Comparison of the perceived image quality between two digital imaging systems for neonatal bedside radiography – a case study

Shevaun Annique van Zyl, Ramadimetja Mable Kekana*

Department of Radiographic Sciences, University of Pretoria, Pretoria, South Africa

Email addresses: shevaun.vanzyl@yahoo.com (S. van Zyl), mable.kekana@up.ac.za (M. Kekana)

Corresponding author: M Kekana, Tel: +2711 354 4371,

ABSTRACT

Background: Chest x-rays are performed daily in the neonatal intensive care and high care units. The skill of the radiographer is critical for obtaining the best image quality and limiting the patient's radiation exposure. The literature states that indirect flat panel detectors produce images of superior quality in comparison to computed radiography systems. At Steve Biko Academic Hospital a decision was made to revert from the direst digital radiography (DR) system to the computed radiography (CR) system, due to poor image quality experienced.

Method: The case study objective was to conduct a comparative analysis describing key technical factors contributing to image quality. The analysis entailed retrospectively comparing the images obtained during 2010 and 2011. An image analysis form was utilised in evaluating the technical aspects of the image. A total of 160 images were viewed by 16 participants sampled from the radiography, radiology and paediatric departments. The participants were asked to re-evaluate two of their allotted images after five days to determine their reliability.

Results: Findings were that the DR system provides significantly better image quality than the CR system ($p < 0.05$) for all the technical factors evaluated. However technical improvements are recommended. A wide variance in intra-observer reliability was also been found.

Conclusion: This case study demonstrated that DR images were considered to be superior to CR images. Recommendations include: a standardised technique for imaging the neonates; optimisation of the imaging software for the digital detectors, improved feedback systems in terms of exposure index values, and the training of radiographers and referring physicians in technical image analysis.

Key words: Image quality, technical factors, computed radiography, digital radiography, exposure index.

Introduction

Chest x-rays are taken on a daily basis in the neonatal intensive care units (NICUs) and high care units (HCUs). Doctors rely on images of optimal quality to determine the diagnosis and to monitor the neonate's treatment. Research conducted confirms the importance of the relationship between image quality and radiation dose for radiographic investigations performed in the wards with the use of a mobile x-ray unit.[1-4]. Comparisons between photostimulable phosphor (computed radiography (CR) imaging systems) and indirect flat panel detectors (IFPDs) have been made with regard to image quality and radiation dose.[1-4] Theoretically, the IFPDs should produce images of superior quality, as the detector quantum efficiency (DQE) and modulation transfer function (MTF) are superior to those of CR imaging systems.[1-4]

Steve Biko Academic Hospital (SBAH) is a tertiary academic institution situated in Pretoria, South Africa. This hospital serves as a referral hospital for most district and regional hospitals. The hospital manages a great variety of clinical conditions and also has operating theatres, intensive and high care units where mobile radiographic examinations are performed. Digital radiography using portable indirect flat panel detector technology was implemented in 2006. The assumption made by radiographers in this institution was that the introduction of digital imaging systems would aid in reducing the radiation dose to the neonates.[3] However, the overexposure in digital radiography is rewarded by high quality images, which then subjects neonates to high doses of radiation.[4] Failure to observe the exposure index values is one area that need to be drawn to the attention of the radiographers continuously, especially when imaging children, to aid in prevention of unnecessary higher exposure to radiation.

Towards the end of 2010 a decision was made to revert to CR cassettes where 18X24cm cassette size was used. The chief radiographers outlined a number of problems they had experienced with the DR system. The problems ranged from the size of the direct flat panel detector (35x43cm) in relation to the incubators, the presence of severe image noise, and an image that was consistently of reduced brightness prior to post-processing. Prior to the images being uploaded onto the picture archiving and communication system (PACS), the images experienced a loss of resolution due to the automatic magnification, which could be attributed to the large detector size. There were also no standardised image quality criteria at SBAH, according to which the images are evaluated. No research was conducted in this department to establish the factors contributing to the deterioration of the image quality, however it was assumed to be the result of the use of the flat panel detectors.

This assumption, without scientific evidence, resulted in the department reverting to the use of CR system when performing bedside radiography on neonates. It should be noted that acceptability of image quality is not a set international standard[5]. The Commission of the European Communities (CEC) gives recommendations on image quality criteria and that each radiology department must establishes its own quality standards as part of their quality assurance.[5-7] Quality assurance is a joint effort by the radiographers, radiologists, referring doctors, medical physicists and technicians.[6, 8]

The DR images were obtained by the use of the GE Definium 800 AMX digital mobile x-ray unit with permanent filtration of 1.3 aluminium equivalence, and a 32 kilowatt generator. The CR system utilised was the Agfa NX 3.0.8300, and the same mobile unit was used to obtain the images for the DR and CR systems. Each mobile unit had technique charts and radiographers were encouraged to use these exposure factors to obtain the images.

The aim of the study was to compare the perceived quality of the neonatal chest images obtained with the CR and DR systems during bedside radiography at SBAH. The objective was to make the evaluation on technical image quality only, which includes factors such as brightness, contrast, penetration, noise and resolution. These technical factors were assessed to establish which factors contributed to the poor image quality experienced, and which system was superior.

Materials and methods

Image quality produced by the two digital imaging systems was compared. The data collection instrument used for this study was a self-designed image analysis form. To ensure validity, this

form was derived from the image analysis protocol set by Mcquillen-Martenson[9] as well as the standardised technique recommended by the Commission of the European Communities[10]. The designed image analysis form provided the participants with the imaging criteria to guide them during the evaluation sessions.

The data collection form consisted of a:

- Visual analogue scale (VAS) to determine the overall image quality. Six scales in the first section of the analysis form consisted of 100mm lines with descriptors below the lines to indicate to the participants the ranges of acceptability. Each scale was preceded with a positive statement concerning technical image quality factors to be assessed (including density, contrast, resolution, penetration and noise). A key (seen in Figure 1) provided an indication of the location of the ranges of acceptability on the scale. The descriptors included: unacceptable, sub-optimal but acceptable, acceptable and optimal.

<50	50 - 70	70 - 90	>90
Unacceptable	Suboptimal acceptable	Acceptable	Optimal

Figure 1: Visual analogue scale key

The ranges of acceptability were established in a pre-test conducted prior to the commencement of the study, as well as the use of literature.[5] The VAS was chosen as the receiver operating characteristic (ROC) curve has various criteria or categories, but does not allow one to choose in between the categories when one is unsure of an answer. The VAS allows the participant to overcome the uncertainty, as there are no specific criteria to choose from.[11,12] The methodology used was adopted from Balassy et.al.[11] No post processing by manipulation of the image brightness or contrast was performed during the study. Participants were asked to rate the following technical qualities, which were posed as positive questions, on the VAS, namely; 1) overall image quality, 2) bony cortical outlines sharply defined, 3) soft tissue structures appear sharply defined, 4) brightness, 5) penetration sufficient to demonstrate the cortical outlines and 6) image contrast sufficient to demonstrate bony and soft tissue structures.

- An ordinal technical factor table listing five technical factors in the second section of the analysis form is presented in Table 1. The form presented guidance for the participants to indicate to which degree they perceived the image quality factors listed.

Table 1: Ordinal technical factor table

Technical factor	Unacceptable	Suboptimal acceptable	Acceptable
1) Density/brightness	Too low density and image is grainy. Image cannot even be manipulated to the useful range.	Too high density. The image maybe manipulated towards the useful range.	Useful density range that need no manipulation.
2) Contrast	Narrow scale with lungs appearing too white and heart shadow cannot be outlined	Narrow gray scale – lungs are too dark and spine seen through the heart. May be manipulated	Wide scale of contrast that allows visualisation of bony and soft-tissue with subtle differentiation.
3) Penetration	Penetration either insufficient to see spine through heart shadow, or burned out appearance of lungs	Penetration could be improved upon	Penetration adequate – no adjustment required
4) Resolution	Resolution poor – pixelated appearance seen	Resolution could be improved upon, though adequate for diagnostic purposes	Resolution is adequate – no improvement required
5) Noise	Image severely grainy	Image slightly grainy – could be improved upon	Image not grainy

The researcher collected over 500 images from the PACS at SBAH that were taken during January to August 2010, and January to August 2011. The collected images were saved in the Joint Photographic Experts Group (JPEG) format. The collected images included normal chest images and images of a variety of disease states. The radiographers at that stage were not monitoring the exposure index values but relying on their subjectivity in analysing the quality of the images produced, a practice which should be discouraged.[13] The collected images were labelled with specific image codes, which made the DR and CR images discernable to the researchers only. They were then randomly grouped so that each participant could evaluate ten images given to them. The images were stored on compact discs (CD) that were not labelled, which ensured that the researcher could not know which group of images were selected by each participant.

The images were displayed on a computer monitor that is similar to the viewing computer monitors in SBAH – this was an NEC MultiSync LCD monitor (model: 1980 SXi), with 96 DPI and a resolution of 1280x1024. The brightness was set to 100% and the contrast to 50%, which are the same percentages as the computer monitors utilised in the various departments in SBAH are set at. The participants were then asked to return for a re-analysis session five days after the main analysis session (day 5). The re-analysis session was done so that they could re-analyse two images (one from each imaging system) previously analysed, to establish intra-reader reliability.

For this study convenience sampling was used, because the researchers used subjects that were available to participate in the research study.[14] A total of 16 participants were invited. They were five paediatricians, nine radiographers and two radiologists. By including representatives of various departments in this study, a better consensus of the perceived image quality was expected.

The reproducibility within systems was assessed utilising the Bland-Altman limits of agreement[16] and also the intra-class correlation, when the readings on days zero and five by the same participant were being considered. The agreement between DR and CR systems was also assessed, using the Bland-Altman Limits of Agreement along with the intra-class correlation by question and day. Fisher's Exact test was employed to assess the relationship between each system and image categories[17] (from Section two on the analysis form). Testing was done at the 0.05 level of significance ($p < 0.05$). Data analysis was done utilising Stata: Release 11 Statistical software.[18] It is important to note that neonates had not been imaged with both imaging systems, as the imaging systems were not utilised simultaneously during the same time period. Thus a direct comparison was not possible.

Results and discussion

When reviewing the presented results, it is important to note that Day zero refers to the main analysis session or the first reading of an image, and Day five refers to the analysis undertaken five days later. The data mean and standard deviation of image quality were assessed on the VAS. The DR and CR systems were compared with respect to mean image quality for each question. Acting on statistical advice, the random effects of the Generalised Least Squares (GLS) regression analysis[15] was used to ensure intra-reader correlation between the first analysis on Day one and the second analysis on Day five.

- Results from the VAS

  Table 2 shows results obtained from the Fisher's Exact test. At the 5% level of significance, significant differences are characterized by P-values that are less than 0.05. As such, it can be seen from the table that there are statistically significant differences between the two groups with regards to each one of the 6 variables of comparison.

Table 2: Comparison of the mean perceived image quality values between DR and CR systems as indicated on VAS

Question	DR System		CR System		p-Value
	Mean	SD	Mean	SD	
1	64.9	27.54	56.3	22.34	0.029
2	76.5	20.61	65.5	18.15	0.000
3	68.1	25.65	56.5	22.82	0.001
4	67.5	26.24	55.5	23.59	0.002
5	70.6	26.08	61.04	20.56	0.007
6	68.7	25.34	55.03	23.51	0.000

Table 3 shows results obtained from the Bland-Altman limits of agreement[16]. At the 5% level of significance, significant differences are characterized by 95% confidence intervals for true differences between means that do not contain zero. As such, it can be seen from the table that there are no statistically significant differences between the means being compared with each other with regards to each one of the 12 variables of comparison.

Table 3: Bland-Altman Limits of Agreement between Day 0 and Day 5 readings by system

Comparison*	System	Question	Mean difference	SD difference	95% Limits of Agreement	
					Lower limit	Upper limit
Day0 - 5	DR	1	-10.6	19.72	-50.04	28.84
Day0 - 5	DR	2	-3.1	16.23	-35.56	29.36
Day0 - 5	DR	3	-2.73	20.95	-44.63	39.17
Day0 - 5	DR	4	-1.2	27.24	-55.68	53.28
Day0 - 5	DR	5	-5.07	22.01	-49.09	38.95
Day0 - 5	DR	6	2.27	19.57	-36.87	41.41
Day0 - 5	CR	1	-13.33	27.52	-68.37	41.71
Day0 - 5	CR	2	-2.93	18.32	-39.57	33.71
Day0 - 5	CR	3	-5.13	27.48	-60.09	49.83
Day0 - 5	CR	4	-19	29.74	-78.48	40.48
Day0 - 5	CR	5	-9.07	25.45	-59.97	41.83
Day0 - 5	CR	6	-8.67	30.15	-68.97	51.63

*Comparison Day zero value, minus Day five value

The Limits of Agreement in Table 4 show the difference in answers given between the DR and CR systems by day and question. Here the mean values from the CR system were subtracted from the DR system, to show how much the systems differed in quality on Day zero and Day five. Once again there was poor agreement between the answers given as indicated by the wide range within the lower and upper limits.

Table 4: Bland-Altman Limits of Agreement between DR and CR systems, by day and question

Comparison*	Day	Question	Mean difference	SD difference	95% Limits of Agreement	
					Lower limit	Upper limit
DR - CR	0	1	-0.33	28.92	-58.17	57.51
DR - CR	5	1	-3.07	34.73	-72.53	66.39
DR - CR	0	2	5	20.98	-36.96	46.96
DR - CR	5	2	5.2	26.07	-46.94	57.34
DR - CR	0	3	5.13	36.01	-66.89	77.15
DR - CR	5	3	2.73	34.04	-65.35	70.81
DR - CR	0	4	14.07	37.53	-60.99	89.13
DR - CR	5	4	-3.73	33.08	-69.89	62.43
DR - CR	0	5	1.6	28.75	-55.9	59.1
DR - CR	5	5	-2.4	25.35	-53.1	48.3
DR - CR	0	6	8.4	39.69	-70.98	87.78
DR - CR	5	6	-2.53	25.1	-52.73	47.67

*Comparison DR value, minus CR value

The intra-class correlation between the Day zero and Day five readings within images is presented in Table 5. Intra-class correlation as utilised here refers to the individual participants' readings between Day zero and Day five. From Table 5, it can be seen that there are overall higher correlation values for the DR system than the CR system. These correlation values indicate that the agreement between readings on Day zero and Day five are more reliable for the DR system.

Table 5: Intra-class correlation between Day 0 and Day 5 readings within images

Question	Imaging System	Intra-class Correlation	95% Confidence Interval
1	Pooled	0.538	(0.282 ; 0.795)
	DR	0.753	(0.530 ; 0.976)
	CR	0.194	(0.000 ; 0.690)
2	Pooled	0.650	(0.442 ; 0.858)
	DR	0.708	(0.452 ; 0.965)
	CR	0.595	(0.262 ; 0.928)
3	Pooled	0.545	(0.291 ; 0.799)
	DR	0.724	(0.480 ; 0.970)
	CR	0.312	(0.000 ; 0.777)
4	Pooled	0.313	(0.000 ; 0.638)
	DR	0.582	(0.242 ; 0.923)
	CR	0.000	(0.000 ; 0.515)
5	Pooled	0.629	(0.411 ; 0.847)
	DR	0.762	(0.547 ; 0.978)
	CR	0.436	(0.019 ; 0.853)
6	Pooled	0.563	(0.316 ; 0.809)
	DR	0.808	(0.630 ; 0.987)
	CR	0.178	(0.000 ; 0.677)

- <u>Results from the technical factor table</u>

In Table 6 the p-values for Day zero and Day five are shown. The p-values seen indicate that there were significant differences between the two imaging systems for all the questions on Day zero. Technical factors one (density/brightness) and three (penetration) did not have significant differences on Day five as those values exceed the 0.05 limit.

Table 6: Fisher's exact values for DR and CR systems comparing Day 0 and Day 5 values from the technical factor table

Question	Day	p-Value
1	0	0.001
	5	0.139
2	0	0.001
	5	0.003
3	0	0.025
	5	0.091
4	0	0.001
	5	0.025
5	0	0.001
	5	0.001

The column graphs in figures 2 and 3 show the frequency distribution of answers given for technical factor on Day zero and Day five for the DR and CR systems respectively. Figure 2 shows a graphical depiction for five sets of comparisons made with regards to five technical factors that affect image quality. It can be seen from the figure that there is a significant difference between days (0) and (5) with regards to the frequency of unacceptable contrasts. The same is true between days (0) and (5) with regards to the frequency of unacceptable noise. The answers given for the CR system were more in the range of "sub-optimal acceptable" to "unacceptable" quality.
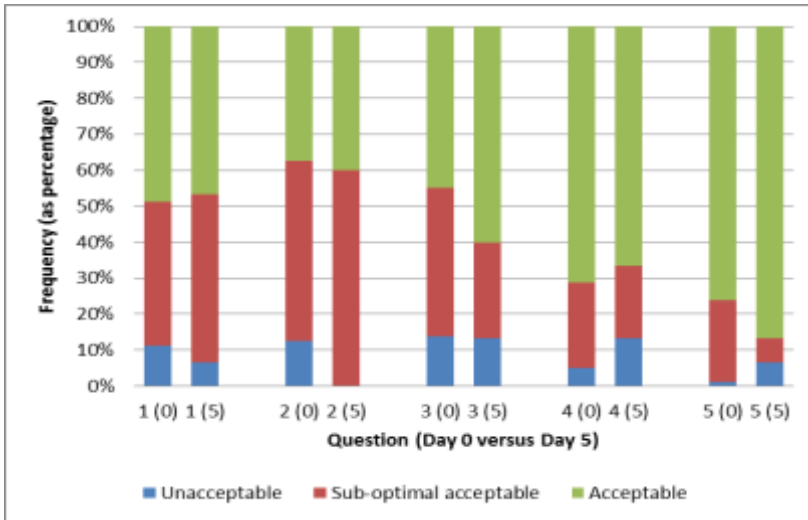
Figure 2: Frequency of answers given for the DR system on the technical factor table, on Day 0 and Day 5

Figure 3 shows a graphical depiction for five sets of comparisons made with regards to five technical factors that affect image quality. It can be seen from the figure that there is a significant difference between days (0) and (5) with regards to the frequency of unacceptable penetration.
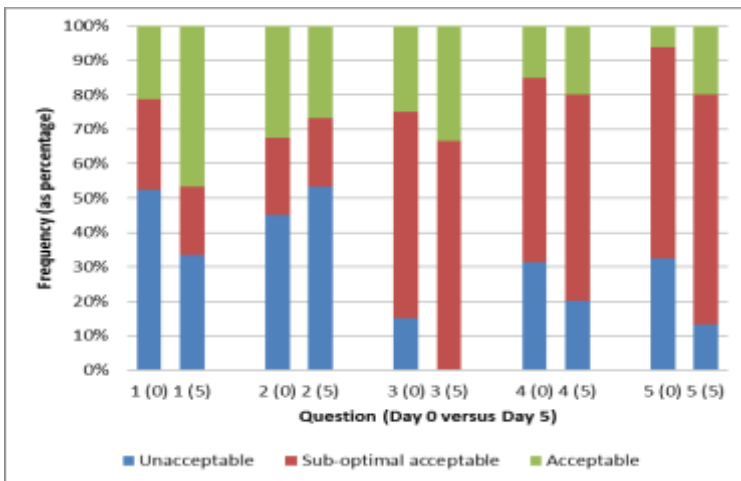

Figure 3: Frequency of the answers given for the CR system on the VAS, on Day 0 and Day 5

The results of this study seem to be in agreement with most of the literature on image quality of CR and DR imaging systems. Clinical studies, according to Korner[3], show that the flat-panel detectors utilised in DR imaging have the best image quality and low-contrast performance of all the digital detectors. In this study there were significant differences in the perceived image quality between the CR and DR imaging systems. However, when looking at the average values obtained from the VAS, it can be seen that the DR system still needs to be optimised.

The wide variance in answers given between Day zero and Day five is also explained in the literature, as perceived approaches have a lack of reproducibility. However, this limitation can be overcome when a large number of images are analysed to obtain accurate results.[19] Martin also mentions that perceived image quality is limited by an observer's visual threshold, which could also be overcome if the observer is trained in viewing many images.[19] The observer also requires a set of imaging criteria to be able to make a decision regarding the quality of the image.[19] The image analysis form for this provided guidance for participants in evaluating the image.

Conclusions

Image quality by definition is a means to visually represent the clinical information obtained by x-ray equipment.[13,19] According to Martin, this statement implies that at some stage image quality will have to be based on the judgement of the human observer.[13,19] This study utilised this approach in allowing the participants to evaluate the technical quality of neonatal chest images. The results demonstrated that participants were more satisfied with the image quality obtained with the DR imaging system than with the CR imaging system. This finding is in agreement with the current literature on this topic.[13]

Although the DR system proved to provide better technical quality images, there were still some technical factors (e.g. brightness/density, penetration and contrast) that were evaluated as "sub-optimal acceptable". The technical factors described may not have been clear to some of the participants, since there were three different categories of professionals with different expertise in image evaluation. Another limitation was that a direct comparison of the images was not possible as the two imaging systems were not used at the same time.

Some recommendations are that: acceptable image criteria for neonatal chest images should be established; a standardised technique should be used when neonates are being imaged[4]; and the imaging software should be optimised for each system as this can aid in improving the image quality even more.[2] In addition all radiographers should comply with the set standards when producing, evaluating and accepting the images. Radiographers In this hospital should be trained in analysing the images for technical quality and the observation of the exposure index values. The final recommendation is that this study could be expanded to include more images and more participants from the relevant departments or to other hospitals. Technical image analysis must form part of the quality assurance of the department.

Conflict of interest:

The authors declare that there was no financial assistance of gain that accrued from conducting this study. There is further no personal relationship with any institution or organisation that could have influenced the study and its findings.

REFERENCES

1. Rapp-Bernhardt U, Bernhardt TM, Lenzen H, Esseling R, Roehl FW, Schiborr M, Theobald-Hormann I, Heindel W. Experimental evaluation of a portable indirect flat panel detector for the pediatric chest: Comparison with storage phosphor radiography at different exposures by using a chest phantom. *Radiology.* 2005; 237[2]:485-491

2. Korner M, Weber CH, Wirth S, Pfeifer KJ, Reiser MF, Treitl M. Advances in digital radiography: Physical principles and system overview. *Radiographics.* 2007; 27[3]:675-686

3. Don S. Radiosensitivity of children: Potential for overexposure in CR and DR and magnitude of doses in ordinary radiographic examinations. *Pediatr Radiol.* 2004; 34[3]:167-172

4. Seibert JA. Trade-offs between image quality and dose. *Pediatr Radiol.* 2004; 34[3]:183-195

5. Carlton RR, Adler AM. *Principles of radiographic imaging – an art and a science.* 4th edition. New York; Clifton Park: Thompson Delmar Learning, 2006.

6. FDA News Release. *FDA proposal aims to help reduce unnecessary radiation exposure for children.* Page Last Updated: 14 May 2012. Available from: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm303386.htm Accessed: 06 May 2014

7. Bushong SC. *Radiologic science for technologists. Physics, Biology and Protection.* 7th edition. St Louis, Mo. Elsevier Mosby. 2013.

8. Kruskal JB, Anderson S, Yam CS, Sosna J. Strategies for establishing a comprehensive quality and performance improvement program in a radiology department. *Radiographics.* 2009;[29]:315-329

9. McQuillen Martenson K. *Radiographic image analysis.* 2nd edition. St Louis; Missouri: Elsevier Saunders, 2006.

10. McParland BJ, Gorka W, Lee R, Lewall DB, Omojola MF. Radiology in neonatal intensive care unit: dose reduction and image quality. *Br. J. Radiol.* 1996;[69]:929-937

11. Balassy C, Prokop M, Weber M, Sailer J, Herold CJ, Schaefer-Prokop C. Flat-panel display (LCD) versus high-resolution gray-scale display (CRT) for chest radiography: an observer preference study. *AJR.*2005;[184]:752-756.

12. Jensen MP, Chen C, Brugger AM. Interpretation of visual analog scale ratings and change scores: a reanalysis of two clinical trials of post-operative pain. *The J Pain.* September 2003;4[7]:407-414.

13. Moore QT et al. *Image gently: Using exposure indicators to improve pediatric digital radiography.* Radiol Tech. 2012[84]1:93-99

14. Burns N, Grove SK. *The practice of nursing research: Conduct, critique and utilization.* St Louis, Mo: WB Saunders/Elsevier. 2009.

15. Fox J. *Time-Series Regression and Generalized Least Squares. Appendix to An R and S-PLUS Companion to Applied Regression.* 2002. Available from http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-timeseries-regression.pdf. Accessed on 05 May 2014.

16. Myles PS, Cui J. Editorial I – Using the Bland-Altman to measure agreement with repeated measures. *Br J of Anaest.* 2007; 99[3];309-311.

17. Samuels ML, Witmer JA. *Statistics for the Life Sciences.* 3rd ed. New Jersey; Upper Saddle River: Pearson Education Inc. 2003.

18. StataCorp. *Stata: Release 11.* Statistical software. Texas: College Station, StataCorp LP. 2009.

19. Martin CJ. The importance of radiation quality for optimisation in radiology. *Biomed Imaging Interv J.* 2007;3[2]e38.