (NPS)

NPS is a function occurring in the frequency domain rather than an absolute measure, however examination of the NPS can increase understanding of the effect of increasing amperage (mA) on images, and variation that can occur depending on the region imaged. To facilitate display and analysis of the NPS results, NPS variance data was plotted using a pseudospectrum estimate via the Eigenvector method (from image analysis carried out using MATLAB®) (MATLAB, 1994-2020). Examples of resultant graphs are seen in *figure 80* and magnified in *figure 81*. NPS values presented in terms of 'Power' have units in decibels (dB) which were plotted against normalized frequency values to facilitate data comparison. Comparative display and graph analyses were carried out in Excel®, and information on the peaks detected was produced in the form of peak values, kurtosis and skewness.

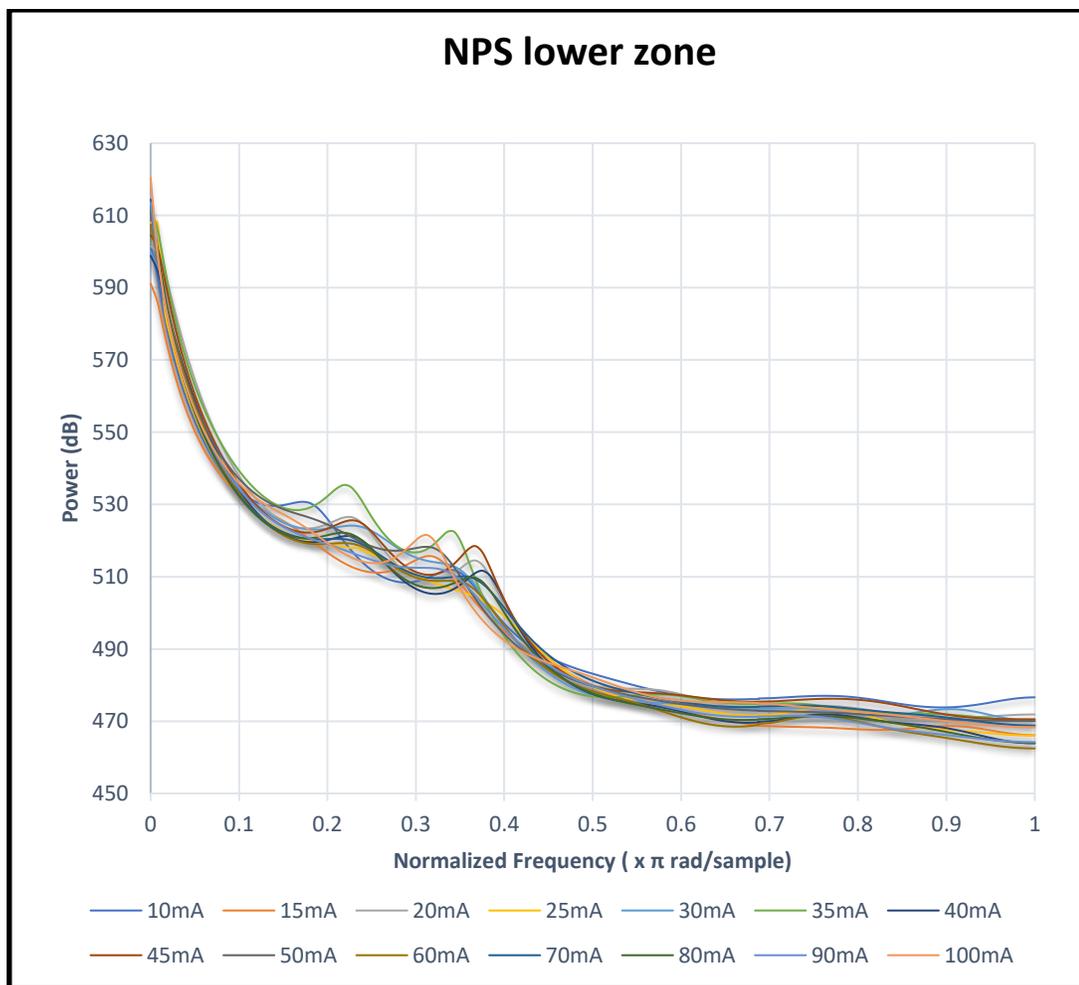


Figure 80. NPS: lower zone.

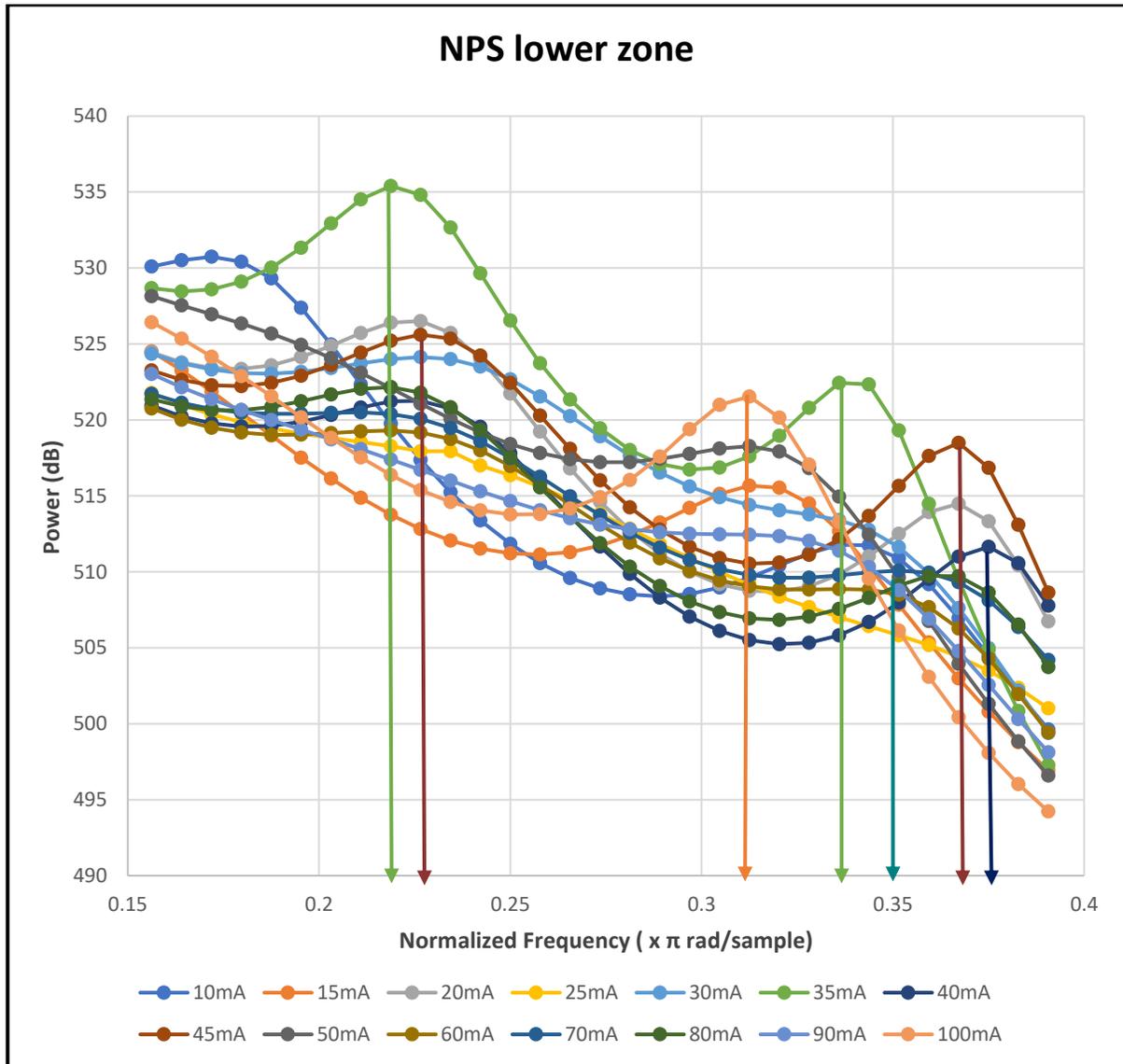


Figure 81. NPS: lower zone magnified (normalized frequency range 0.15 - 0.4), with examples of peak locations marked by arrows.

Magnified in *figure 81*, the individual peaks from *figure 80* were more readily identified. To appreciate how the NPS function was affected by varying amperage, there were aspects that required examination in detail. In addition, as with previous IQMs, the use of attenuation jackets impacted on the results obtained. Examples of resultant NPS graphs, from images obtained using attenuation jackets, as seen in *figure 82*, were magnified in *figure 83* to enable improved

visualisation of peaks. Smaller objects in the spatial domain correspond to higher frequencies in the frequency domain (Bushberg et al., 2012b).

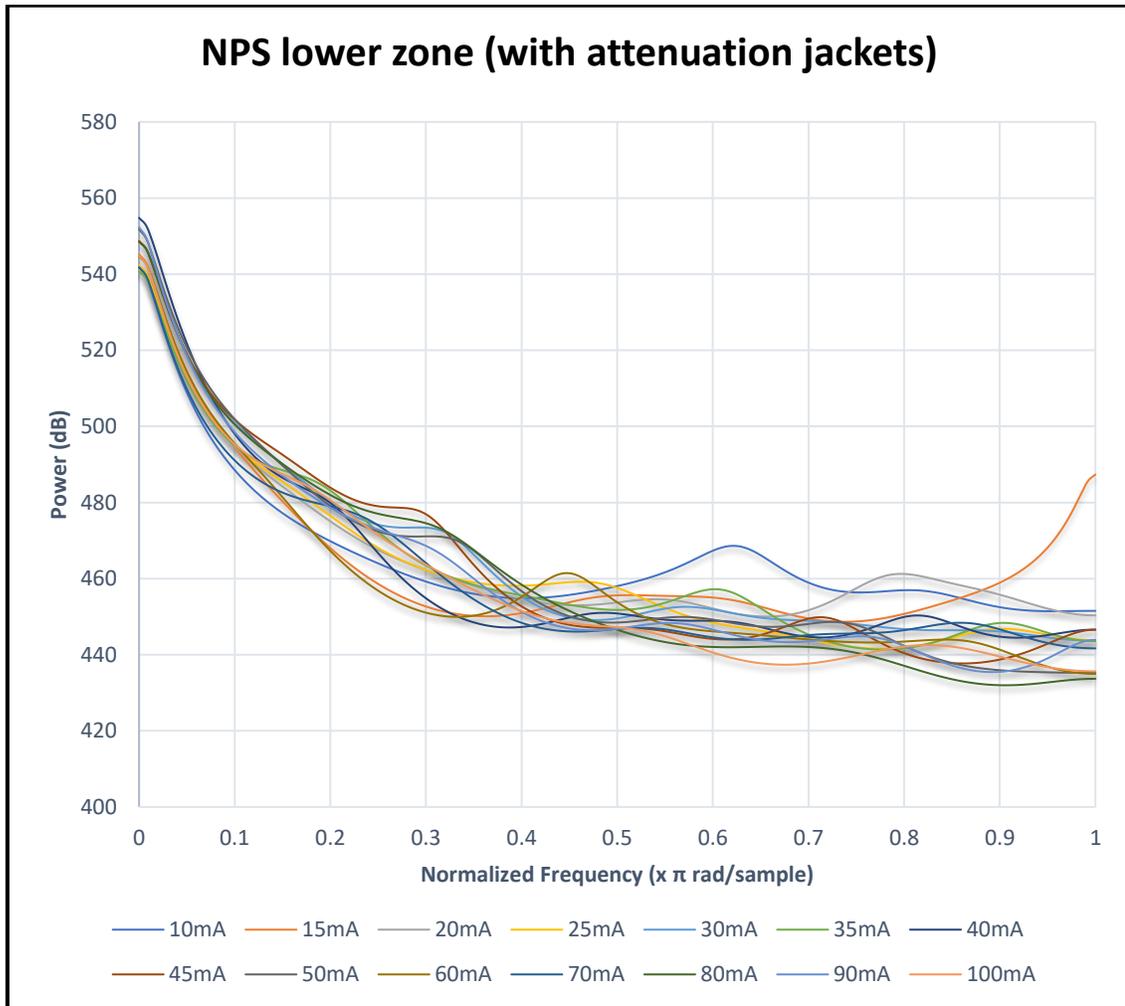


Figure 82. NPS: lower zone (with attenuation jackets).

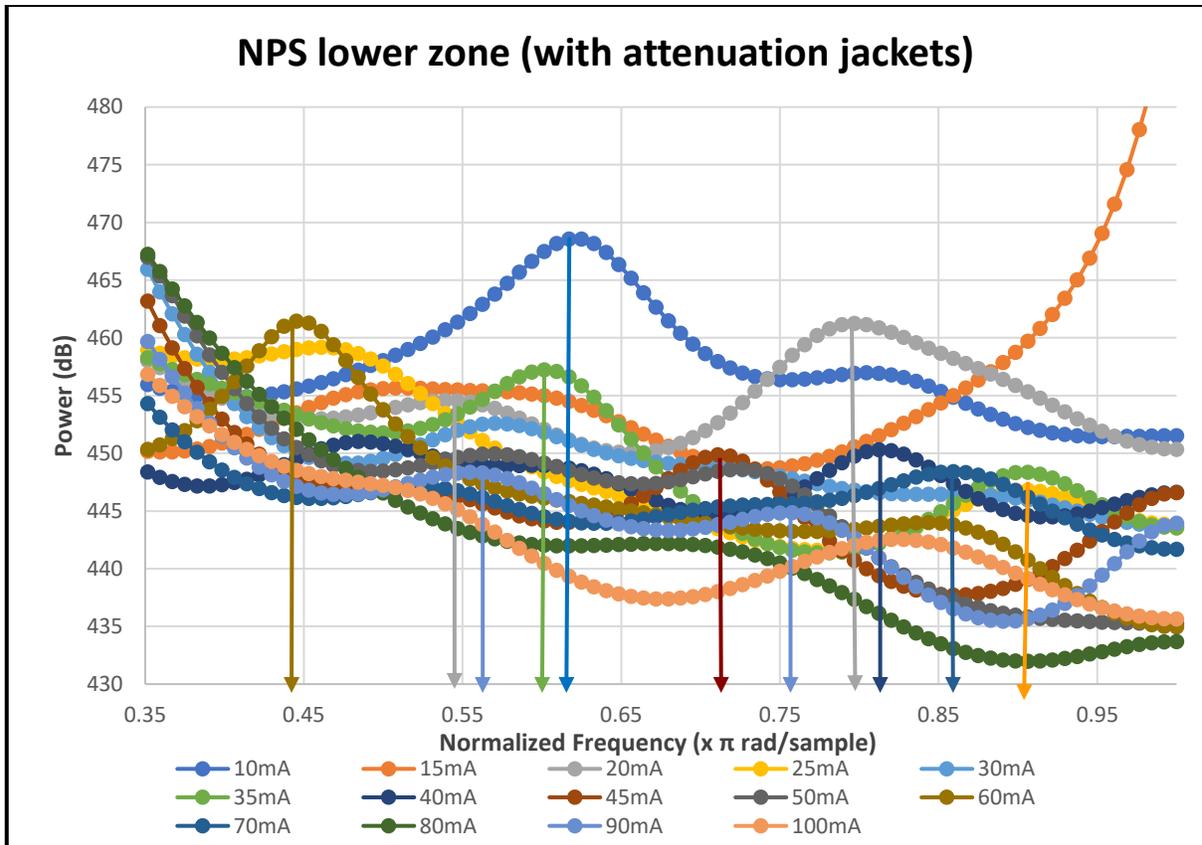


Figure 83. NPS: lower zone (with attenuation jackets) magnified (normalised frequency range 0.35 – 1), with examples of peak locations marked by arrows.

The peaks tended to group into six sets over the normalized frequency range, as listed in *table 26*. While these groupings were then set arbitrarily, they provided a way to better understand the contents of the graphed data.

Set	Normalized frequency range
1	0.1 - 0.25
2	0.25 - 0.35
3	0.35 - 0.4
4	0.4 - 0.5
5	0.5 - 0.75
6	0.75 - 1

Table 26. Peak groups: normalized frequency ranges.

When attenuation jackets were used, peaks were more likely to occur in the upper end of the normalized frequency range. This is demonstrated clearly in the data sets represented in *figure 84* and *figure 85*. This is possibly due to an increase in image noise contributing to the resultant

frequency data. While some peaks moved into the next set, some peaks flattened and others arose, as visualised when comparing *figure 80* and *figure 82*.

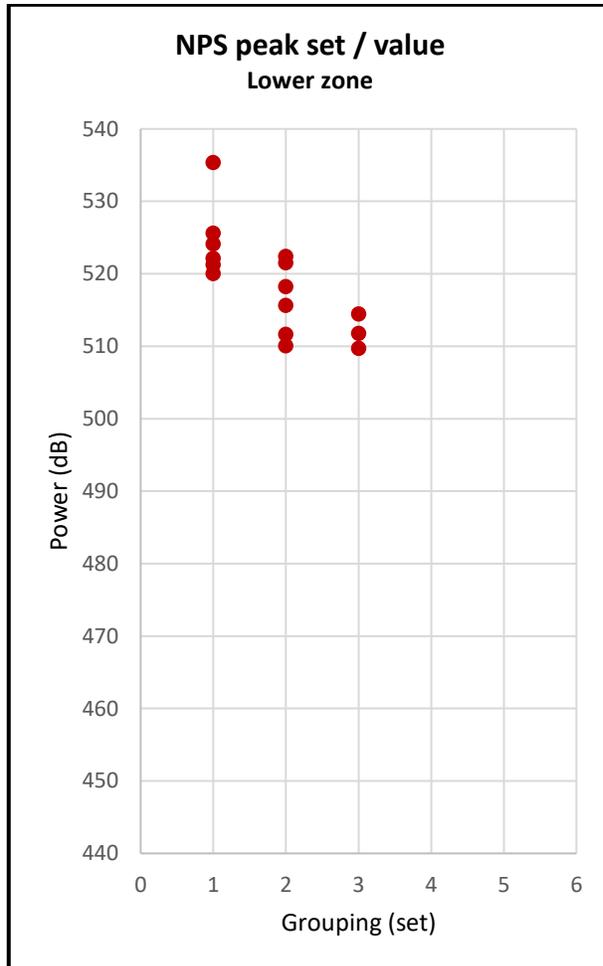


Figure 84. NPS peak grouping (lower zone).

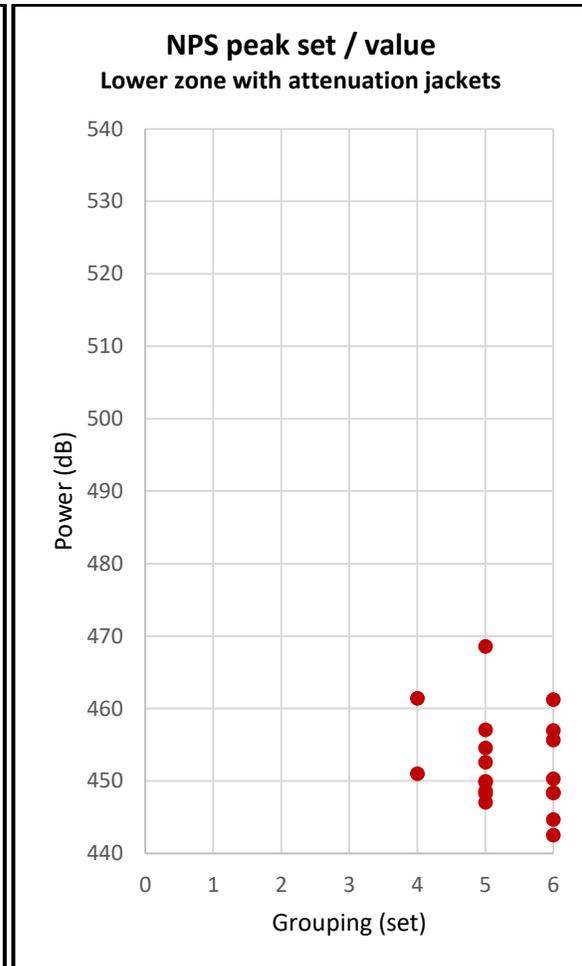


Figure 85. NPS peak grouping (lower zone with attenuation jackets).

As amperage increased, peaks tended to move to the right. Demonstrated in *figures 86 and 87*, these simplified graphs contain only datasets from images acquired at 10mA, 50mA and 100mA. In addition, there was a decrease in peak values when attenuation jackets were used, along with a decrease in peak size. A downward linear trend was identified when attenuation jackets were used, with NPS peak values (dB) plotted against amperage (mA) seen in the example in *figure 88*, where regression analysis demonstrated some relation to increasing amperage ($R^2 = 0.518$).

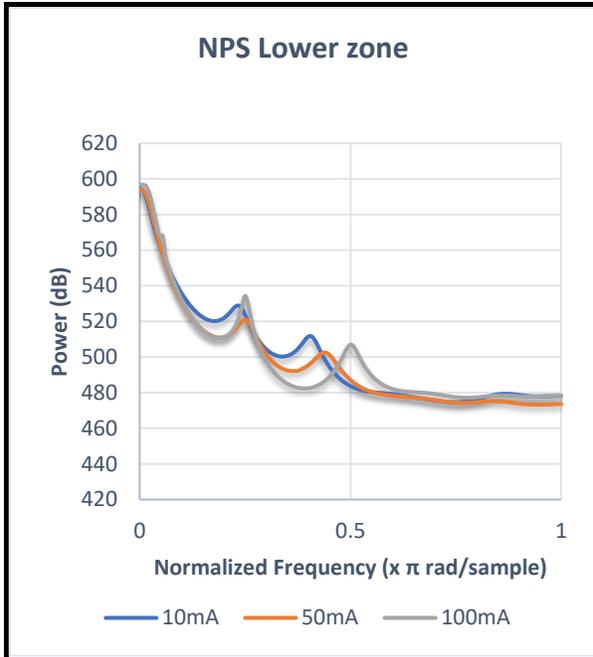


Figure 86. Simplified NPS lower zone.

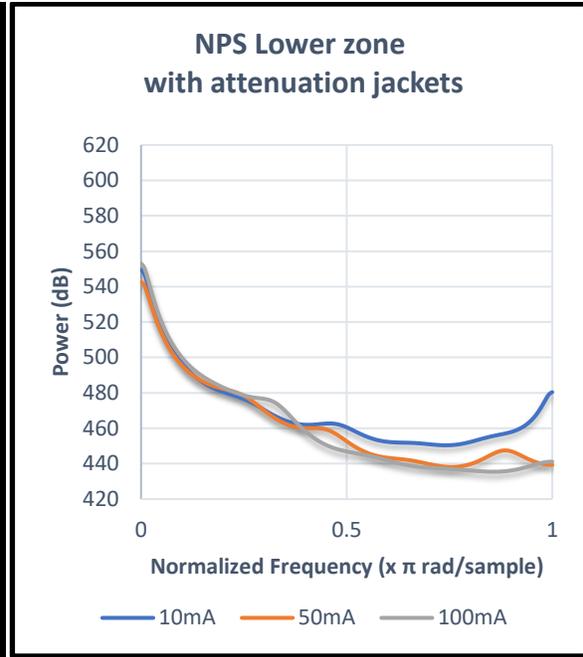


Figure 87 Simplified NPS lower zone with attenuation Jackets.

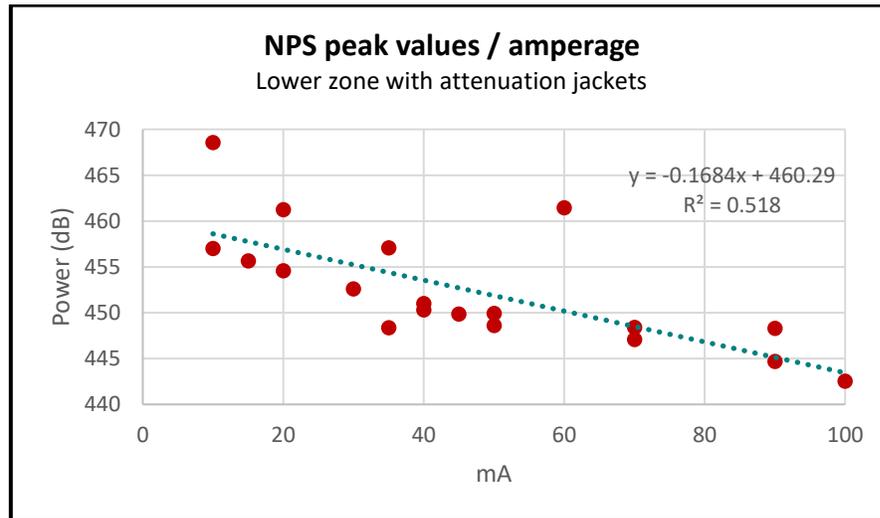


Figure 88. NPS peak values / amperage, with linear trend line.

As image data was converted from the spatial domain to the frequency domain, changes in image content affected the NPS function. This is demonstrated in *figure 89* where a difference in shape of the graphs occurred, with variation in structures depending on the region imaged.

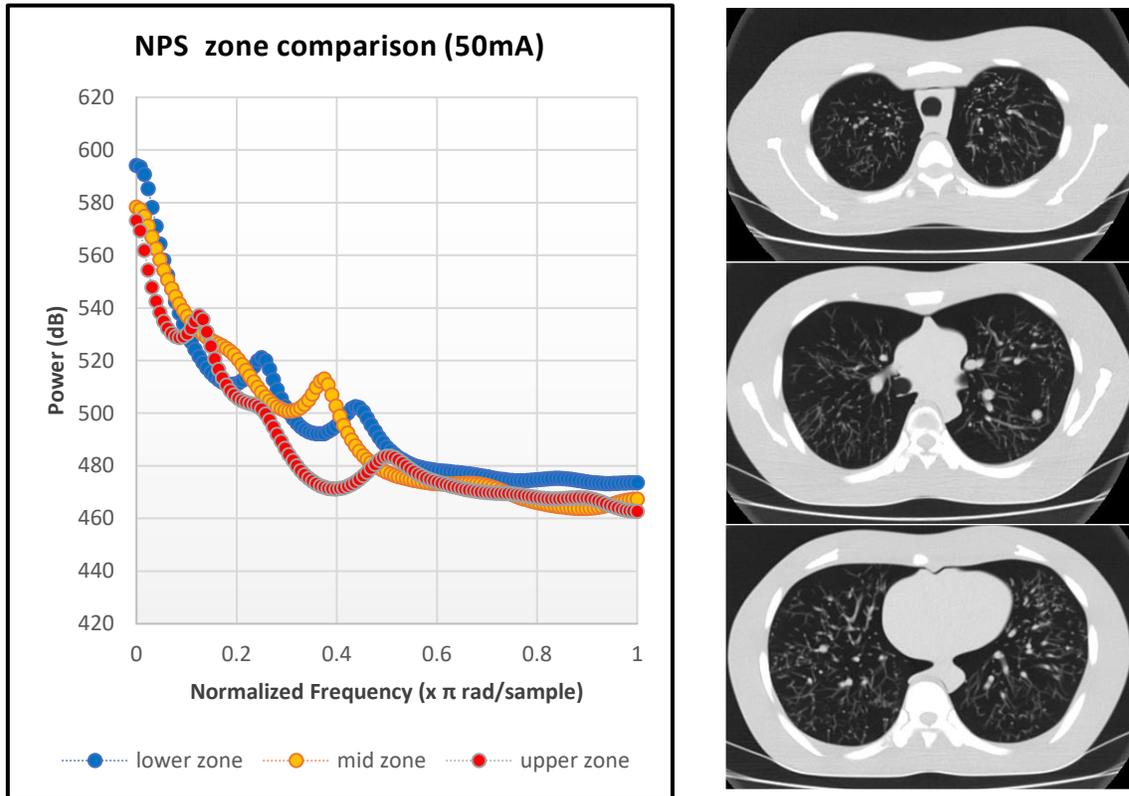


Figure 89. NPS zone comparison with datasets obtained at 50mA. Image display from top to bottom for upper zone, middle zone and lower zone respectively.

In addition to peak values, peak shapes were further defined with kurtosis and skewness. Kurtosis, while sometimes used to describe the flatness of a peak (Microsoft, n.d.), it more accurately describes the tails of a distribution rather than describing the peak itself, where a positive value indicates tails that contain more values and a negative value indicates tails that contain less values (Graphpad, 2020). Skewness describes how symmetrical a distribution is, with negative values indicating more tail values to the left, and positive values indicating more tail values to the right (Graphpad, 2020).

When averaged over a complete data set from the middle zone, kurtosis values were 0.11 and 0.60, without and with attenuation jackets, respectively. This is consistent with the shapes of the graphs as the tails of distributions within peaks widen with the use of attenuation jackets (as seen in the simplified graphs in *figures 86 and 87*). In addition, analysis of kurtosis produced an

increased number of positive values at lower amperages in comparison to higher amperages where the values approached zero or became negative. An example of this is seen in *figure 90 (graph A)*.

Conversely, analysis of skewness produced an increased number of negative values at lower amperages which then approached zero at higher amperages. An example of this is seen in *figure 90 (graph B)*. This implies a greater degree of symmetry with increasing amperage, for distribution of values within the peaks, where the shape of the peak visualised became increasingly upright or shifted to the left (as seen in *figure 86*).

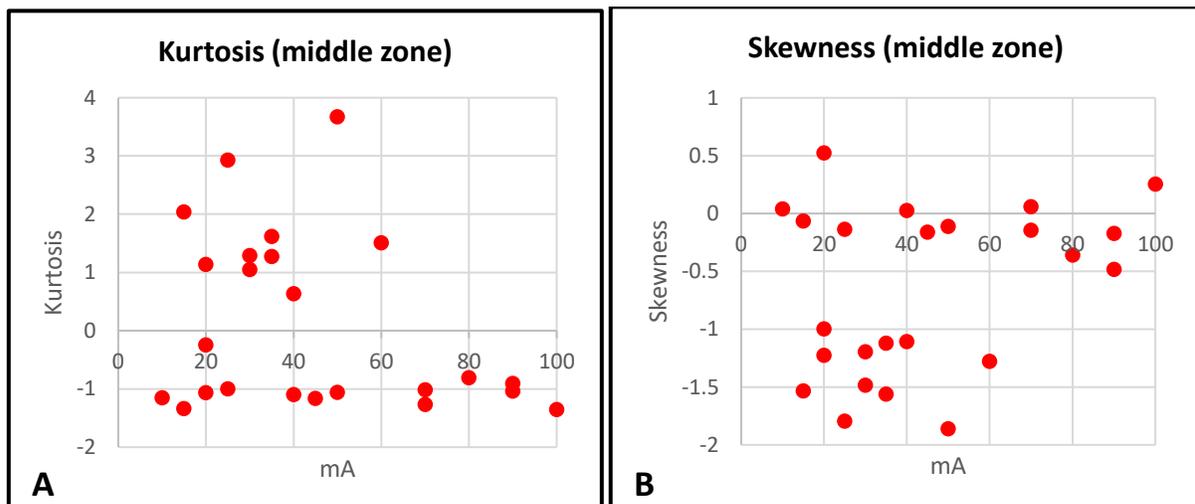


Figure 90. NPS Peak distributions for middle zone (A) Kurtosis and (B) Skewness.

In summary, NPS is a function in the frequency domain rather than an absolute measure. Peaks tended to group into six sets over the normalized frequency range, which facilitated demonstration of a clear shift to the right when attenuation jackets are used. In addition, as amperage increased, peaks tended to move to the right. There was a decrease in peak values when attenuation jackets were used, along with a decrease in peak size. A downward linear trend of peak values was identified, with increasing amperage when attenuation jackets were used.

Changes in image content affected the NPS function. A difference in shape of the graphs occurred with variation in structures depending on the region imaged. Kurtosis increased with the use of attenuation jackets. Kurtosis decreased at higher amperage where the values approached zero or became negative. Conversely, skewness increased at higher amperages where the values approached zero or became positive.

NPS enabled the effect of increasing amperage to be visualised, as well as the effect of using attenuation jackets. It also demonstrated differences in NPS between regions that were imaged. NPS is useful as an IQM to visualise the effects of amperage on image quality, however it would need to be used in conjunction with other IQMs to quantify these effects.

5.10 Texture analysis

Texture analysis uses spatial variation in pixel intensities to examine aspects such as smoothness or coarseness and randomness or regularity in relation to image structure (Mathworks, 1994-2020). Image analysis was performed using LIFEx[®] software (Nioche et al., 2018). IQMs examined were from those produced by LIFEx[®] using a Gray-Level-Co-occurrence-Matrix (GLCM). These were contrast, correlation, energy (uniformity), homogeneity (relating to structures) and entropy.

5.10.1 Contrast

Contrast (sometimes called variance or inertia), in this context is a measure of the intensity Contrast between a pixel and its neighbour which equals zero for a constant image (namely equal values for the whole image) (Mathworks, 1994-2020). Resultant graphs for Contrast are shown in *figure 91* (mid zone) and *figure 92* (upper and lower zones).

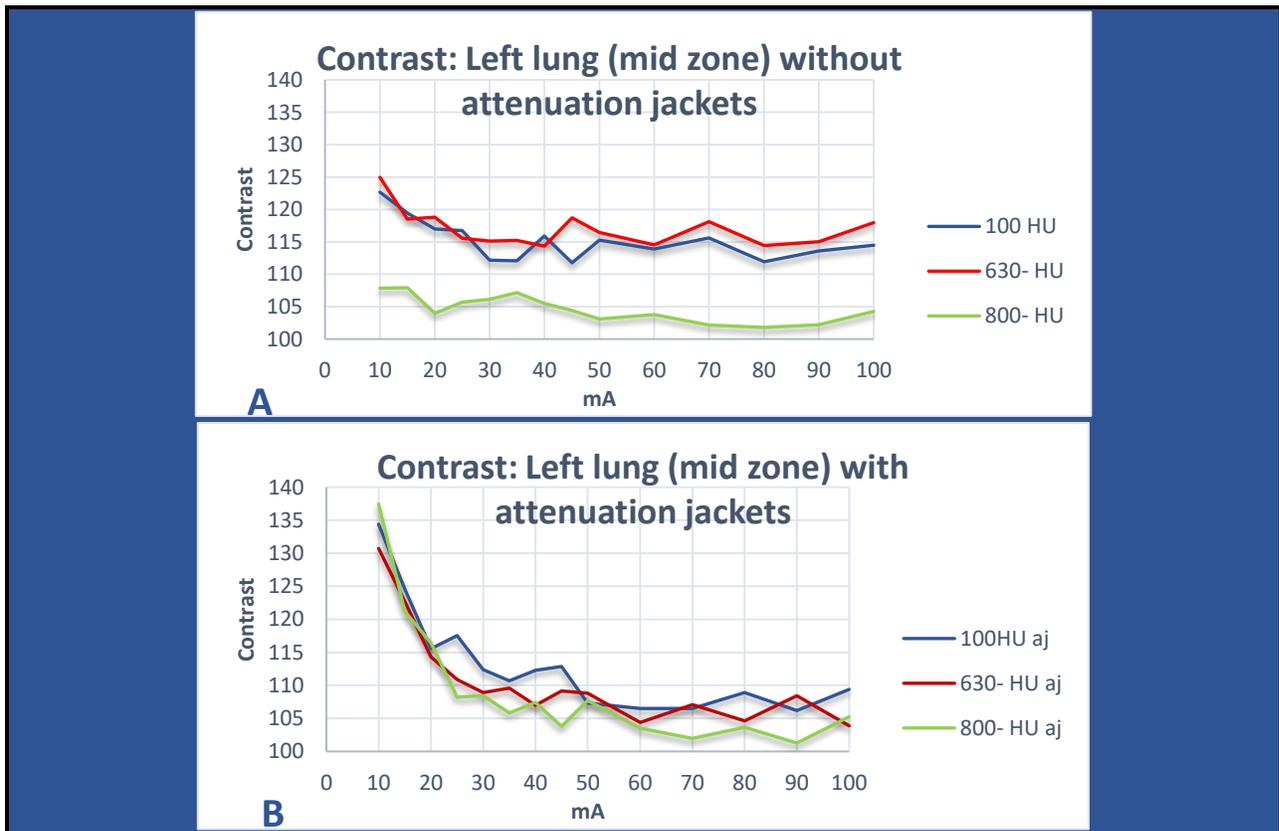


Figure 91. Contrast: Graphs for mid zone (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets.

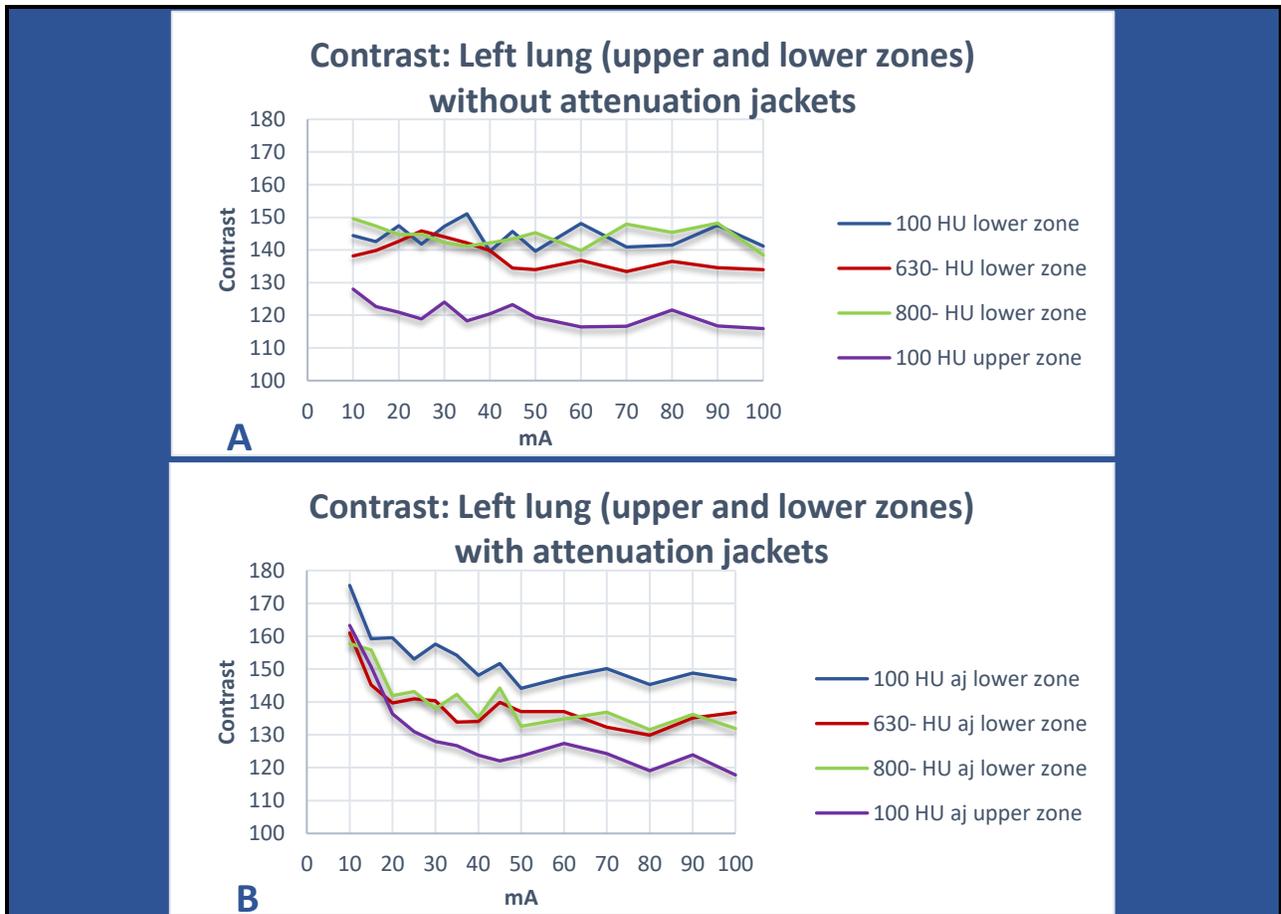


Figure 92. Contrast: Graphs for upper and lower zones (where HU refers to the simulated lesion density in that image series): (A) without attenuation jackets, (B) with attenuation jackets.

There is very little relation within the Contrast data series without attenuation jackets, with Pearson correlation values (r) for these data series, that ranged from 0.28 to 0.77 when compared to overall averaged data series. When the graphs in *figures 91* and *92* were visually assessed, however, they demonstrated an increasing gradient in the lower amperage range when attenuation jackets had been used, therefore an impact is still detected in this circumstance. In addition, the lung region analysed impacted on Contrast values, and the use of attenuation jackets had varying impact with the greatest impact demonstrated in the mid zone (*figure 91*).

In graphs of individual Contrast data series acquired over upper zone, mid zone and lower zones, peaks and plateaus were identified, where Contrast values were similar or less at lower amperage. Examples of this process for Contrast are seen in *figure 93*, where a red line (mA1)

indicates a value that was then improved upon at a lower amperage indicated by a green line (mA2).

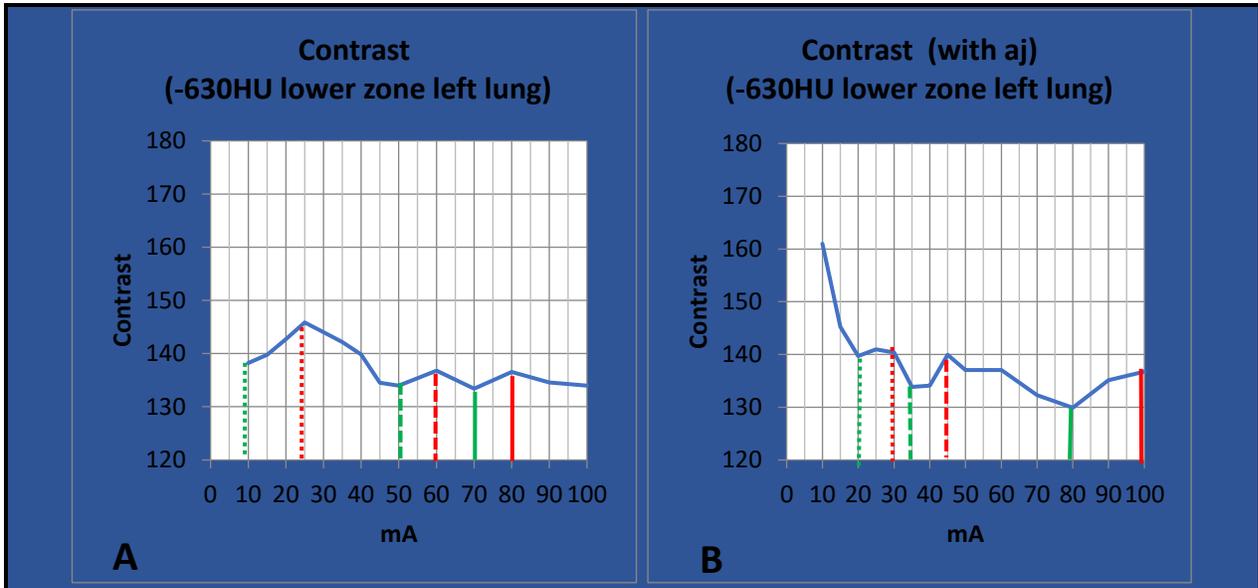


Figure 93. Contrast: Lower zone zones (where HU refers to the simulated lesion density related to that image series) (A) without attenuation jackets, with improvement (red line to green line) from: 80mA to 70mA, 60mA to 50mA, 25mA to 10mA; (B) with attenuation jackets, with improvement (red line to green line) from 100mA to 80mA, 45mA to 35mA, 30mA to 20mA

The potential usefulness of this information is seen in *table 27*, where Contrast values are similar or less at reduced amperage. This is then translated into potential for reduced radiation exposure. In these examples, the potential for reduction in radiation exposure ranged from 12.2% to 56.5%. When looking at data from all acquisitions in this study, potential for reduction in radiation exposure ranged from 19.6% to 35.7%. However, given the lack of correlation seen when attenuation jackets were not used, and varying values between regions analysed, the usefulness of this information is limited for this IQM.

mA1	mA2 (lower)	Contrast(mA1)	Contrast(mA2)	Contrast % difference	CTDI _{vol} mA1	CTDI _{vol} mA2	CTDI _{vol} % difference
Without attenuation jackets							
80	70	136.534	133.4	2.30	3.7	3.25	12.16
60	50	136.784	133.986	2.798	2.8	2.3	17.86
25	10	145.851	138.197	7.654	1.15	0.5	56.52
With attenuation jackets							
100	80	136.738	129.906	5.00	4.6	3.7	19.57
45	35	139.924	133.876	4.32	2.1	1.65	21.43
30	20	140.371	139.731	0.46	1.4	0.9	35.71

Table 27. Contrast: Lower zone left lung for example in *figure 93*, where improved values at lower amperage potentially translate into reduced radiation exposure (CTDI_{vol}).

However, it is still worth noting the frequency of occurrence at the amperages of where peaks, plateaus and troughs occur, for comparison with other IQMs. To identify any pattern and if this occurred at any particular amperage, frequency analyses were carried out.

Overall, the most frequent amperage where an improvement was possible, was 100mA (25%), as seen in *figure 94*. Overall, the most frequent amperage where improvement occurred, was 80mA (13%), also seen in *figure 94*.

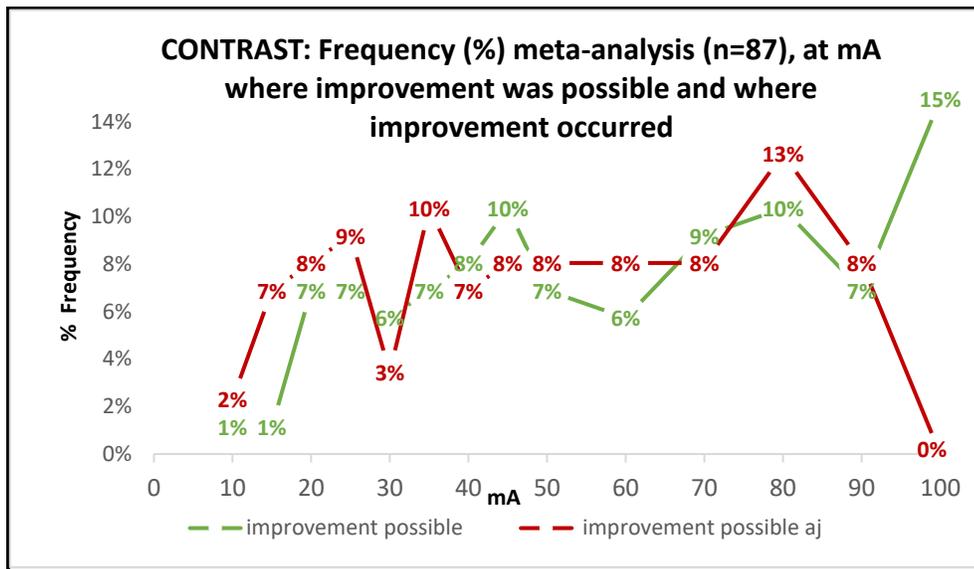


Figure 94. Contrast: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperages where improvement was possible, without attenuation jackets, were: 100mA (14%) and 70mA (12%), as seen in *figure 95*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 100mA (14%), 80mA (14%), 45mA (11%) and 40mA (11%), also seen in *figure 95*.

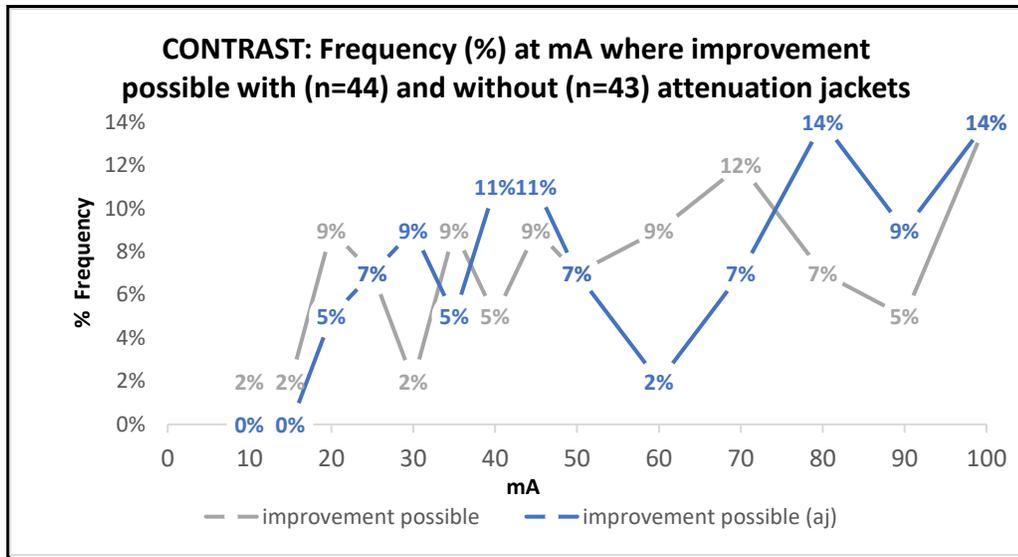


Figure 95. Contrast: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.

The most common amperages where improvement occurred were: 80mA (14%), 60mA (12%), as seen in figure 96. The most frequent amperages where improvements occurred, with attenuation jackets, were: 35mA (14%), 90mA (11%) and 80mA (11%), also seen in figure 96.

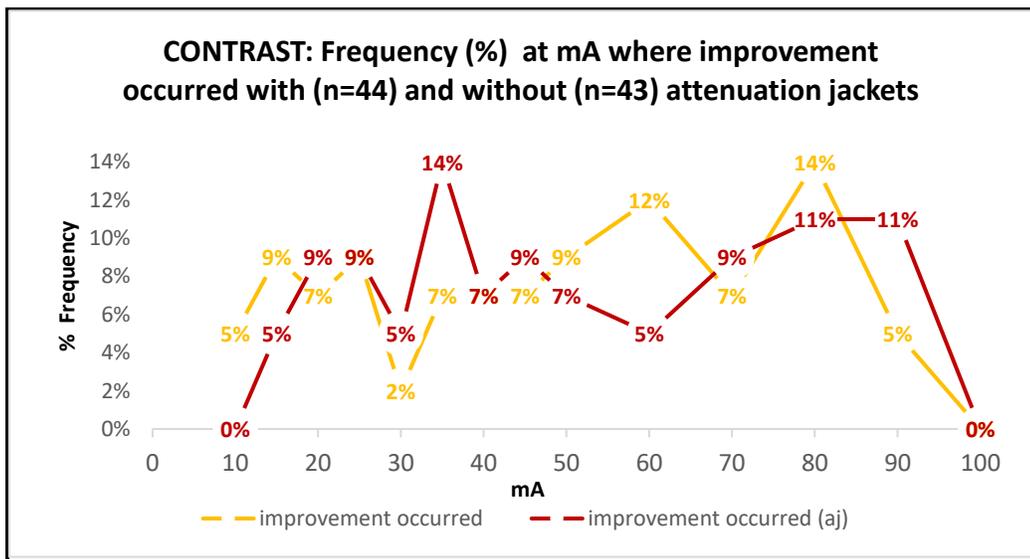


Figure 96. Contrast: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.

There is a very good relation of Contrast data series with overall average values when attenuation jackets have been used, with correlation values (r) ranging from 0.93 to 0.98. However, looking at the lung zones separately enabled peaks to be identified more readily. An example of this is demonstrated in *figure 97*, where there are consistent peaks demonstrated at 45mA and 90mA (which then have improved values at lower amperages).

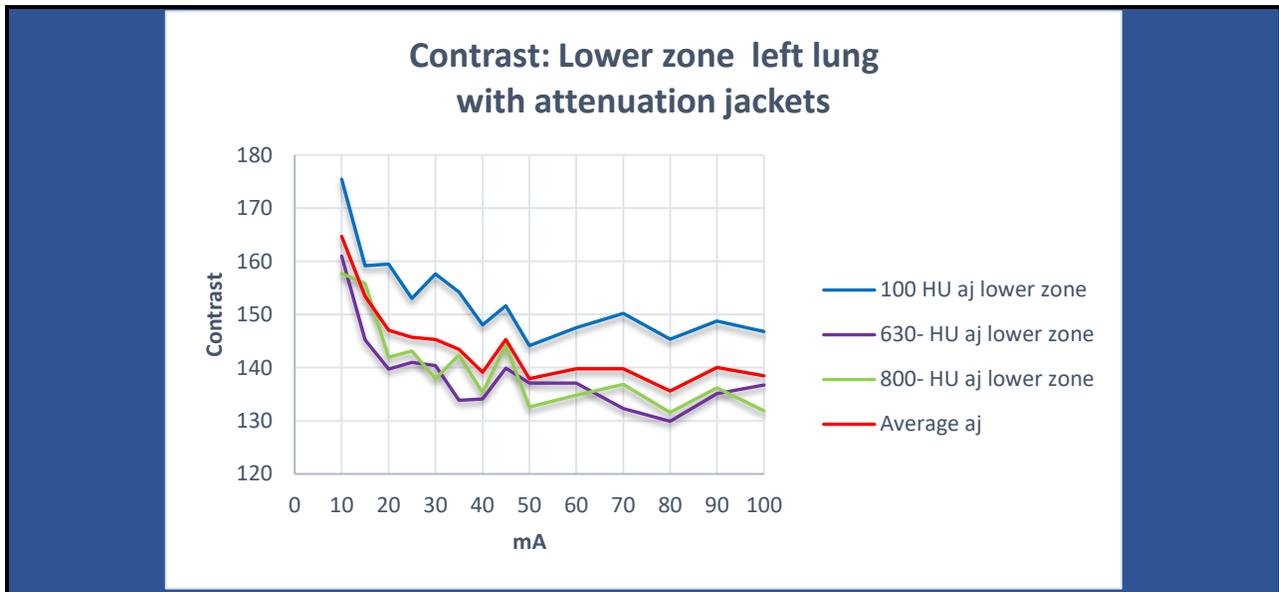


Figure 97. Contrast: Lower zone left lung (where HU refers to the simulated lesion density relating to each image series, and Average aj refers to the average of these data series only). Consistent peaks at 45mA and 90mA.

Contrast values of the simulated lesions were also examined. Values were affected more by lesion density (*figure 98*) than lesion size, although as the lesion density decreased the effect became more apparent with lesion size as seen in *table 28*.

Contrast: Δ 10mm-8mm	Without attenuation jackets	With attenuation jackets
100HU	-5.6%	12.9%
-630HU	31%	20.2%
-800HU	73.7%	76.5

Table 28. Contrast: Difference in values between 10mm and 8mm simulated lesions at varying density, expressed as a percentage.

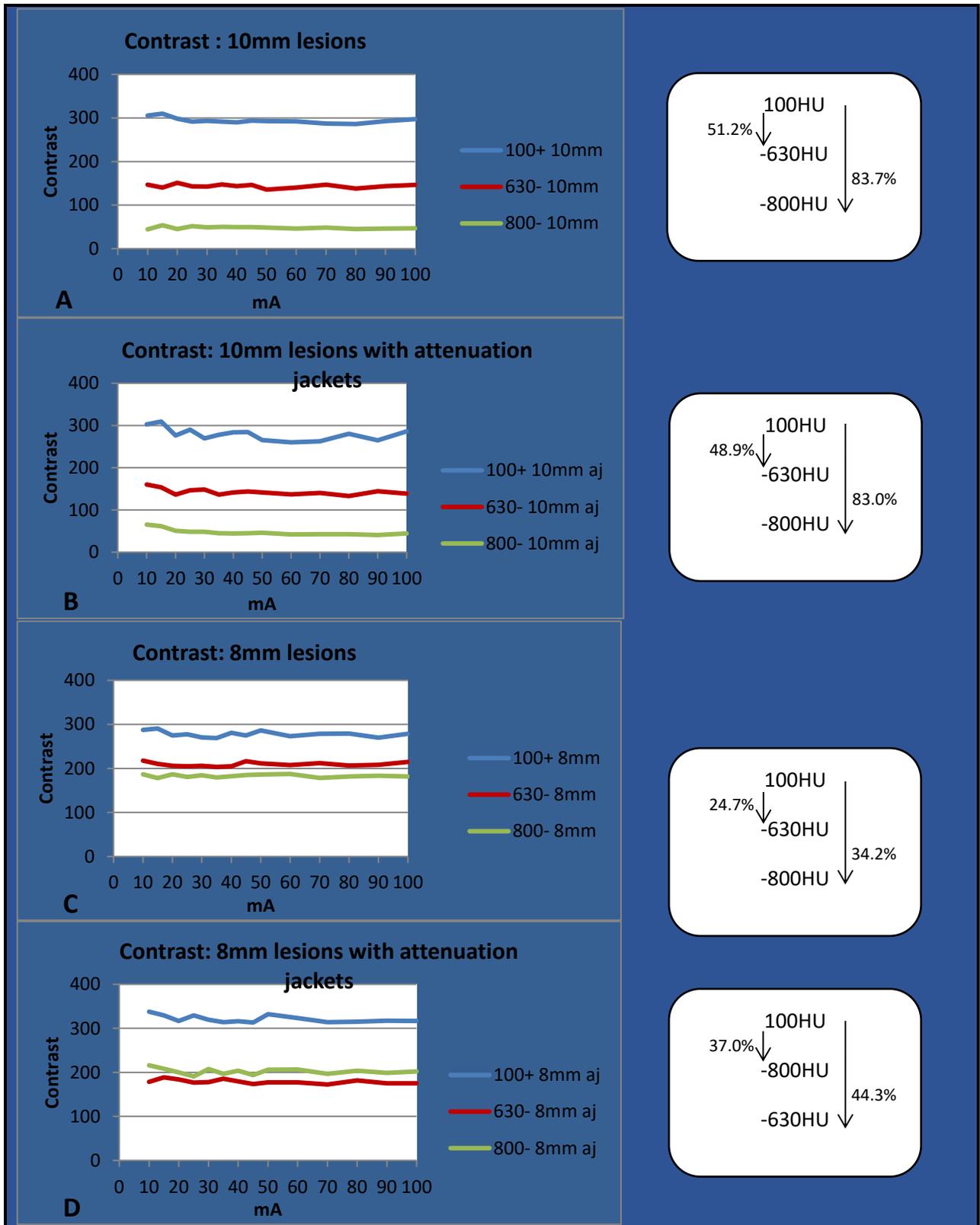


Figure 98. Contrast: Graphs for 10mm and 8mm simulated lesions showing differences in values relating to density (HU): (A) 10mm without attenuation jackets, (B) 10mm with attenuation jackets, (C) 8mm without attenuation jackets, (D) 8mm with attenuation jackets; NB. % differences are expressed in comparison to 100HU data series.

In summary, Contrast is an IQM produced during image texture analysis, which is a measure of the intensity contrast between a pixel and its neighbour. As an IQM it has some limitations, and values were affected by the region of the lung being analysed, particularly in the mid zone. This, combined with very little relation within Contrast data series when attenuation jackets were not used, means the use of this IQM is limited. In clinical practice, inter-patient variability in attenuating tissue would mean that this IQM would be inconsistent if used when establishing a measure of image quality.

When visually assessed, graphs of the Contrast data series demonstrated an increasing gradient in the lower amperage range when attenuation jackets had been used. The use of attenuation jackets had varying impact with the greatest impact demonstrated in the mid zone.

Analysis, looking at frequency of occurrence at amperages of where peaks, plateaus and troughs occurred, revealed the most frequent amperage where an improved contrast value then occurred at lower amperage, was 100mA. The most frequent amperage where an improvement occurred, was 80mA.

There was a very good relation of Contrast data series with overall average values when attenuation jackets had been used. However, examining the lung zones separately enabled peaks to be identified more readily, where an example of contrast values in the lower zone of the left lung demonstrated consistent peaks at 45mA and 90mA, where improvement was possible at lower amperages.

In addition to looking at lung fields of view, contrast values of the simulated lesions were examined. Values were affected more by lesion density than lesion size, although as the lesion density decreased the effect became more apparent with lesion size as well.

5.10.2 Correlation

Correlation is a measure of how correlated a pixel is to its neighbour (Mathworks, 1994-2020). This measure has a range from -1 to 1, depending on whether negatively or positively correlated (Mathworks, 1994-2020). The Resultant graphs for correlation are shown in *figure 99*.

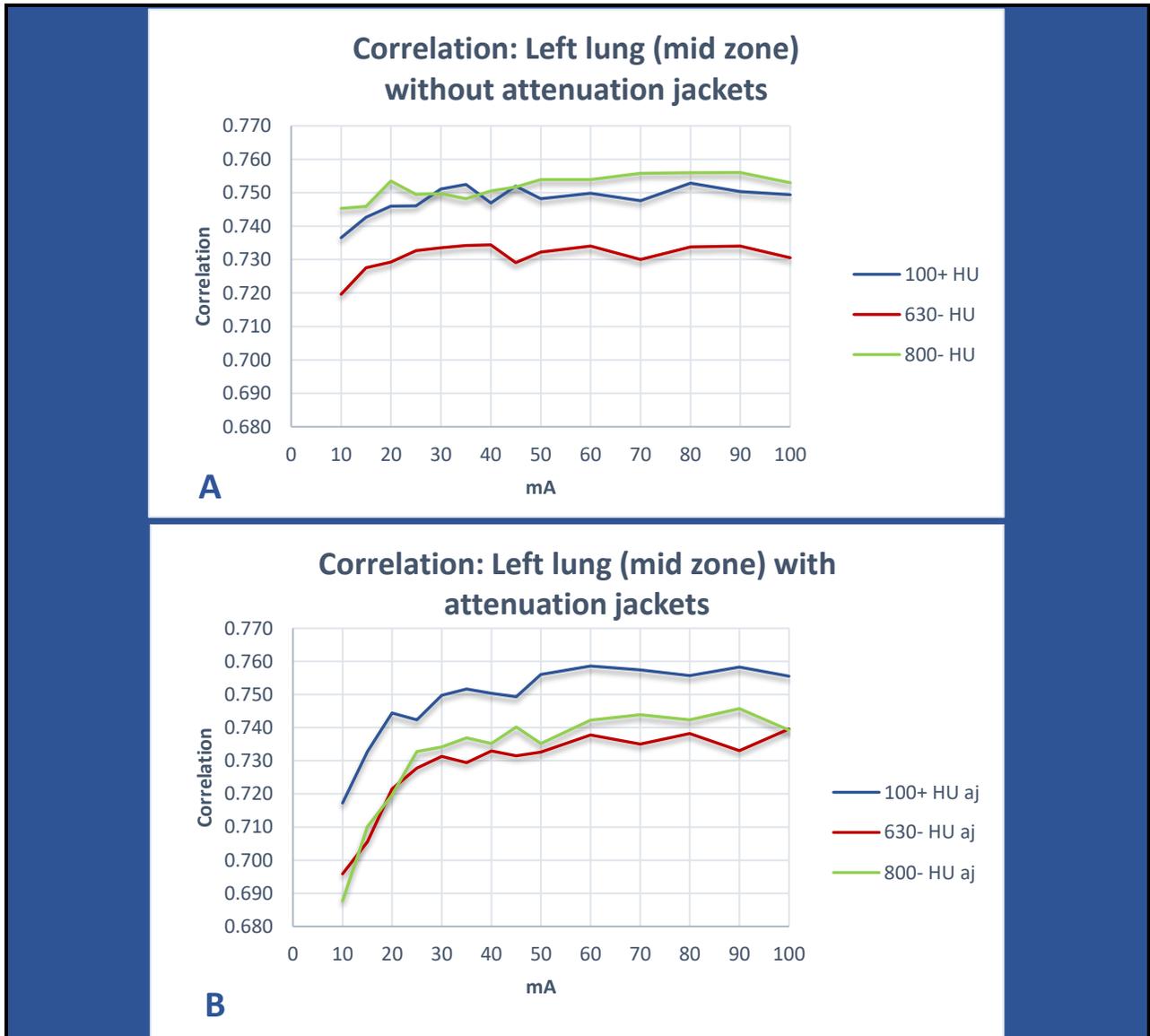


Figure 99. Correlation: Graphs for mid zone (where HU refers to the simulated lesion density relating to that image series): (A) without attenuation jackets, (B) with attenuation jackets.

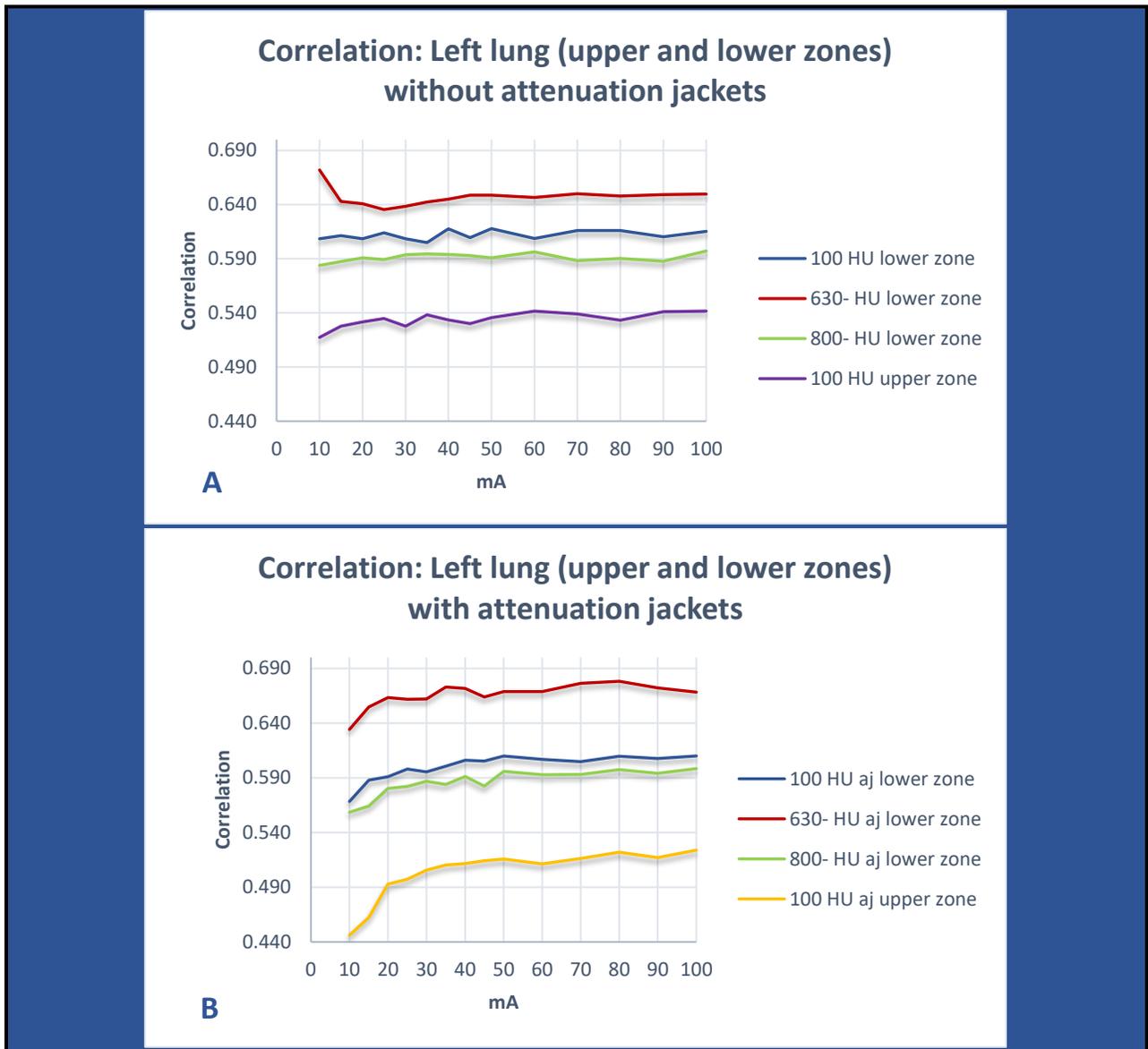


Figure 100. Correlation: Graphs for upper and lower zones (where HU refers to the simulated lesion density related to that image series): (A) without attenuation jackets, (B) with attenuation jackets.

There is little relation within the Correlation data series without attenuation jackets, particularly in the lower zone, with Pearson correlation values (r) ranging from 0.03 to 0.96 when compared to overall averaged data series. When visually assessed, however, the graphs in figures 99 and 100 demonstrated an increasing gradient in the lower amperage range when attenuation jackets had been used. In addition, the lung region analysed impacted on Correlation values, and the use

of attenuation jackets had varying impact with the greatest impact demonstrated in the upper zone (*figure 100*).

In graphs of individual Correlation data series acquired over upper zone, mid zone and lower zones, peaks and plateaus were identified, where Correlation values were similar or higher at lower amperage. Examples of this process for Correlation are seen in *figure 101*, where a red line (mA1) indicates a value that was then improved upon at a lower amperage indicated by a green line (mA2).

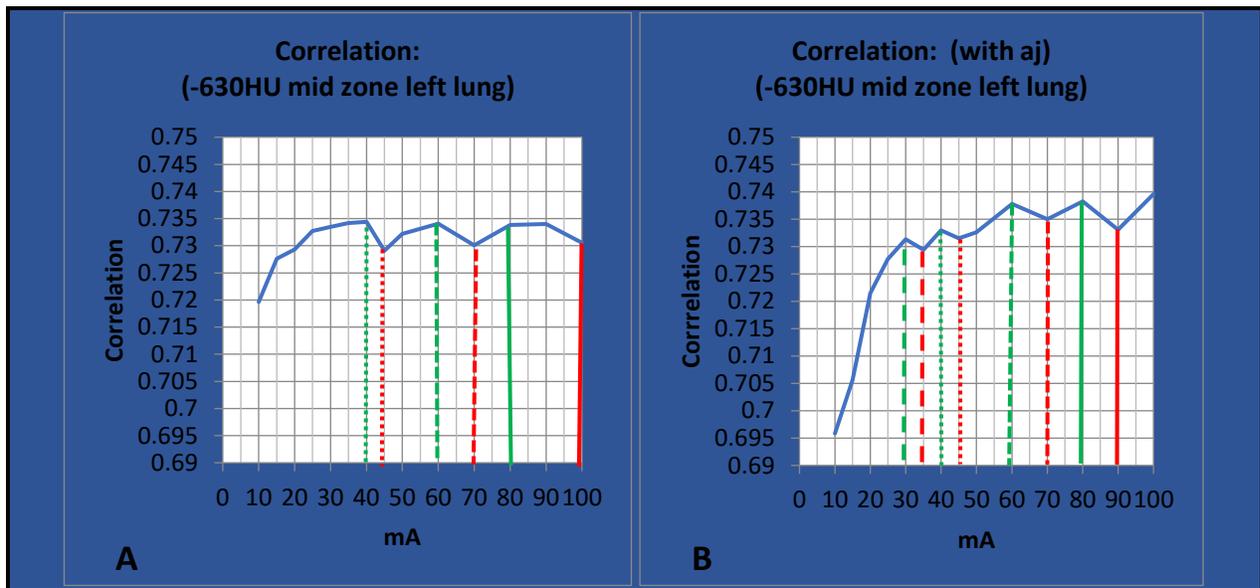


Figure 101. Correlation: Mid zone (where HU refers to the simulated lesion density related to that image series) (A) without attenuation jackets, with improvement (red line to green line) from: (A) 100mA to 80mA, 70mA to 60mA, 45mA to 40mA; (B) with attenuation jackets, with improvement (red line to green line) from 90mA to 80mA, 70mA to 60mA, 45mA to 40mA, 35mA to 30mA.

The potential usefulness of this information is seen in *table 29*, where Correlation values are similar or higher at reduced amperage. This is then translated into potential for reduced radiation exposure. In these examples, the potential for reduction in radiation exposure ranged from 9.5% to 20%. When looking at data from all acquisitions in this study, potential for reduction in radiation exposure ranged from 8.7% to 56.5%. However, given the lack of statistical correlation seen when attenuation jackets were not used and varying values between regions analysed, the usefulness of this information is limited for this IQM.

mA1	mA2 (lower)	Contrast(mA1)	Contrast(mA2)	Contrast % difference	CTDI _{vol} mA1	CTDI _{vol} mA2	CTDI _{vol} % difference
Without attenuation jackets							
100	80	0.7305	0.7338	0.45	4.60	3.70	19.57
70	60	0.7301	0.7341	0.54	3.25	2.80	13.85
45	40	0.7291	0.7344	0.72	2.10	1.90	9.52
With attenuation jackets							
90	80	0.7331	0.7382	0.69	4.15	3.70	10.84
70	60	0.735	0.7378	0.38	3.25	2.80	13.85
45	40	0.7315	0.733	0.20	2.10	1.90	9.52
35	30	0.7294	0.7313	0.26	1.65	1.40	15.15

Table 29. Correlation mid zone left lung for example in figure 101, where improved values at lower amperage potentially translate into reduced radiation exposure (CTDI_{vol}).

Although, it is still worth noting the frequency of occurrence at the amperages of where peaks, plateaus and troughs occur, for comparison with other IQMs. To identify any pattern and if this occurred at any particular amperage, frequency analyses were carried out.

Overall, the most frequent amperage, where an improvement was possible, was 100mA (17%), as seen in figure 102 Overall, the most frequent amperages where improvement occurred, were 80mA (13%) and 50 (11%), also seen in figure 102.

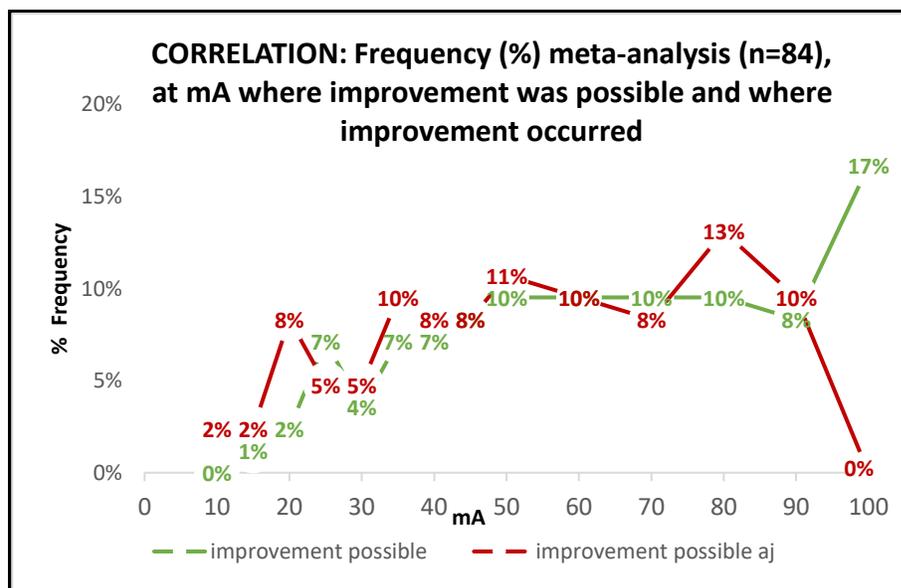


Figure 102. Correlation: Frequency meta-analysis of occurrence at mA where improvement was possible, and where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperages where improvement was possible, without attenuation jackets, were: 100mA (17%),

80mA (14%) and 45mA (12%), as seen in *figure 103*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 100mA (17%), 80mA (14%) and 45mA (12%), also seen in *figure 103*.

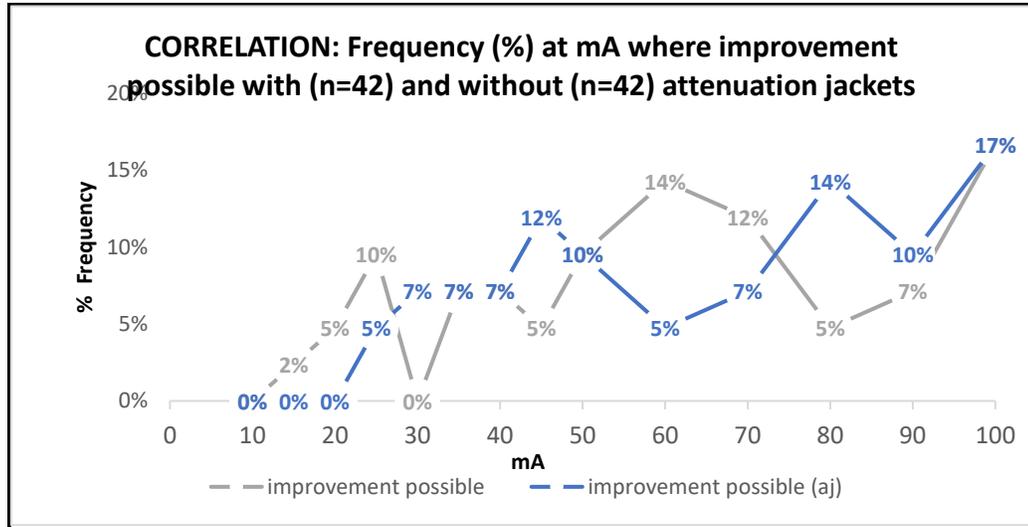


Figure 103. Correlation: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.

The most common amperages where improvement occurred, without attenuation jackets, were: 80mA (14%), 60mA (12%) and 50mA (12%), as seen in *figure 104*. The most frequent amperages where improvement occurred, with attenuation jackets, were: 90mA (14%), 80mA (12%) and 35mA (12%), also seen in *figure 104*.

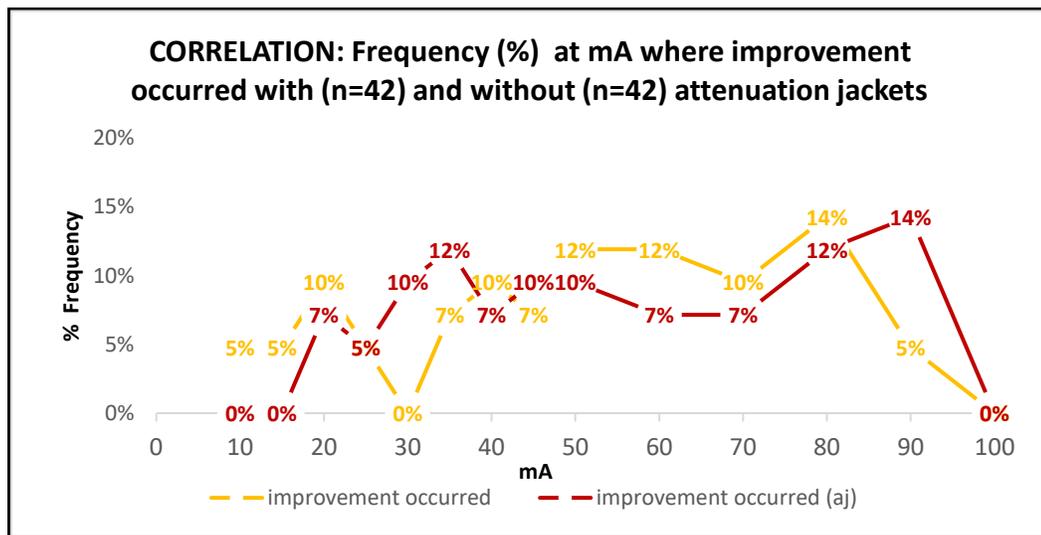


Figure 104. Correlation: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.

There is a very good relation of Correlation data series with the overall averaged data series, when attenuation jackets have been used, with Pearson correlation values (r) ranging from 0.978 to 0.995. However, looking at the lung zones separately enabled peaks to be identified more readily. An example of this is demonstrated in *figure 105*, where there are consistent peaks demonstrated at 45mA and 90mA at which Correlation values had improved.

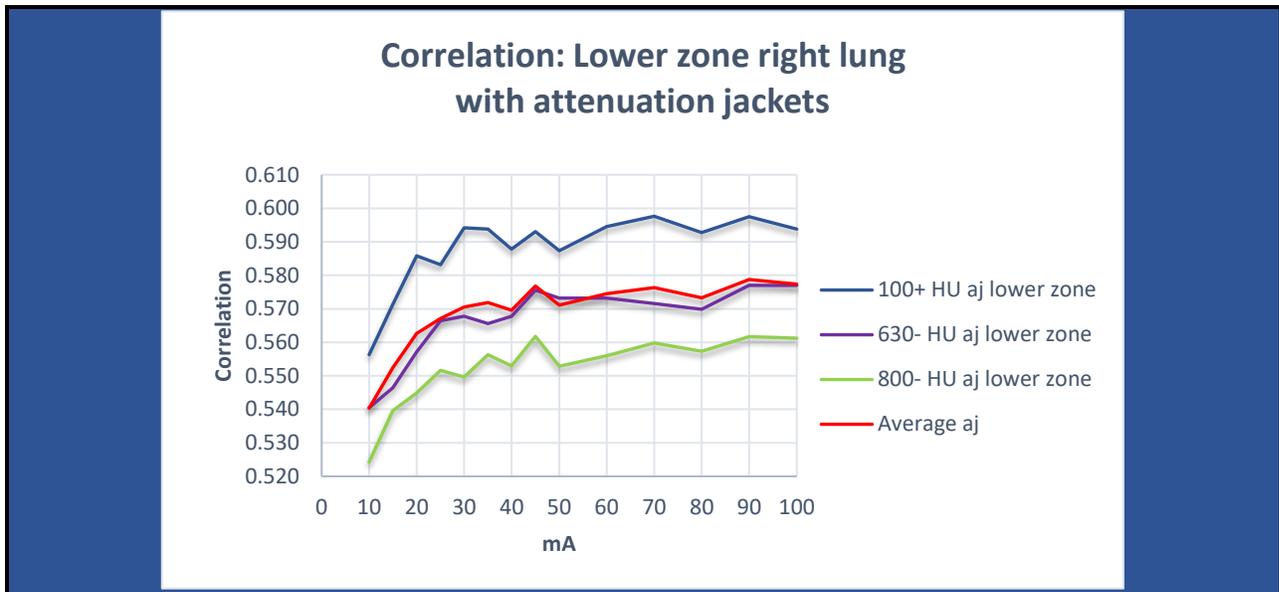


Figure 105. Correlation: Lower zone right lung (where HU refers to the simulated lesion density relating to each image series, and Average aj refers to the average of these data series only). Consistent peaks at 45mA and 90mA.

In addition to looking at Correlation values from fields of view over the lungs, Correlation values of the simulated lesions were examined. Values were affected more by lesion density (*figure 106*) than lesion size, particularly between 100HU and -800HU. As the lesion density decreased, however, the effect became more apparent with lesion size combined with the use of attenuation jackets, as seen in *table 30*.

Correlation: Δ 10mm-8mm	Without attenuation jackets	With attenuation jackets
100HU	15.3%	15.1%
-630HU	15.2%	17.2%
-800HU	16.8%	19.0%

Table 30. Correlation: Difference in values between 10mm and 8mm simulated lesions at varying density, expressed as a percentage.

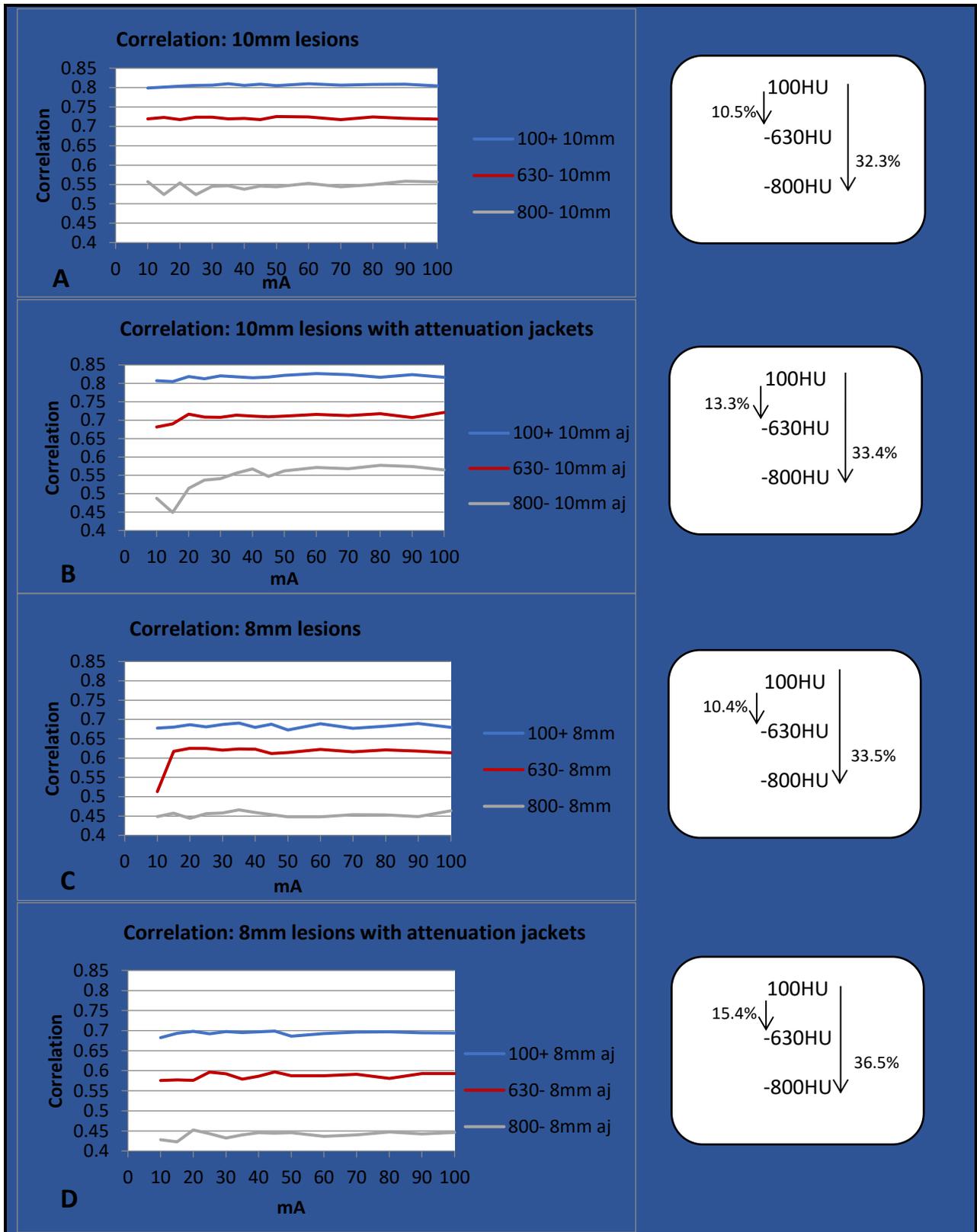


Figure 106. Correlation: Graphs for 10mm and 8mm simulated lesions showing differences in values relating to density (HU): (A) 10mm without attenuation jackets, (B) 10mm with attenuation jackets, (C) 8mm without attenuation jackets, (D) 8mm with attenuation jackets; NB. percentage differences are expressed in comparison to 100HU data series.

In summary, Correlation is an IQM produced during image texture analysis, which is a measure of how correlated a pixel is to its neighbour (Mathworks, 1994-2020). As an IQM it has some limitations, and values were affected by the region of the lung being analysed, particularly in the lower zone without attenuation jackets and upper zone with attenuation jacket. This, combined with very little relation within Correlation data series when attenuation jackets were not used, means the use of this IQM is limited and had little predictive value for radiation dose reduction. The effect of increasing amperage on Correlation values at the lower amperage range was apparent when attenuation jackets were used.

Frequency analysis of occurrence for amperages where peaks, plateaus and troughs occurred, revealed the most frequent amperage where an improvement was possible, was 100mA. The most frequent amperage where an improved Correlation value occurred was 80mA.

There was a very good relation of Correlation data series with overall averaged data series, when attenuation jackets had been used. However, examining the lung zones separately enabled peaks to be identified more readily. The example of Correlation values in the lower zone of the right lung demonstrated consistent peaks at 45mA and 90mA, where the improvement in values occurred.

In addition to examining Correlation values from fields of view over the lungs, Correlation values of the simulated lesions were examined. Values were affected more by lesion density than lesion size, particularly between 100HU and -800HU. As the lesion density decreased, however, the effect of lesion size became more apparent when combined with the use of attenuation jackets.

5.10.3 Energy

Energy (or uniformity) provides values that are the sum of squared elements in the GLCM, and has a range from 0 to 1 where the Energy value equals 1 for a constant image (Mathworks, 1994-2020). It was expected that the values of this IQM would increase with increasing amperages, as the images became less noisy. Resultant graphs for energy are shown in *figure 107*.

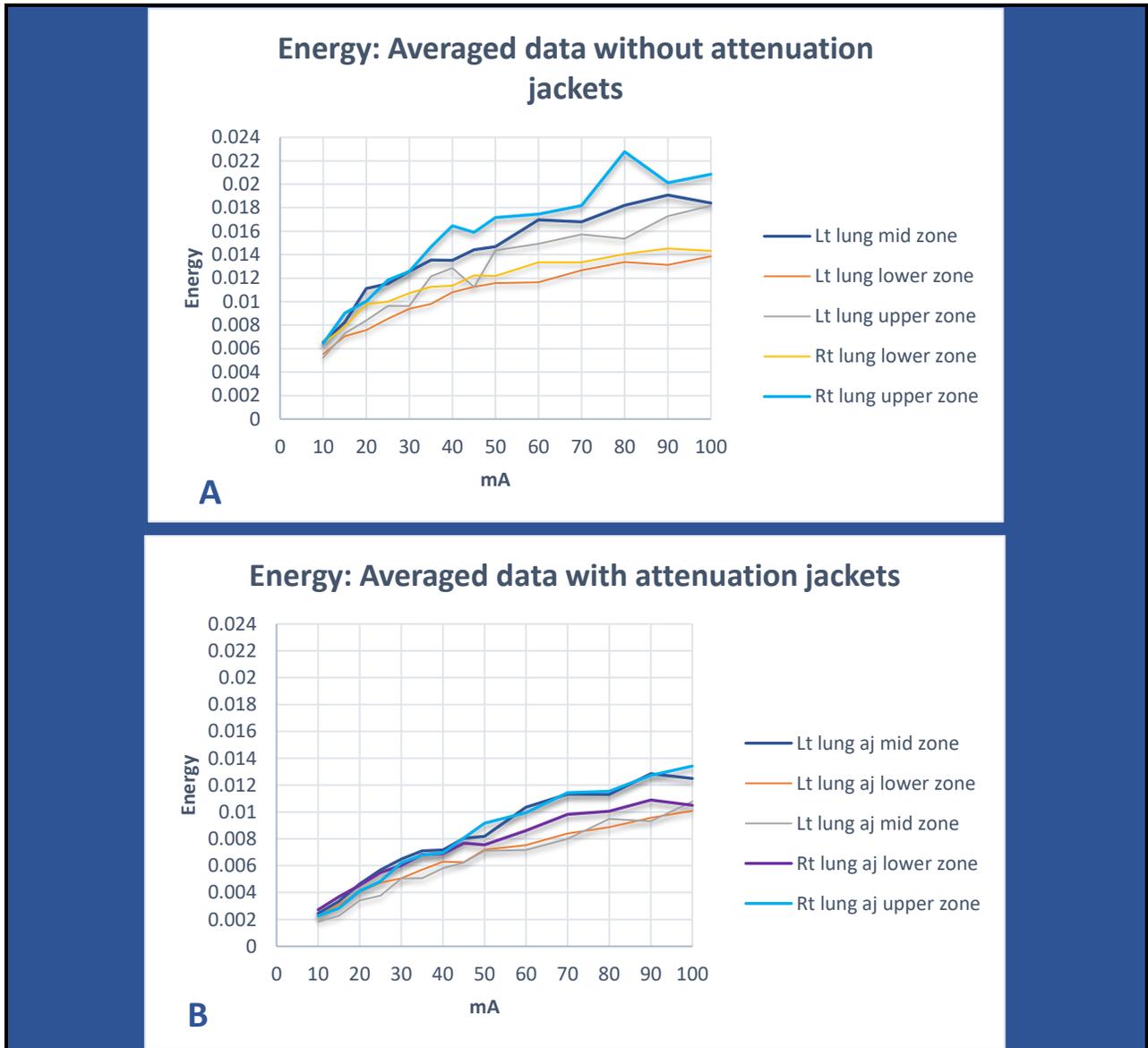


Figure 107. Energy: Averaged data graphed (A) without attenuation jackets, (B) with attenuation jackets.

Good correlation occurred both with and without attenuation jackets, when compared to Energy values in the averaged data series for each, although correlation was less without attenuation jackets. The results, as seen in *table 31*, show Pearson correlation values (r) for the data series, that ranged from 0.982 to 0.993 without attenuation jackets, and correlation values (r) that ranged from 0.991 to 0.998 when attenuation jackets were used. Consequently, using averaged data series was reasonable when examining these Energy results.

Energy: (r)	Without attenuation jackets	With attenuation jackets
Averaged data series	1	1
Lt lung mid zone	0.991467	0.995593
Lt lung lower zone	0.993252	0.997164
Lt lung upper zone	0.982462	0.990959
Rt lung lower zone	0.990098	0.99509
Rt lung upper zone	0.987582	0.998184

Table 31. Correlation of Energy graph values with averaged data series (Pearson correlation coefficient values r).

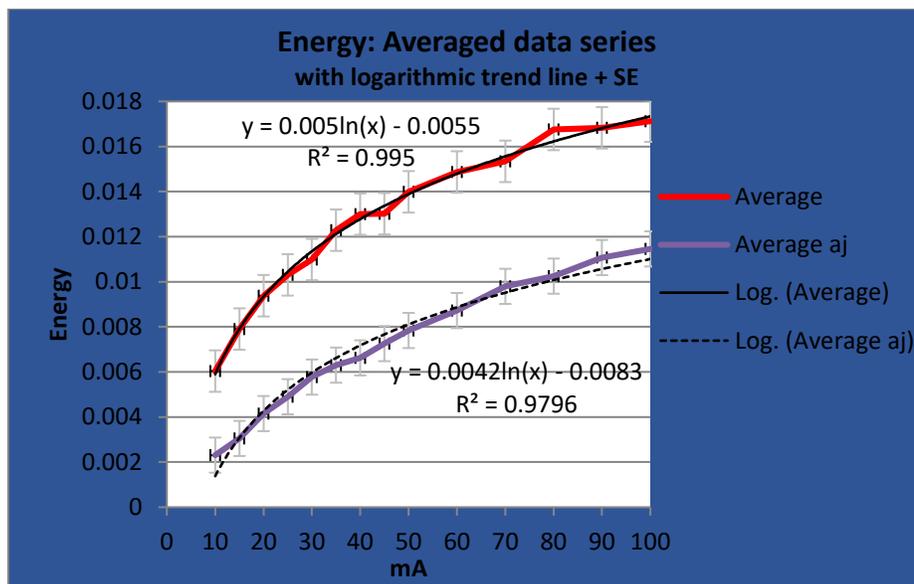


Figure 108. Energy: Averaged data graphs with and without attenuation jackets showing logarithmic trend lines and standard error bars (SE).

Energy values of averaged data sets, both with and without attenuation jackets, are graphed in *figure 108*, where logarithmic trend lines have been applied. Regression analysis produced an R^2 value that decreased slightly when attenuation jackets were used ($R^2=0.995$ decreased to $R^2=0.9796$) and demonstrated good correlation of Energy values with a logarithmic trend. When

attenuation jackets were used, the slope of the graph decreased by 16%, and the y-intercept decreased by 34%. When comparing results with and without attenuation jackets, the Energy results were statistically significant ($p < 0.05$). Results of the T-test, assuming equal variances, are detailed in *table 32* with the T-stat ($T = 4.93$) higher than the two-tail critical value ($T = 2.06$).

t-Test: Two-Sample Assuming Equal Variances	
alpha	0.05
df	25
t Stat	4.925621964
P(T<=t) two-tail	4.51837E-05
t Critical two-tail	2.059538553

Table 32. T-test for averaged Energy data series with and without attenuation jackets.

In graphs of individual data series acquired over upper zone, mid zone and lower zones, peaks and plateaus were identified, where energy values were similar or higher at lower amperage. Examples of this process for energy are seen in *figure 109*, where a red line (mA1) indicates a value that was then improved upon at a lower amperage indicated by a green line (mA2). In this instance when looking at overall averages for energy, there were no peaks or plateaus identified when attenuation jackets were used.

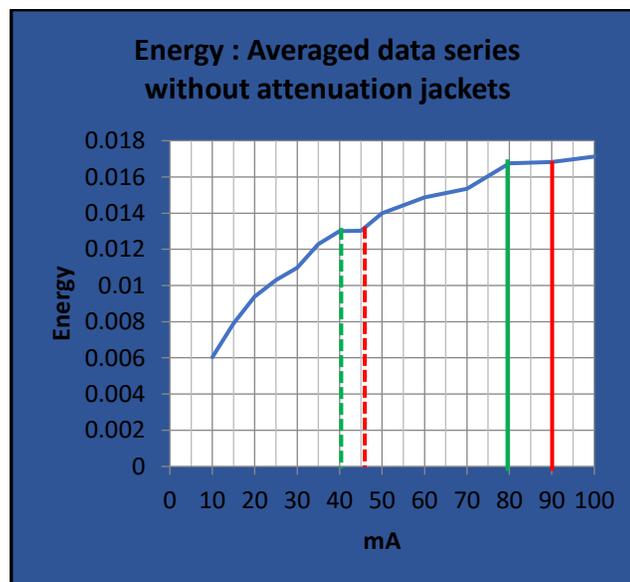


Figure 109. Energy: Averaged data series, with improvement (red line to green line) from: 90mA to 80mA, 45mA to 40mA (without attenuation jackets).

The potential usefulness of this information is seen in *table 33*, where Energy values are similar or higher at reduced amperage (mA2). This is then translated into potential for reduced radiation exposure. In this example, the potential for reduction in radiation exposure ranged from 9.5% to 10.8%. When looking at Energy data from all acquisitions in this study, potential for reduction in radiation exposure ranged from 8.7% to 21.7%.

mA1	mA2 (lower)	Energy(mA1)	Energy(mA2)	Energy % difference	CTDI _{vol} mA1	CTDI _{vol} mA2	CTDI _{vol} % difference
Without attenuation jackets							
90	80	0.016829012	0.016755952	0.43	4.15	3.7	10.8
45	40	0.013016349	0.012998392	0.14	2.1	1.9	9.5

Table 33. Energy: for averaged data series where improved values at lower amperage translate into potential reduction in radiation exposure (CTDI_{vol}).

To identify any pattern and if this occurred at any particular amperage, frequency analyses were carried out. Overall, the most frequent amperages, where improvement was possible, were: 100mA (19%) and 50mA (16%), as seen in *figure 110*. Overall, the most frequent amperages where improvement occurred were: 90mA (19%), 45mA (12%) and 35mA (12%), also seen in *figure 110*.

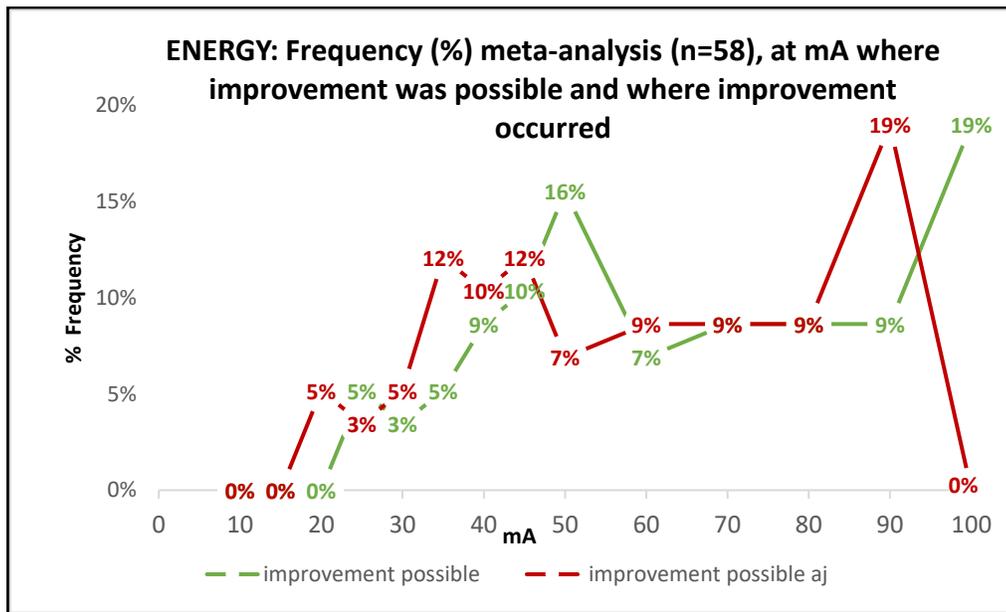


Figure 110. Energy: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperages, where improvement was possible, without attenuation jackets, were: 100mA (19%), 50mA (16%) and 70mA (13%), as seen in *figure 111*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 100mA (19%), 80mA (15%), 50mA (15%), 60mA (11%), 45mA (11%) and 40mA (11%), also seen in *figure 111*.

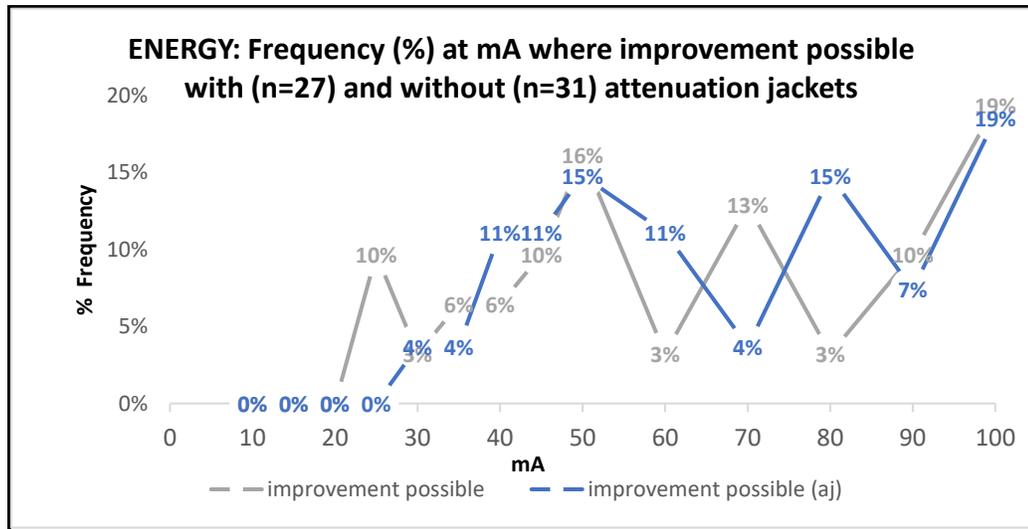


Figure 111. Energy: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.

The most common amperages where improvement occurred, without attenuation jackets, were: 90mA (19%), 60mA (13%) and 45mA (13%), as seen in *figure 112*. The most frequent amperages where improvement occurred, with attenuation jackets, were: 90mA (19%), 70mA (15%) and 35mA (15%) with 11% for 50mA, 45mA and 40mA, also seen in *figure 112*.

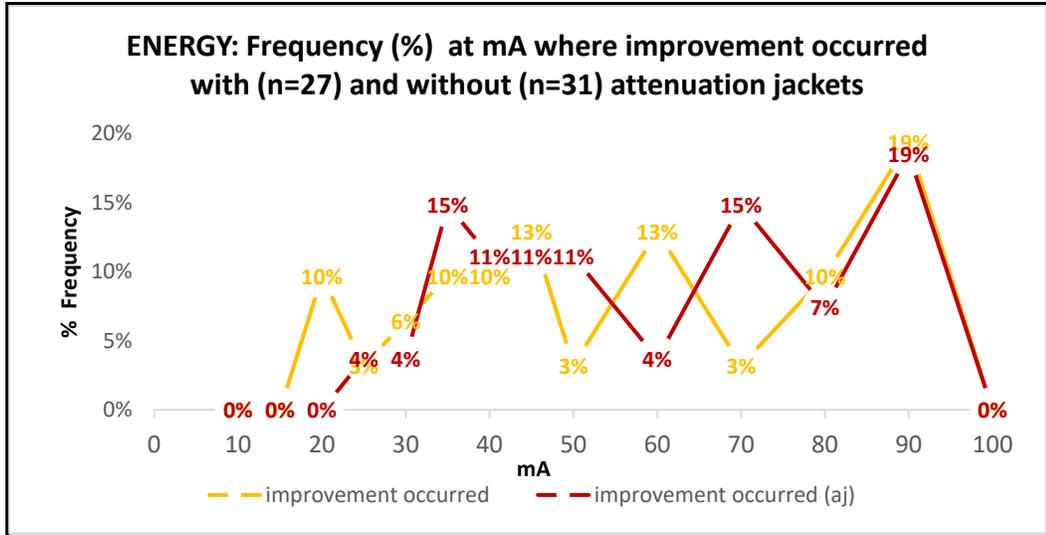


Figure 112. Energy: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.

In addition to looking at lung fields of view, Energy values of the simulated lesions were examined. Energy was affected by both lesion density (*figure 113*) and lesion size. The effect of lesion density was particularly noticeable with the 10mm simulated lesions, with Energy values improving as lesion density decreased. As the lesion density decreased the effect of lesion size became more apparent, as did the use of attenuation jackets, as seen in *table 34*.

Energy: Δ 10mm-8mm	Without attenuation jackets	With attenuation jackets
100HU	1.6%	10.4%
-630HU	15.1%	26.5%
-800HU	75.6%	73.0%

Table 34. Difference in Energy values between 10mm and 8mm simulated lesions at varying density, expressed as a percentage.

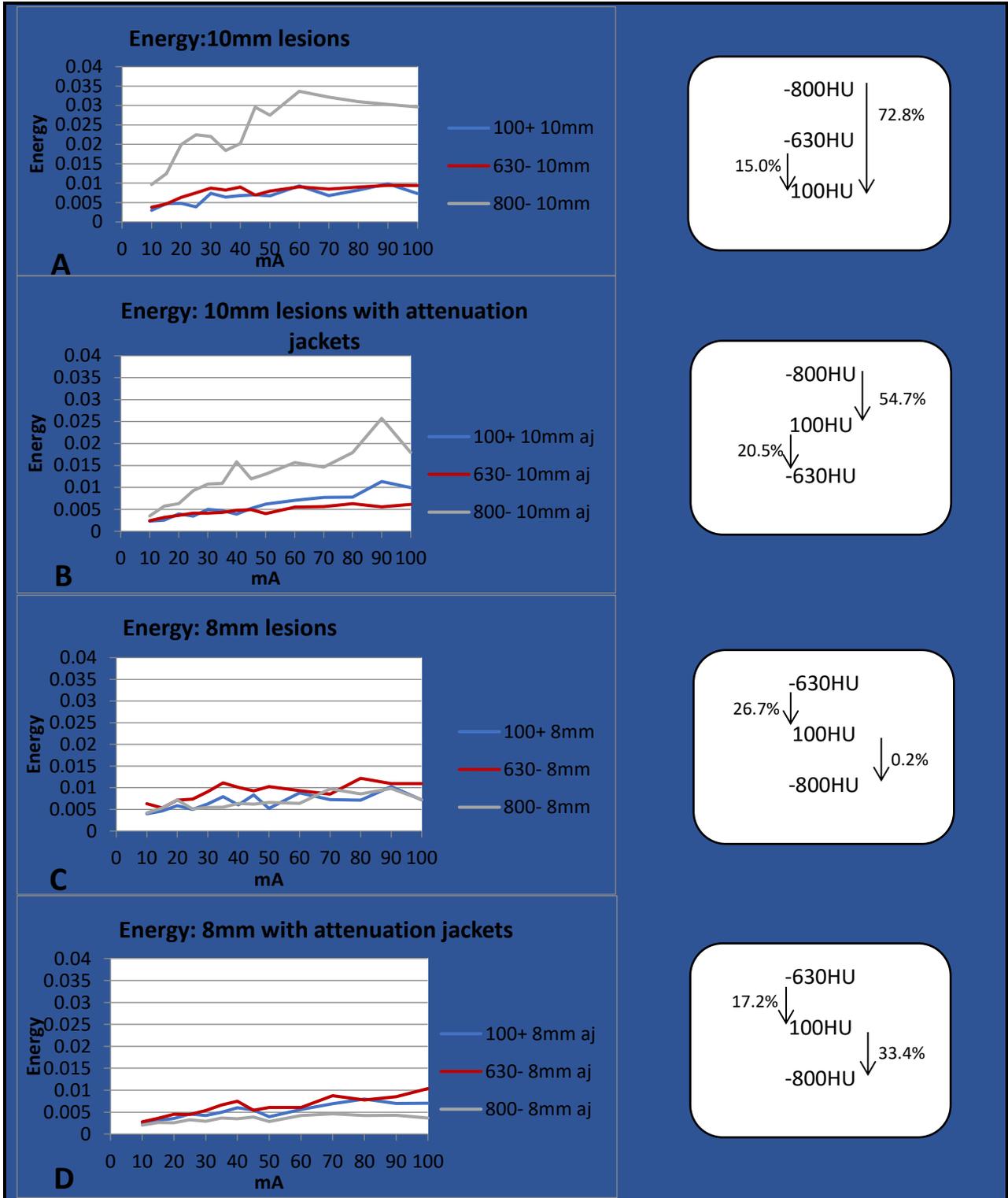


Figure 113. Energy: Graphs for 10mm and 8mm simulated lesions showing differences in values relating to density (HU): (A) 10mm without attenuation jackets, (B) 10mm with attenuation jackets, (C) 8mm without attenuation jackets, (D) 8mm with attenuation jackets; NB. percentage differences are expressed in comparison to 100HU data series.

In summary, Energy (or uniformity) provides values that are the sum of squared elements in the GLCM, where a value of 1 is that of a constant image (Mathworks, 1994-2020). This IQM has provided useful results, with very good correlation between data series for each set, both with and without attenuation jackets.

Regression analysis produced R^2 values which demonstrated good relation of Energy values with a logarithmic trend. In addition, the difference between Energy values obtained with and without the use of attenuation jackets was statistically significant ($p < 0.01$).

Analysis of peaks and plateaus of the individual data series highlighted amperages at 100mA and 50mA (both with and without attenuation jackets) where the Energy then improved at lower amperage. These improvements were then commonly seen at 90mA and 45mA (both with and without attenuation jackets). This therefore identified a pattern where there is potential for use of lower amperages while maintaining image quality.

In addition to looking at lung fields of view, Energy values of the simulated lesions were examined. Energy was affected by both lesion density and lesion size, with Energy values improving as lesion density decreased. As the lesion density decreased the effect of lesion size became more apparent, as did the use of attenuation jackets.

5.10.4 Homogeneity

Homogeneity (looking at spatial variations) returns a value that measures the closeness of how the elements in the GLCM are distributed in relation to the GLCM diagonal, and has values that range from 0 to 1, where a diagonal GLCM has a value of 1 (Mathworks, 1994-2020). Resultant graphs for homogeneity are shown in *figure 114*.

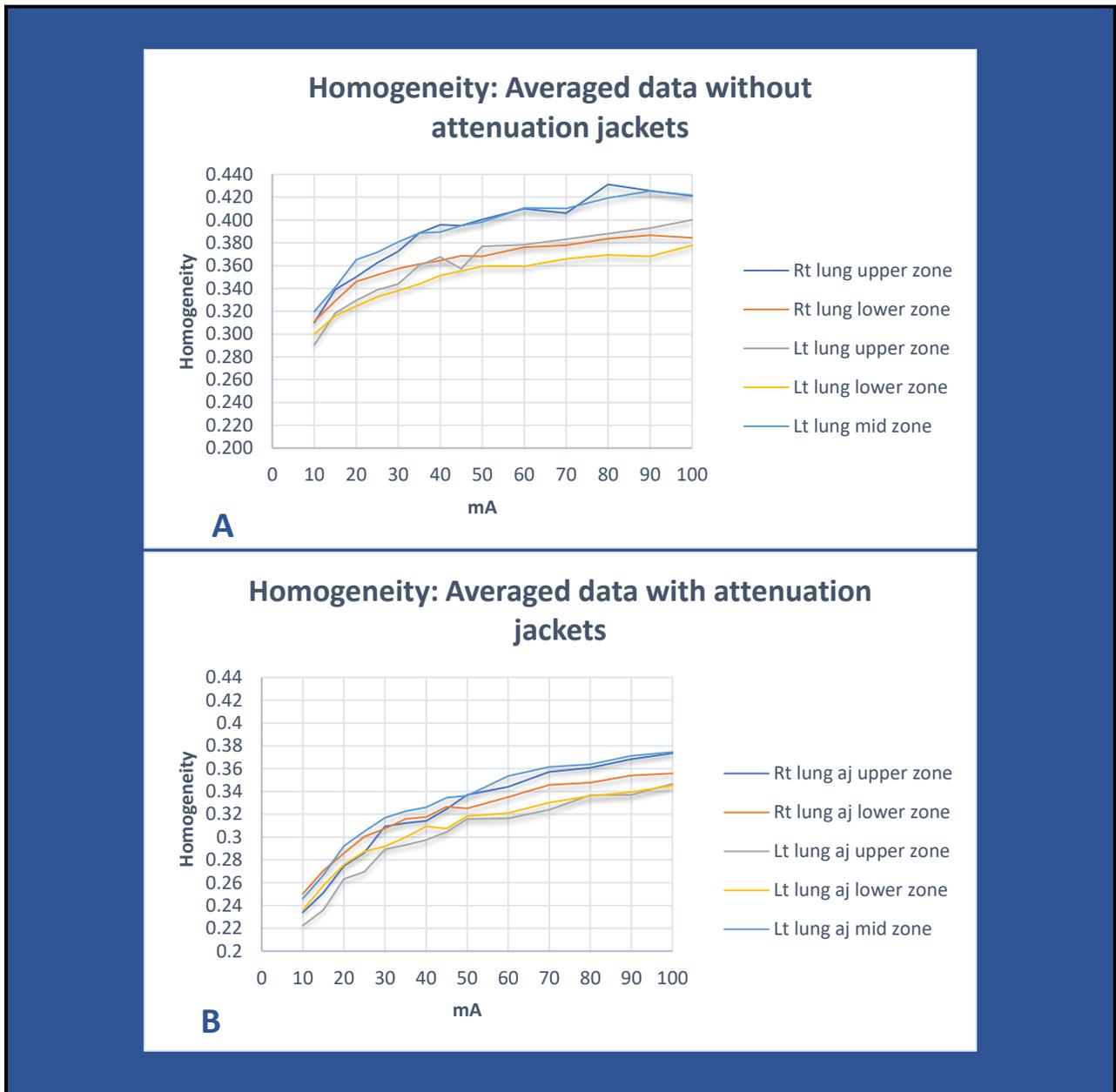


Figure 114. Homogeneity: Averaged data series graphed (A) without attenuation jackets, (B) with attenuation jackets.

Very good correlation occurred both with and without attenuation jackets, when compared to Homogeneity values in the averaged data series for each. The results, as seen in *table 35*, show Pearson correlation values (r) for these data series, that ranged from 0.993 to 0.996 without attenuation jackets, and correlation values (r) that ranged from 0.997 to 0.998 when attenuation jackets were used. Consequently, using averaged data series was reasonable when examining these Homogeneity results.

Homogeneity: (r)	Without attenuation jackets	With attenuation jackets
Averaged data series	1	1
Rt lung upper zone	0.994601	0.998066
Rt lung lower zone	0.994088	0.997554
Lt lung upper zone	0.993095	0.99714
Lt lung lower zone	0.992516	0.996893
Lt lung mid zone	0.995722	0.998357

Table 35. Correlation of Homogeneity values with averaged data series (Pearson correlation coefficient values r).

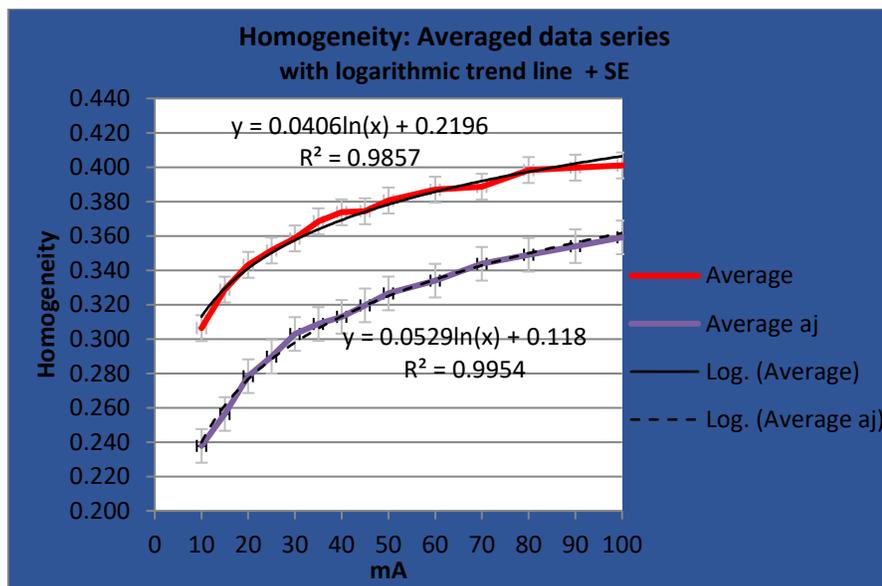


Figure 115. Homogeneity: Graphs of averaged data series, with and without attenuation jackets, showing logarithmic trend lines and standard error bars (SE).

Homogeneity values of averaged data series, both with and without attenuation jackets, are graphed in *figure 115*, where logarithmic trend lines have been applied. Regression analysis produced high R^2 values and demonstrated good correlation of Homogeneity values with a logarithmic trend. When attenuation jackets were used, the slope of the graph increased by 23%,

and the y-intercept decreased by 46%. Comparison between the two data series show the difference to be statistically significant ($p < 0.05$). Results of the T-test, assuming equal variances, are detailed in *table 36* with the T-stat ($T = 4.55$) higher than the two-tail critical value ($T = 2.06$).

t-Test: Two-Sample Assuming Equal Variances	
alpha	0.05
df	24
t Stat	4.550484
P(T<=t) two-tail	0.00013
t Critical two-tail	2.063898562

Table 36. T-test for averaged homogeneity data series with and without attenuation jackets.

In graphs of individual data series acquired over upper zone, mid zone and lower zones, peaks and plateaus were identified, where Homogeneity values were similar or higher, at lower amperage. Examples of this process for Homogeneity are seen in *figure 116*, where a red line (mA1) indicates a value that improved (or was similar) at a lower amperage indicated by a green line (mA2).

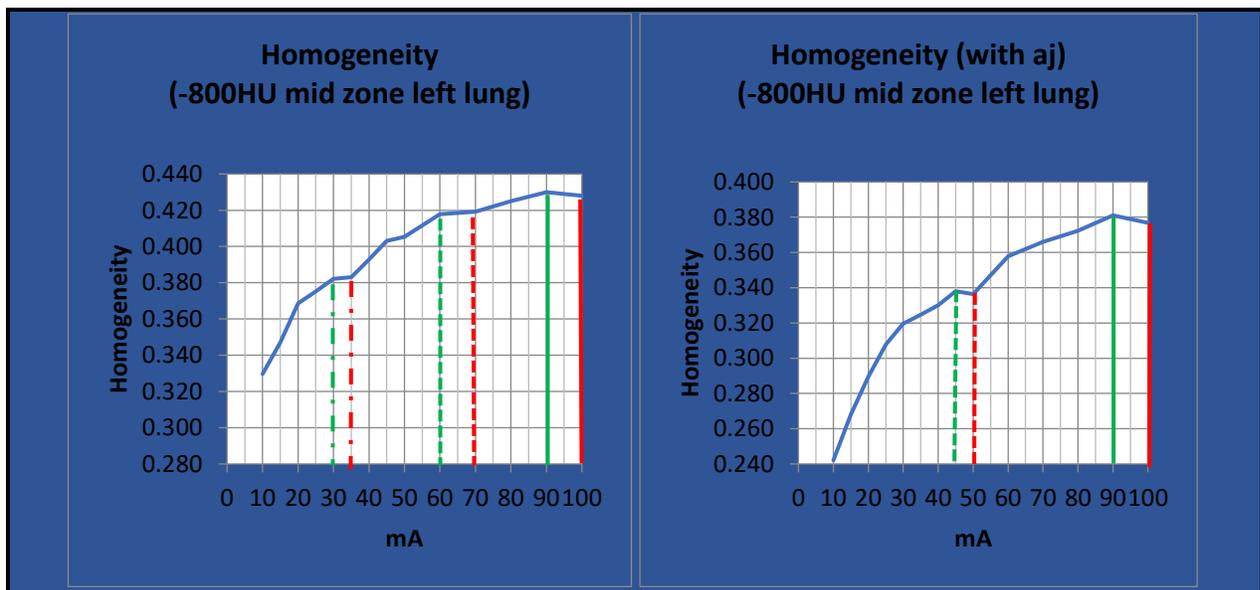


Figure 116. Homogeneity mid zone (where HU refers to the simulated lesion density related to that image series) (A) without attenuation jackets, with improvement (red line to green line) from: 100mA to 90mA, 70mA to 60mA, 35mA to 30mA; (B) with attenuation jackets, with improvement (red line to green line) from 100mA to 90mA, 50mA to 45mA.

The potential usefulness of this information is seen in *table 37*, where Homogeneity values are similar or higher at reduced amperage. This is then translated into potential for reduced radiation exposure. In this example, the potential for reduction in radiation exposure ranged from 8.7% to 15.2%. When looking at Homogeneity data from all acquisitions in this study, potential for reduction in radiation exposure ranged from 8.7% to 32.5%.

mA1	mA2 (lower)	Homogeneity (mA1)	Homogeneity (mA2)	Homogeneity % difference	CTDI _{vol} mA1	CTDI _{vol} mA2	CTDI _{vol} % difference
Without attenuation jackets							
100	90	0.428	0.430	0.47	4.6	4.15	9.8
70	60	0.419	0.418	-0.24	3.25	2.8	13.8
35	30	0.383	0.382	-0.26	1.65	1.4	15.2
With attenuation jackets							
100	90	0.377	0.381	1.06	4.6	4.15	9.8
50	45	0.336	0.338	0.60	2.3	2.1	8.7

Table 37. Homogeneity: for averaged data series where improved values at lower amperage translate into potential reduction in radiation exposure (CTDI_{vol}).

To identify any pattern and if this occurred at any particular amperage, frequency analyses were carried out. Overall, the most frequent amperages, where improvement was possible were: 100mA (19%), 60mA (15%) and 45mA (14%), as seen *figure 117*. Overall, the most frequent amperages where improvement occurred were: 90mA (17%), 50mA (15%), 60mA (12%) and 40mA (12%), also seen in *figure 117*.

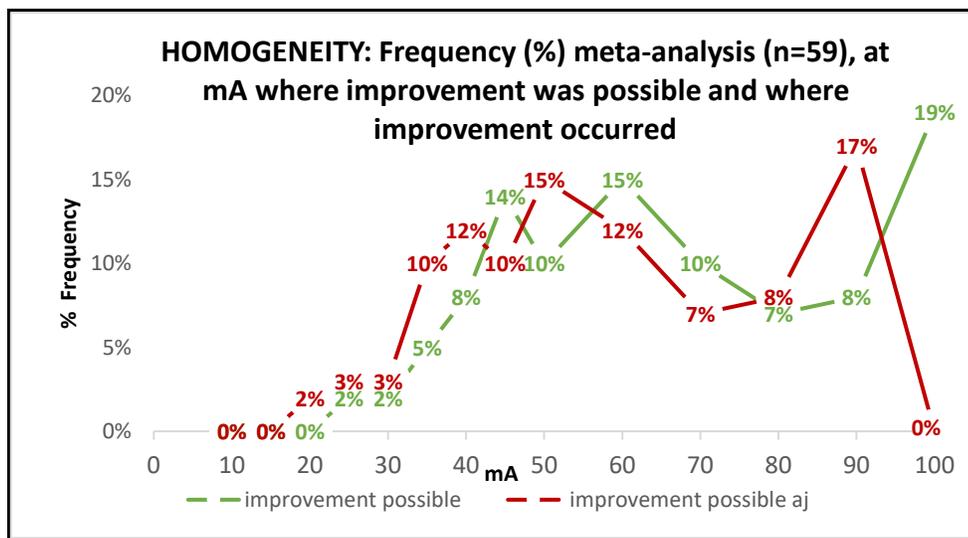


Figure 117. Homogeneity: Frequency meta-analysis, of occurrence at mA where improvement was possible, and at mA where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperages, where improvement was possible, without attenuation jackets, were: 100mA (20%), 70mA (17%), 60mA (14%) and 45mA (11%), as seen in *figure 118*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 100mA, 60mA and 45mA at 17%, with 80mA, 50mA, and 40mA at 13%, also seen in *118*.

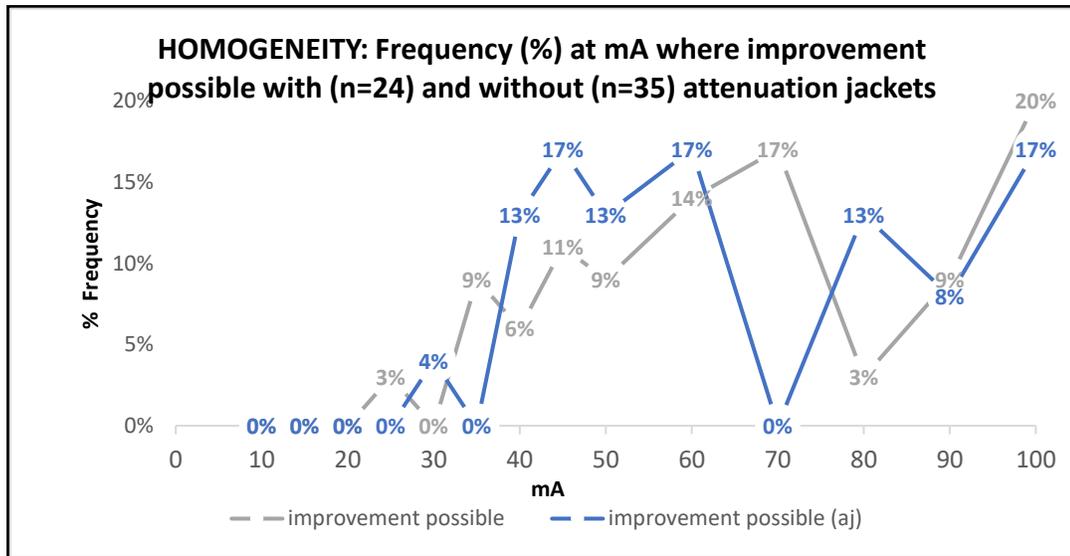


Figure 118. Homogeneity: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.

The most common amperages where improvement occurred, without attenuation jackets, were: 60mA (20%), 90mA (17%) and 50mA (14%), as seen in *figure 119*. The most frequent amperages where improvement occurred, with attenuation jackets, were: 90mA, 50mA and 40mA at 17%, with 70mA, 45mA and 35mA at 13%, also seen in *figure 119*.

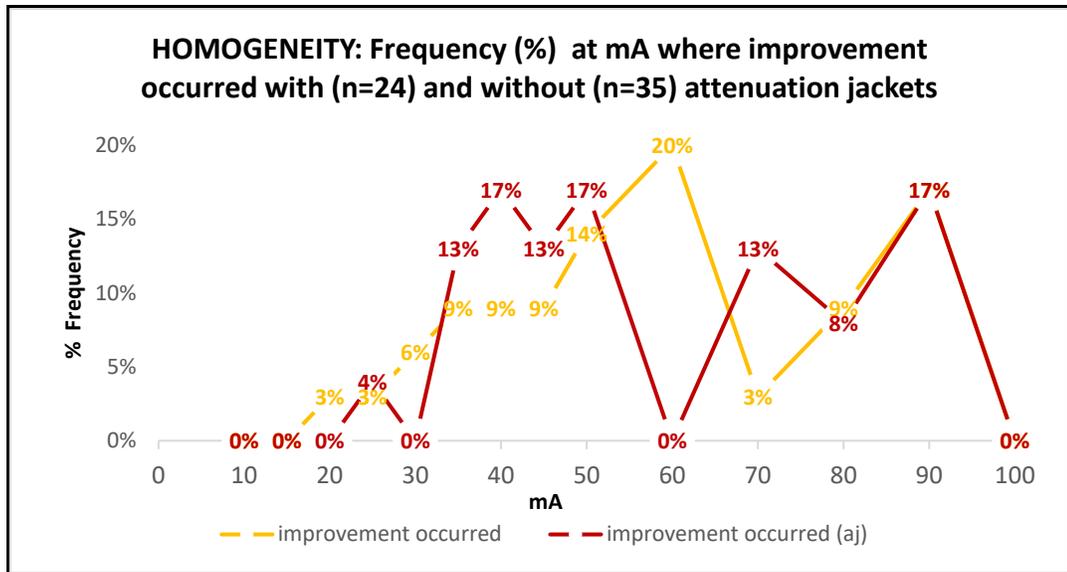


Figure 119. Homogeneity: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.

In addition to looking at lung fields of view, homogeneity values of the simulated lesions were examined. Homogeneity was affected by both lesion density (*figure 120*) and lesion size. The effect of lesion density was particularly noticeable with the 10mm simulated lesions, with homogeneity values improving as lesion density decreased (*graph A, figure 120*). As lesion size decreased so did the Homogeneity values, but these were then less affected by changes in density. The use of attenuation jackets had the effect of decreasing Homogeneity values, however there was little effect on differences in Homogeneity between the 10mm and 8mm simulated lesions at each density, as seen in *table 38*.

Homogeneity: Δ 10mm-8mm	Without attenuation jackets	With attenuation jackets
100HU	7.8%	9.0%
-630HU	9.3%	4.0%
-800HU	37.6%	36.2%

Table 38. Difference in Homogeneity values between 10mm and 8mm simulated lesions at varying density, expressed as a percentage.

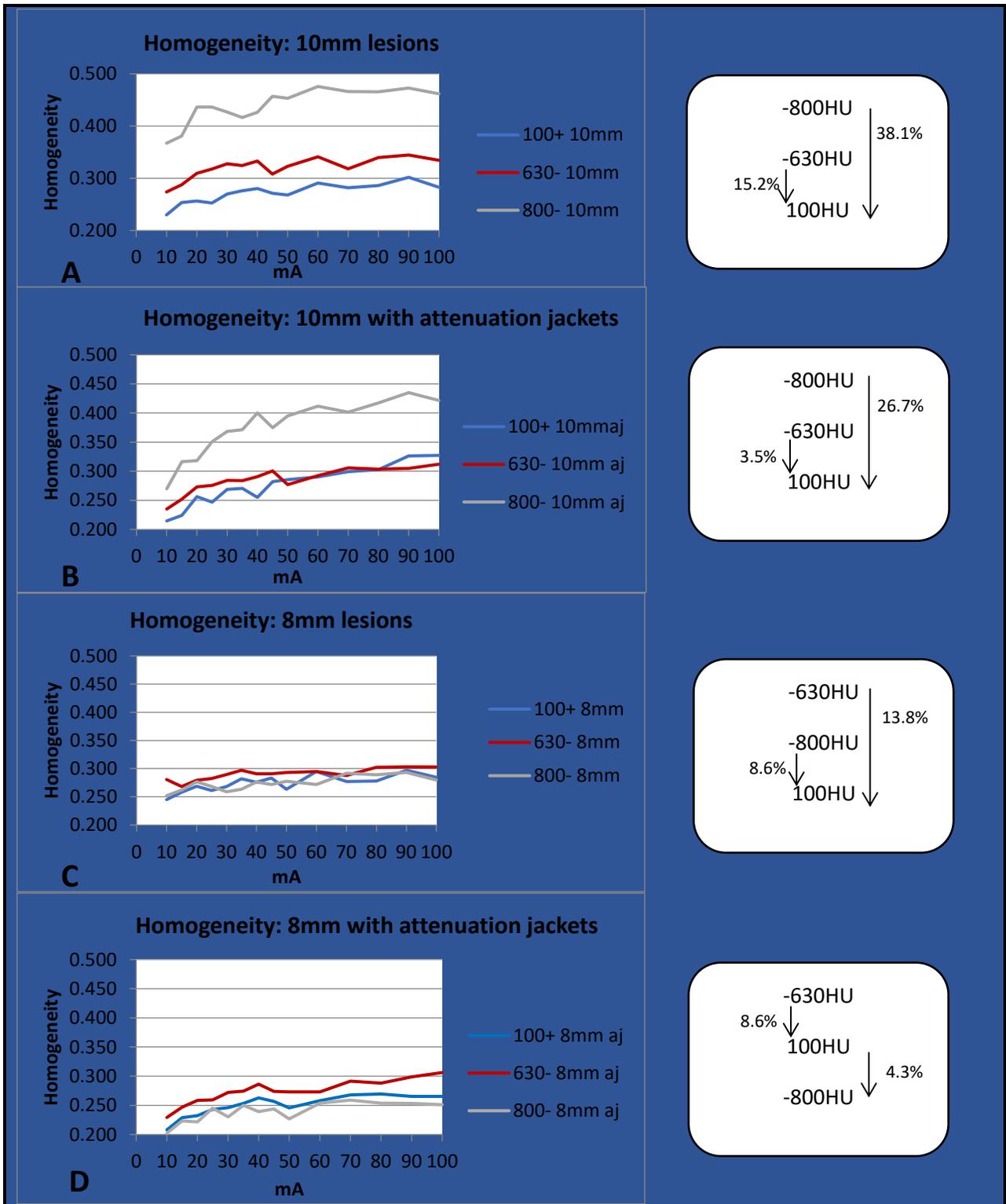


Figure 120. Homogeneity: Graphs for 10mm and 8mm simulated lesions showing differences in values relating to density (HU): (A) 10mm without attenuation jackets, (B) 10mm with attenuation jackets, (C) 8mm without attenuation jackets, (D) 8mm with attenuation jackets; NB. percentage differences are expressed in comparison to 100HU data series.

In summary, Homogeneity (looking at spatial variations) returns a value that measures the closeness of how the elements in the GLCM are distributed in relation to the GLCM diagonal, where a diagonal GLCM has a value of 1. This IQM has also provided useful results, with very good relation between data series acquired both with and without attenuation jackets.

Regression analysis produced R^2 values which demonstrated good correlation of Homogeneity values with a logarithmic trend. In addition, the difference between Homogeneity values obtained with and without the use of attenuation jackets was statistically significant ($p < 0.05$). The use of attenuation jackets impacted on Homogeneity by lowering the values produced.

Analysis of peaks and plateaus of individual data series highlighted amperages at 100mA, 60mA and 45mA (both with and without attenuation jackets) where the Homogeneity then improved at lower amperage. These improvements were then commonly seen at 90mA, 50mA and 40mA (both with and without attenuation jackets). An emerging pattern was identified with some potential for use of lower amperages while maintaining image quality.

In addition to looking at lung fields of view, Homogeneity values of the simulated lesions were examined. Homogeneity was affected by both lesion density and lesion size, with Homogeneity values improving as lesion density decreased. As lesion size decreased so did the Homogeneity values, but these were then less affected by changes in density. The use of attenuation jackets had the effect of decreasing Homogeneity values for simulated lesions.

5.10.5 Entropy

Entropy is measure of variability or randomness where a constant image has a value of 0 (Mathworks, 1994-2020). It was expected that the values of this IQM would decrease with increasing amperages, as the images became less noisy and therefore less disordered. Resultant graphs for Entropy are shown in *figure 121*.

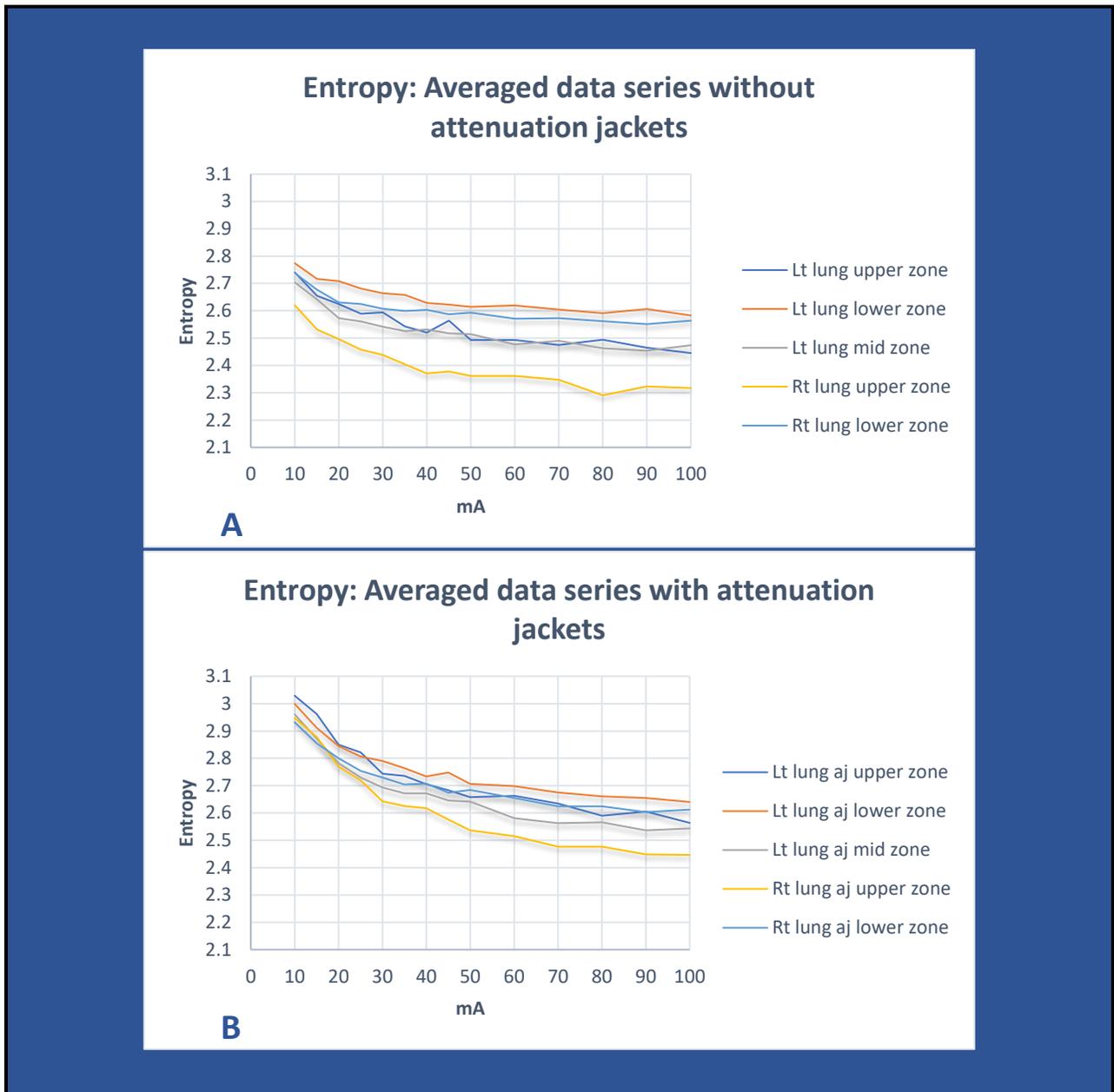


Figure 121. Entropy: Averaged data series graphed (A) without attenuation jackets, (B) with attenuation jackets.

Very good correlation occurred both with and without attenuation jackets, when compared to Entropy values in the averaged data series for each. The results, as seen in *table 39*, show Pearson correlation values (r) for these data series, that ranged from 0.982 to 0.993 without attenuation jackets, and correlation values (r) that ranged from 0.995 to 0.998 when attenuation jackets were used. Consequently, using averaged data series was reasonable when examining these Entropy results.

Entropy: (r)	Without attenuation jackets	With attenuation jackets
Averaged data series	1	1
Lt lung mid zone	0.986769	0.996772
Lt lung lower zone	0.986941	0.995724
Lt lung upper zone	0.982026	0.995184
Rt lung lower zone	0.982581	0.996581
Rt lung upper zone	0.993253	0.997768

Table 39. Correlation of Entropy values with averaged data series (Pearson correlation coefficient values r).

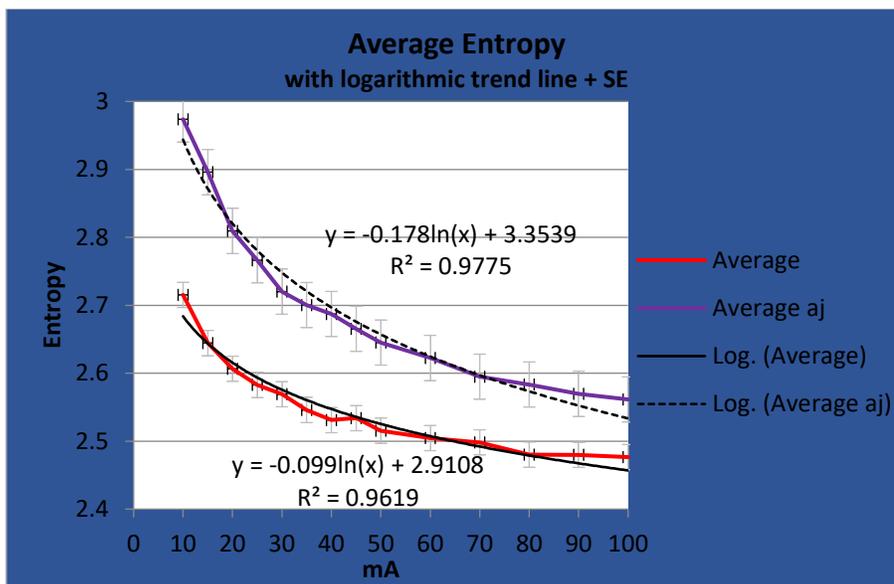


Figure 122. Entropy: Averaged data series graphed with and without attenuation Jackets, showing logarithmic trend lines and standard error bars (SE).

Entropy values for averaged data sets, both with and without attenuation jackets, are graphed in *figure 122*, where logarithmic trend lines have been applied. Regression analysis produced high R^2 values and demonstrated good correlation of Entropy values with a logarithmic trend. When attenuation jackets were used, the slope of the graph increased by 44%, and the y-intercept

increased by 13%. When comparing results with and without attenuation jackets, the difference in Entropy values was statistically significant ($p < 0.05$). Results of the T-test, assuming equal variances, are detailed in *table 40* with the T-stat ($T = -3.96$) higher than the two-tail critical value ($T = 2.06$).

t-Test: Two-Sample Assuming Equal Variances	
alpha	0.05
df	26
t Stat	-3.96008
P(T<=t) two-tail	0.000519
t Critical two-tail	2.05529

Table 40. T-test for averaged Entropy data series with and without attenuation jackets.

In graphs of individual data series acquired over upper zone, mid zone and lower zones, peaks and plateaus were identified, where Entropy values were similar or higher at lower amperage. Examples of this process for Entropy are seen in *figure 123*, where a red line (mA1) indicates a value that was then improved upon at a lower amperage indicated by a green line (mA2).

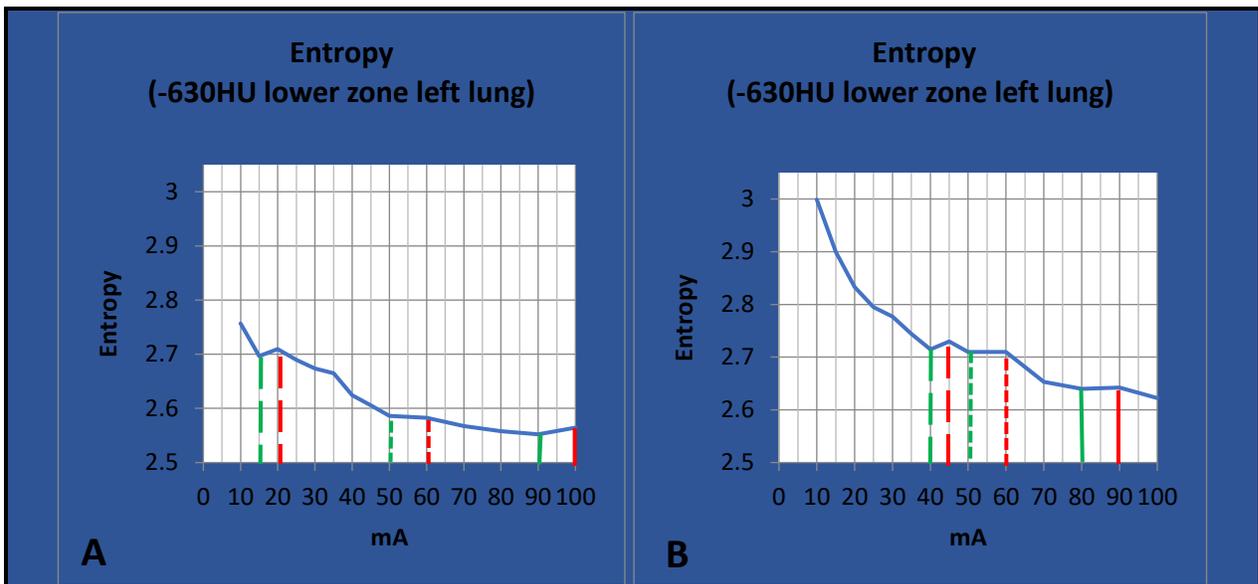


Figure 123. Entropy lower zone (where HU refers to the simulated lesion density related to that image series) (A) without attenuation jackets, with improvement (red line to green line) from: 100mA to 90mA, 60mA to 50mA, 20mA to 15mA; (B) with attenuation jackets, with improvement (red line to green line) from 90mA to 80mA, 60mA to 50mA, 45mA to 40mA.

The potential usefulness of this information is seen in *table 41*, where entropy values are similar or higher at reduced amperage (mA2). This is then translated into potential for reduced radiation exposure. In this example, the potential for reduction in radiation exposure ranged from 9.5% to 22.2%. When looking at entropy data from all acquisitions in this study, potential for reduction in radiation exposure ranged from 8.7% to 45.5%.

mA1	mA2 (lower)	Entropy (mA1)	Entropy (mA2)	Entropy % difference	CTDI _{vol} mA1	CTDI _{vol} mA2	CTDI _{vol} % difference
<i>Without attenuation jackets</i>							
100	90	2.564	2.552	0.468	4.6	4.15	9.78
60	50	2.582	2.586	-0.155	2.8	2.3	17.86
20	15	2.709	2.697	0.443	0.9	0.7	22.22
<i>With attenuation jackets</i>							
90	80	2.642	2.64	0.076	4.15	3.7	10.84
60	50	2.71	2.71	0.000	2.8	2.3	17.86
45	40	2.73	2.715	0.549	2.1	1.9	9.52

Table 41. Entropy: for averaged data series where improved values at lower amperage translate into potential reduction in radiation exposure (CTDI_{vol}).

To identify any pattern and if this occurred at any particular amperage, frequency analyses were carried out. Overall, the most frequent amperages where improvement was possible were: 100mA (15%), 70mA (13%), 45mA (13%), 90mA (11%) and 60mA (11%), as seen in *figure 124*. Overall, the most frequent amperages where improvement occurred were: 90mA (15%), 60mA (13%), 40mA (13%), 80mA (11%) and 50mA (11%), also seen in *figure 124*.

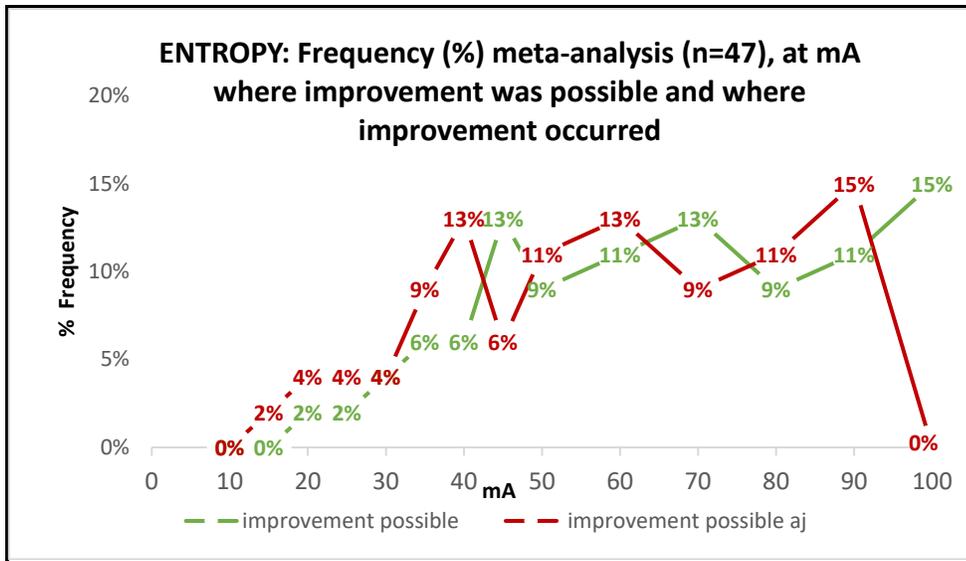


Figure 124. Entropy: Frequency meta-analysis, of occurrence at mA where improvement possible, and at mA where improvement occurred.

Looking separately at data acquired with and without attenuation jackets, the most frequent amperages where improvement was possible, without attenuation jackets, were: 100mA (17%), 70mA (17%), 60mA (13%) and 45mA (13%), as seen in *figure 125*. The most frequent amperages where improvement was possible, with attenuation jackets, were: 90mA (17%), 100mA (13%), 80mA (13%) and 45mA (13%), also seen in *figure 125*.

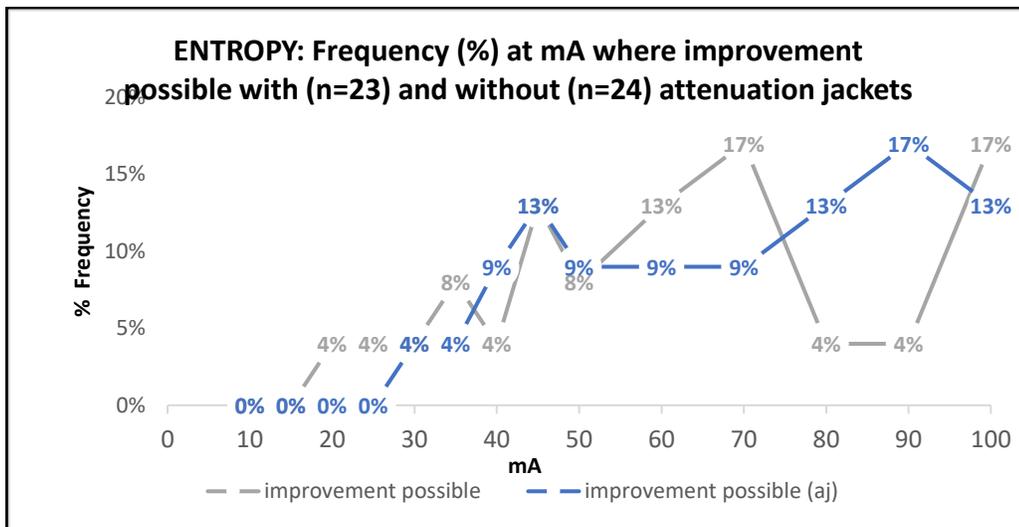


Figure 125. Entropy: Frequency of occurrence at mA where improvement was possible, with and without attenuation jackets.

The most common amperages where improvement occurred, without attenuation jackets, were: 90mA (17%), 60mA (17%), 50mA (13%) and 40mA (13%), as seen in *figure 126*. The most frequent amperages where improvement occurred, with attenuation jackets, were: 80mA (17%), 90mA (13%), 70mA (13%) and 40mA (13%), also seen in *figure 126*.

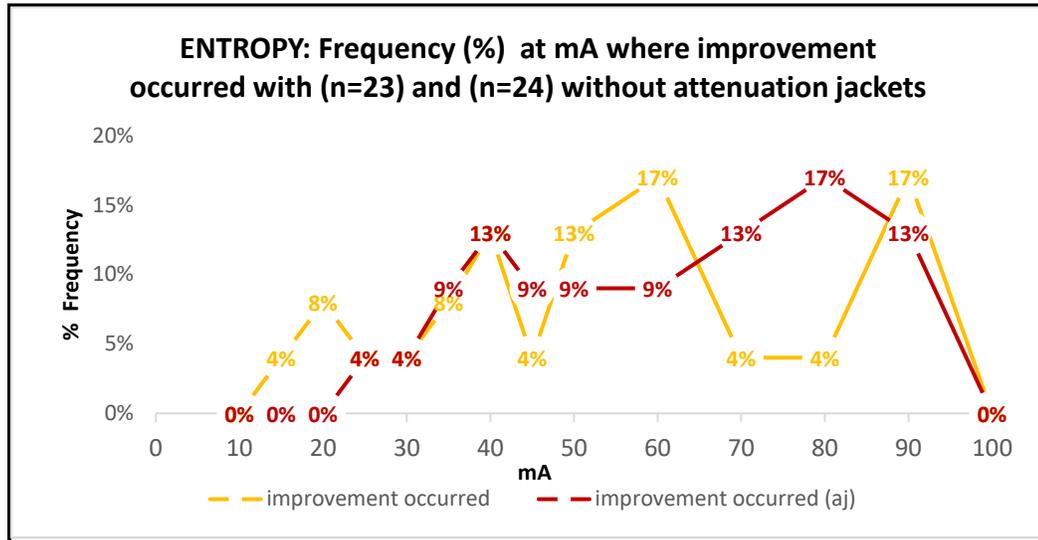


Figure 126. Entropy: Frequency of occurrence at mA where improvement occurred, with and without attenuation jackets.

In addition to looking at lung fields of view, Entropy values of the simulated lesions were examined. Entropy was higher at a low lesion density with a simulated lesion size of 10mm (*figure 127*). Apart from this, Entropy values of simulated lesions remained similar. The use of attenuation jackets had the effect of raising simulated lesion Entropy values very slightly, with little effect on differences between the 10mm and 8mm simulated lesions at densities of 100HU and -630HU, and slightly more noticeable at -800HU, as seen in *table 42* and *figure 127*.

Entropy: Δ 10mm-8mm	Without attenuation jackets	With attenuation jackets
100HU	0.4%	2.0%
-630HU	1.3%	0.4%
-800HU	18.9%	16.3%

Table 42. Difference in Entropy values between 10mm and 8mm simulated lesions at varying density, expressed as a percentage.

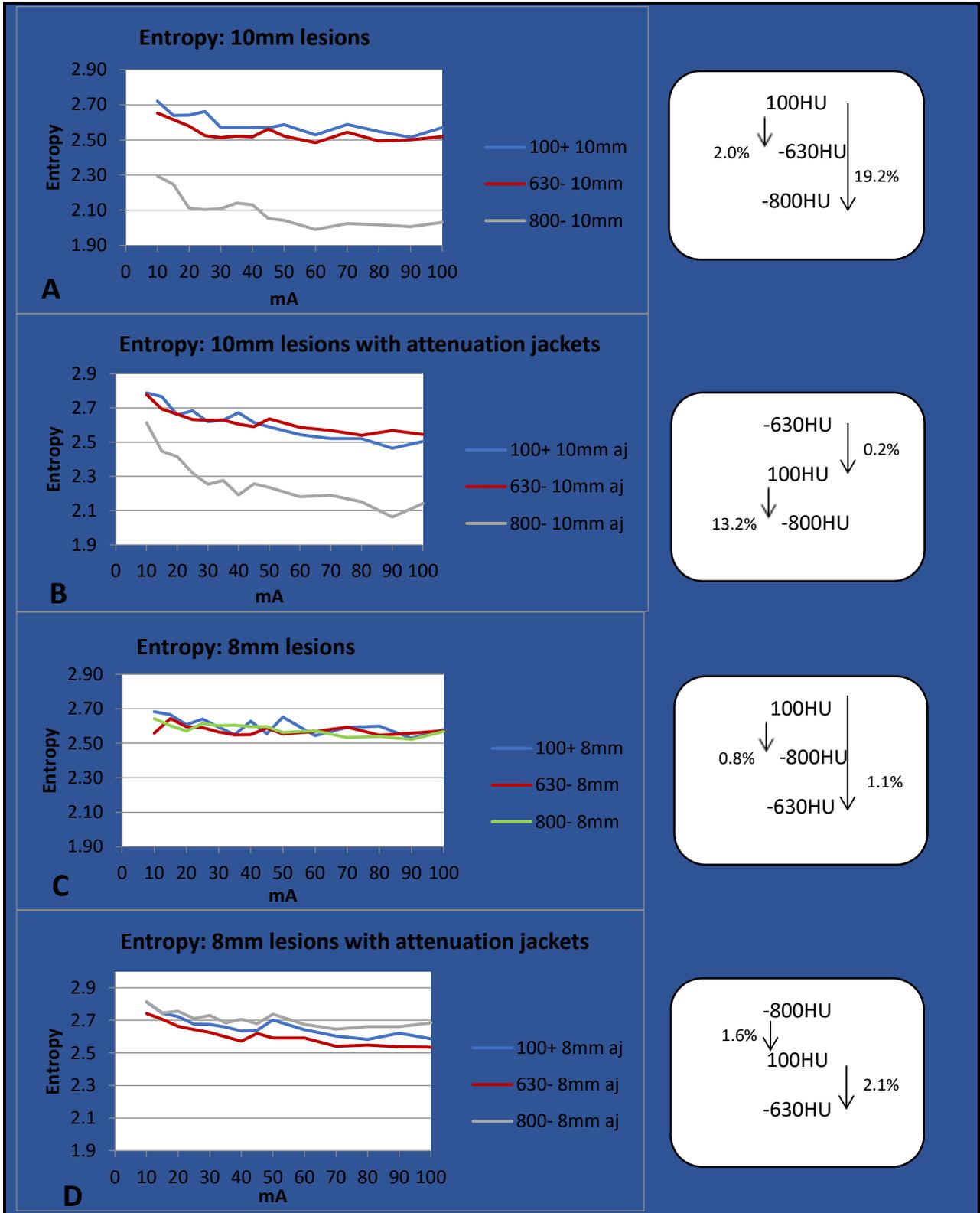


Figure 127. Entropy: Graphs for 10mm and 8mm simulated lesions showing differences in values relating to density (HU): (A) 10mm without attenuation jackets, (B) 10mm with attenuation jackets, (C) 8mm without attenuation jackets, (D) 8mm with attenuation jackets; NB. percentage differences are expressed in comparison to 100HU data series.

In summary, Entropy is a measure of variability or randomness where a constant image has a value of 0 (Mathworks, 1994-2020). This IQM has also provided useful results, with very good correlation between data series, acquired both with and without attenuation jackets.

Regression analysis produced R^2 values which demonstrated a good relation of Entropy values with a logarithmic trend. In addition, the difference between Entropy values obtained with and without the use of attenuation jackets was statistically significant ($p < 0.05$). The use of attenuation jackets impacted on Entropy by increasing the values produced. Interestingly, this effect was far less noticeable when examining Entropy values of the simulated lesions.

Analysis of peaks and plateaus of the individual data series highlighted amperages at 100mA, 70mA and 45mA, where the Entropy value then improved at lower amperage. These improvements were then commonly seen at 90mA, 60mA and 40mA. This emerging pattern was identified, in addition to a good relation with logarithmic trending, indicating some predictive potential for use of lower amperages while maintaining image quality.

In addition to looking at lung fields of view, Entropy values of the simulated lesions were examined. Entropy was higher at a low lesion density with a simulated lesion size of 10mm. Apart from this, Entropy values of simulated lesions remained similar, with the use of attenuation jackets raising lesion Entropy values slightly.

5.11 Best performing image quality metrics (IQM)

The best performing IQMs were mean squared error (MSE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), non-shift-edge ratio (NSER) and texture analysis IQMs looking at Energy, Homogeneity and Entropy. These IQMs had good correlations between the individual datasets, with and without attenuation jackets separately. (Although, NSER demonstrated little difference whether attenuation jackets were used or not). The impact of the use of attenuation jackets is seen with statistically significant differences between data sets. In addition, changes in gradient and y-intercept of graphed data are seen in *table 43*. Where regional information was available, these IQMs showed little regional variation. This enabled averaged data to be used which provided robust measurable information on how increasing amperage affected image quality.

	<i>gradient</i>	<i>y-intercept</i>
MSE	↑ 80%	↑ 77%
PSNR	↑ 37%	↓ 76%
SSIM	↑ 7%	↓ 13%
NSER	↑ 28%	↓ 4%
Energy	↓ 16%	↓ 34%
Homogeneity	↑ 23%	↓ 46%
Entropy	↑ 44%	↑ 13%

Table 43. Effect of using attenuation jackets seen in graphed data with changes in gradient and y-intercept.

6.0 Chapter Six: Results - Observer Performance

The results presented in this chapter are from the perspective of image perception. Quantification of responses from observers, and the statistical significance of those responses, is presented. Visual grading characteristics (VGC) analysis looks at the impact of image noise (created by changes in amperage) on visualisation tasks (Bath & Hansson, 2016). The effect of changing amperage is examined using median gradings given by the observers, in relation to fulfilment of image quality criteria (EC., 2000). Receiver operating characteristic (ROC) analysis looks at the impact of amperage on a lesion detection task (Chakraborty, 2015). Missed lesion detection is analysed in relation to simulated lesion size, density, use of attenuation jackets and amperage at which the images were acquired. A time-case-lesion analysis looks for any effect of amperage or simulated lesion density on the time taken to complete the observer performance tasks (Bath & Hansson, 2016).

6.1 Participant demographics

Taking part in the Observer Performance studies using ViewDEX were participants with a wide range of experience and from a range of backgrounds. This was to remove bias as much as possible, to facilitate assessment of actual image quality. These included a Consultant, Radiology registrars, medical student, radiotherapy Radiographer and Nuclear Medicine Technologists. Information regarding their CT viewing experience is detailed in *table 44* (legend in *table 45*).

<i>Role code</i>	<i>Experience CT images (yrs)</i>	<i>Annual case load CT</i>
4	1	low
2	5	low
4	1	moderate
1	2	moderate
4	5	moderate
1	14	moderate
3	9	high

Table 44. Participant backgrounds and CT viewing experience.

Legend (Table 1):
Role code
Nuclear Medicine Technologist = 1
Radiographer = 2
Consultant Nuclear Medicine Physician = 3
Medical student / Radiology Registrar (ST1) = 4
Annual case load CT
Low = <100
moderate = 100 -300
High = >300

Table 45. Legend (*Table 44*).

6.2 VGC analysis

ViewDEX output data was analysed dividing the observer tasks into two sets: questions 1 and 2 relating to the visualisation of proximal bronchi in the left lung, and the difficulty created by the amount of noise content to that decision; and questions 3 and 4 relating to the peripheral vessels in the left lung, and the difficulty created by the amount of noise content to that decision. Analysis on the paired data was carried out by 'VGC Analyzer' (version 1.0, release 2), which performed a 'Multi-Reader- Multiple-Case' (MRMC) statistical analysis, and determined the *area under the curve* (AUC_{VGC}) (Bath & Hansson, 2016). The AUC_{VGC} uncertainty was determined using 'non-parametric resampling' techniques to produce *confidence intervals* (CI) and a *p-value* (Bath & Hansson, 2016). The reference condition was defined as the response for the confidence level regarding fulfilment of image criteria, and the test condition defined as the response to how much image noise interfered with that decision.

'Fixed-Reader' analysis was chosen due to the limited number of participants which are designated as *readers* in this method of analysis. Results based on the binormal VGC curve analyses were used as this is recommended, rather than the trapezoid curve which can underestimate the confidence interval (CI) (Bath & Hansson, 2016).

The first set of tasks relating to visualisation of proximal bronchi, produced results for the AUC_{VGC} giving a value of 0.480045, and an asymmetric 95% CI (0.462296, 0.501008) (*figure 128*). As the confidence interval includes 0.5 (demarcation of identical distribution), statistical significance cannot be stated (Hansson et al., 2016).

The second set of tasks relating to visualisation of the peripheral lung vessels, produced results for the AUC_{VGC} giving a value of 0.560481, with a standard deviation of 0.0101285 and an asymmetric 95% CI (0.520180, 0.560538) (*figure 128*). This was statistically very significant with $p = 0.0005$.

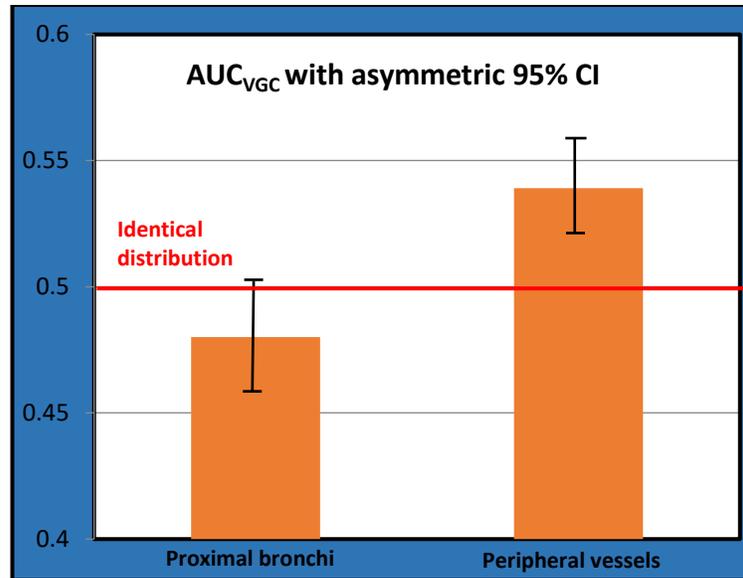


Figure 128. AUC_{VGC} with asymmetric 95% confidence intervals.

The resultant VGC curves are seen in *figures 129 and 130* where the closer the curve is to the diagonal line, the more identical the distributions between the reference condition (fulfilment of criterion response) and the test condition (how much noise interfered with that decision) (Hansson et al., 2016). In addition, For the first question set (*figure 129*), there is only very slight curving from the diagonal noted and this is slightly s-shaped. In the VGC curve for the second question set (*figure 130*), any deviation from the diagonal is very slight and with the curve appearing slightly above the diagonal.

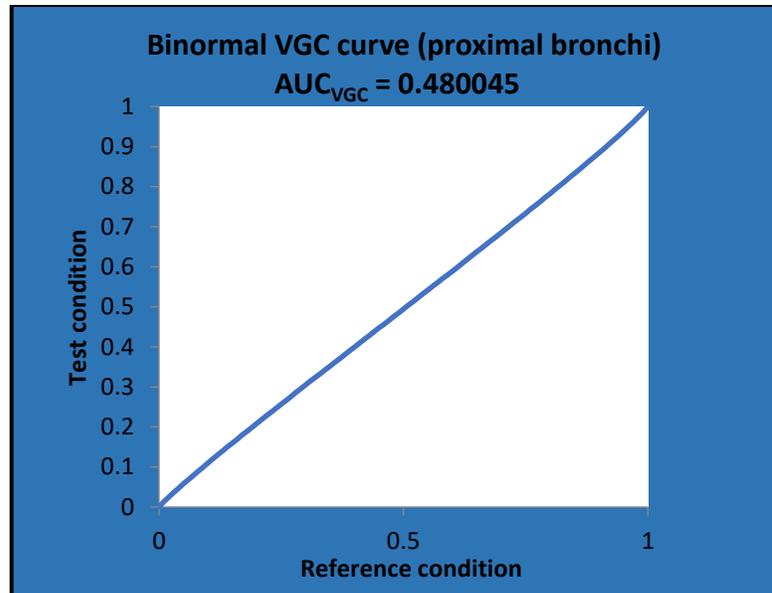


Figure 129. Binormal VGC curve for first question set (proximal bronchi). [Reference condition = fulfilment of image quality criterion; Test condition = how much noise interfered].

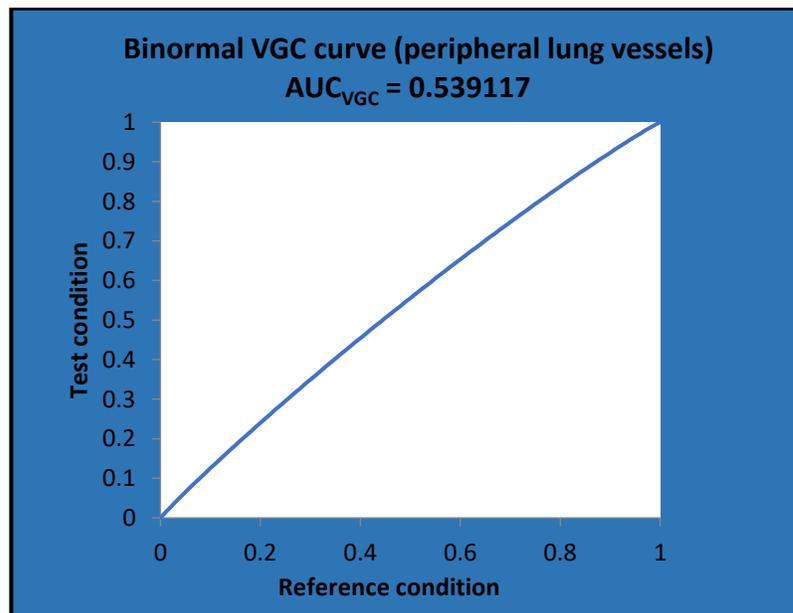


Figure 130. Binormal VGC curve for second question set (peripheral lung vessels). [Reference condition = fulfilment of image quality criterion; Test condition = how much noise interfered].

These statistically significant results show strong correlation between the impact of image noise and the fulfilment of visualisation criteria in CT images, particularly in relation to visualisation of peripheral lung vessels. In addition, the AUC_{VGC} values demonstrated in *figures 6.2.1*, showed a

clear difference between the observer responses relating to the sharpness of proximal bronchi and the visualisation of peripheral vessels, where sharpness of the proximal bronchi was less affected by noise (test condition).

6.3 Fulfilment of Image Quality Criteria

Observers graded images presented to them in ViewDEX according to image criteria as stated in the European guidelines for CT quality (EC., 2000). These guidelines relate to image quality in terms of visualisation, critical reproduction and visual sharpness. The results from the observer performance gradings ranged from where they were confident that the criterion was fulfilled, to where they were confident that the criterion was not fulfilled. Lower grading indicated higher image quality. Effects on observer performance data relating to the use of attenuation jackets was carried out by performing Mann-Whitney significance testing using the Real Statistics resource Pack® add-in for Excel® (Bobbitt, 2018; Zaiontz, n.d.).

Regional variations of image gradings between images from upper, middle and lower zones (Radiology-Masterclass, 2020), meant that it was necessary to look at the results from each region separately. These variations can be seen in *tables 43 and 44*, where Mann-Whitney significance testing ($\alpha=0.05$), as a non-parametric test for ordinal data, demonstrated whether the use of attenuation jackets caused a significant difference in the results, and whether there was a significant difference between regions. P-values for each region from image evaluation of proximal structures and peripheral structures, looking at the significant difference with the use of attenuation jackets, can be seen in *table 46* ($p = 0.01 - 0.19$). The only region where the use of attenuation jackets was not significant, was in the upper zone when peripheral structures were evaluated ($p = 0.19$).

p-exact	Proximal structures	Peripheral structures
Upper zone	0.05	0.19
Middle zone	0.04	0.07
Lower zone	0.03	0.01

Table 46. Image quality criteria: p-values for Mann-Whitney significance testing within regions.

Inter-regional variation is exemplified in *table 47* with p-values demonstrating the significant difference between each region of image evaluation for peripheral structures, both with and without the use of attenuation jackets. Regional variation between the upper zone and other zones was significant (<0.05). However, regional variation between middle and lower zones was not significant.

p-exact	No aj	With aj
Upper zone v Middle zone	0.00	0.00
Upper zone v Lower zone	0.00	0.00
Middle zone v Lower zone	0.10	0.54

Table 47. Image quality criteria: p-values for Mann-Whitney significance testing between regions (with and without attenuation jackets).

When examining the effect of increasing amperage on fulfilment of image quality criteria, the use of attenuation jackets impacted at the lower range of amperages (10mA – 40mA). Without the use of attenuation jackets, fulfilment of image quality criteria was not as good in the upper zones, when compared to middle and lower zones. In addition, the fulfilment of image criteria was generally not as good for visualisation of peripheral structures, when compared to proximal structures. Demonstrated in *figures 131 and 132*, median gradings are graphed where 1 = image criterion was fulfilled, and 5 = image criterion was not fulfilled (indicating poorer quality images). These median values were calculated for scores from the seven observers for each zone. However, due to only single image sets obtained for each of the zones with and without attenuation jackets, and intra-operator variability, there was commonly a difference in the scoring of the fulfilment of image criteria by a range of two (and sometimes more).

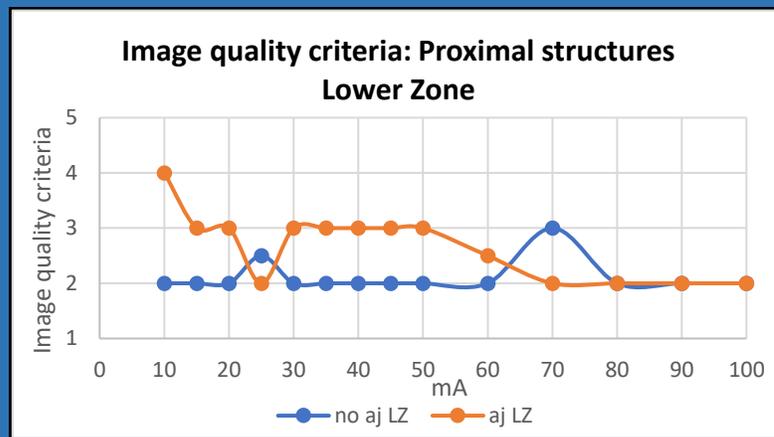
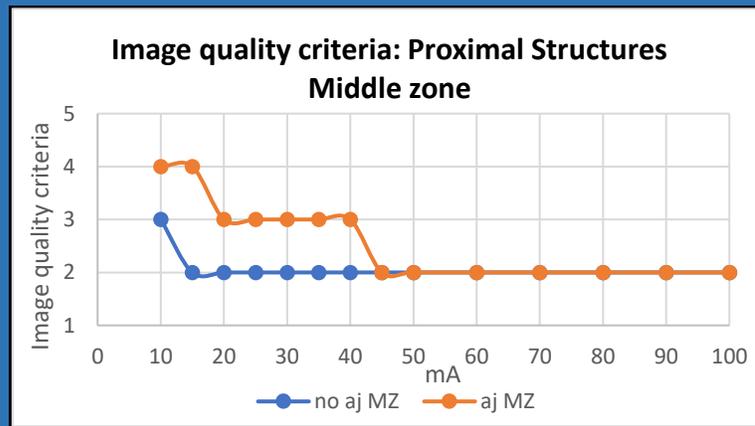
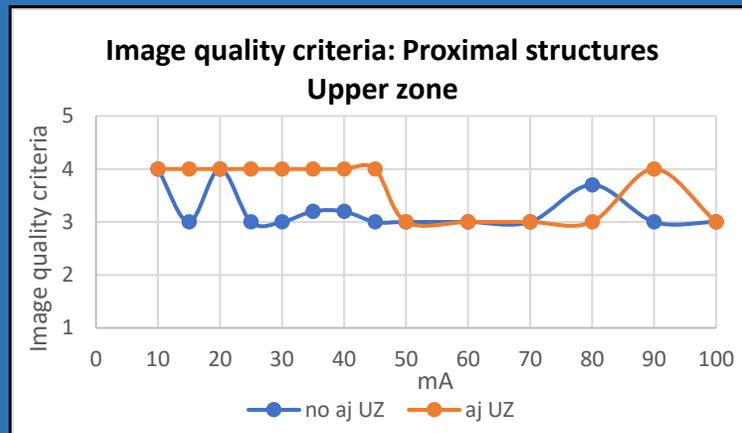


Figure 131. Fulfilment of image quality criterion (median values): Proximal structures in upper, middle and lower zones. [1 = criterion fulfilled; 5 = criterion not fulfilled].

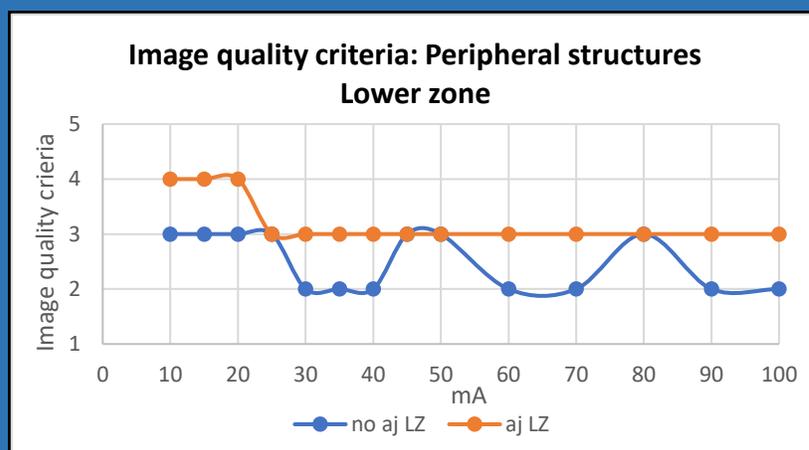
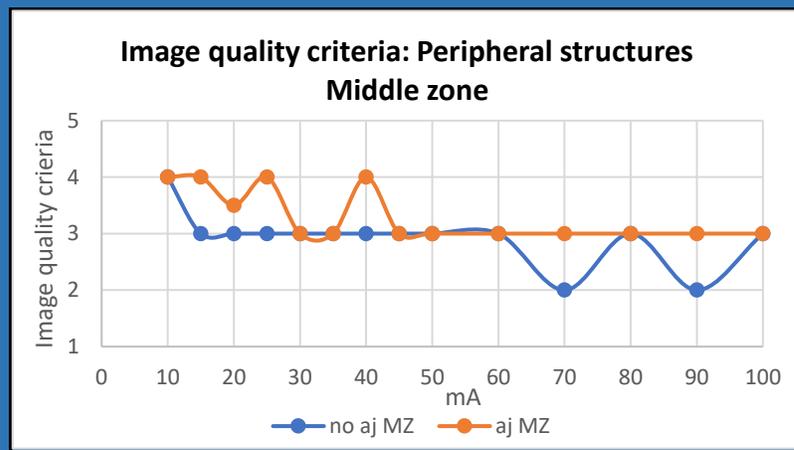
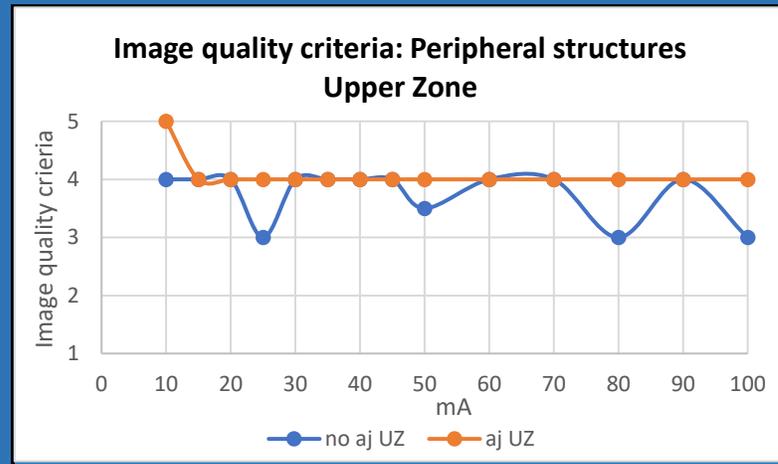


Figure 132. Fulfilment of image quality criterion (median values): Peripheral structures in upper, middle and lower zones [1 = criterion fulfilled; 5 = criterion not fulfilled].

While these results demonstrate that image quality showed some response to increasing amperage, this occurred with some undulating patterns. However, assessment was limited due to limitations on how ordinal data can be analysed. At an amperage of 10mA, fulfilment of the criterion implied poorer image quality, with improvement usually occurring at increased amperage. In addition, if the proximal structures in the mid zone were the focus of an examination, these results (as seen in *figure 131*) would indicate little value in increasing the amperage beyond 45mA as there is no further improvement in fulfilment of image quality criteria.

6.4 ROC analysis

ViewDEX output data for the localisation task, and the confidence with which these localisations were made, was then analysed by RJAFROC software (Chakraborty, 2015). During these analyses, the paired response data underwent Dorfman-Berbaum-Metz Multi-Reader- Multiple-Case' (DBM-MRMC) Hillis significance testing and analysis of variance (ANOVA) (Chakraborty, 2015). The type I error rate was set at 0.05 therefore only p-values less than 0.05 are statistically significant.

The input data was split into two separate analyses of five 'modalities' each (10mA to 30mA; 35mA to 60mA), where each mA setting was described as a *modality*. Case numbers were assigned where each *case* related to the contents of the image in terms of lesion size and density or whether it was 'normal', and whether it was acquired with an attenuation jacket. There were 7 *readers* (or participants). False positive (FP) localisations were identified when a non-lesion localisation occurred in a normal image, or when there was a difference of greater than 10-pixel values from truth coordinates, in any direction.

Each of the analyses involved 595 cases in total. In the first analysis (modalities 10mA to 30mA) 73% of lesions were localised. False positive localisations occurred in 22% of normal images, with 5% incorrect localizations made in the abnormal images. In the second analysis (modalities 35mA

to 60mA) 71% of lesions were localised. False positive localisations occurred in 23% of normal images, with 6% incorrect localisations made in abnormal images.

Output data from the 'random readers and fixed cases' option was used. This was the most appropriate option as 'fixed cases' should be used when images, from which the input data were created, were acquired using an imaging phantom. Both analyses had the significance level set at $\alpha=0.05$. The first analysis (modalities 10mA to 30mA) produced results that were non-significant ($p = 0.4020$ and $F(4, 24) = 1.05$, ns). The second analysis (modalities 35mA to 60mA) also produced results that were non-significant ($p = 0.9146$ and $F(4, 24) = 0.24$, ns).

These results demonstrate that neither analysis showed any significant difference in lesion localisations at different amperage. There is even less significance, however, in the analysis involving the medium range amperage (35mA to 60mA).

6.5 Missed lesion detection

For all the image views containing simulated lesions, data from the lesion detection task revealed that 27% (369/1372) were missed. A breakdown of these numbers is seen in *table 48*. All 5mm simulated lesions were only of a density equivalent to 100HU, and 93% (183/196) of these were missed.

HU / aj	Lesion size	Incidence of missed lesion detection	% increase of missed lesion detection with attenuation jackets
100	5	88	
100 aj	5	95	7%
100	8	50	
100 aj	8	53	6%
630	8	6	
630 aj	8	4	
800	8	22	
800 aj	8	11	
100	10	19	
100 aj	10	21	10%

Table 48. Incidence of missed lesion detection relating to lesion density, use of attenuation jackets and size.

From a possible 588 image views containing 8mm simulated lesions, 25% (146/588) were missed. Further analysis of lesion detection in images containing an 8mm simulated lesion, revealed simulated lesion densities with the following breakdown: 71% (100HU), 7% (-630HU) and 23% (-800HU). All of these simulated lesions were placed in the lower lung zone. Of the 10mm simulated lesions (located in the middle lung zone), the only lesion detections that were missed, 7% (40/588), were at a density of 100HU.

Unexpectedly, the use of attenuation jackets did not contribute greatly to missed lesion detection (*table 48*). In fact, in some instances (-630HU and -800HU), there were proportionately less missed lesion detections as demonstrated by a 50% reduction at -800HU when attenuation jackets were used.

In summary, 100HU simulated lesions had the highest incidence of missed lesion detection, which illustrates that higher lesion density does not necessarily mean a higher detection rate, suggesting that a similarity of density in surrounding structures may have contributed to missed lesion detection, in addition to size. Low lesion density also contributed to missed lesion detection, where this increased with simulated lesions of 800HU (*table 6.5.1*). The use of attenuation jackets with subsequent scatter in an image does not necessarily imply a lower rate of lesion detection. Finally, missed lesion detection occurred over the full range of amperages at which the images were acquired, which was slightly increased over the middle range of amperages (*figure 133*).

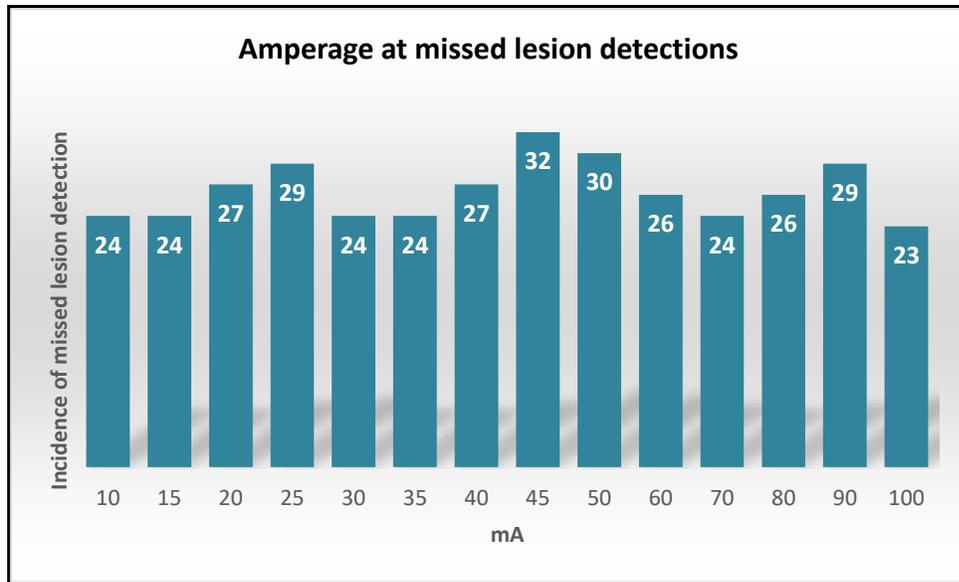


Table 133. Incidence of missed lesion detection at varying amperage.

6.6 Time-case-lesion analysis

For each of the participants, information was analysed from peaks when looking at the time taken to look at each image and complete the observer tasks (referred to as a *case*). The purpose was to determine if there was any commonality or trend able to be observed. An example of case versus time taken is seen in *figure 134*. Major peaks in each of the participants’ sessions were identified, and related to specific case details.

The first one to two cases took longer, as participants settled into the routine of performing the observer study, despite practice immediately prior to this. As the cases were randomised individually for each participant, this did not impact on the time taken for any case in particular. In addition, when a new session was started after a break, a peak occurred for the first case. An example of this is seen in *figure 134*, from data for one of the participants.

The total time for completion of image viewing by each of the participants ranged from 72 minutes to 111 minutes, with an average of 85 minutes. The average time per image ranged from 30 seconds to 47 seconds.

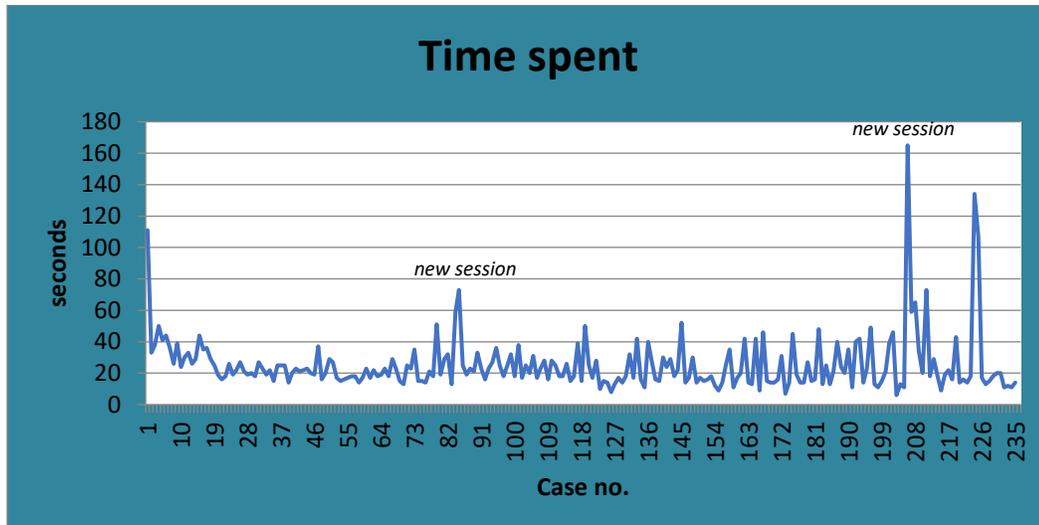


Figure 134. Example of case versus time spent.

In 3/7 participants, scatter in the images, when attenuation jackets were used on the phantom, had a particular impact on the time taken, over a range of amperage and simulated lesion sizes, and images acquired using attenuation jackets caused 28% of the peaks in time taken combining all participant responses (*figure 135*). In a different combination, the higher density simulated lesion (100HU), which was of similar density to the structural surrounds, particularly impacted on the time taken for 3/7 participants, and images containing simulated lesions of 100HU caused 18% of the peaks in time taken combining all participant responses (*figure 135*).

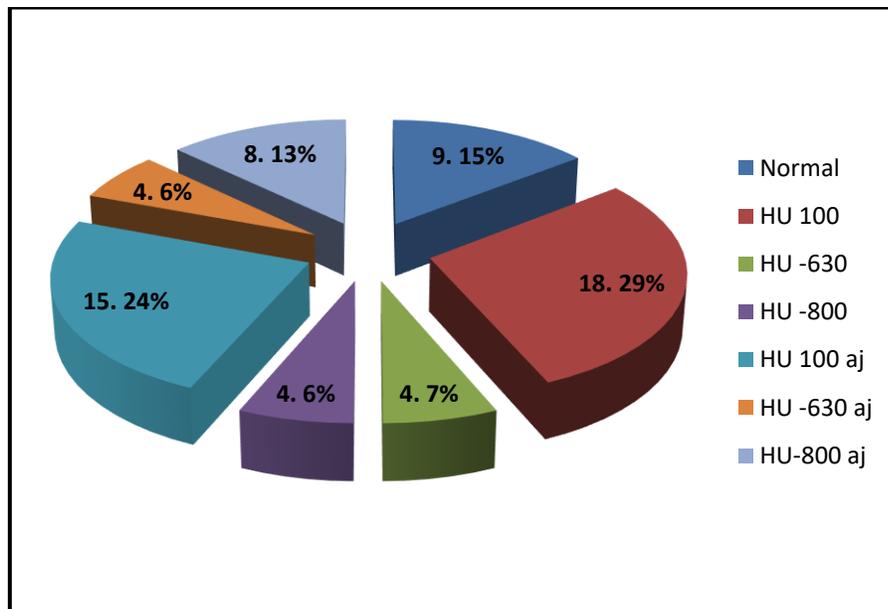


Figure 135. Peaks in time taken related to images containing simulated lesions of different density.

There were 62 cases identified over all (excluding the peak at the beginning and at the start of a new session) which caused participants to ponder for longer. The largest segments of data in the context of time taken (as seen in *figure 135*) are related to images containing 100HU simulated lesions, both with and without attenuation jackets. This may have been due to the simulated lesion density being of a similar density to the surrounding structures (Kagaku, n.d.). When related to amperage, the largest segments of data in the context of time taken (as seen in *figure 136*) occurred at 15mA (16%), 30mA (11%), 35mA (11%), 80mA (10%) and 100mA (11%).

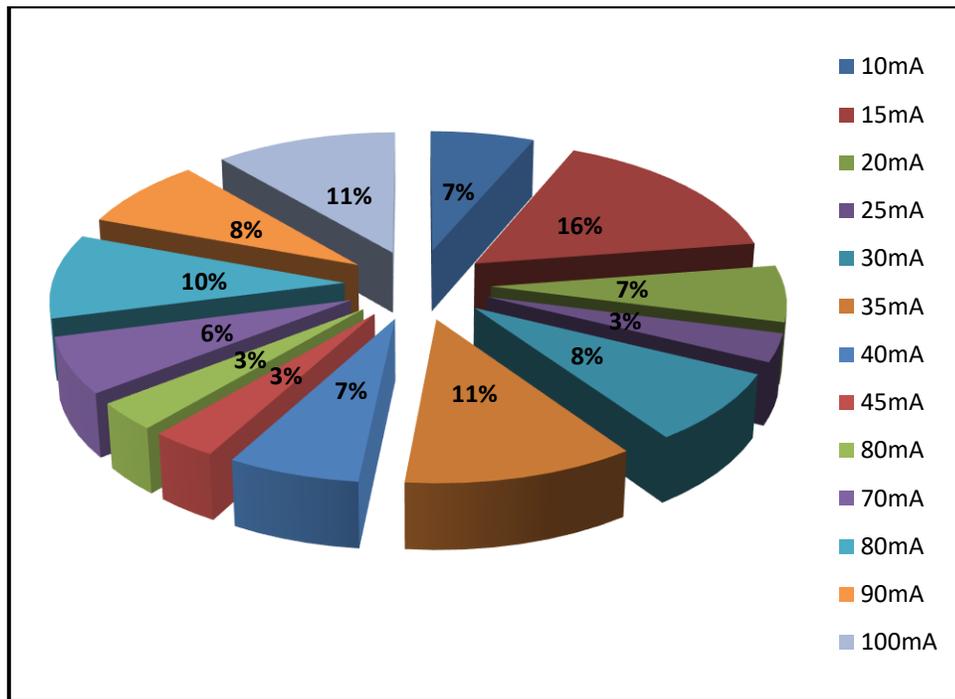


Figure 136. Peaks in time taken related to images acquires at varying amperage.

7.0 Chapter Seven: Correlations and Associations

7.1 Introduction

Multivariate correlations were performed on the best performing image quality metrics (IQM), as described in the Results chapter. The IQMs were Mean squared error (MSE), Peak signal-to-noise ratio (PSNR), Structural similarity index (SSIM), Non-shift edge ratio (NSER), and texture analysis (specifically Energy, Homogeneity and Entropy). These IQMs demonstrated good correlation with values in averaged data series (see Results chapter) which facilitated examination of the results. It is worth noting that MSE, PSNR, SSIM, and NSER were obtained from the whole of the image matrix in Matlab® (Mathworks, 1994-2020). However, texture analysis was able to be performed over different regions of the thorax in LIFEx, enabling accurate data over the specific organ or region that was being assessed (Nioche et al., 2018). Energy, Homogeneity and Entropy showed little variation between regions.

Parametric and nonparametric measures of association for the IQMs, were performed in JMP® statistics software (alpha value set to $\alpha=0.05$) to obtain results for Pearson's r coefficient and Spearman's ρ coefficient (JMP, 2020 a). Nonparametric methods, which make fewer assumptions, may be useful in circumstances where the situation is less known (Linebach et al., 2014).

Non-parametric measures of association were performed between Observer Performance data (confidence that 'European CT image quality guidelines for critical reproduction' were fulfilled), and the IQMs mentioned above (EC., 2000). Spearman's rho correlation values were obtained in SPSS®, as the observer performance data for fulfilment of image quality criteria was ordinal (SPSS, 2021). Analysis with Observer Performance data was carried out on a regional basis (upper, middle, and lower zones) as there were regional differences.

7.2 Multivariate correlations and associations

7.2.1 Pearson product-correlation (Pearson’s r coefficient)

To examine the strength of linear relationships between variables, a parametric measure of association, Pearson’s r coefficient, has values from +1 to -1 (JMP, 2020 a). The closer to +1 or -1, the stronger the linear relationship (positive or negative respectively) (JMP, 2020 a). When the r value approximates zero, this implies that there is no linear relationship indicated by the correlation (Bewick et al., 2003). Bewick et al. (2003) also remind us that this does not rule out a relationship that is non-linear.

The scatterplot in *figure 137* shows relationships between IQM’s (without attenuation jackets) enabling negative and positive relationships to be easily visualised. Those relationships with strong correlations ($r > 0.9$, $r < -0.9$) have been indicated by yellow boxes.

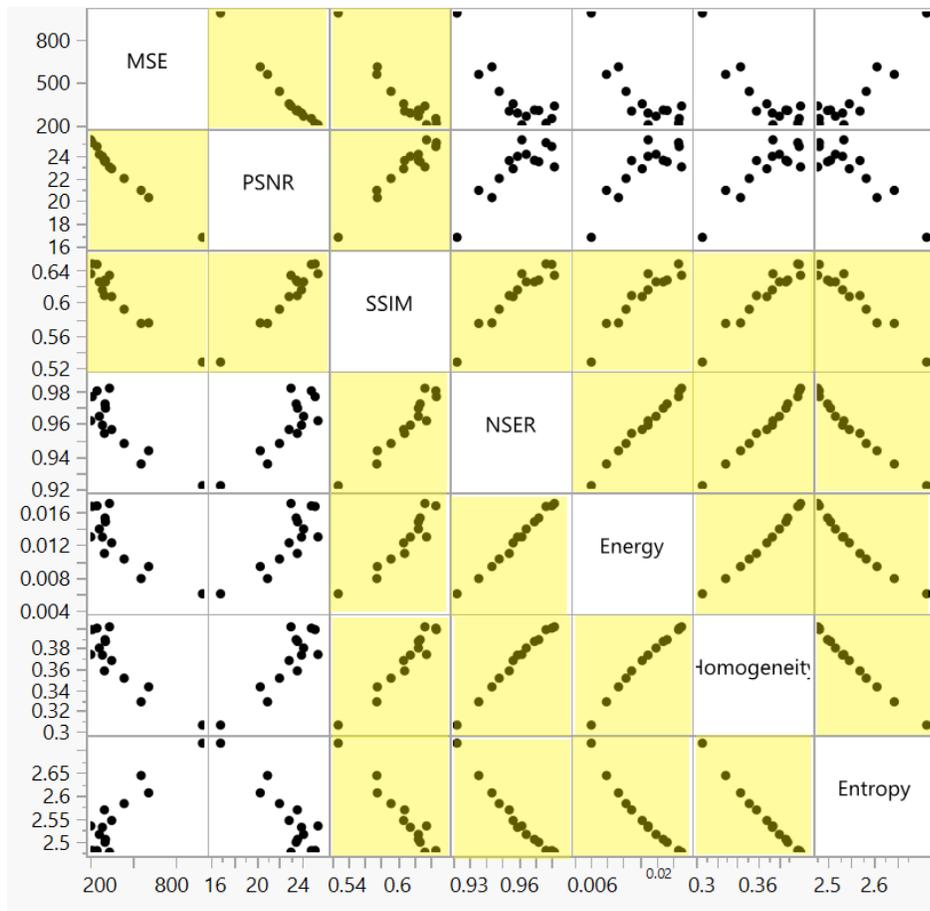


Figure 137. Scatter plot matrix of relationships between IQMs (no attenuation jackets). Yellow squares indicate those relationships with Pearson’s r values > 0.9 or < -0.9 . (Values for individual IQMs on both axes).

Pearson’s r values for IQM’s, without attenuation jackets, with the strongest correlation, both positive and negative, ($r > 0.9, r < -0.9$) listed in *tables 49 and 50*, include 95% confidence limits and all have a probability $p < 0.0001$. The strongest positive correlations demonstrated, were between Homogeneity, NSER and Energy (*table 49*). The strongest negative correlations demonstrated, were mainly between Entropy and other IQMs, along with a negative correlation between PSNR and MSE (*table 50*).

variable	by Variable	Correlation	Lower 95%	Upper 95%
Homogeneity	NSER	0.9952	0.9844	0.9985
Energy	NSER	0.9948	0.9831	0.9984
Homogeneity	Energy	0.9927	0.9763	0.9977

Table 49. Pearson’s r values with strongest positive correlations (without attenuation jackets).

variable	by Variable	Correlation	Lower 95%	Upper 95%
Entropy	Homogeneity	-0.9962	-0.9988	-0.9876
PSNR	MSE	-0.989	-0.9966	-0.9646
Entropy	NSER	-0.9864	-0.9958	-0.9564
Entropy	Energy	-0.9793	-0.9936	-0.9341
Entropy	SSIM	-0.971	-0.991	-0.9084

Table 50. Pearson’s r values with strongest negative correlations (without attenuation jackets).

The scatterplot in *figure 138* shows relationships between IQM’s (with attenuation jackets) enabling negative and positive relationships to be easily visualised. Those relationships with strong correlations, both positive and negative ($r > 0.9, r < -0.9$) indicated by yellow boxes, can be seen to include all the IQM relationships analysed, with exception of a negative relationship between MSE and Energy ($r = -0.8438$; $CI_{95\%} = -0.9494$ to -0.5672).

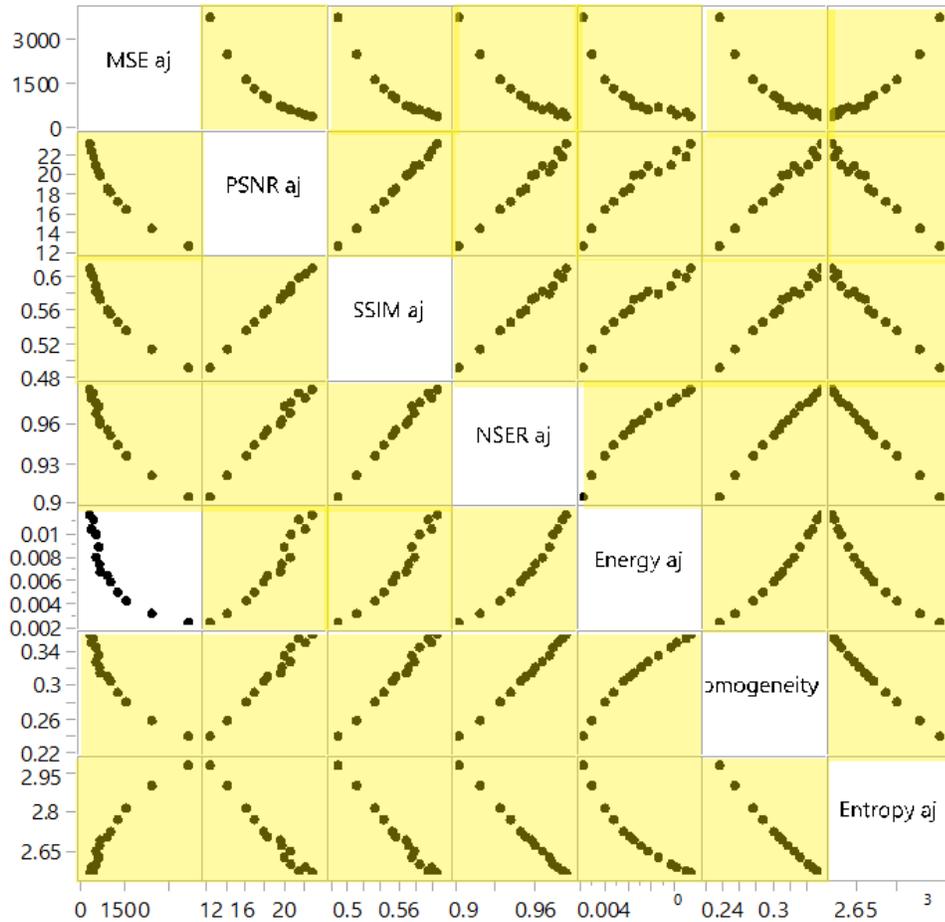


Figure 138. Scatter plot matrix of relationships between IQMs (with attenuation jackets). Yellow squares indicate those relationships with Pearson’s r values > 0.9 or < -0.9 . (Values for individual IQMs on both axes).

Pearson’s r values for IQM’s, with attenuation jackets, with the strongest correlations ($r > 0.97$, $r < -0.97$) listed in *tables 51 and 52* include 95% confidence limits and all have a probability $p < 0.0001$. The strongest positive correlations demonstrated, were between SSIM and other IQMs (Homogeneity, NSER, PSNR) and between Homogeneity and NSER (*table 51*). The strongest negative correlations demonstrated, were between Entropy and other IQMs (NSER, Homogeneity, SSIM, PSNR) (*table 52*).

Variable	by Variable	Correlation	Lower 95%	Upper 95%
SSIM aj	PSNR aj	0.9988	0.9962	0.9996
Homogeneity aj	NSER aj	0.9972	0.991	0.9992
NSER aj	SSIM aj	0.994	0.9806	0.9982
Homogeneity aj	SSIM aj	0.9928	0.9768	0.9978
NSER aj	PSNR aj	0.9896	0.9666	0.9968
Homogeneity aj	PSNR aj	0.9881	0.9618	0.9963
Homogeneity aj	Energy aj	0.9828	0.9448	0.9947

Table 51. Pearson's *r* values with strongest positive correlations (with attenuation jackets).

Variable	by Variable	Correlation	Lower 95%	Upper 95%
Entropy aj	NSER aj	-0.9987	-0.9996	-0.9958
Entropy aj	Homogeneity aj	-0.9956	-0.9986	-0.9857
Entropy aj	SSIM aj	-0.9924	-0.9977	-0.9756
Entropy aj	PSNR aj	-0.9875	-0.9962	-0.9599

Table 52. Pearson's *r* values with strongest negative correlations (with attenuation jackets).

Comparison of Pearson's *r* values for IQMs, enabled quantification of the effect on linear relationships for individual IQMs when attenuation jackets were used. The correlation values with 95% confidence limits in *table 53*, all have a probability $p < 0.0001$. Of these best performing IQM that were analysed, the linear relationships least affected by use of attenuation jackets were NSER, Homogeneity and Entropy. The IQM where the linear relationship was most affected by use of attenuation jackets, was PSNR.

Variable	by Variable	Correlation	Lower 95%	Upper 95%
NSER aj	NSER	0.9956	0.9856	0.9986
Homogeneity aj	Homogeneity	0.9947	0.9828	0.9984
Entropy aj	Entropy	0.994	0.9807	0.9982
Energy aj	Energy	0.9882	0.9622	0.9964
SSIM aj	SSIM	0.961	0.8781	0.9879
MSE aj	MSE	0.9476	0.8386	0.9836
PSNR aj	PSNR	0.8564	0.5971	0.9536

Table 53. Pearson's *r* values with correlations for IQMs with and without attenuation jackets.

7.2.2 Spearman’s ρ correlation coefficient

Spearman’s ρ correlation is a nonparametric measure of association, and ranks the data variables instead of using the data values themselves (JMP, 2019, 2020 a). With values ranging from +1 to -1, a stronger monotonic relationship (going in the same relative direction) is indicated as the values approximate ± 1 (JMP, 2020 a; SSPS, 2021).

Spearman’s ρ values for IQM’s, without attenuation jackets, with the strongest correlation ($r > 0.9, r < -0.9$) listed in *tables 54 and 55*, all have a probability $p < 0.0001$. The strongest positive correlations demonstrated, were between Homogeneity, Energy and NSER (*table 54*). The strongest negative correlations demonstrated, were mainly between Entropy and other IQMs, along with a negative correlation between PSNR and MSE (*table 55*).

Variable	by Variable	Spearman ρ
Homogeneity	Energy	1
Energy	NSER	1
Homogeneity	NSER	1
NSER	SSIM	0.921
Energy	SSIM	0.921
Homogeneity	SSIM	0.921

Table 54. Spearman’s ρ values with strongest positive correlations (without attenuation jackets).

Variable	by Variable	Spearman ρ
Entropy	NSER	-0.996
Entropy	Homogeneity	-0.996
Entropy	Energy	-0.996
PSNR	MSE	-0.987

Table 55. Spearman’s ρ values with strongest negative correlations (without attenuation jackets).

Spearman’s ρ values for IQM’s, with attenuation jackets, with the strongest correlation ($r > 1.0, r < -1.0$) listed in *tables 56 and 57*, all have a probability $p < 0.0001$. The strongest positive correlations demonstrated, were between SSIM and PSNR, in addition to Energy, NSER and Homogeneity correlating with one another (*table 56*). The strongest negative correlations demonstrated, were mainly between Entropy and the remainder of the IQMs, along with a negative correlation between PSNR and MSE, and SSIM and MSE (*table 57*).

Variable	by Variable	Spearman ρ
SSIM aj	PSNR aj	1
Energy aj	NSER aj	1
Homogeneity aj	NSER aj	1
Homogeneity aj	Energy aj	1

Table 56. Spearman’s ρ values with strongest positive correlations (with attenuation jackets).

Variable	by Variable	Spearman ρ
PSNR aj	MSE aj	-1
SSIM aj	MSE aj	-1
Entropy aj	NSER aj	-1
Entropy aj	Energy aj	-1
Entropy aj	Homogeneity aj	-1

Table 57. Spearman’s ρ values with strongest negative correlations (with attenuation jackets).

Comparison of Spearman’s ρ values for IQMs, enabled quantification of the effect on monotonic relationships for individual IQMs, when attenuation jackets were used, but this time computed on ranked data. The correlation values in *table 58*, all have a probability $p < 0.0001$. Of these best performing IQM that were analysed, the monotonic relationships least affected by use of attenuation jackets, were NSER, Energy, and Homogeneity. The IQM where the monotonic relationship was most affected by use of attenuation jackets, was MSE.

Variable	by Variable	Spearman ρ
NSER aj	NSER	1
Energy aj	Energy	1
Homogeneity aj	Homogeneity	1
SSIM aj	SSIM	0.923
PSNR aj	PSNR	0.697
MSE aj	MSE	0.688

Table 58. Spearman’s ρ values with correlations for IQMs with and without attenuation jackets.

7.2.3 Summary IQM multivariate correlations and associations

Without the use of attenuation jackets, SSIM, NSER, Energy, Homogeneity, and Entropy, all demonstrated a strong correlation with one another, whether positively or negatively, as visually seen in *figure137*. However, with the use of attenuation jackets, all seven of the IQMs in this multivariate analysis demonstrated strong correlation, as seen in *figure 138*. In the clinical setting, with patient variation in body habitus, this would facilitate consistency when selecting an IQM. Equally important, however, this demonstrated the importance of using attenuation jackets (or equivalent) for consistency in phantom work, when establishing acceptable parameters of image quality suitable for the task intended.

It is also useful to note that when associations looking at the strength of monotonic relationships are examined, texture analysis IQMs (Homogeneity, Energy, and Entropy) do not necessarily have the strongest strength of association with each other. This would indicate that most of these top performing IQMs could be considered independently useful, in the clinical context for assessing image quality (with the exception of PSNR and MSE which are the most affected by the use of attenuation jackets, demonstrated in *table 58*).

However, in the clinical setting, these IQM's have to be linked with the human visual system (HVS) and what the clinical observer is seeing. These relationships are examined in the following section (7.2.4).

7.2.4 Correlation between IQMs and Observer Performance

Comparison of correlation coefficients, between IQMs and observer performance data for image quality criteria, enabled quantitative correlation of physical and psychophysical assessment. There were two groups of data for image quality criteria, which related to visual sharpness of proximal structures and peripheral structures. Due to regional differences for image quality criteria in Observer Performance data, the results are presented separately. In addition, the effect when attenuation jackets were used was apparent.

Spearman's ρ , as a non-parametric measure of strength of monotonic relationships using ranked data, between IQMs and observer performance quality criteria, demonstrated the strongest monotonic relationships approximating ± 1 (JMP, 2020 a; SSPS, 2021). Observer performance image quality criteria looking at proximal structures (upper zone), had the strongest correlation (ρ) with NSER, Energy, and Homogeneity without attenuation jackets, and with MSE, PSNR and SSIM with attenuation jackets (*figure 139 and table 59*). Weakest correlations (ρ) were with MSE and PSNR (without attenuation jackets) and NSER, Energy, Homogeneity and Entropy (with attenuation jackets) also seen in *figure 139 and table 59*.

Using Spearman's ρ , observer performance image quality criteria looking at proximal structures (middle zone), showed no difference in strength of correlation with all the IQMs ($\rho= 0.447$ and $\rho= -0.447$), without attenuation jackets. This was the same for strength of correlation with all the IQMs, with attenuation jackets ($\rho= 0.911$ and $\rho= -0.911$).

Using Spearman's ρ , observer performance image quality criteria looking at proximal structures (lower zone), had the strongest correlation with PSNR without attenuation jackets, and NSER, Energy, Homogeneity and Entropy, with attenuation jackets (*figure 140 and table 60*). The weakest correlation (ρ) was seen with NSER, Energy, Homogeneity and Entropy without attenuation jackets, and MSE, PSNR and SSIM with attenuation jackets, also seen in *figure 140 and table 60*.

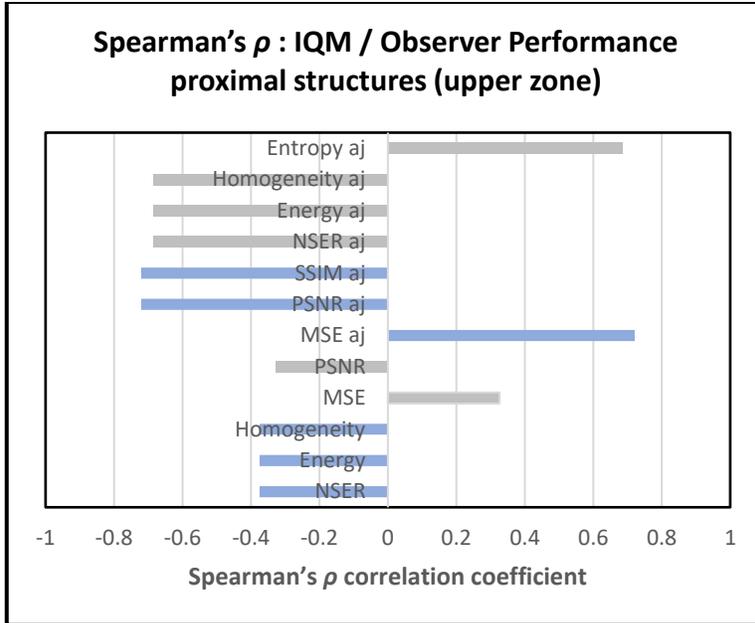


Figure 139. Spearman's ρ values with strongest positive and negative correlations (with and without attenuation jackets), for proximal structures (upper zone). Weakest correlations highlighted in grey.

Upper zone (proximal):	
Without attenuation jackets	
NSER	-0.375
Energy	-0.375
Homogeneity	-0.375
MSE	0.326
PSNR	-0.326
With attenuation jackets	
MSE aj	0.721
PSNR aj	-0.721
SSIM aj	-0.721
NSER aj	-0.684
Energy aj	-0.684
Homogeneity aj	-0.684
Entropy aj	0.684

Table 59. Spearman's ρ values with strongest positive and negative correlations (with and without attenuation jackets), for proximal structures (upper zone). Weakest correlations highlighted in grey.

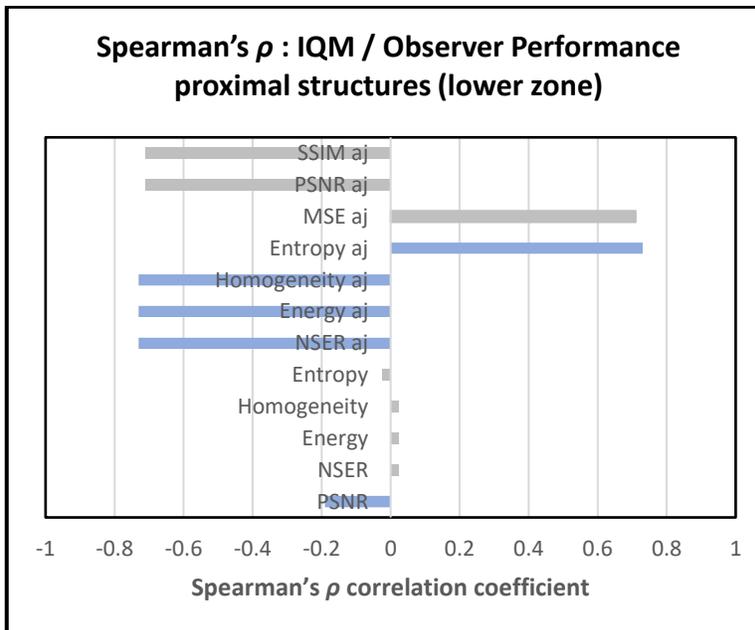


Figure 140. Spearman's ρ values with strongest positive and negative correlations (with and without attenuation jackets), for proximal structures (lower zone). Weakest correlations highlighted in grey.

Lower zone (proximal):	
Without attenuation jackets	
PSNR	-0.191
NSER	0.025
Energy	0.025
Homogeneity	0.025
Entropy	-0.025
With attenuation jackets	
NSER aj	-0.730
Energy aj	-0.730
Homogeneity aj	-0.730
Entropy aj	0.730
MSE aj	0.710
PSNR aj	-0.710
SSIM aj	-0.710

Table 60. Spearman's ρ values with strongest positive and negative correlations (with and without attenuation jackets), for proximal structures (lower zone). Weakest correlations highlighted in grey.

Using Spearman’s ρ , observer performance image quality criteria looking at peripheral structures (upper zone), had the strongest correlations with Energy, Homogeneity and Entropy, without attenuation jackets (*figure 141 and table 61*). The weakest correlation (ρ) was observed with MSE and PSNR, without attenuation jackets, also seen in *figure 141 and table 61*. Observer performance image quality criteria looking at peripheral structures (upper zone) showed no difference in strength of correlation with all the IQMs ($\rho= 0.447$ and $\rho= -0.447$), with attenuation jackets.

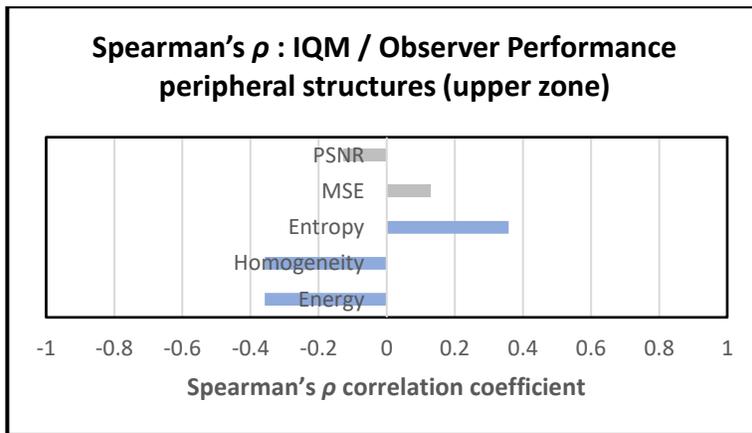


Figure 141. Spearman’s ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (upper zone). Weakest correlations highlighted in grey.

Upper zone (peripheral)	
Without attenuation jackets	
Energy	0.358
Homogeneity	0.358
Entropy	-0.358
MSE	0.128
PSNR	-0.128

Table 61. Spearman’s ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (upper zone). Weakest correlations highlighted in grey.

Using Spearman’s ρ , observer performance image quality criteria looking at peripheral structures (middle zone), had the strongest correlations with NSER, Energy, Homogeneity and Entropy, without attenuation jackets (*figure 142 and table 62*). The weakest correlation (ρ) was observed with PSNR, without attenuation jackets, also seen in *figure 142 and table 62*. Observer performance image quality criteria looking at peripheral structures (middle zone) showed no difference in strength of correlation with all the IQMs ($\rho= 0.447$ and $\rho= -0.447$), with attenuation jackets.

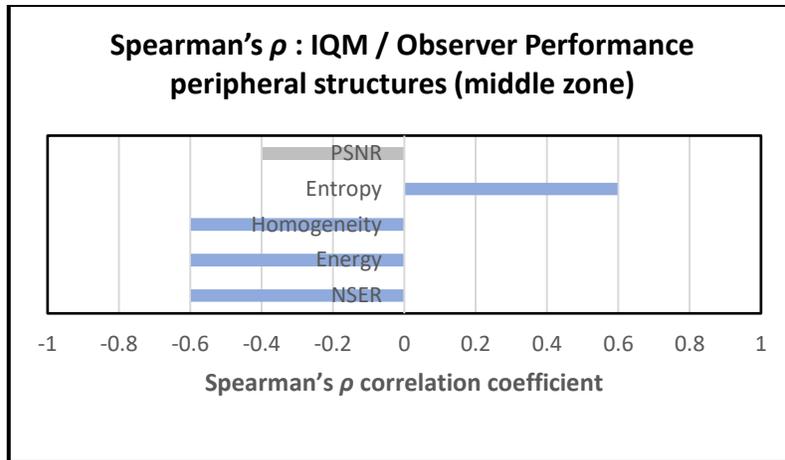


Figure 142. Spearman's ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (middle zone). Weakest correlations highlighted in grey.

Middle zone (peripheral)	
Without attenuation jackets	
NSER	-0.598
Energy	-0.598
Homogeneity	-0.598
Entropy	0.598
PSNR	-0.398

Table 62. Spearman's ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (middle zone). Weakest correlations highlighted in grey.

Using Spearman's ρ , observer performance image quality criteria looking at peripheral structures (lower zone), had the strongest correlation with Entropy without attenuation jackets, and the weakest correlation was observed with MSE (figure 143 and table 63). Observer performance image quality criteria looking at peripheral structures (lower zone) showed no difference in strength of correlation with all the IQMs ($\rho= 0.713$ and $\rho= -0.713$), with attenuation jackets.

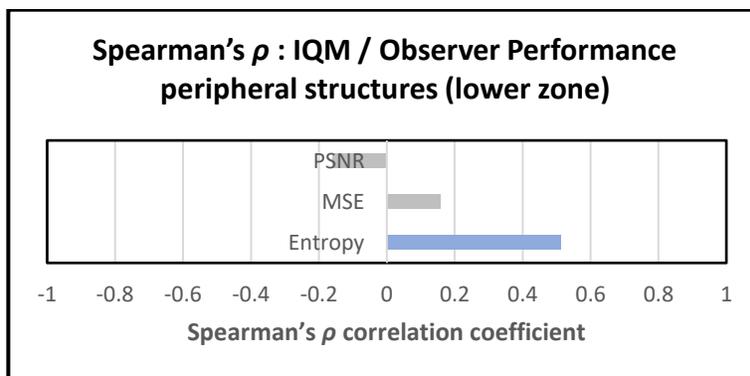


Figure 143. Spearman's ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (lower zone). Weakest correlations highlighted in grey.

Lower zone (peripheral)	
Without attenuation jackets	
Entropy	0.514
MSE	0.159
PSNR	-0.159

Table 63. Spearman's ρ values with strongest positive and negative correlations (without attenuation jackets), for peripheral structures (lower zone). Weakest correlations highlighted in grey.

In summary, correlation of IQMs and observer performance data for image quality criteria, demonstrated variations in values when looking at observer performance image quality criteria for the different regions, as well as variations relating to whether attenuation jackets were used. Confounding comparisons further, were variations in the way IQM analyses dealt with the image data, and the significance of contrast resolution, spatial resolution, and the surrounds relating to image features. However, when a meta-analysis was performed of the frequency at which correlation coefficients (ρ) indicated how well the IQMs performed (best or worst), against how observers perceived the image quality, there have emerged some IQMs that performed better than the rest. In *figure 144*, it is seen that Entropy, Homogeneity and Energy and NSER performed the best. In addition, these four IQMs had the lowest frequency for performing the worst. In contrast, the worst performing IQMs were MSE and PSNR, also seen in *figure 144*.

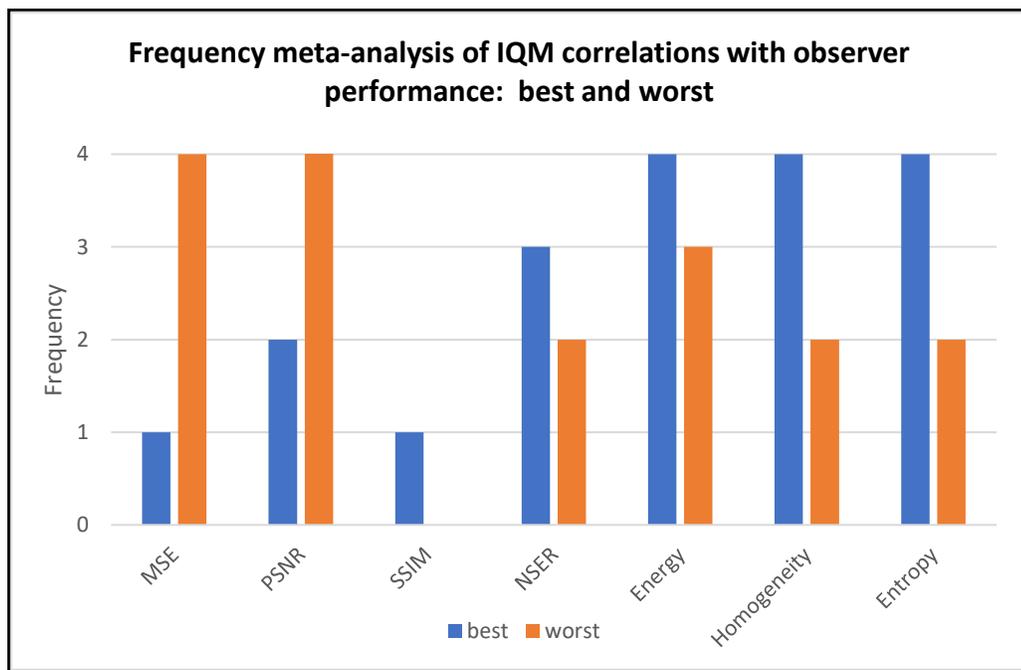


Figure 144. Frequency meta-analysis, for Spearman’s ρ correlation coefficients for IQM performance when correlated with observer performance data for fulfilment of image quality criteria.

8.0 Chapter Eight: Discussion

The aim of this thesis was to evaluate task-specific observer performance against a wide variety of image quality metrics to establish whether they can be used to predict CT image quality, as required by human observers for a specific diagnostic task. Images were acquired to simulate the lung using an anthropomorphic chest phantom. This is a radiographic high-contrast region of the body therefore research findings are specific for this region.

Low-dose CT (LDCT) is not only used in hybrid imaging, but in CT screening programs as well. With increasing use of LDCT and subsequent increase in radiation dose to the overall population as well as to the individual, establishing suitable IQMs is essential (UNSCEAR, 2010). With no single IQM able to capture all aspects of image quality, this research looked at correlation between the IQMs that were used, and then looked at their correlation with observer performance.

There are multiple factors affecting CT image quality, and various acquisition parameters which can be altered including amperage and voltage, all impacting on the radiation dose required to obtain images of suitable quality for their intended purpose. Determining the resolution or clarity of features required in an image, and the level of artefact (such as streaking or partial volume effects), utilises a balance of slice thickness and acceptable noise level, while selection of a reconstruction filter or kernel produces the suitable smoothing or sharpness (Bushberg et al., 2012b; Romans, 2013; Seeram, 2009d). The CT radiation dose to a patient can be reduced by increasing the pitch or lowering the amperage, but with less photons used to produce the image over a particular area, there are less resultant signals and the image is noisier therefore of poorer image quality (Sil et al., 2012). In addition, selection of the correct voltage is essential to ensure optimal penetration of x-rays for the intended task (Seeram, 2009g).

Image acquisition:

The method chosen required all image acquisition parameters to be fixed, except amperage. This was because amperage, as a singular component, impacted on the number of electrons produced in the x-ray tube and subsequent radiation exposure (Seeram, 2009g). Changing amperage was easily achieved with discrete interval changes, and not derived from a combination of factors, as

in milliamperere seconds for example. An anthropomorphic chest phantom was used to obtain CT images with increasing amperage from 10mA to 100mA. Simulated lesions of varying density were placed in the phantom, and images were obtained both with and without attenuation jackets.

Image quality metrics (IQM):

IQM were chosen to assess a spectrum of image characteristics looking at image noise, image contrast, edge definition and surrounding structures. These IQMs ranged from signal-based to those that were more structure-based therefore approximating perception of the human visual system (HVS) more closely.

The visual quality of the images improved with increasing amperage as expected, and this was borne out in most IQMs used. However, the response to increasing amperage was not linear and in some cases varied between regions of the lung being examined. Patterns of undulations (peaks, troughs, and plateaus) were examined in graphed data, as commonalities were sought after to indicate any potential for reduction in radiation dose at certain amperages, where the same or improved quality of information was potentially available at a lower amperage. There was no pattern detected overall, except where the level of response plateaued between 90mA and 100mA. There were, however, patterns that started to emerge specific to some of the individual IQMs.

Results of most IQMs demonstrated a difference that was statistically significant when an increased thickness of simulated tissue was introduced. NSER was a particular exception, which provided values that demonstrated no significant difference when using regression analysis comparing data obtained with and without the use of attenuation jackets, although the slope of the graph increased by 28% when attenuation jackets were used indicating an increased influence of amperage in the lower range. A change in gradient and y-intercept on graphed data were features of the use of attenuation jackets, with changes in gradient ranging from 7% to 80%, and changes in y-intercept ranging from 4% to 77%. This does raise the question of whether phantoms without annuli attachments, should be used in determining image quality at a given dose (at least for some parts of the body for which the CT settings were intended). The use of

anthropomorphic phantoms is a step closer to the reality of image viewing (for which image data in the clinical context is intended), but even then, the use of attenuation jackets impacts on the outcomes of the IQMs used. This could be of particular importance when translating between CT scanner platforms, especially if protocol translation between the systems has a dependency on patient size (McKenney et al., 2014).

As the IQMs are looking at different characteristics of the images, they are affected by spatial resolution, contrast resolution, and image features and their surrounds. It would be expected that there would be some variation between them in their response to increasing amperage. In addition, some of the IQMs are obtained from an analysis that involves the whole image, while other IQMs provide the opportunity for regional and even lesion analysis. To assess the suitable quality of an image for the task at hand, the need for an IQM using regional/lesion analysis would need to be determined.

Due to the structural variation within the region of the phantom being examined in this research, and its radiographic high-contrast nature, SNR and CNR proved not to be useful IQMs in this context. These IQMs were designed for use with homogenous backgrounds, particularly SNR (Bushberg et al., 2012a). As noise texture is what a person viewing an image is sensitive to, then, how noise is described in SNR is too simplistic (Bochud et al., 1999). In addition, CNR is not the same as detectability by a person viewing the image because it does not take in to account the structural complexity surrounding an image feature such as a lung nodule, therefore it is not representative of viewing an image in the clinical context.

MSE and PSNR were obtained using analysis involving the whole image, but as purely signal-based IQMs they did not take in to account any structural complexity or regional variation. However, SSIM and NSER did take in to account structural complexity and NSER took this a step further, where edge locations were determined during analysis, which then described loss of edge definition and image blur (Zhang et al., 2011). Nevertheless analysis obtained using these two IQMs still involved only the whole image and not regional analysis. There was an outlier for the PSNR dataset acquired without attenuation jackets at 10mA, which included the -630HU

simulated lesion. A transient equipment issue may have accounted for this, and highlights the need for multiple acquisitions which could have supported exclusion of this data.

NPS, as a function in the frequency domain providing a visualization of noise texture in an image, this IQM could be used in conjunction with other IQMs which obtain discrete values for image data comparison (Bushberg et al., 2012a). As the shape of the NPS curves showed change with image content, comparison would have to be made over exactly the same region when visualising the differences between two images. NPS produced lower power values when attenuation jackets were used, with noise contributing to a greater degree in the higher frequencies. Increased amperage affected the NPS, where, without the use of attenuation jackets, peak location increased in frequency, potentially as a result of increasing detail available in the source image.

Texture analysis involved five IQMs (Contrast, Correlation, Energy, Homogeneity and Entropy) as descriptors to assess the effect of amperage on image quality. Gray-Level-Co-occurrence-Matrix (GLCM) data was used rather than histogram features because histogram features, as first-order features, do not take into account spatial relationships (Mayerhoefer et al., 2020). GCLM data is second-order, which includes multi-directional spatial relationships of pixels or voxels (Mayerhoefer et al., 2020). Texture analysis afforded the advantage of performing regional analyses, however, not all of these IQMs proved useful in the context of this research. Contrast and Correlation, looking at intensity and correlation with neighbouring pixels respectively, were affected by the lung region analysed which caused a lack of consistency in the data obtained (Mathworks, 1994-2020). This is not surprising given the structural variation in the region being imaged. Values for Energy (or uniformity) and Homogeneity (looking at spatial variations) were expected to increase as amperage increased, due to images becoming less noisy, and this did indeed happen. While useful in a comparative assessment of image quality, absolute values would not be particularly useful for a large region of the lungs which is radiographically high-contrast and non-uniform. Entropy (looking at variability or randomness) had an inverse relationship to Energy, as one would expect, where values decreased as amperage increased. This IQM would also be useful in a comparative assessment of image quality. The use of

attenuation jackets proved to have a significant impact on the results of IQM used in the texture analysis.

Observer Performance:

Observer performance studies were carried out with participant observers having a range of imaging expertise and backgrounds, where a combination of years of experience and annual caseload impacted on their CT viewing experience. This was intended to facilitate assessment of image quality without bias relating to individual clinical ability. Using fulfillment of international image quality criteria (EUR16262), visual grading characteristics (VGC) analysis demonstrated that image noise impacted less on fulfilment of criteria for the sharpness of proximal structures, and was statistically significant for visualization of peripheral vessels (Bath & Hansson, 2016; EC., 2000). Proximal structures are more substantial in nature, therefore the impact of noise in detecting degradation in the sharpness of these structures has possibly been underestimated because of this. As expected, fulfilment of image quality criteria did show an improvement with increasing amperage, particularly at lower amperage, however this was not linear in nature and occurred with some undulating patterns as with the IQMs. In addition, the profile of this response could vary depending on the region and structures being assessed, for example in this research, fulfilment of the image quality criterion for clear visualization of peripheral vessels was less satisfactory in the upper zone, and more satisfactory in the lower zone.

Receiver operating characteristic (ROC) analysis for a lesion detection task did not demonstrate any significant difference in lesion localisations at increasing amperage for varying densities. This could be due in part to the adaptability of the human visual system (HVS) (Gonzalez & Woods, 2008a). However, further analysis of the data from this detection task revealed that a high percentage of missed lesions had a lesion density of 100HU, which is similar in density to the multiple surrounding structures. In addition, VGC time-case-lesion analysis showed that while there was not one image that particularly caused extra time to be taken, five out of seven observers recorded increased time taken for images with 100HU simulated lesions. In addition, images with 100HU simulated lesions caused the highest percentage of time taken overall, with percentages of 18.3% (without attenuation jackets) and 15.2 % (with attenuation jackets). The

next group, for time taken overall, was 9.1% for normal images. Three out of seven observers recorded increased time taken when attenuation jackets were used, for images containing a range of simulated lesion densities and sizes, over a range of amperages. These findings also relate to the concept of conspicuity, which, while not analysed specifically, is inherent in this research which assessed clear visualization of structures or simulated lesions in the context of the complexity of their surroundings (Kundel & Revesz, 1976).

Correlations:

Statistical correlations were carried out between the best performing IQM (MSE, PSNR, SSIM, NSER and texture analysis IQMs looking at Energy, Homogeneity and Entropy). This performance was based on whether the IQM had good correlation in regression analysis between the individual datasets, with and without attenuation jackets separately, which meant that their consistent performance made them more useful. Pearson's r values described the strength of linear relationships that were least affected by use of attenuation jackets, which were NSER, Homogeneity and Entropy. The IQM where the strength of the linear relationship was most affected by use of attenuation jackets, was PSNR. Spearman's ρ , using ranked data, described the strength of monotonic relationships least affected by the use of attenuation jackets, as being NSER, Energy, and Homogeneity. The IQM where the strength of the monotonic relationship was most affected by use of attenuation jackets, using Spearman's ρ , was MSE. It is worth noting, however, that these correlations do not rule out non-linear relationships.

When correlating IQM and observer performance data, a frequency meta-analysis of Spearman's ρ correlation coefficients, Homogeneity, Entropy, NSER and Energy showed the strongest correlation with observer performance. As these are structurally based analytics this is as one would expect. MSE showed the worst correlation with observer performance. There is no individual IQM that stands out from this group of leading IQMs, however the advantage of the texture analysis IQMs (Homogeneity, Entropy and Energy) readily facilitating regional analysis, may make them more suitable when assessing the image quality required for specific tasks. In addition, even strong correlation does not necessarily indicate

equivalency, however it is a step closer to identifying a suitable IQM (Shaughnessy et al., 2012).

Evaluating visualisation of peripheral vessels, was an observer task directed at the quality of an image allowing detail in an image to be visualised. Spearman's ρ coefficients, when comparing the results of this observer task to the group of leading IQMs, the values ranged from $\rho = 0.36$ to 0.74 or $\rho = -0.36$ to -0.74 . These values indicate moderate to strong monotonic relationships between these IQMs and human observers.

When comparing the results of the observer task, evaluating sharpness of proximal structures with the group of leading IQMs, the coefficient values had a far broader range (Spearman's $\rho = 0.03$ to 0.91 , or $\rho = -0.03$ to -0.91). These values indicate a strength of monotonic relationships between these IQMs and human observers, ranging from virtually non-existent to strong. The lack of agreement may be in part due to the task itself, where, as already mentioned, there may be difficulties in assessing degradation of larger structures in an image.

Limitations:

Simulated lesion placement was intended to blend with structures in the upper, middle, and lower lung zones, therefore they were not taped, and they were changed between image acquisition sets of each simulated lesion density. Subsequently, they may not have been in precisely the same position between each set of images due to the fragile nature of the lung inserts in the imaging phantom. In addition, movement may have occurred during the placement of attenuation jackets. However, the exact slice containing each simulated lesion was selected for image analysis and image viewing.

In addition to the findings for this research being specific for the radiographic high-contrast and multi-structured lung fields, the simulated lesions were of a spherical nature. In the clinical environs, this is not necessarily the case. Gestalt theories of perception, particularly the aspects of closure and continuity, would suggest that these simulated lesions were more easily detected because of their spherical nature (Sabih et al., 2011).

Only single image data sets were obtained for each of the lung zones (upper, middle and lower) which included each density of the simulated lesions, plus 'normal' sets of image data, with and without attenuation jackets. This limits assessment of the degree of uncertainty of the data presented.

Statistical correlations, so far, have only examined monotonic relationships between IQMs and human observers. Further work on this could potentially be valuable and contribute to the increasing international body of work. In addition, inter -correlations between IQMs did not rule out non-linear relationships.

Clinical context:

In clinical practice, establishing acquisition parameters to ensure suitable image quality for the purpose intended is essential to optimize radiation dose, while at the same time allowing the clinician to carry out their diagnostic task with confidence and avoidance of error. This facilitates justification of the radiation dose, and further informs any changes to settings if required for specific patient or study requirements. Image acquisition using dose modulation was not used for the purposes of this research, however this is used in the clinical context, and mitigates the impact of patient size on lesion localization and image artefact, to some degree.

This research demonstrated the effect of attenuation jackets, which translates into body habitus in the clinical setting, which may vary with patient size, clinical treatment, or drugs. In addition, the region of the body being examined and the complexity of lesion surrounds was demonstrated in the IQM analysis. This highlights the importance of using appropriate phantoms with attenuation tissue mimicked in some way. Proper calibration and optimisation of the imaging equipment, is of course a prerequisite. Some IQMs for assessment of image quality have been demonstrated to mimic the HVS more closely, which would indicate suitability for assessing image quality when establishing acquisition parameters for clinical use.

Future work:

Future work could look at the challenges associated with positioning of simulated lesions for consistency in lesion localization tasks, that would be undetectable to human observers, and to prevent movement when placing attenuation jackets. In addition, using different shaped lesions with less defined borders, and randomly placed, would better mimic clinical reality. Changing the variable acquisition parameter could add further depth to results obtained. Not only looking at LDCT, would give a broader perspective. In addition, performing acquisitions across a number of imaging platforms could assess the generalisability of results across CT scanners. Using dose modulation would make further work more clinically relevant, and altering other parameters would add greater depth to the body of knowledge.

Performing a large number of acquisitions producing multiple image data sets over lung zones and each density of simulated lesions, would increase the degree of certainty for the data analysis. An increased number of human observers may be useful in determining whether patterns in undulations of the data can give more definitive indications for dose reduction. Further work in observer performance looking at how a lesion in an area affects image perception and further work on how different surroundings impact on diagnosis, may be useful.

Observer performance studies, using sliding scales that produce continuous data for recording fulfillment of image quality criteria, would be advantageous. This would overcome the limitations of using ordinal data and allow higher levels of analysis, facilitating greater sensitivity in results. Caution is required to ensure the sliding scale is not 'binning' the data in the background so that resultant values are ordinal, and not continuous as desired.

Development of new IQMs, perhaps using radiomics and artificial intelligence, could open new frontiers in image quality measurement. Consequently, increased accuracy could be applied to radiation dose reduction in image acquisition.

A multidisciplinary approach would be invaluable in future work, involving technologists and radiographers, mathematicians and statisticians, physicists, computer programmers, and image observers from a broad range of imaging expertise and background. This is important because obtaining images, reporting images, and analysing image content is multi-faceted. People from

all these specialties working together could contribute their depth of knowledge and expertise on a larger scale, from image acquisition and observer performance through to comprehensive results analysis and implications. As a number of software packages were used for analysis in this research, it would also be useful to develop integrated software performing a range of IQM's, to ensure optimized equipment performance, and a range of signal-based IQMs along with more perceptually based IQM's.

9.0 Chapter Nine: Conclusion

The object of this research was to examine image quality in low-dose CT (LDCT), using task-specific observer performance and image quality metrics, while identifying relationships between image quality and human perception. The international dilemma of how to assess image quality in medical imaging has been contemplated for decades, as has the controversy surrounding how to relate image quality to human observers. This challenge is increasingly apparent in the continuing effort to lower radiation dose to patients, while producing images that are suitable for the specific task for which they were intended. Image quality is not just about noise, it is also about edge definition, resolution of detail, the character of physical features in the images and their surroundings. All this, combined with the adaptability of the human visual system, contributes to the difficulties faced when attempting to find suitable image quality measures (IQM).

While these research findings are specific to the radiographic high-contrast region of the thorax, the findings have wider implications for image quality measurement in general, and impact in the clinical setting. Results of most IQMs demonstrated a significant difference when attenuation jackets were used. This highlights the importance of considering body habitus when performing image quality assessment, and whether annuli or attenuation jackets should be a standard attachment when performing these measurements, particularly when translating between CT scanner platforms.

A variety of IQMs were used to assess image characteristics such as image noise, image contrast, edge definition and surrounding structures. These IQMs ranged from those that were more signal-based such as SNR, CNR and PSNR, to those that were more structure-based and closer to approximating the human visual system, such as NSER and texture analysis. Texture analysis provided another dimension to the possibilities of available image quality metrics (IQM), but not all texture analysis IQMs behaved in the same way. The traditional IQMs of SNR and CNR were not useful in this research, particularly because of the high level of structural variability and the high-contrast nature of the region being assessed. However, this does not exclude their

usefulness in optimising equipment performance, and they may be suitable in regions of the body where homogenous background regions can be obtained, such as the abdomen.

The quality of the images did indeed improve with increasing amperage as expected, although this response was non-linear and there were undulations in the graphed data. These undulations were examined for any emerging patterns, and while no pattern was detected overall, except where the level of response plateaued between 90mA and 100mA, there were some patterns starting to emerge when using some IQMs. Commonalities in these patterns were sought after with a view to indicating any potential for radiation dose reduction at specific amperage.

Observer performance studies were carried out with two tasks for visual grading characteristics (VGC) using international CT quality criteria (EUR16262), and a lesion detection task for receiver operating characteristic (ROC) analysis. The participating observers had a wide range of expertise and image viewing backgrounds. The lesion localization task did not indicate significant differences for increasing amperage or varying densities, potentially due to the adaptability of the human visual system, and the spherical nature of the simulated lesions. The visual grading task for clear visualisation of peripheral vessels showed that noise impacted significantly on fulfilment of image criteria, with varying amperage, but noise did not impact as much on the other visual grading task relating to sharpness of proximal structures - the observer responses were more varied, however.

The leading IQMs showing the strongest inter-correlations for linear relationships, were NSER, Energy (uniformity) and Homogeneity and an inverse relationship with Entropy. When correlated with the responses of human observers, this same group of IQMs showed the strongest monotonic relationships. In addition, the responses for the fulfilment of image quality criteria for visualisation of peripheral vessels looking at image detail, showed in particular a moderate to strong correlation with these IQM.

The research has brought to light these points which may be useful to follow when assessing image quality: (1) it is important to define the task for which the image is needed, with the detail required in an image and the level of image artefact that would be acceptable; (2) obtain a profile of specific equipment response; (3) it may be advantageous to use anthropomorphic phantoms,

however for all phantom work, when assessing image quality, use attachments that simulate increased body habitus; and (4) determine appropriate IQMs for image analysis, including the need for regional/lesion analysis or not, and one which can then be related to human observer responses.

The information gained from this process can inform clinicians about viewing scans of different quality and their expectations of the images. This would impact on decision making of clinicians when determining CT acquisition parameters for the required diagnostic task, and on their confidence levels when reporting, as well as ongoing quality assessment of images. When changing acquisition parameters, they can then ensure that the radiation exposure has been increased enough to have diagnostic benefit, or it has been able to be reduced without loss of diagnostic utility. There is potential to distinguish between image dependent errors and non-image dependent errors. There would also be an impact on training regimens which could be established to recognise the significance of lesion detection at lower radiation doses.

In conclusion, this research demonstrated that NSER, Energy (uniformity), Homogeneity and Entropy had moderate to strong correlation with human observers. In addition, the importance of simulating body habitus was highlighted, when assessing image quality for the task intended. In a radiographic high-contrast region with a complex structural background, using an IQM that takes in to account spatial correlations would seem essential. Going forward, an integrated approach is needed to optimise radiation dose, correctly simulate the region of the body to be examined, and choose appropriate IQMs to approximate the human visual system.

References:

- Abbey, C. K. (2013). *Classification images aid understanding of visual task performance and diagnosis*. SPIE Newsroom. Retrieved 23/02/2014 from <http://spie.org/x92089.xml?pf=true&ArticleID=x92089>
- Al-Hinnawi, A. R., & Daear, M. (2012). Image Texture Descriptors to Quantify Bilateral Filter on Low Dose Computerized Tomography. *International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 5, No. 3, September, 2012, 5(3)*, 123-135.
- ALA. (2017). *Basic Types of Lighting | American Lighting Association*. Retrieved 19/11/2017 from <https://www.americanlightingassoc.com/Lighting-Fundamentals/3-Types-of-Lighting.aspx>
- Aldrich, J. (1995). Correlations Genuine and Spurious in Pearson and Yule. *Statistical Science, 10(4)*, 364-376.
- Altman, D. (1991). *Practical Statistics for Medical Research* (First ed.). Chapman & Hall. <https://doi.org/10.1002/sim.4780101015>
- Ashman, C. J., Yu, J. S., & Wolfman, D. (2000). Satisfaction of search in osteoradiology. *AJR Am J Roentgenol, 175(2)*, 541-544. <https://doi.org/10.2214/ajr.175.2.1750541>
- Aziz, Z., Padley, S., & Hansell, D. (2004). CT techniques for imaging the lung: recommendations for multislice and single slice computed tomography. *European Journal of Radiology, 52*, 119-136. <https://doi.org/10.1016/j.ejrad.2004.01.005>
- Bankier, A., MacMahon, H., Goo, J.M., Rubin, G., Schaefer-Prokop, C., Naidich, M. (2017). Recommendations for Measuring Pulmonary Nodules at CT: A Statement from the Fleischner Society. *Radiology, 285(2)*, 584-600. <https://doi.org/10.1148/radiol.2017162894>
- Barnes, E. (2013a). *CT screening: Few ground-glass nodules go rogue, but 10% grow*. Auntminnie.com. Retrieved 31/12/2013 from http://www.auntminnie.com/index.aspx?sec=rca&rsna_2013&pag=dis&ItemID=105908
- Barnes, E. (2013b). *Data sampling technique cuts CT lung dose to 0.25 mSv*. Auntminnie.com. Retrieved 31/12/2013 from <http://www.auntminnie.com/index.aspx?sec=sup&cto&pag=dis&ItemID=106074>
- Barrett, H. (n.d.) *An Introduction To Objective Assessment Of Image Quality*. Retrieved 27/12/2013 from http://optics.nuigalway.ie/people/barrett/Lecture_1.pdf
- Barrett, H., & Myers, K. (2003). Image Quality. In *Foundations of Image Science*. (pp. 913-1000). John Wiley & Sons, Inc.
- Barrett, H., & Myers, K. (2003b). Image Quality. In *Foundations of Image Science*. (pp. 925). John Wiley & Sons, Inc.
- Bath, M., & Hansson, J. (2016). VGC Analyzer: A Software For Statistical Analysis Of Fully Crossed Multiple-Reader Multiple-Case Visual Grading Characteristics Studies. *Radiation Protection Dosimetry, 169(1-4)*, 46-53. <https://doi.org/10.1093/rpd/ncv542>
- Bell, D., & Niknejad, M. (2005-2021). *Contrast resolution*. Radiopaedia.org. Retrieved 14/01/2021 from <https://radiopaedia.org/articles/contrast-resolution?lang=us>
- Berbaum, K., Franken Jr, E., Dorfman, D., Barloon, T., Ell, S., Lu, C., Smith, W., & Abu-Yousef, M. (1986). Tentative diagnoses facilitate the detection of divers lesions in chest radiographs. *Radiology, 21*, 532-539.
- Berland, L. L., Silverman, S. G., Gore, R. M., Mayo-Smith, W. W., Megibow, A. J., Yee, J., Brink, J. A., ...& Taylor, A. J. (2010). Managing Incidental Findings on Abdominal CT: White Paper of the ACR Incidental Findings Committee. *JACR, 7(10)*, 754-773. <https://doi.org/10.1016/j.jacr.2010.06.013>

- Berta, L., Mascaro, L., Feroldi, P., & Maroldi, R. (2014). Optimisation of an MDCT abdominal protocol: Image quality assessment of standard vs. iterative reconstructions. *Phys Med*, *30*(3), 271-279. <https://doi.org/10.1016/j.ejmp.2013.07.126>
- Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical care*, *7*(6), 451-459. <https://doi.org/10.1186/cc2401>
- Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 13: Receiver operating characteristic curves. *Critical Care*, *8*(6), 508-512. <https://doi.org/10.1186/cc3000>
- Bhosale, P., Wagner-Bartak, N., Wei, W., Kundra, V., & Tamm, E. (2015). Comparing CNR, SNR, and image quality of CT images reconstructed with soft kernel, standard kernel, and standard kernel plus ASIR 30% techniques. *International Journal of Radiology*, *2*(2), 60-65.
- Bobbitt, Z. (2018). *Mann-Whitney U test*. Statology.org. Retrieved 10/01/2021 from <https://www.statology.org/mann-whitney-u-test/>
- Bochud, F. O., Valley, J. F., Verdun, F. R., Hessler, C., & Schnyder, P. (1999). Estimation of the noisy component of anatomical backgrounds. *Medical Physics*, *26*(7), 1365-1370. <https://doi.org/10.1118/1.598632>
- Boedeker, K., & McNitt-Gray, M. (2007). Application of the noise power spectrum in modern diagnostic MDCT: part II. Noise power spectra and signal to noise. *Physics in Medicine and Biology* *52*, 4047-4061. <https://doi.org/10.1088/0031-9155/52/14/003>
- Burgess, A. (2010a). Signal detection in radiology. In E. Samei & E. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (First published 2010; Paperback 2014 ed., pp. 47-72). Cambridge University Press. ISBN: 978-1-107-42463-0
- Burgess, A. (2010b). Signal detection theory - a brief history. In E. Samei & E. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (First published 2010; Paperback 2014 ed., pp. 26-46). Cambridge University Press. ISBN: 978-1-107-42463-0
- Burgess, A. (2011). Visual Perception Studies and Observer Models in Medical Imaging. *Seminars in Nuclear Medicine*, *41*, 419-435. <https://doi.org/10.1053/j.semnuclmed.2011.06.005>
- Bushberg, J., Seibert, J. A., Leidholdt, E., & Boone, J. (2012a). Image quality. In *The Essential Physics of Medical Imaging* (3rd ed., pp. 60-100). Walters Kluwer - Lippincott Williams & Wilkins.
- Bushberg, J., Seibert, A., Leidholt, E., & Boone, J. (2012b). Computed tomography. In *The Essential Physics of Medical Imaging* (3rd ed., pp. 312-374). Walters Kluwer - Lippincott Williams & Wilkins. ISBN: 978-1-4511-1810-0
- Bushberg, J., Seibert, A., Leidholt, E., & Boone, J. (2012c). Computed tomography. In *The Essential Physics of Medical Imaging* (3rd ed., p. 91). Walters Kluwer - Lippincott Williams & Wilkins. ISBN: 978-1-4511-1810-0
- Bushberg, J., Seibert, A., Leidholt, E., & Boone, J. (2012d). Computed tomography. In *The Essential Physics of Medical Imaging* (3rd ed., p. 364). Walters Kluwer - Lippincott Williams & Wilkins. ISBN: 978-1-4511-1810-0
- Bushberg, J., Seibert, A., Leidholt, E., & Boone, J. (2012e). Computed tomography. In *The Essential Physics of Medical Imaging* (3rd ed., p. 92). Walters Kluwer - Lippincott Williams & Wilkins. ISBN: 978-1-4511-1810-0
- Båth, M. (2010). Evaluating imaging systems: practical applications. *Radiation Protection Dosimetry*, *139*(1-3), 26-36. <https://doi.org/10.1093/rpd/ncq007>
- Båth, M., Håkansson, M., Börjesson, S., Kheddache, S., Grahn, A., Ruschin, M., Tingberg, A., Mattsson, S., & Månsson, L. G. (2005). Nodule detection in digital chest radiography: introduction to the RADIUS chest trial. *Radiation Protection Dosimetry*, *114*(1-3), 85-91. <https://doi.org/10.1093/rpd/nch575>

- Båth, M., & Månsson, L. (2007). Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation. *BJR*, *80*(951), 169-177. <https://doi.org/http://dx.doi.org/10.1259/bjr/35012658>
- Börjesson, S., Håkansson, M., Båth, M., Kheddache, S., Svensson, S., Tingberg, A., Grahn, A., Ruschin, M., Hemdal, B., Mattsson, S., & Månsson, L. (2005). A software tool for increased efficiency in observer performance studies in radiology. *Radiat Prot Dosimetry*, *114*(1-3), 45-52. <https://doi.org/10.1093/rpd/nch550>
- Carrino, J. (2002). Digital image quality: a clinical perspective. In B. Reiner, E. Siegel, & J. Carrino (Eds.), *Quality Assurance Challenges in the Digital Medical Enterprise*. Society for Computer Applications in Radiology. Retrieved 31/08/2014 from <http://www.aapm.org/meetings/02am/pdf/8469-75819.pdf>
- Casey, B. (2017). *Study: CT screening would cut lung cancer deaths in China*. Auntminnie.com. Retrieved 31/07/2017 from <http://www.auntminnie.com/index.aspx?sec=supc=cto&pag=dis&ItemID=117838&wf=7773>
- Catphan 500 and 600 Manual. (2015). In: The Phantom Laboratory. Retrieved 22/09/2019 from <https://www.phantomlab.com/>
- CATPHAN 600 [Brochure 'CATPHAN 500/600']. (n.d.). In: The Phantom Laboratory. Catphan accessories. (2018). Retrieved 17/11/2017 from <https://www.phantomlab.com/catphan-accessories>
- Chakraborty, D. (2010). A Status Report On Free-Response Analysis. *Radiation Protection Dosimetry*, *139*(1-3), 20-25.
- Chakraborty, D. (2011a). *The ROC task; The free-response task; JAFROC. (Updated 2/3/11)*. Retrieved 11/11/2017 from www.devchakraborty.com
- Chakraborty, D. P. (2011b). New Developments in Observer Performance Methodology in Medical Imaging. *Seminars in Nuclear Medicine*, *41*(6), 401-418. <https://doi.org/10.1053/j.semnuclmed.2011.07.001>
- Chakraborty, D. P. (2013). A brief history of FROC paradigm data analysis. *Acad Radiol*, *20*(7), 915-919. <https://doi.org/10.1016/j.acra.2013.03.001>
- Chakraborty, D. P. (2015). *Package 'RJafroc'. Analysis of Data Acquired Using the Receiver Operating Characteristic Paradigm and Its Extensions*. Retrieved 11/11/2017 from <https://cran.r-project.org/web/packages/RJafroc/RJafroc.pdf>
- Chen, H. (2017). *What are 'frequency bins'*. Mathworks. Retrieved 14/01/2021 from www.mathworks.com/matlabcentral/answers/322142-what-are-frequency-bins
- Chen, W., Petrick, N. A., & Sahiner, B. (2012). Hypothesis Testing in Noninferiority and Equivalence MRMC ROC Studies. *Acad Radiol*, *19*, 1158-1165. <https://doi.org/10.1016/j.acra.2012.04.011>
- Chesters, M. S. (1992). Human visual perception and ROC methodology in medical imaging. *Physics in Medicine and Biology*, *37*(7), 1433-1476. <https://doi.org/10.1088/0031-9155/37/7/001>
- Choi, H., & Lee, C. (2011). No-reference image quality metric based on image classification. *EURASIP Journal on Advances in Signal Processing*, *2011*(65). Retrieved 25/08/2014 from <https://asp-urasipjournals.springeropen.com/articles/10.1186/1687-6180-2011-65>
- CMS. (2009). *Decision Memo for Screening Computed Tomography Colonography (CTC) for Colorectal Cancer (CAG-00396N)*. Centers for Medicare and Medicaid services. Retrieved 03/08/2017 from [https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=220&ver=14&NcaName=Screening+Computed+Tomography+Colonography+\(CTC\)+for+Colorectal+Cancer&bc=BEAAAAAAEAAA&&fromdb=true](https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=220&ver=14&NcaName=Screening+Computed+Tomography+Colonography+(CTC)+for+Colorectal+Cancer&bc=BEAAAAAAEAAA&&fromdb=true)
- Cody, D., Kim, H.-J., Cagnon, C., Larke, F., McNitt-Gray, M., Kruger, R., Flynn, M., Seibert, A., Judy, P., & Wu, X. (2010). Normalized CT Dose Index of the CT Scanners Used in the National Lung

- Screening Trial. *American Journal of Roentgenology*, 194(6), 159-166.
<https://doi.org/10.2214/AJR.09.3268>
- Computed Tomography [RadiologyKey]. (2016). Retrieved 02/08/2017 from
<https://radiologykey.com/computed-tomography-3/>
- Cox, D. R. (1984). Interaction. *International Statistical Review / Revue Internationale de Statistique*, 52(1), 1-24. <https://doi.org/10.2307/1403235>
- Dendy, P., & Heaton, B. (2013). *Optimising Perception [in Physics for Diagnostic Radiology]* (3rd ed., pp. 249-250).
- Deng, F., & Murphy, A. (2005-2021). *Radiographic contrast*. Radiopaedia.org. Retrieved 14/01/2021 from <https://radiopaedia.org/articles/radiographic-contrast?lang=us>
- Dirks, S., Yu, L., Vrieze, T., Favazza, C., Ferrero, A., Leng, S., Weaver, J., & McCollough, C. (2017). Efficient and Quantitative Measurement of Task-Based Image Quality Metrics in CT. AAPM 2017 59th Annual Meeting (American Association of Physicists in Medicine), Denver, Colorado.
- Dougherty, G. (2009a). *Digital Image Processing for Medical Applications*. Cambridge University Press. (p. xi). Retrieved 30/07/2017 from
<http://ultra.sdk.free.fr/docs/DxO/Digital%20Image%20Processing%20for%20Medical%20Applications.pdf>
- Dougherty, G. (2009b). *Digital Image Processing for Medical Applications*. Cambridge University Press. (p. 3). Retrieved 30/07/2017 from
<http://ultra.sdk.free.fr/docs/DxO/Digital%20Image%20Processing%20for%20Medical%20Applications.pdf>
- Duan, X., Wang, J., Leng, S., Schmidt, B., Allmendinger, T. h., Grant, K., Flohr, T., & McCollough, C. H. (2013). Electronic Noise in CT Detectors: Impact on Image Noise and Artifacts. *AJR*, 201, W626-W632. <https://doi.org/10.2214/AJR.12.10234>
- EC. (2012). *RADIATION PROTECTION N° 162: Criteria for Acceptability of Medical Radiological Equipment used in Diagnostic Radiology, Nuclear Medicine and Radiotherapy*. European Commission: Quality Assurance Reference Centre. <https://doi.org/10.2768/22561>
- Eckert, M., & Bradley, A. (1998). Perceptual quality metrics applied to still image compression. *Signal Processing*, 70, 177-200.
- Engstrom, E., Reiser, I., & Nishikawa, R. (2009). Comparison of power spectra for tomosynthesis projections and reconstructed images. *Medical Physics*, 36(5), 1753-1758.
<https://doi.org/10.1118/1.3116774>
- ESR. (2011). Usability of irreversible image compression in radiological imaging. A position paper by the European Society of Radiology. *Insights Imaging*, 2, 103-115. <https://doi.org/10.1007/s13244-011-0071-x>
- EC. (2000). *European Guidelines Quality Criteria Computed Tomography EUR16262*. European Commission's Radiation Protection Actions. Retrieved 30/11/2014 from
<https://op.europa.eu/en/publication-detail/-/publication/d229c9e1-a967-49de-b169-59ee68605f1a>
- Fanous, R., Kashani, H., Jimenez, L., Murphy, G., & Paul, N. S. (2012). Image quality and radiation dose of pulmonary CT angiography performed using 100 and 120 kVp. 99(5), 990-996.
<https://doi.org/10.2214/AJR.11.8208>
- Farlex_Partner_Medical_Dictionary. (2012). *Psychophysical*. Thefreedictionary.com. Retrieved 14/01/2021 from <https://medical-dictionary.thefreedictionary.com/psychophysical>
- Fischella, V., Båth, M., Allansdotter Johnsson, A., Jäderling, F., Bergsten, T., Persson, U., Mellingen, K., & Hellström, M. (2010). Evaluation of image quality and lesion perception by human readers on 3D CT colonography: comparison of standard and low radiation dose. *Eur Radiol.*, 20, 630-639.
<https://doi.org/10.1007/s00330-009-1601-5>

- Funama, Y., Awai, K., Nakayama, Y., Kakei, K., Nagasue, N., Shimamura, M., Sato, N., Sultana, S., Morishita, S., & Yamashita, Y. (2005). Radiation dose reduction without degradation of low-contrast detectability at abdominal multisection CT with a low-tube voltage technique: phantom study. *Radiology*, *237*, 905-910. <https://doi.org/http://dx.doi.org/10.1148/radiol.2373041643>
- Funama, Y., Taguchi, K., Utsunomiya, D., Oda, S., Katahira, K., Tokuyasu, S., & Yamashita, Y. (2014). Image quality assessment of an iterative reconstruction algorithm applied to abdominal CT imaging. *Phys Med*, *30*(4), 527-534. <https://doi.org/10.1016/j.ejmp.2014.02.005>
- Gaillard, F., Yeung, J., & et al. (n.d.). *Signal-to-noise ratio*. Radiopaedia. Retrieved 09/01/2020 from <https://radiopaedia.org/articles/signal-to-noise-ratio-1?lang=gb>
- Gayou, O. (2012). Influence of acquisition parameters on MV-CBCT image quality. *Journal of Applied Clinical Medical Physics*, *13*(1), PACS numbers: 87.57.C-; 87.57.nf. <https://doi.org/10.1120/jacmp.v13i1.3638>
- Gebejes, A., & Huertas, R. (2013). Texture Characterization based on Grey-Level Co-occurrence Matrix. In *Proceedings of the 2nd Information and Communication Technologies-International Conference ICTIC. Žilina, Slovakia: 25-29 March 2013* (pp. 375–378). Retrieved 02/06/2015 from <http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=0CDQQFjAD&url=http%3A%2F%2Fwww.ictic.sk%2Farchive%2F%3Fvid%3D1%26aid%3D3%26kid%3D50201-78%26q%3Df1&ei=QGJzVajNOYvuUvaHgKgL&usq=AFQjCNFlodJRf5G5FwJXc315Z2Wx9meelQ>
- Gervaise, A., Osemont, B., Louis, M., Lecocq, S., Teixeira, P., & Blum, A. (2014). Standard dose versus low-dose abdominal and pelvic CT: comparison between filtered back projection versus adaptive iterative dose reduction 3D. *Diagn Interv Imaging*, *95*(1), 47-53. <https://doi.org/10.1016/j.diii.2013.05.005>
- Geyer, L. L., Schoepf, U. J., Meinel, F. G., Nance, J. W. J., Bastarrika, G., Leipsic, J. A., Paul, N. S., Rengo, M., Laghi, A., & De Cecco, C. N. (2015). State of the art: iterative CT reconstruction techniques. *Radiology*, *276*(2), 339-357. <https://doi.org/10.1148/radiol.2015132766>
- Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, *25*(2), 141-151. <https://doi.org/10.11613/BM.2015.015>
- Gierada, D. S., Bierhals, A. J., Choong, C. K., Bartel, S. T., Ritter, J. H., Das, N. A., Hong, C., Pilgram, T. K., Bae, K. T., Whiting, B. R., Woods, J. C., Hogg, J. a. C., Lutey, B. A., Battafarano, R. J., Cooper, J. D., Meyers, B. F., & Patterson, G. A. (2010). Effects of CT section thickness and reconstruction kernel on emphysema quantification: relationship to the magnitude of the CT Emphysema Index. *Acad Radiol.*, *17*(2), 146(141-124). <https://doi.org/10.1016/j.acra.2009.08.007>
- Gonzalez, R., & Woods, R. (2008a). Digital image fundamentals. In *Digital Image Processing* (3rd Edit., pp. 36-43). Pearson Prentice Hall. ISBN-13: 9780135052679
- Gonzalez, R., & Woods, R. (2008b). Representation and description. In *Digital Image Processing* (3rd Edit., pp.795-856). Pearson Prentice Hall. ISBN-13: 9780135052679
- Graphpad. (2020). *Graphpad statistics guide: Interpreting skewness and kurtosis*. www.graphpad.com. Retrieved 31/08/2014 from https://www.graphpad.com/guides/prism/8/statistics/stat_skewness_and_kurtosis.htm
- Gurney, J. (1996). Missed lung cancer at CT: imaging findings in nine patients. *Radiology*, *199*(1), 117-122. <https://doi.org/org/10.1148/radiology.199.1.8633132>
- Hanson, K. M. (1981). Noise and contrast discrimination in computed tomography. In T. Newton & G. Potts (Eds.), *Radiology of the Skull and Brain* (Vol. 5, pp. 3941-3955). C.V. Mosby. ISBN: 9780801636622
- Hansson, J., Månsson, L. G., & Båth, M. (2016). The validity of using ROC software for analysing visual grading characteristics data: an investigation based on the novel software VGC Analyzer. *Radiat Prot Dosimetry*, *169*(1-4), 54-59. <https://doi.org/https://doi.org/10.1093/rpd/ncw035>

- Hoeschen, C., Tischenko, O., Buhr, E., & Illers, H. (2005). Comparison of technical and anatomical noise in digital thorax X-ray images. *Rad. Prot. Dosim.*, *114*, 75-80. <https://doi.org/doi:10.1093/rpd/nch543>
- Hsieh, J. (1995). Image artifacts, causes and correction. In L. W. Goldman & J. B. Fowlkes (Eds.), *Medical CT and ultrasound: current technology and applications*. American Association of Physicists in Medicine. ISBN: 1883526035
- Hunt, B., Lipsman, R., & Rosenberg, J. (2006). *A guide to MATLAB® for Beginners and Experienced Users* (2nd Edition ed., pp.xi-xv). Cambridge University Press. ISBN: 978-0-521-61565-5
- Håkansson, M., Svensson, S., Zachrisson, S., Svalkvist, A., Båth, M., & Månsson, L. (2010). VIEWDEX: an efficient and easy-to-use software for observer performance studies. *Radiation Protection Dosimetry*, *139*(1-3), 42-51. <https://doi.org/10.1093/rpd/ncq057>
- Håkansson, M., Svensson, S., Zachrisson, S., Svalkvist, A., Båth, M., & Månsson, L. G. (2009). ViewDEX 2.0: a Java-based DICOM-compatible software for observer performance studies. In *Proceedings of SPIE - The International Society For Optical Engineering, Medical Imaging, March 12, 2009*, 7263, 72631G 72631-72610. <https://doi.org/10.1117/12.811511>
- ICRU. (1996). *Medical Imaging - The Assessment of Image Quality (Report 54)*. Issued 08/04/1996. International Commission on Radiation Units and Measurement, Bethesda, USA.
- ICRU. (2012). Report 87. Radiation Dose and Image-Quality Assessment in Computed Tomography. *Journal of the International Commission on Radiation Units and Measurements*, *12*. Sage publishing. <https://doi.org/10.1093/jicru/ndt007>
- Itoh, S., Ikeda, M., Arahata, S., Kodaira, T., Isomura, T., Kato, T., Yamakawa, K., Maruyama, K., & Ishigaki, T. (2000). Lung cancer screening: minimum tube current required for helical CT. *Radiology*, *215*(1), 175-183. <https://doi.org/10.1148/radiology.215.1.r00ap16175>
- Itti, L. (2007). Visual salience. *Scholarpedia*, *2*(9), 3327. <https://doi.org/10.4249/scholarpedia.3327>
- Jessop, M. (2015). *Task-specific observer performance and image quality: direct and indirect relationships in low-dose CT images*. [Research summary: REC reference 15/LO/0739]. NHS Health Research Authority. Retrieved 14/01/2021 from <https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/observer-performance-and-image-quality-in-low-dose-ct/>
- Jessop, M., Thompson, J. D., Coward, J., Sanderud, A., Jorge, J., de Groot, M., Lanc, a, L. s., & 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. <https://doi.org/10.2967/jnmt.114.147447>
- JMP. (2019). *Nonparametric correlations*. Retrieved 30/11/2020 from https://www.jmp.com/en_us/learning-library/correlation-and-regression.html
- JMP. (2020 a). *Correlations and multivariate techniques*. Retrieved 22/11/2020 from <https://www.jmp.com/support/help/en/15.2/index.shtml#page/jmp/correlations-and-multivariate-techniques.shtml>
- Kagaku, K. (n.d.). Chest phantom N1 'LUNGMAN' [product information]. In. Kyoto, Japan. Retrieved 08/06/2015 from https://www.kyotokagaku.com/en/products_data/ph-1_01/
- Kalra, M., Maher, M., Toth, T., Hamberg, L., Blake, M., & Saini, S. (2004). Strategies for CT radiation dose optimization. *Radiology*, *230*, 619-628. <https://doi.org/10.1148/radiol.2303021726>
- Kamath, S., Song, W., Chvetsov, A., Ozawa, S., Lu, H., Samant, S., Liu, C., Li, J. G., & Palta, J. R. (2011). An image quality comparison study between XVI and OBI CBCT systems. *J Appl Clin Med Phys*, *12*(2), 3435.
- Keimel, C., & Diepold, K. *Improving the prediction accuracy of PSNR by simple temporal pooling*. Technische Universität München, Institute for Data Processing, Arcisstr. 21, 80333, Munich, Germany. Retrieved 26/12/2013 <http://mediatum.ub.tum.de/doc/965979/965979.pdf>

- Kijewski, M. F., & Judy, P. F. (1987). The noise power spectrum of CT images. *Physics in Medicine & Biology*, 32(5), 565.
- Kim, A. (2018). *CT lung screening cuts cancer mortality 51% in Japan*. Retrieved 17/12/2018 from <https://www.auntminnie.com/index.aspx?sec=supc=cto&pag=dis&ItemID=124046>
- Kim, B., Lee, K. H., Kim, K. J., Mantiuk, R., Hahn, S., Kim, T. J., & Kim, Y. H. (2008). Prediction of perceptible artifacts in JPEG 2000-compressed chest CT images using mathematical and perceptual quality metrics. *AJR Am J Roentgenol*, 190(2), 328-334. <https://doi.org/10.2214/AJR.07.2502>
- Kitami, A., Kamio, Y., Hayashi, S., & al., e. (2012). One-dimensional mean computed tomography value evaluation of ground-glass opacity on high-resolution images. *Gen Thorac Cardiovasc Surg*, 60, 425-430. <https://doi.org/10.1007/s11748-012-0066-7>
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4, 219-227.
- Krupinski, E. (2000). The importance of perception research in medical imaging. *Radiation Medicine*, 18(6), 6.
- Krupinski, E. (2010a). Current perspectives in medical image perception. *Attention, Perception and Psychophysics*, 72(5), 1205-1217. <https://doi.org/10.3758/APP.72.5.1205>
- Krupinski, E. (2010b). Perceptual factors in reading medical images. In E. Samei & E. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (First published 2010; First paperback edition 2014 ed., pp. 81-90). Cambridge University Press. ISBN: 978-1-107-42463-0
- Krupinski, E. (2011). The role of perception in imaging: past and future. *Seminars in Nuclear Medicine*, 41, 392-400. <https://doi.org/10.1053/j.semnuclmed.2011.05.002>
- Krupinski, E., & Samei, E. (2010). Future of medical image perception. In E. Samei & E. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (First published 2010; First paperback edition 2014 ed., pp. 413-416). Cambridge University Press. ISBN: 978-1-107-42463-0
- Kundel, H. L., & Revesz, G. (1976). Lesion conspicuity, structured noise, and film reader error. *AJR Am J Roentgenol*, 126(6), 1233-1238. <https://doi.org/10.2214/ajr.126.6.1233>
- Medical_Image_Perception_Laboratory (2017). *Satisfaction of Search in Diagnostic Radiology*. Retrieved 10/11/2017 from <http://perception.radiology.uiowa.edu/Research/SatisfactionofSearchinDiagnosticRadiology/tabid/112/Default.aspx>
- Lai, K., & Frush, D. (2006). Managing the radiation dose from pediatric CT. *Applied Radiology*, 35(4), 13-20.
- Ledenius, K., Svensson, E., Lhammar, F. S., Wiklund, L.-M., & Thilander-Klang, A. (2010). A method to analyse observer disagreement in visual grading studies: example of assessed image quality in paediatric cerebral multidetector CT images. *BJR*, 83, 604-611. <https://doi.org/10.1259/bjr/26723788>
- Lee, C. H., Goo, J. M., Ye, H. J., Ye, S.-J., Park, C. M., Chun, E. J., & Im, J.-G. (2008). Radiation dose modulation techniques in the multidetector CT era: from basics to practice. *RadioGraphics*, 28, 1451-1459. <https://doi.org/10.1148/rg.285075075>
- Leitz, W., Manson, L., Hedberg-Vikstrom, B., & Kheddache, S. (1993). In search of optimum chest radiography techniques. *The British Journal of Radiology*, 66, 314-321.
- Levi, D., & Klein, S. (1990). Equivalent intrinsic blur in spatial vision. *Vision Research*, 30, 1971 -1992.
- Li, F., Sone, S., Abe, H., MacMahon, H., Armato, S. r., & Doi, K. (2002). Lung cancers missed at low-dose helical CT screening in a general population: comparison of clinical, histopathologic, and imaging findings. *Radiology*, 225(3), 673-683. <https://doi.org/10.1148/radiol.2253011375>

- Li, Q., Yu, H., Zhang, L., Fan, L., & Liu, S. (2013). Combining low tube voltage and iterative reconstruction for contrast-enhanced CT imaging of the chest-initial clinical experience. *Clin Radiol*, *5*, e249-253. <https://doi.org/10.1016/j.crad.2012.12.009>
- Lifex.org (n.d.). License. Retrieved 30/09/2018 from <https://www.lifexsoft.org/index.php/product/license>
- Linebach, J., Tesch, B., & Kovacsiss, L. (2014). *Nonparametric Statistics for Applied Research*. Springer Science+Business Media. <https://doi.org/10.1007/978-1-4614-9041-8>
- Liu, M., Wang, Q., Yang, X., & Zhn, M. (2011). Application of objective image quality measures on CT image. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, *28*(2), 357-364.
- Ludewig, E., Richter, A., & Frame, M. (2010). Diagnostic imaging-evaluating image quality using visual grading characteristic (VGC) analysis. *Vetinary Research Communications*, *34*(5), 473-479. <https://doi.org/10.1007/s11259-010-9413-2>
- Lundin, M. (2012). *Aspects on image quality in radiologic evaluation of the urinary tract* (Publication Number Medical Dissertations No. 1298) Linköping University Sweden]. Linköping, Sweden. Retrieved 23/02/2014 from <http://liu.diva-portal.org/smash/get/diva2:516406/FULLTEXT01.pdf>
- Lundström, C. (2013). vPSNR: a visualization-aware image fidelity metric tailored for diagnostic imaging. *Int J Comput Assist Radiol Surg*, *8*(3), 437-450. <https://doi.org/10.1007/s11548-012-0792-4>
- MacRedmond, R., Logan, P., Lee, M., Kenny, D., Foley, C., & Costello, R. (2004). Screening for lung cancer using low dose CT scanning. *Thorax*, *59*(3), 237-241. <https://doi.org/10.1136/thx.2003.008821>
- Magnotta, V. A., Lee, F., & Birn, F. (2006). Measurement of Signal-to-Noise and Contrast-to-Noise in the fBIRN Multicenter Imaging Study. *Journal of Digital Imaging*, *19*(2), 140-147. <https://doi.org/10.1007/s10278-006-0264-x>
- Manning, D. (2005). Medical Imaging - the interpretation of visual information (Inaugural lecture). In (pp. 1-20). Researchgate; University of Cumbria website.
- Manning, D. (2010). Cognitive factors in reading medical images. In E. Samei & E. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (First published 2010; First paperback edition 2014 ed., pp. 91-106). Cambridge University Press. ISBN: 978-1-107-42463-0
- Manning, D., Ethell, S., & Donovan, T. (2004). Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph. *British Journal of Radiology*, *77*, 231-235.
- Manning, D., Gale, A., & Krupinski, E. (2005). Perception research in medical imaging [Commentary]. *BJR*, *78*, 683-685. <https://doi.org/10.1259/bjr/72087985>
- Marin, D., Nelson, R., Schindera, S., Richard, S., Youngblood, R., Yoshizumi, T., & Samei, E. (2010). Low-tube-voltage, high-tube-current multidetector abdominal CT: improved image quality and decreased radiation dose with adaptive statistical iterative reconstruction algorithm—initial clinical experience. *Radiology*, *254*, 145-153. <https://doi.org/http://dx.doi.org/10.1148/radiol.09090094>
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. WH Freeman. ISBN: 9780262514620
- Mathworks. (1994-2020). *MATLAB*. Retrieved 7/06/2015 from <http://uk.mathworks.com/products/matlab/>
- Mathworks. (2014a). *PSNR*. Mathworks. Retrieved 31/08/2014 from www.mathworks.co.uk/help/images/ref/psnr.html
- Mathworks. (2014b). *Structural Similarity (SSIM)*. Mathworks. Retrieved 31/08/2014 from www.mathworks.co.uk/help/images/ref/ssim.html
- Mathworks. (2019). *Texture Analysis*. Retrieved 08/01/2020 from https://uk.mathworks.com/help/images/texture-analysis-1.html?searchHighlight=Texture%20analysis&s_tid=doc_srchtile

- Mayerhoefer, M. E., Materka, A., Langa, G., Häggström, I., Szczypiński, P., Gibbs, P., & Cook, G. (2020). Introduction to Radiomics. *Journal of Nuclear Medicine*, 61(4), 488. <https://doi.org/10.2967/jnumed.118.222893>
- McEntee, M. F., Ryan, J., Evanoff, M. G., Keeling, A., Chakraborty, D., Manning, D., & Brennan, P. C. (2007). Ambient lighting: setting international standards for the viewing of softcopy chest images. In *SPIE proceedings - The International Society for Optical Engineering: Medical Imaging, 8 March 2007, San Diego, CA, United States*. <https://doi.org/10.1117/12.706772>
- McKenney, S. E., Seibert, J. A., Lamba, R., & Boone, J. M. (2014). Methods for CT automatic exposure control protocol translation between scanner platforms. *J Am Coll Radiol*, 11(3), 285-291. <https://doi.org/10.1016/j.jacr.2013.10.014>
- Medical_Dictionary. (2008). *Visual acuity*. Gale Encyclopaedia of Medicine (2008), found in thefreedictionary.com. Retrieved 15/01/2021 from <https://medical-dictionary.thefreedictionary.com/visual+acuity>
- Medical_Dictionary. (2009). *Farlex*. Retrieved 14/01/2021 from <https://medical-dictionary.thefreedictionary.com/perturbation>
- Merriam-Webster. (n.d.). *Merriam-Webster on-line dictionary*. <https://www.merriam-webster.com/dictionary/>
- Metz, C. (2006). Receiver Operating Characteristic Analysis: A tool for the quantitative evaluation of observer performance and imaging systems. *J Am. Coll. Radiol.*, 3(6), 413-422. <https://doi.org/DOI: 10.1016/j.jacr.2006.02.021>
- Microsoft. (n.d.). *Excel for Microsoft 365*. In (Version MSO (16.0.13127.20402))
- Mirchandani, G. (2012). Lab 3. Spectral Analysis in Matlab - Part II. In *EE289 Lab*. University of Vermont.
- Miéville, F., Bolard, G., Bulling, S., Gudinchet, F., Bochud, F., & Verdun, F. (2012). Effects of computing parameters and measurement locations on the estimation of 3D NPS in non-stationary MDCT images. *Physica Medica*, 29, 684-694. <https://doi.org/10.1016/j.ejmp.2012.07.001>
- Moore, C., & Murphy, A. (2005-2021). *Digital Imaging and Communications in Medicine (DICOM)*. Retrieved 14/01/2021 Radiopaedia.com. <https://radiopaedia.org/articles/digital-imaging-and-communications-in-medicine-dicom?lang=us>
- Morgan, C. L., & Miller, M. D. (1983). *Basic principles of computed tomography*. University Park Press. ISBN: 0839117051.
- Murphy, A. (2005-2021). *Aliasing*. Radiopaedia.org. Retrieved 14/01/2021 from <https://radiopaedia.org/articles/aliasing-artifact-ct>
- Murphy, A., & Goel, A. (2020). *Kilovoltage peak*. Radiopaedia.org. Retrieved 09/01/2021 from <https://radiopaedia.org/articles/kilovoltage-peak?lang=us>
- National_Instruments. (2015). *Structural similarity index*. Retrieved 20/08/2020 from http://zone.ni.com/reference/en-XX/help/372916T-01/nivisionconcepts/structural_similarity_index/
- NHSE-National Cancer Programme. (2019). Targeted Screening for Lung Cancer with Low Radiation Dose Computed Tomography. Retrieved 28/07/2021 from <https://www.england.nhs.uk/wp-content/uploads/2019/02/targeted-lung-health-checks-standard-protocol-v1.pdf>
- NIH. (2020). *ImageJ*. NIH (US National Institutes of Health). Retrieved 30/07/2017 from <https://imagej.nih.gov/ij/features.html>
- Nioche, C., Orhac, F., Boughdad, S., Reuzé, S., Goya-Outi, J., Robert, C., Pellot-Barakat, C., Soussan, M., Frouin, & Buvat, I. (2018). LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Research (AACR)*, 78(16), 4786-4789. <https://doi.org/10.1158/0008-5472.CAN-18-0125>
- Obuchowski, N. (2000). Sample Size Tables For Receiver Operating Characteristic Studies. *AJR*, 175, 603-608. <https://doi.org/10.2214/ajr.175.3.1750603>

- Obuchowski, N. A. (1997). Testing for equivalence of diagnostic tests. *Am. J. Roentgenol.*, *168*, 13-17. <https://doi.org/10.2214/ajr.168.1.8976911>
- Oxford_Dictionaries. (2021). *Luminous*. Lexico.com. Retrieved 14/01/2021 from <https://www.lexico.com/definition/luminous>
- Oxford_Dictionary_of_Statistical_Terms (2006). *Interaction*. (The International Statistical Institute). ISBN13 : 9780199206131
- Pahn, G., Skornitzke, S., Schlemmer, H., Kauczor, H., & Stiller, W. (2016). Toward standardized quantitative image quality (IQ) assessment in computed tomography (CT): A comprehensive framework for automated and comparative IQ analysis based on ICR. *Physica Medica*, *32*(1), 104-115. <https://doi.org/10.1016/j.ejimp.2015.09.017>.
- Park, S., Kupinski, M., Clarkson, E., & Barrett, H. (2003). Ideal-observer performance under signal and background uncertainty. *Inf Process Med Imaging*, *18*, 342-353.
- Pasquier, H., Gardavaud, F., Rahmouni, A., & Luciani, A. (2017). Impact of iterative reconstructions on image quality in clinical CT images demonstrated by a novel noise power spectrum measurement tool. In *European Congress of Radiology (ECR), March 1-5, 2017, Vienna*. <https://dx.doi.org/10.1594/ecr2017/C-3121>
- Pedersen, J. H., Saghir, Z., Wille, M. a. M. W., Thomsen, L. H., Skov, B. G., & Ashraf, H. (2016). Ground-glass opacity lung nodules in the era of lung cancer CT screening: Radiology, Pathology, and Clinical management [Review]. *Oncology, Cancer Network*.
- Pedersen, M., Cherepkova, O., & Mohammed, A. (2017). Image quality metrics for the evaluation and optimization of capsule video endoscopy enhancement techniques. *Journal of Imaging Science and Technology*, *61*(4), 040402-040401–040402-040408. <https://doi.org/10.2352/J.ImagingSci.Technol.2017.61.4.040402>
- Pelli, D. G., & Bex, P. (2013). Measuring contrast sensitivity. *Vision Research*, *90*(Supplement C), 10-14. <https://doi.org/https://doi.org/10.1016/j.visres.2013.04.015>
- Peskin, A., & Dima, A. (2010). Modeling clinical tumors to create reference data for tumor volume measurement. *Lecture notes in computer science*. (Vol. 6454, pp. 736-746).
- Pickhardt, P. J., Correale, L., Delsanto, S., Regge, D., & Hassan, C. (2018). CT colonography performance for the detection of polyps and cancer in adults ≥ 65 years old: systematic review and meta-analysis. *American Journal of Roentgenology*, *211*(1), 40-51. <https://doi.org/10.2214/AJR.18.19515>
- Pickhardt, P. J., Lubner, M. G., Kim, D. H., Tang, J., Ruma, J. A., Muñoz del Rio, A., & Chen, G.-H. (2012). Abdominal CT with model-based iterative reconstruction (MBIR): initial results of a prospective trial comparing ultralow-dose with standard-dose imaging. *American Journal of Roentgenology*, *199*, 1266-1274. <https://doi.org/10.2214/AJR.12.9382>
- Pineda, A. R., Tward, D. J., Gonzalez, A., & Siewerdsen, J. H. (2012). Beyond noise power in 3D computed tomography: The local NPS and off-diagonal elements of the Fourier domain covariance matrix. *Medical Physics*, *39*(6Part1), 3240-3252. <https://doi.org/10.1118/1.4705354>
- Pollard, B., Chawla, A, D elong, D., Hashimoto, N., & Samei, I. (2008). Object detectability at increased ambient lighting conditions. *Medical Physics*, *35*, 2204-2213.
- Pollard, B., Samei, E., Chawla, A., Beam, C., Heyneman, L., Koweek, L., Martinez-Jimenez, S., Washington, L., Hashimoto, N., & McAdams, H. (2012). The effects of ambient lighting in chest radiology reading rooms. *Journal of Digital Imaging*, *25*(4), 520-526. <https://doi.org/10.1007/s10278-012-9459-5>
- Radiology-Masterclass. (2020). *Chest X-ray Anatomy - Lung zones*. Retrieved 18/12/2020 from https://www.radiologymasterclass.co.uk/tutorials/chest/chest_home_anatomy/chest_anatomy_page3

- Radiology-tip-database. (2018). *Low-contrast detectability*. Radiology-tip.com. Retrieved 01/01/2021 from <https://www.radiology-tip.com/serv1.php?type=db1&dbs=Low%20Contrast%20Detectability>
- Rangaraju, K. S., Kumar, K., & Renumadhavi, C. H. (2012). Review paper on Quantitative Image Quality Assessment - Medical Ultrasound Images. *International Journal of Engineering Research & Technology*, 1(4), 6.
- Revesz, G. (1985). Conspicuity and uncertainty in the radiographic detection of lesions. *Radiology*, 154, 625-626.
- Ricci, N., Kobi, M., & Yee, J. (2020). CT colonography for colorectal cancer screening. *Radiology Nursing*, 39, 185-193. <https://doi.org/10.1016/j.jradnu.2020.04.005>
- Roberson, W. (2015). *Why an image is converted to double using I=im2double(I); in matlab? - MATLAB Answers - MATLAB Central*. MATLAB Answers. Retrieved 18/11/2017 from <https://uk.mathworks.com/matlabcentral/answers/233674-why-an-image-is-converted-to-double-using-i-im2double-i-in-matlab>
- Romans, L. (2013). *CT Image Quality*. CEwebservice. Retrieved 03/08/2017 from <http://www.cewebservice.com/coursePDFs/CTImageQuality.pdf>
- Rossman, K., & Wiley, B. (1970). The central problem in the study of radiographic image quality. *Radiology*, 96, 113-118.
- Royal_College_of_Radiologists. (2019). Picture archiving and communication systems (PACS) and guidelines on diagnostic display devices. Third Edition. Ref No. BFCO(19)2. London. Retrieved 10/01/2020 from https://www.rcr.ac.uk/system/files/publication/field_publication_files/bfcr192_pacs-diagnostic-display.pdf
- Sabih, D., Sabih, A., Sabih, Q., & Khan, A. N. (2011). Image perception and interpretation of abnormalities; can we believe our eyes? Can we do something about it? In *Insights Imaging* (Vol. 2, pp. 47-55). <https://doi.org/10.1007/s13244-010-0048-1>
- Samei, E., & Krupinski, E. (2010). Medical image perception. In E. Samei & E. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (First published 2010; First paperback 2014 ed., pp. 1-8). Cambridge University Press. ISBN: 978-1-107-42463-0
- Sasi varnan, C., Jagan, A., Kaur, J., Jyoti, D., & Rao, D. S. (2011). Image Quality Assessment Techniques on Spatial Domain. *International Journal of Computer Science and Technology*, 2(3), 177-184.
- Saunders, R., & Samei, E. (2010). Image quality and its perceptual relevance. In E. Samei & E. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (First published 2010; First paperback 2014 ed., pp. 157-164). Cambridge University Press. ISBN: 978-1-107-42463-0
- Schindera, S. (2017). *EuroSafe imaging on "appropriate image quality": Introduction*. Paper presented at European Congress of Radiology (ECR) March 1-5, 2017, Vienna. Retrieved 30/07/2017 from <http://ecronline.myesr.org/ecr2017/index.php?p=recorddetail&rid=0673b4c0-ddf5-4bc9-a3b9-6edb1b087ebe>
- Seeram, E. (2009a). Digital image processing. In *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control*. (3rd ed., pp. 56-83). Saunders Elsevier. ISBN: 978-1-4160-2895-6
- Seeram, E. (2009b). Computed tomography: an overview. In *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control*. (3rd ed., pp. 1-28). Saunders Elsevier. ISBN: 978-1-4160-2895-6
- Seeram, E. (2009c). Image reconstruction. In *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control*. (3rd ed., pp. 134-150). Sanders Elsevier. ISBN: 978-1-4160-2895-6

- Seeram, E. (2009d). Single slice spiral/helical computed tomography: physical principles and instrumentation. In *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control* (3rd ed. pp. 246-265). Saunders Elsevier. ISBN: 978-1-4160-2895-6
- Seeram, E., & Lapointe, J. (2009e). Computed tomography of the head, cerebral vessels, neck and spine. In *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control*. (3rd ed., pp. 395-420). Saunders Elsevier. ISBN: 978-1-4160-2895-6
- Seeram, E., & Hsieh, J. (2009f). Image quality. In *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control* (3rd ed., pp. 118-217). Saunders Elsevier. ISBN: 978-1-4160-2895-6
- Seeram, E. (2009g). Radiation dose in computed tomography. In *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control*. (3rd ed., pp. 218-245). Saunders Elsevier. ISBN: 978-1-4160-2895-6
- Seeram, E. (2009h). *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control*. (3rd ed., pp. 241). Saunders Elsevier. ISBN: 978-1-4160-2895-6
- Shaughnessy, J., Zechmeister, E., Zechmeister, J. (2012). Survey Research. In *Research Methods in Psychology*. (9th ed. pp. 137-182). McGraw-Hill. ISBN: 978-0-07-803518-0
- Seibert, J. A. (2014). Iterative reconstruction: how it works, how to apply it. *Pediatric Radiology*, 44(Supplement 3), 431-439. <https://doi.org/10.1007/s00247-014-3102-1>
- Seshadrinathan, K., Pappas, T., Safranek, R., Chen, J., Wang, Z., Sheik, H., & Bovik, A. (2009). Image Quality Assessment. In A. Bovik (Ed.), *The Essential Guide to Image Processing* (pp 553-595). Academic Press. <https://doi.org/10.1016/B978-0-12-374457-9.00021-4>
- Siemens. (2007). *SOMATOM Emotion 6/16-slice configuration Application Guide*. Retrieved 17/11/2017 from pdfstream.manualsonline.com/5/5e918bd2-8a20-4879-bde2-9ca93db9cc02.pdf
- Siemens. (2009). *CARE Dose4D*. From www.siemens.com/healthcare magazine. Retrieved 06/08/2017 from <https://image.slidesharecdn.com/somatomsessions25-00079318-00270155-140821164033-phpapp01/95/somatom-sessions-25-9-638.jpg?cb=1408639448>
- Siewerdsen, J. H., Cunningham, I. A., & Jaffray, D. A. (2002). A framework for noise-power spectrum analysis of multidimensional images. *Medical Physics*, 29(11), 2655-2671. <https://doi.org/10.1118/1.1513158>
- Sil, J., Lawson, R., Hogg, P., Elias, M., Kane, T., & Taylor, R. (2012). Reducing dose in CT Imaging: theory, practice, and lessons learnt from very low-dose CT. *eRadimaging*. Retrieved 02/08/2017 from <http://www.eradimaging.com/site/article.cfm?ID=785>
- SSPS. (2021). *Spearman Rank Correlations – Simple Tutorial*. www.ssps-tutorials.com. Retrieved 12/01/2021 from <https://www.spss-tutorials.com/spearman-rank-correlation/>
- Stern, M. K., & Johnson, J. H. (2010). Just Noticeable Difference. In *The Corsini Encyclopedia of Psychology*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470479216.corpsy0481>
- 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. *Eur Radiol*, 14(48), 48-58. <https://doi.org/10.1007/s00330-003-1971-z>
- Svalkvist, Angelica, Svensson, Sune, Håkansson, Markus, Båth, Magnus, Månsson, & Gunnar, L. (2016). Viewdex: a status report. *Radiat Prot Dosimetry* 169(1-4), 38-45. <https://doi.org/10.1093/rpd/ncv543>
- Tammemagi, M. C., Schmidt, H., Martel, S., McWilliams, A., Goffin, J. R., Johnston, M. R., Nicholas, G., Tremblay, A., Bhatia, R., Liu, G., Soghrati, K., Yasufuku, K., Hwang, D. M., Laberge, F., Gingras, M., Pasian, S., Couture, C., Mayo, J. R., Fauerbach, P. V. N., Atkar-Khattra, S., ..., & Lam, S. (2017). Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): a single-arm, prospective study - The Lancet Oncology. *Lancet Oncology*, 18, 1523-1531. [https://doi.org/doi:10.1016/S1470-2045\(17\)30597-1](https://doi.org/doi:10.1016/S1470-2045(17)30597-1)

- Tapiovaara, M. (2008). Review of relationships between physical measurements and user evaluation of image quality. *Radiat aprot Dosimetry*, 129(1-3), 244-8. <https://doi.org/10.1093/rpd/ncn009>
- Tingberg, A., Herrmann, C., Lanhede, B., Almen, A., Besjakov, J., Mattsson, S., Sund, P., Kheddache, S., & Månsson, L. G. (2000). Comparison of two methods for evaluation of the image quality of lumbar spine radiographs. *Radiation Protection Dosimetry*, 90(1-2), 165-168.
- Tischenko, O., Hoeschen, C., Effenberger, O., Reissberg, S., Buhr, E., & Doehring, W. (2003). Measurement of the noise components in the medical x-ray intensity pattern due to overlaying nonrecognizable structures. In *Proceedings of SPIE 5030, Medical Imaging, February 15-20, 2003, San Diego, CA*. <https://doi.org/doi:10.1117/12.480227>
- Tourassi, G. (2010). Receiver operating characteristic analysis: basic concepts and practical applications. In E. Samei & E. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (First published 2010; First paperback 2014 ed.). Cambridge University Press. ISBN: 978-1-107-42463-0
- Tuddenham, W. (1962). Visual search, image organization, and reader error in roentgen diagnosis. *Radiology*, 78(5), 694-704. <https://doi.org/10.1148/78.5.694>
- UK statutory instruments (2017). Ionising Radiation (Medical Exposure) Regulations 2017. Retrieved 15/01/2021 from <https://www.legislation.gov.uk/uksi/2017/1322/regulation/11/made>
- Ulzheimer, S. (2013). *Image quality in computed tomography*. Siemens. Retrieved 17/12/2013 from http://health.siemens.com/ct_applications/somatomsessions/index.php/image-quality-in-computed-tomography/#
- Ulzheimer, S., & Flohr, T. (2009). Multislice CT: current technology and future developments. In *Multislice CT. Reiser, M., Becker, C., Nikolaou, K., Glazer, G. (Eds.)*. (pp. 3-23). Springer. ISBN: 978-3-540-33124-7
- UNSCEAR. (2010). *Sources and effects of ionising radiation*. United Nations Scientific Committee on the effects of atomic radiation. Retrieved 06/08/2017 from http://www.unscear.org/docs/reports/2008/09-86753_Report_2008_Annex_A.pdf
- USPSTF. (2013). *Lung Cancer Screening: Final recommendation statement*. US Preventive Services Taskforce. Retrieved 10/03/2014 from <https://uspreventiveservicestaskforce.org/uspstf/recommendation/lung-cancer-screening#fullrecommendationstart>
- Vajuhudeen, Z., & Jones, J. (2005-2021). *Spatial resolution*. Radiopaedia.org. Retrieved 14/01/2021 from <https://radiopaedia.org/articles/spatial-resolution?lang=us>
- Veeramallu, B., LavanyaSusanna, C., & Sahitya, S. (2013). Survey on an image quality assessment metric based on early vision features. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(6), 447-449.
- Viera, A., & Garrett, J. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine*, 37(5), 360-363.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172-1217. <https://doi.org/10.1037/a0029333>
- Wagner, R. (2010a). Lessons from dinners with the giants of modern image science. In Samei, Ehsan Krupinski, & Elizabeth (Eds.), *The Handbook of Medical Image Perception and Techniques* (First published 2010; First paperback 2014 ed., pp. 76). Cambridge University Press. ISBN: 978-1-107-42463-0
- Wagner, R. (2010b). Lessons from dinners with the giants of modern image science. In E. Samei & E. Krupinski (Eds.), *The Handbook of Medical Image Perception and Techniques* (First published

- 2010; First paperback edition 2014 ed., pp. 73-78). Cambridge University Press. ISBN: 978-1-107-42463-0
- Wagner, R. F., & Brown, D. G. (1985). Unified SNR analysis of medical imaging systems. *Phys Med Biol*, 30(6), 489-518. <https://doi.org/10.1088/0031-9155/30/6/001>
- Watson, A (Ed.) (1993) Digital images and Human Vision. (pp. 1-78). Massachusetts Institute of Technology. ISBN: 9780262231718
- Welvaert, M., & Rosseel, Y. (2013). On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PLoS ONE*, 8(11), e77089. Retrieved 27/07/2017 from <https://doi.org/https://doi.org/10.1371/journal.pone.0077089>
- Winkler, S., & Mohandas, P. (2008). The evolution of video quality measurement: from PSNR to hybrid metrics. *IEEE Trans. Broadcasting*, 54(3), pp 660-668. <https://doi.org/10.1109/TBC.2008.2000733>
- Xu, Y., He, W., Chen, H., Hu, Z., Li, J., & Zhang, T. (2013). Impact of the adaptive statistical iterative reconstruction technique on image quality in ultra-low-dose CT. *Clin Radiol*, 68(9), 902-908. <https://doi.org/10.1016/j.crad.2013.03.024>
- Xue, W., & Mou, X. (2011). An image quality assessment metric based on Non-shift Edge. In *Proceedings of 18th IEEE International Conference on Image Processing, Brussels, March 11-14, 2011*. Image Processing (ICIP), pp.3309-3312.
- Yee, J., Weinstein, S., Morgan, T., Alore, P., & Aslam, R. (2013). Advances in CT colonography for colorectal cancer screening and diagnosis. *Journal of Cancer*, 4(3), 200-209. <https://doi.org/10.7150/jca.5858>.
- Yee, K. M. (2013). MDCT: Informatics can help evaluate CT image quality. In *Auntminnie.com*. Retrieved 09/05/2019 from <http://www.auntminnie.com/index.aspx?sec=supc=cto&pag=dis&ItemID=103753>
- Yoshihisa, K. T., Mitsudomi. (2013). Management of ground-glass opacities: should all pulmonary lesions with ground-glass opacity be surgically resected? *Translational Lung Cancer Research*, 2(5), 354-363. <https://doi.org/10.3978/j.issn.2218-6751.2013.09.03>
- Zaiontz, C. (n.d.). Real Statistics Resource Pack. In *Real-statistics.com*. Retrieved 02/12/2020 from <https://www.real-statistics.com/free-download/real-statistics-resource-pack>.
- Zarb, F., Rainford, L., & McEntee, M. (2010). Image quality assessment tools for optimization of CT images. *Radiography*, 16(2), 147-155. <https://doi.org/10.1016/j.radi.2009.10.002>
- Zhang, M., Mou, X., & Zhang, L. (2011). Non-Shift Edge Based Ratio (NSER): an image quality assessment metric based on early vision features. *Signal Processing Letters, IEEE*, 18(5), 315-318. <https://doi.org/10.1109/LSP.2011.2127473>
- Zhang, W. (2017). *Visual Saliency in Image Quality Assessment* In *PhD, Online Research @Cardiff (ORCA)*. <http://orca.cf.ac.uk/100239/1/2017zhangwphd.pdf>
- Zhao, S.-J., & Wu, N. (2015). Early detection of lung cancer: Low-dose computed tomography screening in China [Invited review]. *Thoracic Cancer*, 6, 385-389. <https://doi.org/10.1111/1759-7714.12253>
- Zimmerman Jones, A. (2018). *Specific Gravity*. ThoughtCo. Retrieved 14/01/2021 from <https://www.thoughtco.com/specific-gravity-2699007#:~:text=The%20specific%20gravity%20of%20a%20substance%20is%20the,the%20material%20will%20float%20in%20the%20reference%20substance.>
- Zuckier, L. (2011). Guest editorial: perception--the invisible element. *Seminars in Nuclear Medicine*, 41(6), 390-391. <https://doi.org/10.1053/j.semnuclmed.2011.06.006>.

Glossary:

Aliasing : ‘under-sampling (CT)’, ‘causing data superimposition’ (Murphy, 2005-2021).

Anthropomorphic : ‘having human attributes’ (Merriam-Webster, n.d.).

Consistency : ‘ability to be asserted together without contradiction’ (Merriam-Webster, n.d.).

Conspicuity : ‘refers to how well a given feature stands out from surrounding structures’ (Revesz, 1985).

Contrast resolution : ‘ability to distinguish between different intensities’ (Bell & Niknejad, 2005-2021).

Contrast sensitivity function : (CSF) relating to ‘the ability of the human visual system to distinguish an object from its background; it is not the same as visual acuity’ (Seshadrinathan, K. et al., 2009). CSF describes ‘sensitivity to sinusoidal patterns as a function of frequency’ (Barrett & Myers, 2003b)

Contrast-to-noise ratio : (CNR) describes ‘the signal level in relation to ambient noise’ (Bushberg et al., 2012c).

DICOM : ‘Digital Imaging and Communications in Medicine (DICOM) is a worldwide standard for the storage and transmission of medical imaging’ (Moore & Murphy, 2005-2021).

Frequency bins : ‘intervals between samples in the frequency domain’ (Chen, 2017).

Global image analysis: referring to analysis using the whole of an image.

Interpolation : ‘a mathematical method of averaging or smoothing images that are being displayed on a larger number of pixels than that for which they were originally intended’ (EU, 2000).

Luminance : relating to ‘light as it is perceived by the eye’, ‘luminous intensity’ (Oxford_Dictionaries, 2021).

Mean squared error : (MSE) ‘produces values describing distortion’ (Rangaraju et al., 2012).

Noise power spectrum : (NPS) ‘characterises the noise in an image’ ‘describing the noise texture’ (Bushberg et al., 2012d; Seeram & Hsieh, 2009f).

Non-shift-edge ratio : (NSER) looks at ‘edge definition by comparing an image matrix with that of a reference image in the frequency domain’ (Zhang et al., 2011).

Perturbations : ‘the disturbance or a change in a structure or function’ (Medical_Dictionary, 2009).

Psychophysical : relating to ‘the mental perception of physical stimuli’ (Farlex_Partner_Medical_Dictionary, 2012).

Radiographic contrast : ‘the density difference between neighbouring regions on a plain radiograph. High radiographic contrast is observed in radiographs where density differences are notably distinguished (black to white)’ (Deng & Murphy, 2005-2021).

Signal-to-noise ratio : (SNR) relates to the ‘conspicuity of an object’; how well it will be ‘recognised (detected)’ (Bushberg et al., 2012e).

Spatial resolution : refers to the ‘ability of an imaging modality to differentiate two adjacent structures as being distinct from one another’ (Vajuhdeen & Jones, 2005-2021).

Spatiotemporal blur : relating to ‘perception of objects in space and time’ (Merriam-Webster, n.d.); ‘blur involving both spatial and frequency domains’ (Pineda et al., 2012).

Standard error : (SE) indicates the standard deviation divided by the square root of the sample size (s / \sqrt{n}).

Structural similarity index metric : (SSIM) ‘uses a multiplicative combination of structure, contrast and luminance’ ‘performing a comparison with a reference image’ (Mathworks, 1994-2020; National_Instruments, 2015).

Texture analysis : ‘uses spatial variation in pixel intensities to examine aspects such as smoothness or coarseness and randomness or regularity in relation to image structure’ (Mathworks, 1994-2020).

Contrast : ‘sometimes called variance or inertia, is a measure of the intensity contrast between a pixel and its neighbour’ (Mathworks, 1994-2020)

Correlation : ‘is a measure of how correlated a pixel is to its neighbour’ (Mathworks, 1994-2020).

Energy : ‘also called uniformity’ (Mathworks, 1994-2020).

Homogeneity : looks at ‘spatial variations’ (Mathworks, 1994-2020).

Entropy : a measure of ‘variability or randomness’ (Mathworks, 2019).

Visual acuity: ‘the ability to distinguish details and shapes of objects’ (Medical_Dictionary, 2008).

APPENDIX 1

IQM FORMULAS:

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|f(i,j) - g(i,j)\|^2$$

Legend:

f represents matrix data of original image

g represents matrix data of degraded image in question

m represents numbers of rows of pixels of the images and **i** represents the index of that row

n represents number of columns of pixels of the image and **j** represents the index of that column

MAX_r is the maximum signal value that exists in the original 'known to be good' image

Mean squared error (MSE) formula (NI, 2013)

$$\text{PSNR} = 10 \log_{10}(\text{peakval}^2 / \text{MSE})$$

Legend:

MSE = Mean squared error

peakval = peak value

Peak signal-to-noise ratio (PSNR) calculation (Mathworks, 2014a)

$$\text{rNSE}(C,D) = \|E_c \cap E_d\| / \|E_c\|$$

Legend:

r = ratio

NSE = non-shift edges

(C, D) = (original image, distorted image)

$\|O\|$ denotes the number of edge points in an edge map

\cap Used in 'set theory', this means the set created by the two edge maps which contains the elements that both have in common.

Non-shift edge ratio (NSER) calculation (Xue & Mou, 2011)

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1},$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2},$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

Legend:
 l = luminance
 c = contrast
 s = structure

where μ_x , μ_y , σ_x , σ_y , and σ_{xy} are the local means, standard deviations, and cross-covariance for images x , y .

If $\alpha = \beta = \gamma = 1$ (the default for Exponents), and $C_3 = C_2/2$ (default selection of C_3) the index simplifies to:

Structural Similarity Index Metric (SSIM) calculation (Mathworks, 2014b)

$$\text{CNR} = \frac{S(\text{ROI}) - S(\text{BG})}{\sigma}$$

$S(\text{ROI})$ = mean pixel value in object ROI

$S(\text{BG})$ = mean pixel value in object background ROI

σ = standard deviation of pixel values in object background ROI

Contrast-to-noise ratio (CNR) calculation ("Catphan 500 and 600 Manual," 2015; Kamath et al., 2011)

$$\text{SNR} = \frac{\mu_{\text{pixel}}}{\sigma_{\text{pixel}}}$$

μ_{pixel} = mean pixel value in object ROI

σ_{pixel} = standard deviation of pixel values in background ROI

Signal-to-noise ratio (SNR) calculation
(Gaillard et al., n.d.)

APPENDIX 2



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/

23 September 2015

Dear Maryam,

**RE: ETHICS APPLICATION HSCR 15-34 – Task-specific observer performance and image quality:
direct and indirect relationships in low-dose CT images**

Based on the information you provided, I am pleased to inform you that application HSCR15-34 has been approved.

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

Yours sincerely,

A handwritten signature in black ink, appearing to read 'Sue McAndrew'.

Sue McAndrew
Chair of the Research Ethics Panel

APPENDIX 3



Health Research Authority

NRES Committee London - City & East

Bristol Research Ethics Committee Centre

Whitefriars

Level 3, Block B

Lewins Mead

Bristol

BS1 2NT

Telephone: 01173421386

Fax: 01173420445

21 April 2015

Ms LM Maryam Jessop
 Chief Nuclear Medicine Technologist
 Brighton and Sussex University Hospitals NHS Trust
 Nuclear Medicine, RSCH
 Eastern Road
 Brighton
 BN2 5BE

Dear Ms Jessop

Study title: Task-specific observer performance and image quality:
 direct and indirect relationships in low-dose CT images.
REC reference: 15/LO/0739
IRAS project ID: 173535

The Proportionate Review Sub-committee of the NRES Committee London - City & East reviewed the above application on 16 April 2015.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager Mr Rajat Khullar, nrescommittee.london-cityandeast@nhs.net. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Ethical opinion

On behalf of the Committee, the sub-committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

APPENDIX 4

University of
Salford
MANCHESTER

PARTICIPANT INFORMATION
Observer Performance Studies

PROJECT TITLE:

Task-specific observer performance and image quality: direct and indirect relationships in low-dose CT images.

INVITATION:

You are invited to assist in this PhD research involving observer performance studies. Before you decide you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. Ask questions if anything you read is not clear or would like more information.

ABOUT THE PROJECT:

The aim of the research is to determine quantifiable relationships between image characteristics and how well we see image features or lesions in a low-dose CT image. After using appropriate measures of image characteristics and measuring how well an observer sees image features or lesions, the results will then undergo statistical analyses to quantify any direct or indirect relationships.

The importance of this research lies in its potential to increase our understanding of visual perception in the context of image quality. Continuing international emphasis on dose reduction and dose optimization has heightened the need for greater precision that relates to diagnostic accuracy.

WHY YOU HAVE BEEN INVITED:

You have been invited to participate because of your familiarity with viewing radiological images. Approximately 20 volunteers will be invited to participate from groups including Radiologists, Radiographers, Nuclear Medicine Technologists and Physicists.

YOUR ROLE:

When participating in the observer performance studies, you will be presented with a number of CT images within image viewing and analysis software. You will be required to carry out specific tasks such as indicating if quality criteria are met according to the purpose stated, or lesion detection, and indicate how confident you are about your decision.

This will take approximately one hour to complete. You will be given training to assist you in your participation, whether with sample images at the beginning of the session or a *PowerPoint* file.

Risks or benefits:

There are no foreseeable risks to participants. While there is no intended benefit to participants, the information gained will contribute to a greater understanding of visual perception and image quality in low-dose CT.

ABOUT THE IMAGES:

The images (using physics and anthropomorphic phantoms) have been acquired using varying amperage. In some cases further material is then added to surround the object, which absorbs and scatters x-rays, subsequently impacting on image quality. For this research the images are fixed so that windowing cannot be undertaken.

Any anonymised patient images have been acquired as part of their normal imaging procedure.

FACILITATION OF IMAGE VIEWING:

Images will be viewed on monitors that comply with minimum standards, as set out in IT guidance documents for image viewing screens by the Royal College of Radiologists (2012).

Your taking part is entirely voluntary. Should you decide to participate then please e-mail me your registration details on the **Registration Form** attached. You will then be issued with a **participant code** which will be used as a unique identifier.

You will also have to return to me a signed **Consent form**.

While these image viewing sessions will be carried out for this next year, the same participant will not be asked to participate in all the sessions. If agreeable you may be asked to participate in two or three sessions throughout this time, however this is entirely up to you.

CONFIDENTIALITY:

Your participation in the study will be kept confidential, with the participant log kept in a locked office. All data obtained will be presented in such a way that your anonymity will be maintained.

WITHDRAWING FROM THE STUDY:

You may withdraw from the study at any time without consequence. The completed data you have provided, up to the point of withdrawal, will still be used.

ETHICS APPROVAL:

This research has been approved by the Ethics Committee at the University of Salford, the national Research Ethics Committee and by Research and Development for the Brighton and Sussex University Hospitals NHS Trust.

PUBLICATION OF FINDINGS:

It is intended that results of this research will be published in peer reviewed journals and shared in Conference presentations.

CONCERNS OR COMPLAINTS:

If you have a concern about any aspect of this study, you should ask to speak to the researcher (Maryam Jessop) on 01273 696955 ext 4382 / ext 3689 or email (as seen below), and your concerns will be addressed. If you remain unhappy and wish to complain

formally please do so in writing to: Research and Development, Bloomsbury House, Royal Sussex County Hospital, Eastern Road, Brighton BN2 5BE

THANK YOU FOR TAKING THE TIME TO READ THIS INFORMATION SHEET.

Maryam Jessop e-mail: L.M.M.Jessop@edu.salford.ac.uk
or Maryam.Jessop@bsuh.nhs.uk

Research Project for PhD, University of SALFORD.

Version 1.4 September 2015

APPENDIX 5

Research Governance and Ethics Committee (RGEC): College of Health and Social Care, University of Salford

Research Participant Consent Form

Title of Project:

Task-specific observer performance and image quality: direct and indirect relationships in low-dose CT images.

RGEC Ref No: HSCR15/34

Name of Researcher: Linda Maree Maryam Jessop

(Participant initials)

➤ I confirm that I have read and understood the information sheet for the above study (*version 1.4, September 2015*) and what my contribution will be.

➤ I have been given the opportunity to ask questions (via telephone and e-mail, or in person)

➤ I understand that anonymised data collected during the study will be looked at by individuals associated with the University of Salford and Brighton & Sussex University hospitals NHS Trust. I give permission for these individuals to have access to the data.

➤ I understand that participation involves viewing a number of images and assigning my decisions a level of confidence using a confidence scale.

➤ I understand that my participation is voluntary and that I can withdraw from the research at any time **without giving any reason**

➤ **I agree to take part in the above study**

Name of participant

Signature Date

Name of researcher taking consent: : Linda Maree MARYAM JESSOP
Signature and date:

Researchers e-mail address: : L.M.M.Jessop@edu.salford.ac.uk

or Maryam.Jessop@bsuh.nhs.uk

Research Governance and Ethics Committee Consent Form

Modified by researcher: September 2015 V 4.0

APPENDIX 6

ROC input file legend:

<i>Input file 1</i>	
<i>Modality</i>	<i>mA</i>
1	10
2	15
3	20
4	25
5	30

<i>Input file 2</i>	
<i>Modality</i>	<i>mA</i>
1	35
2	40
3	45
4	50
5	60

<i>Case no.</i>		
1	Normal	UR (5mm position)
2	Normal	ML (10mm position)
3	Normal	LR (8mm position)
4	+100	5mm
5	+100	10mm
6	+100	8mm
7	+100aj	5mm
8	+100aj	10mm
9	+100aj	8mm
10	-630aj	10mm
11	-630aj	8mm
12	-630	10mm
13	-630	8mm
14	-800	10mm
15	-800	8mm
16	-800aj	10mm
17	-800aj	8mm

<i>Lesion ID</i>	
0	Normal
1	Abnormal

<i>Weight</i>	
0	All cases

<i>Readers</i>
1
2
3
4
5
6
7

<i>FP (false positive)</i>
•When identified in normal image.
•When difference of > 10 pixel values from truth coordinates, in any direction.