

Chapter 10

Subjective Evaluation – Survey Results

This chapter presents a review and statistical analysis of the data returned by the participants in the radiographic contrast-enhancement mask (RCM) research survey. The survey consisted of RCM survey participants comparing and evaluating eight pairs of clinical digital radiographic (DR) images, one of the images in the pair being the original clinical image received from a clinical site and being deemed suitable for clinical diagnosis, and the other being the same image modified by an appropriate RCM. A pair of images from eight different anatomical regions was reviewed. Each image pair had unique characteristics of patient age, size and presence of pathology.

The images were provided on a CD ROM with a computer program so that participants were able to view the images on their own PC. The participants were given instructions on the use of the image viewing program and means of optimising their computer monitor to provide optimal image comparison. Demographic data as to the participants' occupation and years of experience were also collected.

Four main questions were asked. In these questions participants were required to compare the images and select their preferred image, either the image labelled "Image 1" or that labelled "Image 2". The participants were blinded as to which image was the original image and which was the modified image. The last of these questions asked the participants to identify their overall preferred image. The first three questions had a second associated question which asked for the degree of improvement of the preferred image over the non-preferred image. A scale of 1 to 5 was used, with 1 being a very small and 5 being a very large degree of improvement. The previous chapter described in detail the methodology of the RCM research survey and the images used.

A variety of statistical methods were used to determine whether the preferred images were the modified images. Further statistical tests were conducted to determine whether any factor-related trends existed within the participants' responses.

10.1 Data Collection and Significance

Two hundred CD ROMs, information and response packages were initially prepared for the RCM research survey. Eight different master CD ROMs were produced, each containing a unique set of eight images. Each of the eight CD ROMs was copied 25 times.

Requests for participant involvement in the RCM survey were broadcast via the AIR list-server email message system, AIRNEWS. Potential participants were asked to respond directly to the researcher's email address or other contact numbers provided. Other people were directly approached to be part of the RCM survey's cohort of participants. A total of 207 CD ROM, information and response packages were sent to people who had responded or were approached to be participants in the survey. This required the preparation of additional CD ROM, information and response packages. Packages were sent both within Australia and overseas. Eight packages were posted overseas to people who expressed an interest in being a RCM survey participant.

Participants were asked to copy and/or share the CD ROM and other information with any other willing colleague. This request was made in the information sheet provided with each posted package. Of the 207 and possibly further shared RCM survey packages, 123 were returned. An exact response rate cannot be determined due to the request for sharing of the packages with others. An approximated response rate comparing the initial mailing number of 207 with the final number of 123 returned response sets gives a response rate of 59.4%. Such a high response rate is attributed to the sampling technique used.

External validity of the survey was based on the two main factors of sample size and representation of the population by the sample (Black, 1999; Keppel, 1991; Polgar & Thomas, 1998; Sapsford, 1999). According to Black (1999) the sampling technique used in this survey can be described as both "purposive", being hand-picked subjects, and "volunteers", being those who responded to the AIR list-server request. Both of these sampling techniques have limitations. The purposive technique has the possible

limitation of being non-representative of the population due to potential subjectivity of the researcher (Black, 1999). The volunteer technique can be highly unrepresentative (Black, 1999), due in this case to the volunteer participants' access to and their response via the AIR list-server. It could also be suggested that such a group of volunteers, who used such a list-server, would be more technologically literate. As such, these people could be expected to have a better understanding of DR images and DR imaging technology. It could then be argued that this group of people would be better equipped to evaluate computer-based images. A bias of using people from such a group could then be argued. Such a group of people would, by their increased level of technical understanding, be better equipped to more accurately make computer based image evaluations. Given this, their selection of the best image from the image pair would be more accurate. Inclusion of participants from such a group would further reduce the observer variability discussed in the previous chapter.

Each person receiving a survey package was asked to pass on or copy it to a colleague. Questions were not asked and data were not recorded as to where the participant first gained information about the RCM survey. This further complicated the question of external validity due to the sampling technique, and therefore no formal conclusions about the sampling technique can be drawn.

RCM survey participants were asked their opinions on the comparison of eight sets of original and modified images. The overall sample size of image pairs evaluated in the survey was 984. These were independent samples. The outcome of selection of an image by one participant had no effect on the outcome of a selection of a second image by the same participant.

Keller (2001) provides a means of estimation of minimum sample size required to be representative of the population. The formula for this estimation (Keller, 2001) is given in Equation 10.1.

$$n = \left(\frac{z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})}}{B} \right)^2 \quad \dots\dots\dots 10.1$$

where: n is the sample size;

B is the bound of the error of estimation;

\hat{p} is the proposed sample proportion for an outcome within the population;

$z_{\alpha/2}$ is the z-score resulting from a given two-tailed α .

Using Equation 10.1 with a margin of error of ± 0.05 , \hat{p} assuming a “guess” as to the chosen image which would then equal 0.5 and an $\alpha = 0.05$, the minimum sample size should be 385. Keppel (1991) used a table to estimate sample size for a given power. The power of a statistical test is the probability that the null hypothesis will be rejected if it is false (Kinnear & Grey, 2000). Eng (2003) also describes statistical power as equivalent to sensitivity in medical tests. Increasing the power of statistical test reduces the likelihood of a false-negative test, more correctly known as a Type II error. From Keppel’s (1991) table, with a power of 0.9 and a small effect size, a sample size of 354 would be needed. Eng (2003) stated that it is acceptable to set the power to 0.8. Keppel (1991) also suggested that a power of 0.8 and a medium effect size would be acceptable for most experiments. Using these factors, a sample size of 44 would be the minimum. The RCM survey sample size of 984 is well in excess of both Keller’s and Keppel’s estimates. This sample size can be described as having a high power (power > 0.90) and can then be assumed to be representative of the population.

As discussed in greater detail later, the vast majority of the participants were radiographers. A sample size of 122 radiographers and their evaluation of 976 images can be considered representative of the population of radiographers. A demographic question as to the participants’ country of employment was not asked; consequently representation of the population of radiographers should not be limited to Australia. The results of the data analysis, with a confidence of 95%, can be assumed to be representative of all radiographers.

The RCM survey was also requested by people other than radiographers. The categories of “radiologist” and “other” were used to categorise these people. Of the RCM surveys returned, none was from a radiologist and one of the 123 was classified in the “other” category. No conclusions as to the sample’s representation of any other group other than radiographers can or will be drawn.

10.2 Collation and De-identification of Response Data

Participants were given a choice as to how they recorded their responses to the questions and given a choice of the method of returning the RCM survey data. Response could be recorded electronically either in a Microsoft Excel[®] spreadsheet or in a Microsoft Word[®] document. Alternatively, participants could choose to mark their response on a printed response sheet provided in the package.

Methods of returning the responses were to email the Microsoft Excel[®] spreadsheet or Microsoft Word[®] document to the email address provided, or to save the Microsoft Excel[®] spreadsheet or Microsoft Word[®] document to a 3½ inch floppy disk and post the disk in the reply-paid envelope, or to post the response sheet provided in the reply-paid envelope, or to print the Microsoft Excel[®] or Microsoft Word[®] responses and post them in the reply-paid envelope.

On receipt of the posted responses, survey responses were separated from participants’ consent forms and stored separately. If the responses were received electronically via email, the response file was stored in a separate directory of the researcher’s computer. It was assumed that the participant’s consent form either had already been or would be posted in the reply-paid envelope. These methods of separation and storage of participants’ responses were chosen to ensure anonymity of the participants.

At a later time after the separation and storage of responses, the data were transposed from either the electronic file or the printed response sheet into a Microsoft Excel[®] data spreadsheet. An interval of at least a week between the receipt of a response to

data transposition into the data file was used to further minimise the likelihood of participant identification.

If the data were received electronically in Microsoft Excel[®] format, transposition of data from the response file to the data spreadsheet was done through a computer algorithm. The algorithm was a series of Microsoft Excel[®] instructions to give the participant's response to the questions a number. This value was then electronically copied into the data file. This was done to reduce the likelihood of data input errors. If the responses were received in Microsoft Word[®] or printed format, care was taken to avoid transposition of data errors between the response sheet and the data file. All entries were double-checked in the forward and backward directions to ensure accuracy of data entry.

Following receipt of the final response, the data in the data file were sorted into groups. The groups were given a number between 1 and 8 inclusive according to which CD ROM number the participant had received. For confidentiality purposes, no information was ever stored as to which CD ROM number was sent to a given participant. The data were first sorted into the eight CD ROM disk groups. Further sorting within these groups by responses to the participants' years of experience was then undertaken. There were four options for responses to the question of years of experience. This sorting was done so as to further de-identify the data by not having the data in chronological order of receipt. After completing the collation and sorting, no means existed of identifying a particular participant from any given dataset.

The data in the file were then coded for ease of statistical analysis. For some questions participants were asked to select their preferred image. A dichotomous response was required, that is, to select either image number 1 or image number 2. If the participant preferred and selected the original image, the code for that choice was 0. If the participant preferred and selected the modified image, the code for that choice was 1. If there was no response for a particular question, a code of 9 was used, indicating that no value had been chosen.

The second part to some questions required participants to select a value on a given scale, from 1 to 5. These responses were given a negative value if the answer to the

initial part of the question had been to select the “original” image. If the response to the initial part of the question had been to select the “modified” image, the value was left as a positive value. A non-response to the question was coded with a 0. The use of a positive and negative scale in the coding of participants’ responses allowed easy identification of selection of either the original or the modified image as the preferred image. This facilitated ease of statistical analysis. An example of this coding is that if in the initial part of the question the original image was chosen and in the second part a response of 3 was given, the coding of the two parts to the question would be 0 and -3 respectively.

10.3 Survey Demographics

123 survey responses were returned and collated for analysis. Of the surveys received, 122 were from radiographers and 1 was from a profession classified in the “other” group. Table 10.1 provides an overview of these data.

Table 10.1 Participants’ profession versus years of experience

Profession		Experience				Total
		< 5 years	5 – 10 years	10 – 15 years	> 15 years	
Radiographer	Count	30	28	23	41	122
Other	Count	0	0	0	1	1
Total	Count	30	28	23	42	123
	%	24.4%	22.8%	18.7%	34.1%	100.0%

Participants were also asked to state their years of experience within their profession. The grouping of participants’ experience was explained in Chapter 9 and details are also provided in Table 10.1.

Figure 10.1 provides a plot of the participants’ responses to the question about their years of experience. The largest group, n = 42 (34% of the total number of participants), contained those who had more than 15 years experience within radiography. The responses from other groups were: n = 30 (24%) for less than 5 years experience; n = 28 (23%) for 5 to 10 years experience and n = 23 (19%) for

those participants with 10 to 15 years experience. A cross-tabulation of participant numbers within a profession and their years of experience is shown in Table 10.1.

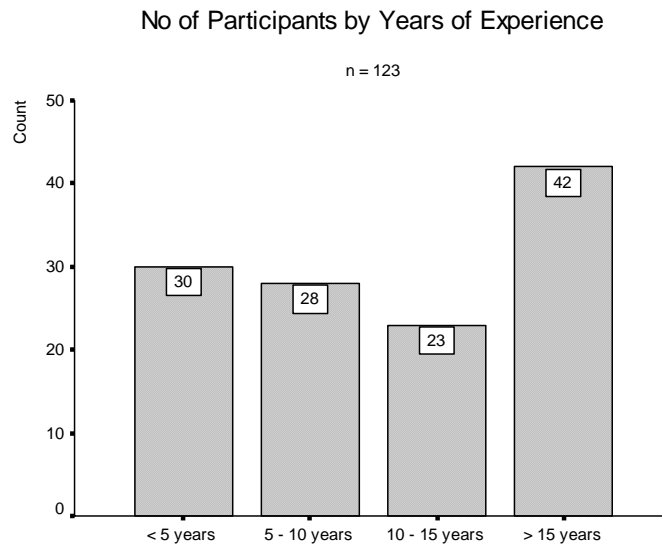


Figure 10.1 Plot of numbers of participants by years of experience

10.4 Participants' Selection of Preferred Image

Participants were asked to review eight pairs of clinical DR images, one of each pair being the original unaltered image and the other being the same image but with an appropriate RCM applied. The total number of image pairs available for response to the questions was 984. The four questions which required a choice between the two images in an image pair were:

- Question A1: Which image shows the best range of optical densities?
- Question B1: In which image is all anatomy easiest to visualise?
- Question C1: Which image allows for simplest contrast and density manipulation for optimal visualisation of all anatomy?
- Question D1: Which of the images has the highest image quality?

For question A1, 81.7% of the participants selected the modified image (n = 804 of 984 images) as showing the best range of optical densities. For question B1, 81.7% of the participants chose the modified image (n = 803 of 984 image pairs) as the image in which all anatomy was easiest to visualise. For Question C1, 81.2% of the

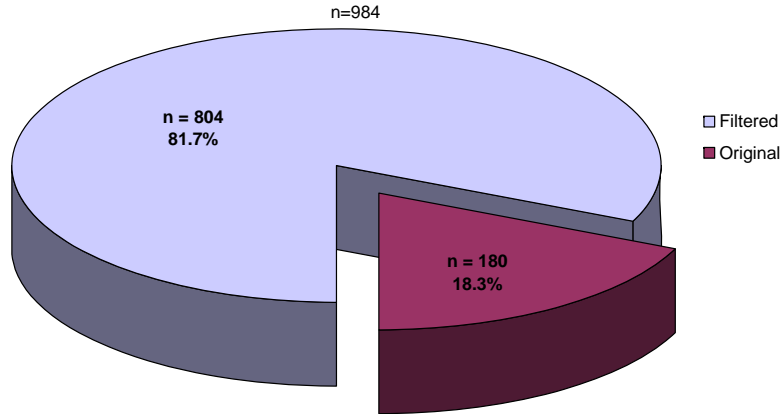
participants selected the modified image (n = 798 of 983 image pairs) as allowing for simplest contrast and density manipulation for optimal visualisation of all anatomy. Overall image quality was assessed by question D1: which of the images has the highest image quality? For this question, 80.1% of the participants selected the modified image (n = 788 of 984 image pairs). Table 10.2 and Figures 10.2 a–d. provide an overview of these responses. Note in Table 10.2 that one response was not provided for question B1 and C1.

Table 10.2 Responses to Questions A1, B1, C1 and D1

	A1: Best range of OD?		B1: Easiest anatomy visualisation?		C1: Simplest contrast manipulation?		D1: Highest image quality?	
	Count	%	Count	%	Count	%	Count	%
Original	180	18.3%	180	18.3%	185	18.8%	196	19.9%
Modified	804	81.7%	803	81.7%	798	81.2%	788	80.1%
Total	984		983		983		984	

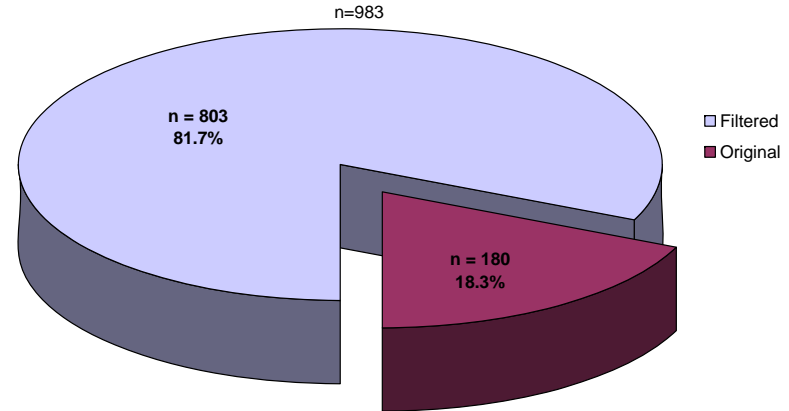
Across the four dichotomous questions, when a choice was required between the original or the modified image, out of the 984 responses the proportion of selecting the modified image was 0.81. This shows an overall strong preference for the modified images over the original images, given that the original images were already deemed suitable for clinical diagnosis.

**Overall Participant Response to Question A1:-
Which image shows the best range of optical densities?**



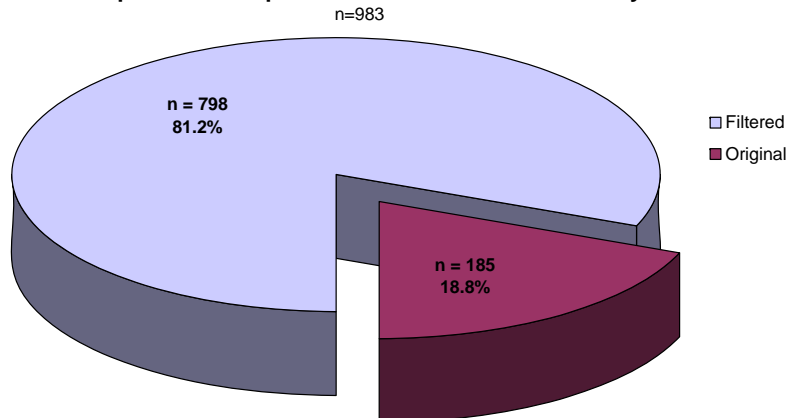
a.

**Overall Participant Response to Question B1:-
In which image is all anatomy easiest to visualise?**



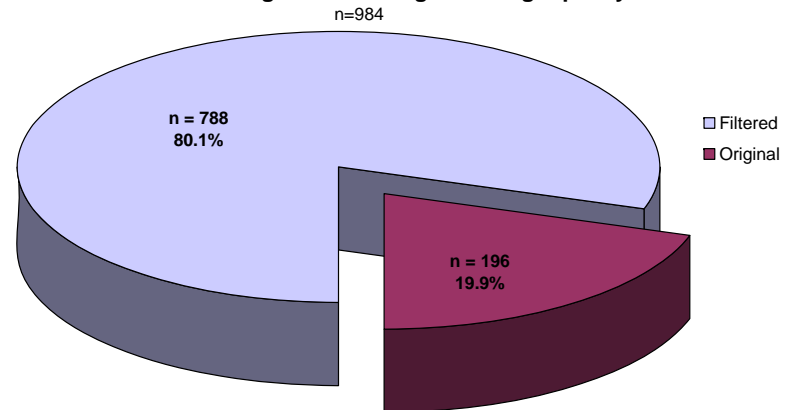
b.

**Overall Participant Response to Question C1:-
Which image allows for simplest contrast and density manipulation for optimal visualisation of all anatomy?**



c.

**Overall Participant Response to Question D1:-
Which of the images has the highest image quality?**



d.

Figure 10.2a, b, c & d

Charts of responses to Questions A1, B1, C1 and D1

10.5 Analysis of Responses to Questions of Preferred Image

Chi squared (χ^2) goodness-of-fit test is designed to determine whether test results are a factor of chance (Black, 1999; Polgar & Thomas, 1998). χ^2 analysis was performed on the above results. The expected frequency of response, given a random choice, would be 50% selection of the original image and 50% of the modified image. The alternative is that the proportion of selection of either image by the participants would not equal 0.5. The calculated χ^2 values for the responses to each dichotomous question are shown in Table 10.3. For responses to all the questions, very high χ^2 values resulted. The p-values of approximating 0.000 for all questions indicate that these responses to all the questions, A1, B1, C1 and D1, can confidently be attributed to non-random choice of the participants.

Table 10.3 Chi squared (χ^2) analysis of responses to Questions A1, B1, C1 and D1

	A1: Best range of OD?		B1: Easiest anatomy visualisation?		C1: Simplest contrast manipulation?		D1: Highest image quality?	
	Obsrvd N	Expect. N	Obsrvd N	Expect. N	Obsrvd N	Expect. N	Obsrvd N	Expect. N
Original	180	492.0	180	491.5	185	491.5	196	492.0
Modified	804	492.0	803	491.5	798	491.5	788	492.0
Total	984		983		983		984	
χ^2 value	395.7		394.8		382.3		356.2	
df	1		1		1		1	
p-value	>0.0001		>0.0001		>0.0001		>0.0001	

Further analysis was undertaken to determine whether the participants who selected the modified image in questions A1, B1 and C1 also selected the modified image as the overall best image in response to question D1. Dobson (2001) suggests the use of generalised linear models when outcomes are measured on a binary or dichotomous scale. A binary response was required by the participants to answer these four questions. The generalised linear method suggested by Dobson (2001) is that of logistic regression.

Logistic regression use has increased dramatically over the last decade (Hosmer & Lemeshow, 2000). Recent medical articles (Antoniades, *et al*, 2000; ; Castro, *et al*,

2000; Henderson *et al*, 1997; Henderson *et al*, 2000) and medical imaging articles (Arenillas *et al*, 2002; Roth *et al*, 2000; Woodburn *et al*, 2002) support the claim of Hosmer & Lemeshow.

Logistic regression is similar to linear regression except that it is used with ordinal or dichotomous data (Hosmer & Lemeshow, 2000; Pampel, 2000). Linear regression is used to analyse relationships between the dependent variable and the independent variables when interval data are used. Linear regression analysis determines the extent that an independent variable affects the dependent variable, through the calculation of that variable's coefficient (Keller, 2001). A typical linear regression equation takes the form $y = B_0 + B_1 x_1 + B_2 x_2 + \dots + B_i x_i$ where y is the dependent variable; x_i are the independent variables and B_i are the regression coefficients.

Responses to four of the questions in the RCM research survey were in binary format, indicating selection of a preferred image. Linear regression is not a valid tool to analyse such dichotomous data (Keller, 2001).

The basis of logistic regression is the calculation of the odds of an event occurring by determining the logit (L). L is the natural logarithm of the odds. The logit can be calculated by a regression process to determine the degree that the independent variables within the model affect the odds (Dobson, 2001; Hosmer & Lemeshow, 2000; Kinnear & Gray, 2000; Pampel, 2000).

Multivariable logistic regression is used when there is more than one independent variable. The method of calculation of the logit is given in Equation 10.2, adapted from Hosmer & Lemeshow (2000) and from Pampel (2000).

$$L = \ln\left(\frac{p}{1-p}\right) = B_0 + B_1x_1 + \dots + B_ix_i \quad \dots\dots\dots 10.2$$

where: L is the logit or natural logarithm of the odds;

p is the probability of the event;

$\frac{p}{1-p}$ is the odds of the event;

B is the logistic regression coefficient;

and x is the independent variable.

The results of the calculations of the independent variable coefficients from the data in Table 10.3 are shown in Table 10.4.

Table 10.4 Multivariable logistic regression analysis of responses to Question D1 given favourable responses to Questions A1, B1 and C1

	B	S.E.	Wald	df	p-values	Exp(B)	R^2
Q. A1	3.600	.388	85.958	1	1.8378E-20	36.596	0.882477
Q. B1	3.149	.406	60.068	1	9.1629E-15	23.323	0.785723
Q. C1	2.229	.395	31.898	1	1.625E-08	9.288	0.508361
Constant	-4.769	.458	108.430	1		.008	0.92277

where:

Q. A1 is: Which image shows the best range of optical densities?

Q. B1 is: In which image is all anatomy easiest to visualise?

Q. C1 is: Which images allows for simplest contrast and density manipulation for optimal visualisation of all anatomy?

The resulting equation, using the coefficients from Table 10.4, is given in the form to show the odds, rather than logit, and hence uses the exponential of B . The equation for determining the odds that a participant will select the modified image as the overall preferred image is given in Equation 10.3. The coefficients of the equation are multiplied by 1 if the participant selected the modified image as a response to that question; and multiplied by 0 if the participant selected the original image as a response to that question. The equation is:

the odds of selecting the modified image as image with the overall highest image quality =

$$\begin{aligned}
 & 0 + (36.6)(\text{if A1 answered Modified}) \\
 & + (23.3)(\text{if B1 answered Modified}) \\
 & + (9.3)(\text{if C1 answered Modified}) \quad \dots\dots\dots 10.3
 \end{aligned}$$

Using Equation 10.3, if the participant selected the modified image when answering questions A1, B1 and C1, the odds then of the participant selecting the modified image as the overall best image were 69.2 to 1. This equates to a probability of 0.986 or 98.6%.

Equation 10.3 predicts that there are very strong positive odds that the modified image would be selected as the image with the overall highest image quality when the participant selected the modified image in answering question A1. It also shows that the odds of selecting the modified image are reduced when selecting the modified image when answering questions B1 and C1.

The Wald statistic, as seen in Table 10.4, is a measure of the significance of the logistic coefficient, *B*. It equals the coefficient divided by the standard error (S.E.) all squared, and has a χ^2 distribution (Pampel 2000). The Wald equation is shown in Equation 10.4

$$\text{Wald} = \left(\frac{B}{\text{S.E.}} \right)^2 \quad \dots\dots\dots 10.4$$

where: Wald is the measure of significance;
B is the logistic regression coefficient;
 and S.E. is the calculated standard error of *B*.

High Wald values with corresponding p-values approaching zero, given an $\alpha = 0.05$, indicate strong significance for the corresponding *B* coefficient (Kinnear & Gray, 2000). The Wald values for question A1 of 85.96, for question B1 of 60.07 and for question C1 of 31.90 suggest strong significance of the coefficient for that variable.

Pampel (2000) and Hosmer & Lemeshow (2000) both explain that the Wald value should not be used as the only measure of significance. Pampel (2000) suggested that the Bayesian information criterion (BIC) is an important indicator of significance with large sample sizes. The sample size of comparison of the overall image quality is 984 and falls under this criterion. The BIC can be calculated using Equation 10.5 (Pampel, 2000).

$$BIC = z^2 - \ln(n) \quad \dots\dots\dots 10.5$$

where: BIC is the Bayesian information criterion measure of significance;

z is the standard score of logistic regression coefficient B ;

and n is the sample size.

BIC values of 0 – 2 are weak, 2 – 6 are positive, 6 – 10 are strong and greater than 10 are very strong grades of evidence for support for the inclusion of that coefficient variable in the resulting predicted logistic regression analysis equation (Raftery, 1995). Table 10.5 shows the BIC values for the questions. The BIC value for question A1 is 79.07, for question B1 is 53.18 and for question C1 is 25.01. All of these BIC values lie within Raftery’s (1995) very strong grade of evidence.

Table 10.5 Bayesian information criterion values calculated using Table 1 and Equation 10.5

	<i>B</i>	S.E.	<i>z</i>	BIC
Q. A1	3.600	.388	9.271365	79.06658
Q. B1	3.149	.406	7.750365	53.17653
Q. C1	2.229	.395	5.647814	25.00618
n =	984			

The coefficient of determination, R^2 , as seen in Table 10.4, represents the proportion of variance in one of the logistic coefficients, B , accounted for by the others (Kinnear & Gray 2000; Polgar & Thomas, 1998). An overall R^2 of 0.923 suggests that 92.3% of the variation in the selection of the modified image as the preferred image with the overall highest quality is accounted for by the variables A1, B1 and/or C1.

Given the strength of the Wald statistic, the BIC and R^2 , it can be confidently stated with a high degree of significance, that Equation 10.3 truly represents the likely outcome. There is a very high likelihood that a participant will select the modified image as being the best overall image after having selected the modified image as a response to being the image:

- showing the best range of optical densities *and*
- in which it is easiest to visualise all the anatomy *and*
- that is simplest in manipulation of contrast and density to visualise the anatomy.

10.6 Participants' Level of Improvement of Their Preferred Image

The participants were asked to provide an indication of the level of improvement of their preferred image over the non-preferred image. The responses were sub-questions to questions A1, B1 and C1. The second part of these questions was:-

A2: Please indicate the level of increase in improvement in range of optical densities.

B2: Please indicate the level of increase in improvement in ease of visualisation.

C2: Please indicate the level of increase in improvement in ease of optimisation.

The scale for the level of improvement is shown in Table 10.6 with the corresponding codes that were used by the participants.

Table 10.6 Scale for improvement levels in the participants' preferred images

<u>Scale Response Amount</u>				
when comparing images				
1	2	3	4	5
very small	small	moderate	Large	very large

The numbers of responses to each question, A2, B2 and C2, at each particular response scale level, are provided in Table 10.7. Plots of the responses in Table 10.7

are seen in Figure 10.3a, b & c. The plots do not include values where there was no response from the participant.

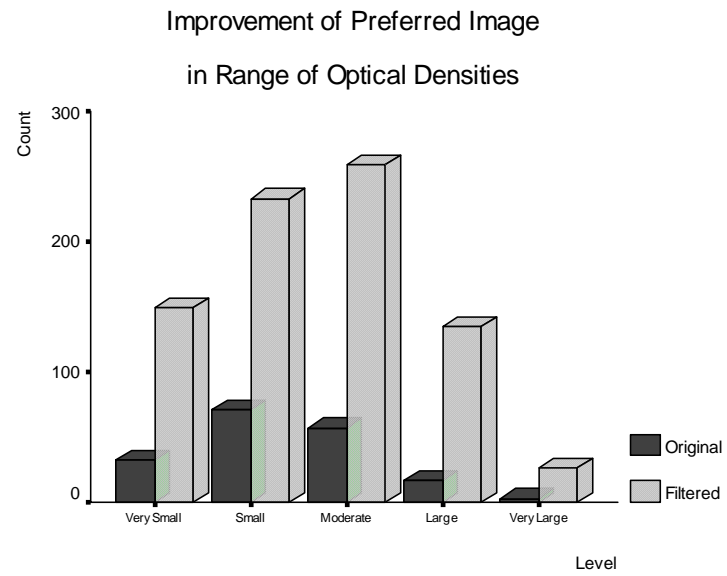
Table 10.7 No. of responses to questions as to improvement levels in the participants' preferred images

	Scale of Response	Q A2	Q B2	Q C2
Original image selected	1	32	46	52
	2	71	68	54
	3	57	53	53
	4	17	9	25
	5	3	3	2
Modified image selected	1	149	138	161
	2	233	268	234
	3	259	250	230
	4	135	134	145
	5	26	11	26
No response	0	2	4	2
Total		984	984	984

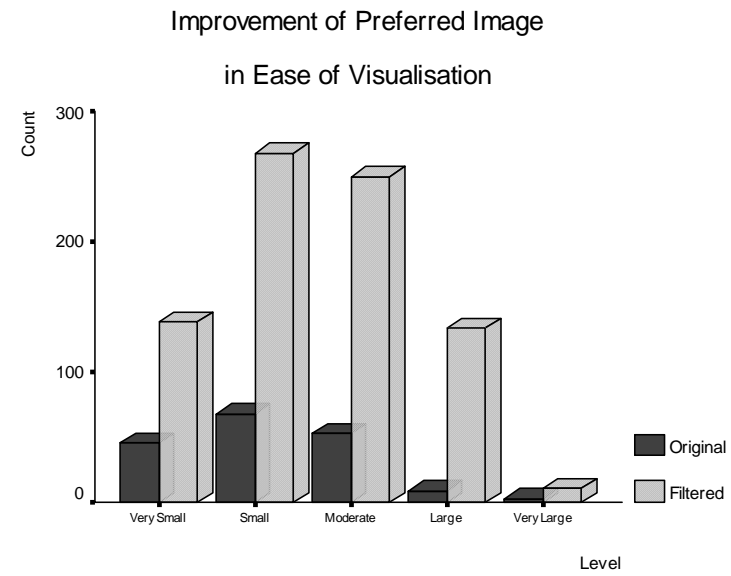
Calculations were made of the means and standard deviations of the responses, when the original and the modified images were selected. Garson (2004) suggests that some scaled data may be treated as interval data and hence such calculations would be a valid means of measuring the central tendency of the data. These measurements can then be used to compare the scale levels of responses from the participants who selected the original image with the scale levels of responses from those who selected the modified image. The means and standard deviations for responses to questions A2, B2 and C2 are given in Table 10.8.

Table 10.8 Weighted means and weighted standard deviations for Questions A2, B2 and C2

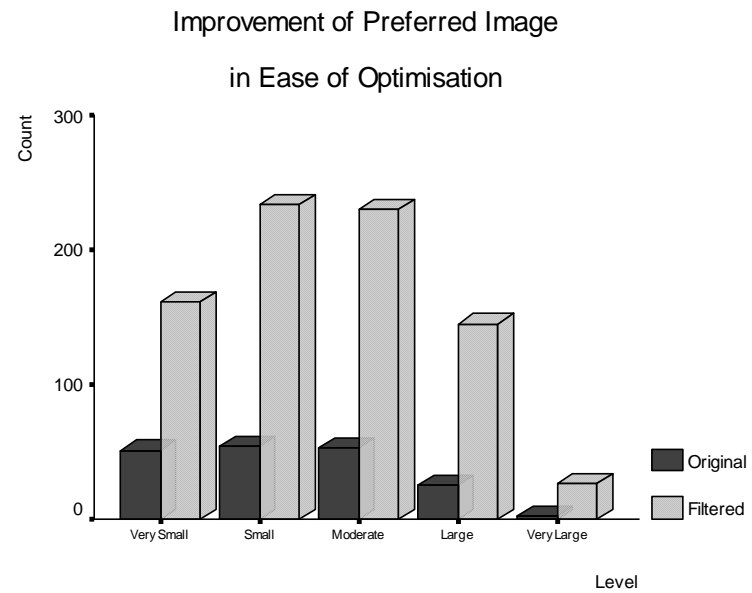
	Q A2		Q B2		Q C2	
	Original Image	Modified Image	Original Image	Modified Image	Original Image	Modified Image
Mean	2.38	2.57	2.19	2.52	2.31	2.55
S.D.	1.05	1.20	1.04	1.12	1.18	1.23



a.



b.



c.

Figure 10.3a, b & c

Plots of responses to Questions A2, B2 and C2

When the original image was selected, the mean responses to the questions were in the range of 2.19 to 2.38, with standard deviations of less than 1.18. The mean of these results would be considered as “moderate” using the scale from Table 10.7.

To test whether there was any significant difference between the scale of responses when either the original image or the modified image was selected, the Friedman test was used. The Friedman test is used to compare two or more populations of ordinal data (Keller, 2001). It is the equivalent test, for ordinal data, to analysis of variance (ANOVA) used with interval data to identify whether differences exist between the population means.

When the original image was selected, the results of comparisons between questions:

A2 indicating the level of increase in improvement in range of optical densities;

B2 indicating the level of increase in improvement in ease of visualisation;

C2 indicating the level of increase in improvement in ease of optimisation;

with an $\alpha = 0.05$, are shown in Table 10.9. When the original image was chosen as the preferred image, there was no significant difference in the way the participants ranked the degree of improvement between their responses to question A2, B2 or C2.

Table 10.9 Friedman Test analysis, when the original image was selected

Question	Rank Sum
A2: <i>improved range of optical densities</i>	11.5
B2: <i>ease of visualisation</i>	9
C2: <i>ease of optimisation</i>	9.5
Fr Stat	0.7
df	2
p-value	0.7047
chi-squared Critical	5.9915

Similarly, the Friedman test was used to compare the means of the responses to the questions on level of improvement when the modified image was selected. In these cases the calculated means, as seen in Table 10.8, for responses to Questions A2, B2

and C2 were 2.57, 2.52 and 2.55 respectively, all with standard deviations of less than 1.23. These mean results would also be considered as “moderate” using the scale from Table 10.7. The results of the Friedman test when the modified image was selected, with an $\alpha = 0.05$, are shown in Table 10.10. When the modified image was chosen as the preferred image, there was no significant difference as to how the participants ranked the degree of improvement between their responses in question A2, B2 or C2.

Table 10.10 Friedman Test analysis, when the modified image was selected

Question	Rank Sum
A2: <i>improved range of optical densities</i>	10.5
B2: <i>ease of visualisation</i>	8
C2: <i>ease of optimisation</i>	11.5
Fr Stat	1.3
df	2
p-value	0.522
chi-squared Critical	5.9915

When either the original image or the modified image was selected as the preferred image, there was no significance difference between responses for questions A2, B2 and C2. It can then be stated that the participants did not respond significantly differently between the three questions when selecting either the original image or the modified image as their preferred image. This implies that generally participants provided a very similar scale of responses, whether they were responding to a question on a range of optical densities, to a question on ease of visualisation, or to a question on ease of optimisation of the image.

Of more interest is whether there was any overall difference in the scale of responses between the participants who selected the original image and those who selected the modified image. Some authors (Black, 1999; Clason & Dormody, 1994; Keller, 2001; Kinnear & Gray, 2000) disagree as to which is the most appropriate method of analysis to compare whether the two samples of non-parametric data are from the same population. The suggested tests are the *t*-test, Mann-Whitney U-tests and Wilcoxon rank-sum tests. *t*-tests are normally associated with interval data. As

discussed above, some scaled data may be treated as interval data (Garson, 2004). All of these tests determine whether the difference between the means of the samples is sufficiently large to state that the samples are from different populations.

All of these tests were performed. The results are shown in Table 10.11. The results of the *t*-tests were also evaluated under Levene's test for equality of variances. Levene's test determines whether the most appropriate *t*-test is to assume equal variances of the samples or to assume unequal variance of the samples (Kinnear & Gray, 2000). *t*-test results will vary between these assumptions. Only the appropriate test results, equal variance assumed or equal variance not assumed, are shown.

Table 10.11 Various tests to compare central tendency of scale when original or modified image was selected for Questions A2, B2 and C2

Q.	Selected Image	t-test Results				Mann-Whitney U Test		Wilcoxon Rank Sum Test	
		\bar{x}	sd	t	p-value (2-tail)	Z stat	p-value (2-tail)	Z stat	p-value (2-tail)
A2: <i>improved range of optical densities</i>	Original n=180	2.38	.940	-2.149	0.032	-2.169	0.030	-2.0886	0.0368
	Modified n=804	2.56	1.078						
B2: <i>ease of visualisation</i>	Original n=180	2.18	.946	-4.065	0.000	-4.028	0.000	-3.8659	0.0002
	Modified n=802	2.51	1.010						
C2: <i>ease of optimisation</i>	Original n=185	2.31	1.052	-2.641	0.009	-2.464	0.014	-2.3619	0.0182
	Modified n=798	2.54	1.108						

The results from the Mann-Whitney U-tests, Wilcoxon rank-sum tests and *t*-tests, all show, with an α of 0.05, that there is significant difference when the modified image was selected compared to when the original image was selected. For each of the three questions:

- A2 indicating the level of increase in improvement in range of optical densities;
- B2 indicating the level of increase in improvement in ease of visualisation;
- C2 indicating the level of increase in improvement in ease of optimisation;

the participants overall responded that there was a greater level of improvement when the modified image was selected.

It can be stated with a high level of confidence that the preferred images were the modified images (as discussed above). From the results in Table 10.11, it can also be stated that there was a greater level of improvement in the factors “range of optical densities”, “ease of visualisation” and “ease of optimisation” of the modified images. Together, both these conclusions strongly indicate that the participants’ preferred image was the modified image.

10.7 Analysis of Participants’ Responses by Anatomical Region

Each participant was asked to evaluate pairs of images from eight different anatomical regions of the body. The selection of anatomical regions and radiographic positions was discussed in Chapter 9, and the categories are listed within Table 10.12.

The anatomical regions were chosen due to their large difference of anatomical thickness or large degree of attenuation of x-ray photons. A wide dynamic range of values exists in the image as a result of these differences. It was felt that anatomical regions with wide dynamic ranges would be most suited for the use of digital RCMs (see Chapters 5, 6 and 7 for a full discussion).

Participant analysis of these anatomical regions resulted in 123 responses being received for each region. Responses were analysed to determine in which anatomical region(s) the participants had the greatest preference for the modified images over the original images. The number and percentage of responses to question C1, and the participants’ preferences for the overall best image, either the original or modified image, for each anatomical region are shown in Table 10.12 and Figure 10.4.

Table 10.12 Participants' responses, by anatomical region, to the question as to which image of a pair is the overall best image

Overall best?		Region								Total
		Shoulder	C Spine	T Spine	T/L Spine	Abdomen	Facial Bones	Feet	Other	
Original	Count	24	33	17	42	35	14	12	19	196
	%	19.5%	26.8%	13.8%	34.1%	28.5%	11.4%	9.8%	15.4%	19.9%
Modified	Count	99	90	106	81	88	109	111	104	788
	%	80.5%	73.2%	86.2%	65.9%	71.5%	88.6%	90.2%	84.6%	80.1%
Total	Count	123	123	123	123	123	123	123	123	984

The anatomical regions for which participants' preferences for the modified images were the greatest were the feet (n=111; 90.2%), facial bones (n=109; 88.6%) and the AP thoracic spine (n=106; 86.2%). The least preferred region for the modified image was thoracic-lumbar spine (n=81; 65.9%). Chi squared analysis as to whether the selections for the thoracic-lumbar spine region differed from a pure guess provided results of $\chi^2 = 12.37$ and a p-value, with 1 degree of freedom, approximating 0.000. At an $\alpha = 0.05$, it can be stated with a high level of confidence that the result of participants' selection of the modified image for the thoracic-lumbar spine was not based on a pure guess. This being the participants' least preferred modified image, the previous comment can also be generalised to the results relating to the other regions, in that those results were also not based on pure guesswork.

Selection % of Best Quality Image by Anatomical Region

Total n = 984 Each Region n = 123

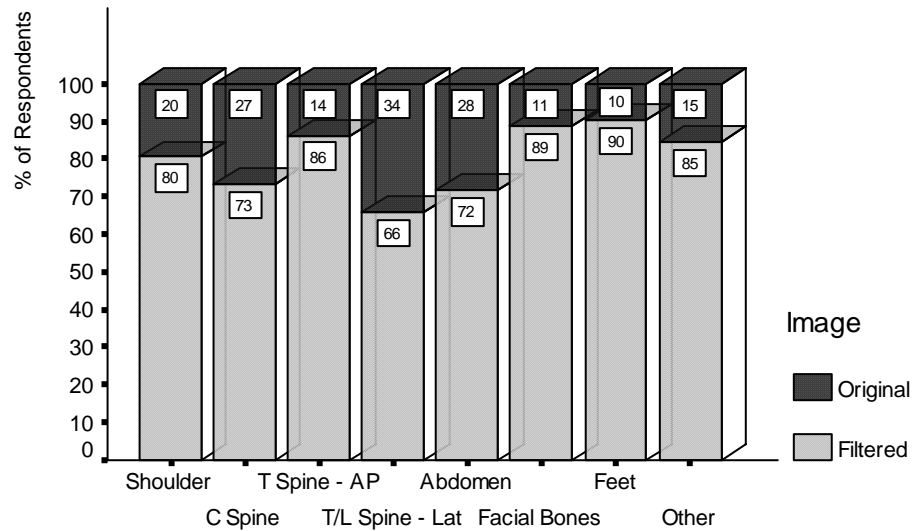


Figure 10.4 Plots of participants' responses, by anatomical region, to the question as to which image of a pair is the overall best image

Participants' responses, by anatomical regions, for Questions A1, B1 and C1 are shown in Tables 10.13–15 and in Figure 10.5a, b & c. Again, participants' responses to questions A1, B1 and C1 showed increased preference for the modified image in the regions of the feet, facial bones and thoracic spine.

Table 10.13 Participants' responses, by anatomical region, to the question as to which image has the best range of optical densities

Range of OD?		Region								Total
		Shoulder	C Spine	T Spine	T/L Spine	Abdomen	Facial Bones	Feet	Other	
Original	Count	20	24	20	42	29	13	14	18	180
	%	16.3%	19.5%	16.3%	34.1%	23.6%	10.6%	11.4%	14.6%	18.3%
Modified	Count	103	99	103	81	94	110	109	105	804
	%	83.7%	80.5%	83.7%	65.9%	76.4%	89.4%	88.6%	85.4%	81.7%
Total	Count	123	123	123	123	123	123	123	123	984

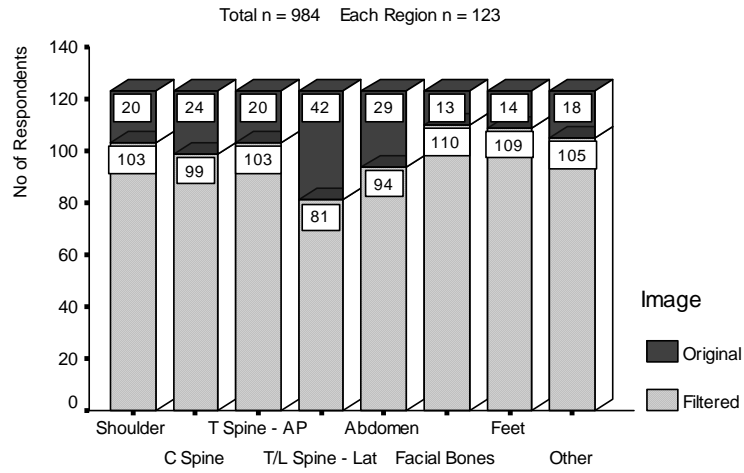
Table 10.14 Participants' responses, by anatomical region, to the question as to in which image anatomy is easiest to visualise

Anat visualise?		Region								Total
		Shoulder	C Spine	T Spine	T/L Spine	Abdomen	Facial Bones	Feet	Other	
Original	Count	24	27	21	32	29	16	8	23	180
	%	19.5%	22.0%	17.1%	26.0%	23.6%	13.0%	6.5%	18.7%	18.3%
Modified	Count	99	95	102	91	94	107	115	100	803
	%	80.5%	77.2%	82.9%	74.0%	76.4%	87.0%	93.5%	81.3%	81.6%
Total	Count	123	122	123	123	123	123	123	123	983

Table 10.15 Participants' responses, by anatomical region, to the question as to which image has simplest visualisation of all anatomy

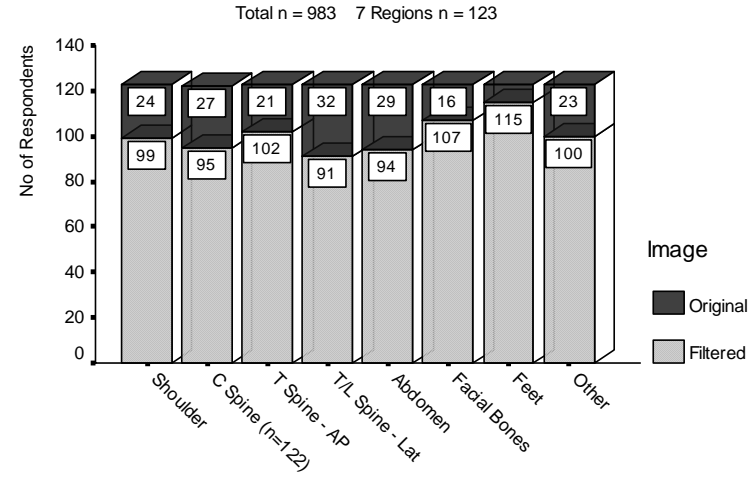
Contrast manipulation?		Region								Total
		Shoulder	C Spine	T Spine	T/L Spine	Abdomen	Facial Bones	Feet	Other	
Original	Count	24	29	13	36	31	15	11	26	185
	%	19.5%	23.6%	10.6%	29.3%	25.2%	12.2%	9.0%	21.1%	18.8%
Modified	Count	99	94	110	87	92	108	111	97	798
	%	80.5%	76.4%	89.4%	70.7%	74.8%	87.8%	91.0%	78.9%	81.2%
Total	Count	123	123	123	123	123	123	122	123	983

Selection of Best Range of OD
by Anatomical Region



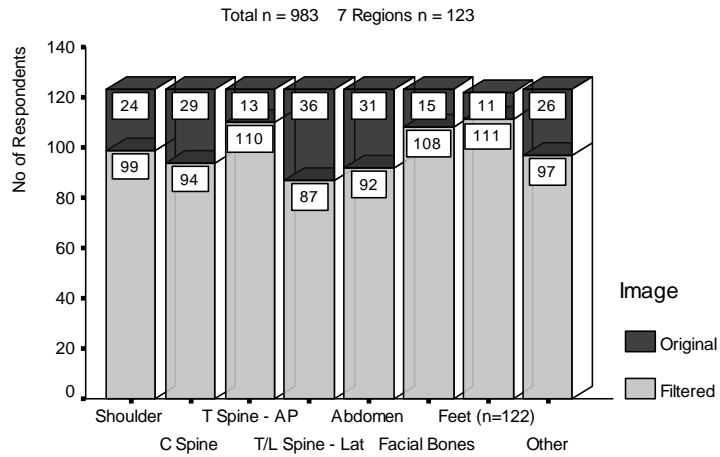
a.

Selection of Ease of Visualisation
by Anatomical Region



b.

Selection of Simplest Display Manipulation
by Anatomical Region



c.

Figure 10.5a, b & c

Plots of responses to Questions A1, B1 and C1, by anatomical region

It has been shown previously (Equation 10.3) that when participants selected the modified image in response to the questions:

A1: Which image shows the best range of optical densities?

B1: In which image is all anatomy easiest to visualise?

C1: Which image allows for simplest contrast and density manipulation for optimal visualisation of all anatomy?

they were more likely to select the modified image as their overall preferred image. Logistic regression was performed to again show the relationship between selecting the modified image in one of these questions and then selecting the modified image as the overall preferred image within each anatomical region.

Each anatomical region was examined to determine the relationship between participants' responses to questions A1, B1 and C1 and selection of the modified image. Only factors with high Wald and strong or greater BIC values were included in the equations. In many instances the BIC values were in the positive range and hence not included.

Shoulder:

No definitive relationship could be determined that was supported by high Wald values and by strong or greater BIC values.

C Spine:

the odds of selecting the modified image as having the overall highest image quality =

$$0 + (104.5)(\text{if A1 answered Modified}) + (59.0)(\text{if B1 answered Modified}) \dots\dots\dots 10.6$$

Using Equation 10.6, if the participant selected the modified image when answering Questions A1 and B1, the odds then of the participant selecting the modified image as the overall best image were 163.5 to 1. This equates to a probability of 0.993 or 99.3%.

The BIC values calculated for Equation 10.6 were 6.4 and 6.1. These values suggest a strong grade of evidence to support the inclusion of the factors A1 and B1 in this relationship.

Thoracic Spine:

No definitive relationship could be determined that was supported by high Wald values and by strong or greater BIC values.

Thoraco-Lumbar Spine:

the odds of selecting the modified image as having the overall highest image quality =

$$0 + (104.5)(\text{if A1 answered Modified}) \dots\dots\dots 10.7$$

The BIC value calculated for Equation 10.7 is 7.4. This suggests a strong grade of evidence to support the inclusion of the factor A1 in this relationship.

Using Equation 10.7, if the participant selected the modified image when answering Question A1, the odds then of the participant selecting the modified image as the overall best image were 104.5 to 1. This equates to a probability of 0.991 or 99.1%.

Abdomen:

No definitive relationship could be determined that was supported by high Wald values and by strong or greater BIC values.

Facial Bones:

No definitive relationship could be determined that was supported by high Wald values and by strong or greater BIC values.

Feet:

No definitive relationship could be determined that was supported by high Wald values and by strong or greater BIC values.

Other:

No definitive relationship could be determined that was supported by high Wald values and by strong or greater BIC values.

10.8 Patient-Related Factors Affecting the Selection of the Modified Image within Anatomical Regions

As discussed above, the images for the data sets were selected from eight different anatomical regions. Within each anatomical region, images were selected with three different patient-related factors. These factors were:

- age – classified as either young or old;
- size – classified as either small to medium or medium to large;
- pathology – classified as either showing pathology within the image (pathology) or not showing pathology within the image (normal)

As indicated, there were dichotomous states within each classification of patient-related factors. There were eight possible different combinations of these factors within an image, as shown in Table 10.17.

Table 10.17 Eight possible combinations of patient-related factors within an image

Age	Size	Path
Young	Small - Med	Normal
Young	Small - Med	Pathology
Young	Med - Large	Normal
Young	Med - Large	Pathology
Old	Small - Med	Normal
Old	Small - Med	Pathology
Old	Med - Large	Normal
Old	Med - Large	Pathology

Participants' selections of their preferred image, given the patient-related factors within each image, are displayed in Table 10.18.

Analysis was needed to determine whether these two nominal variables, the selection of the participant's preferred image and patient-related factors of age, size and

pathology, were related. Data such as that displayed in Table 10.18 is said to form a cross-tabulation table or a contingency table (Keller, 2001). Keller provides a means to determine whether these variables are related. The means of undertaking this is the χ^2 test for a contingency table.

Table 10.18 Contingency table of patient-related factors and selection of modified image as the overall preferred image

Age	Size	Path	No of Selected Image		TOTAL
			Modified	Original	
Young	Small - Med	Normal	114	30	144
Young	Small - Med	Pathology	87	25	112
Young	Med - Large	Normal	90	30	120
Young	Med - Large	Pathology	88	16	104
Old	Small - Med	Normal	101	35	136
Old	Small - Med	Pathology	106	22	128
Old	Med - Large	Normal	105	15	120
Old	Med - Large	Pathology	97	23	120
TOTALS			788	196	984

A measure of strength of the χ^2 test of data in a contingency table should be undertaken (Black, 1999; Kinnear & Gray, 2000). Such a test is the calculation of the Cramér's V. Cramér's V is a coefficient of association similar to Pearson's or Spearman's coefficients of correlation (Black, 1999). As with coefficients of correlation, the value of Cramér's V indicates a lack of association when the value = 0 and perfect association when the value = 1. The χ^2 value and the value of Cramér's V are shown in Table 10.19.

Table 10.19 χ^2 and Cramér's V values of the results in Table 10.18

Test	value
χ^2 value	11.4298
df	7
p-value	0.1209
χ^2 Critical	14.0671
Cramér's V	0.107776

The χ^2 value = 11.43, p-value of 0.12 implies that there is little evidence that the choice of preferred image when answering the question as to overall preferred image was related to the patient factors of age, size and presence of pathology. The Cramér's V value = 0.11 suggest that there was no association between the patient factors of age, size and pathology and the selection of the participants' preferred images.

Logistic regression was performed to determine whether any of the patient factors had a significant influence on the participants' selection of the modified image as their overall preferred image within each anatomical group. In only two anatomical regions could any association be significantly determined. The first of these regions was the shoulder. When there was a pathological condition present in the shoulder images, participants selected the modified image as the overall preferred image. This relationship is described in Equation 10.8.

$$\begin{aligned} &\text{the odds, when a shoulder image was viewed, of selecting the} \\ &\text{modified image as having the overall highest image quality =} \\ &2.6 + (3.3)(\text{if pathology present in the image}) \quad \dots\dots\dots 10.8 \end{aligned}$$

Using Equation 10.8, if there was pathology present in a shoulder image then the odds of the participant selecting the modified image as the overall best image was 5.9 to 1. This equates to a probability of 0.855 or 85.5%.

The Wald statistic provides a p-value for the factors within Equation 10.8 of less than 0.02. With an α of 0.05 this can be considered significant. The BIC value calculated for Equation 10.8 is 7.2. This suggests a strong grade of evidence to support the inclusion of these factors in determining this relationship.

From participants' responses in the "other" anatomical region it was determined that patient size had a significant effect on the selection of the modified image as the overall preferred image. When patient size was medium to large, the modified image was the participants' preferred image. This relationship is described in Equation 10.9.

the odds, when an “other” image was viewed, of selecting the modified image as having the overall highest image quality =
 $0.5 + (2.5)(\text{if patient size was medium to large}) \dots\dots\dots 10.9$

Using Equation 10.9, if the patient’s size was medium to large in the “other” images, then the odds of the participant selecting the modified image as the overall best image was 6.5 to 1. This equates to a probability of 0.866 or 86.6%.

10.9 Participants’ Experience and Selection of the Modified Image within Anatomical Regions

Analysis of whether participants’ years of experience had an effect on the selection of the modified image as the overall preferred image was undertaken. Table 10.20 show the number and percentage of participants’ selection of the modified image as the preferred overall image, subcategorised by participant’s years of experience.

Table 10.20 Selection of the modified image as the overall preferred image by years of experience

Overall best?		Experience				Total
		< 5 years	5 - 10 years	10 - 15 years	> 15 years	
Original	Count	59	46	29	62	196
	%	24.6%	20.5%	15.8%	18.5%	19.9%
Modified	Count	181	178	155	274	788
	%	75.4%	79.5%	84.2%	81.5%	80.1%
Total	Count	240	224	184	336	984

The Kruskal-Wallis test is another means of performing χ^2 analyses. It is used when there are more than two populations, the data are nominal and the samples are independent (Keller, 2001). The data acquired from the participants, when comparing across the different groups based upon experience, met these conditions. When the anatomical regions were combined and when the regions were separated into the different anatomical regions, all p-values calculated using the Kruskal-Wallis test

were greater than 0.05. No significant difference, to an $\alpha = 0.05$, was found between the groups. The results are shown in Table 10.21.

Table 10.21 Evaluation of participants' experience and selection of the modified image as the overall preferred image

Kruskal Wallis Test across Participant's Experience			
Region	χ^2 value	df	p-value
Shoulder	5.827224	3	0.120325
C Spine	1.41934	3	0.701008
T Spine	2.685226	3	0.442744
T/L Spine	3.259554	3	0.353312
Abdomen	2.70117	3	0.440028
Facial Bones	0.142341	3	0.986312
Feet	1.624243	3	0.653905
Other	6.353042	3	0.095642
Combined	5.768468	3	0.123434

Black (1999) and Kinnear & Gray (2000) suggest that to measure the strength of association between nominal data, *Phi* coefficients are the appropriate method. *Phi* coefficients are another means of determining association between nominal data. *Phi* coefficients are preferred when cross-tabulation or contingency tables are not used.

When all the anatomical regions of the images were combined, that is, without subdivision by regions, no significant association, to an α of 0.05, between these factors and selection of the overall preferred image could be determined. The *Phi* coefficients values and their p-values are shown in Table 10.22.

Table 10.22 Tests for association between selection of the modified image as the overall preferred image and years of experience

	Value	p-value.
Phi	.077	.123
No of Valid Cases	984	

10.10 Discussion

A significant number (n=123) of participants evaluated image pairs in the RCM survey. This sample was shown to represent the population of radiographers. The participants selected their preferred image as a response to four questions and provided an indication of the level of improvement in a given factor as a response to three associated questions. Each participant evaluated an image pair for eight different anatomical regions. The sample size of responses for each anatomical region was 123. According to Keppel's (1991) criteria this sample has a power of greater than 0.8 with a medium effect size and hence is a significant number.

The total number of image pairs evaluated was 984. This sample size would have a power greater than 0.90 using Keppel's (1991) criteria for statistical power. Inferences from the results can be made with confidence.

Overall the preferred images were those that had been modified with an appropriate RCM. The participants selected the modified image in 80.1% of the cases. This was shown to be a significant degree of preference over the non-modified original image.

The appropriately modified RCM images were also judged to show a better range of optical densities within 81.7% of the images viewed. This strongly indicated that the RCM had the desired effect of optimising the range of optical densities within the image.

The anatomy within the RCM modified images was judged easiest to visualise in 81.7% of the cases. Ease of visualising anatomy provides an additional indication that the optical densities within the image were optimised.

The final factor that would enhance the viewing of DR images is simplicity in manipulation of the contrast and optical density of the image. Participants selected the modified image in 81.2% of the cases as the image for which it was easiest to optimise contrast and optical density.

Preference for the RCM modified image as the overall preferred image, as the image that had the best range of optical densities, as the image in which it was easiest to visualise all anatomy, and as the image for which it was easiest to optimise display factors, strongly suggests that the appropriately applied RCM was of benefit when viewing the DR images. The DR images that benefited the most from the application of an appropriately applied RCM were those with a large degree of subject contrast in the anatomical area radiographed.

Of the anatomical regions examined, the feet, facial bones and AP projections of the thoracic spine showed the greatest benefit of the applied RCMs. The other anatomical regions, namely the hands and upper femurs in the regions classified as “other”, AP projections of the shoulders, lateral projections of the cervical spine, abdominal images when a horizontal ray was used, and lateral projections of the thoraco-lumbar spine, all had significant associated benefits from the application of an appropriately applied RCM.

When the shoulder images were viewed and the image demonstrated pathology, the benefits of the RCM were noticed to a greater extent than when there was no pathology present. As DR is an imaging modality designed to demonstrate pathology for a diagnosis, the application of a RCM to shoulder DR images would be of increased benefit for the diagnosis of pathology.

The patient-related factor of size was also a determinant of when the RCM was of increased benefit when viewing foot images. When the patient’s size was medium to large, the survey participants selected the modified image of the feet with greater

preference than they did for the original image. An explanation for this increased preference is that when patient size is larger than the average, subject contrast in radiographic examinations of the feet is increased. The size of the patient's foot in the tarsal region increases compared to the increase in size of the toe region. The resultant increase in the density differences between these regions within the image results in an image wherein displayed contrast is increased. It was shown that an appropriately applied RCM will benefit the viewing of foot images under these conditions to a greater extent than when the patient size is smaller.

Radiographer experience in the selections of the modified image as the preferred image was not a significant factor. The results from each group of radiographers with difference years of experience could not be statistically separated from each other. This indicates that the beneficial effect of the RCM on the viewing of the DR images was found across all groups of radiographers with differing years of experience.