Relationship between the visual evaluation of pathology visibility and the physical measure of low contrast detail detectability in neonatal chest radiography.

Abstract

Introduction: The detectability of low contrast detail (LCD) is a method used to assess image quality (IQ) in neonatal radiography; however, there is a lack of data on the relationship between LCD detectability and visual IQ. The study aims at investigating the relationship between the LCD detectability and visual IQ and pathology visibility (PV).

Methods: Several acquisition parameters were employed to obtain a group of images from a neonatal Gammex chest phantom. Three observers applied relative visual grading analysis (VGA) for assessing the IQ and PV. A simulated pneumothorax visibility (PNV) and simulated hyaline membrane disease visibility (HMV) represented PV. Next, a CDRAD 2.0 phantom was radiographed utilising the same acquisition protocols, and several paired images were obtained. With the use of CDRAD analyser software, the detectability of LCD was assessed and expressed by an image quality figure inverse ($IQFi_{inv}$) metric. The correlation between the IQF_{inv} and each of IQ, PNV and HMV was examined.

Results: The physical measure (IQF_{inv}) and the visual assessment of IQ were shown to be strongly correlated ($r=0.95$; $p<0.001$). Using Pearson's correlation, the IQF_{inv}, PNV, and HMV were found to be strongly correlated ($r=0.94$; $p<0.001$) and ($r=0.92$; $p<0.001$), correspondingly.

Conclusion: Results of the study show that physical measures of LCD detectability utilising the CDRAD 2.0 phantom is strongly corelated with visual IQ and PV (PNV and HMV) and can be used to evaluate IQ when undertaking neonatal chest radiography (CXR).

Implications for practice: This study establishes the feasibility of utilising the physical measure (IQF_{inv}) and the CDRAD 2.0 phantom in routine quality assurance and neonatal CXR optimisation studies.

Keywords: CDRAD 2.0 phantom, neonatal CXR, low contrast detail detectability, physical and visual image quality evaluation.

Introduction:

The chest X-ray (CXR) examination is one of the most frequent examinations performed on neonates because of the likelihood of respiratory distress syndrome and pneumothorax[1–3]. Based on the ALARA (as low as reasonably achievable) principle, examinations must be performed with appropriate image quality (IQ) for diagnosis but using the lowest achievable radiation dose[4,5] .This is even more important for neonates who in comparison with adults have higher tissue radiosensitivity[6,7]. Therefore, image quality and dose optimisation studies are an important requirement when seeking to achieve the ALARA principle.

Several studies have reported the use of physical measures for the detectability of low contrast detail (LCD) (which describes the ability to image small objects against a background of low contrast) via the CDRAD 2.0 phantom[8–10]. Such reports are based around optimisation studies where detectability of LCD is the method for IQ evaluation for the resultant CXR images. In this situation, the detectability of LCD and the CDRAD 2.0 phantom is particularly suitable since it has high reliability, it is easy to use and quick. However, to the authors' knowledge the link between the physical measure of the detectability of LCD, via CDRAD 2.0 phantom, and the visual IQ evaluation and pathology visibility (PV) for neonatal chest radiography has not been studied yet.

This study aims to examine the relationship between the physical measure of the detectability of LCD, via CDRAD 2.0 phantom, and the visual evaluations of IQ and PV when undertaking neonatal chest radiography.

Methods

Imaging equipment and technique

The study used a Gammex phantom (Gammex Inc., Middleton, WI)[11] and a CDRAD 2.0 phantom ((Artinis Medical System, The Netherlands)[12]. The first phantom is an anthropomorphic neonatal chest phantom simulating a one to two kg child and is used for assessing IQ and PV, it has used widely in the literature for optimisation studies[13–18]. It has two lungs, one of which has a simulated pneumothorax and the other has simulated hyaline membrane disease (**Figure 1**)[11]. The second

phantom (CDRAD 2.0 phantom, **Figure 2**) was used to evaluate the physical measure of LCD detectability. It is a uniform background phantom which consists of 1 cm thick squares of acyclic plastic plate (26.5 x 26.5 cm) with holes of different depths and diameters[12].

42 radiographic images were generated using Gammex phantom with manual exposure factors (kVp, mAs), a fine focal spot, 115 cm source to image receptor distance, no anti-scatter radiation grid and no additional filtration. During image acquisition different settings of tube potential (55, 60, 65, 70, 75, 80 and 85 kVp) and tube current-time (0.5, 0.7, 1.0, 1.6, 2.0 and 2.5 mAs) were applied.

The same previously specified acquisition parameters were also used to create 42 CDRAD 2.0 radiographic images. To simulate a neonate, a CDRAD 2.0 phantom was positioned between 6 cm of medical grade polymethyl methacrylate (PMMA) slabs. Based on the phantom manufacturer's recommendations, X-ray acquisitions were repeated three times for each exposure factor setting[12]. For all resultant radiographic images a paediatric AP chest post-processing algorithm was applied.

A commercially accessible X-ray machine, supplied by Wolverson X-ray Ltd and equipped with 3 mm Al inherent filtration and a Konica Minolta Aero DR detector from Konica Minolta Medical Imaging USA, Inc., Wayne, NJ, was used for image acquisitions. To produce a wide variety of image qualities, a wide range of acquisition parameters were used to generate phantom images that resembled those commonly acquired in clinical practice.

Image quality evaluation

The CDRAD analyser program computed the detectability of LCD for the CDRAD 2.0 phantom images which the image quality figure inverse (IQF_{inv}) represents [12].

The IQF_{inv} was calculated with Eq. (1).

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

The lowest diameter in column (i) with an accurately recognized visible hole is D(i, th); the contrast of the threshold visible hole in column (i) is Ci.

For the Gammex phantom images, the visual IQ and PV was measured using the relative visual grading analysis (VGA) method by three qualified diagnostic radiographers with clinical experience more than five years. A 5-point Likert scale (worse than, worse, equal to, better, better than) was applied via bespoke Java-based software that displays images to the observers in an arbitrary order[19]. Two 5 mega-pixel (2048 by 2560 pixels) display monitors (DOME E5, NDSsi, Santa Rosa, CA) were used to display the images; the left monitor displayed the reference image, while the right monitor presented the experimental images in a different order. In comparison to the other acquired images, two experienced radiologists chose the reference image as having "average" IQ, pneumothorax visibility (PNV), and hyaline membrane visibility (HMV). **Table 1** illustrates the criteria utilised for the IQ and PV assessment, these were adapted from the European Guidelines and a report by Smet and colleagues[17,20]. The colours highlighted in Figure 1 correspond to the colour-coding for the IQ criteria listed in Table 1.

The sum of the scores from all the IQ criteria was used to calculate the overall visual IQ score for each image and each observer. Then, the mean (and standard deviation [SD]) scores for the three observers were calculated to arrive at the final IQ score. This method was similar to that used to calculate the overall PV (PNV and HMV). For the observational component of the research, the University of Salford Research Ethics Committee approved the study (HSR1617-76).

Figure 1. Result X-ray image using the Gammex phantom.

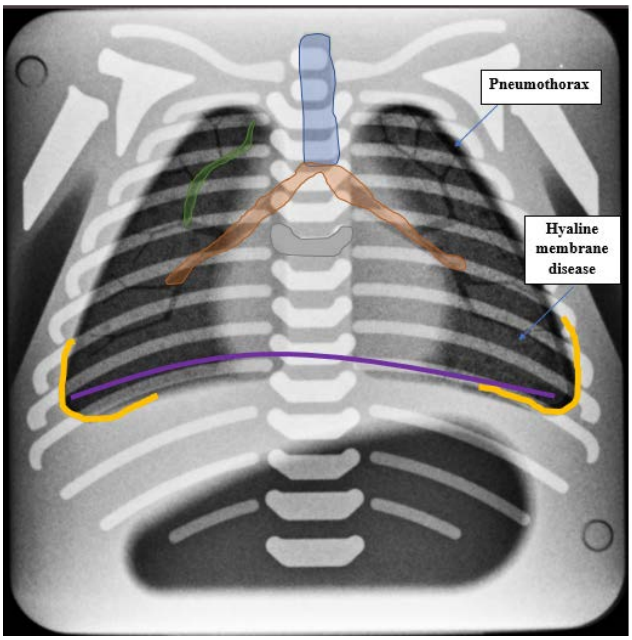

Figure 2 (a) CDRAD 2.0 phantom and (b) resultant its x-ray image.

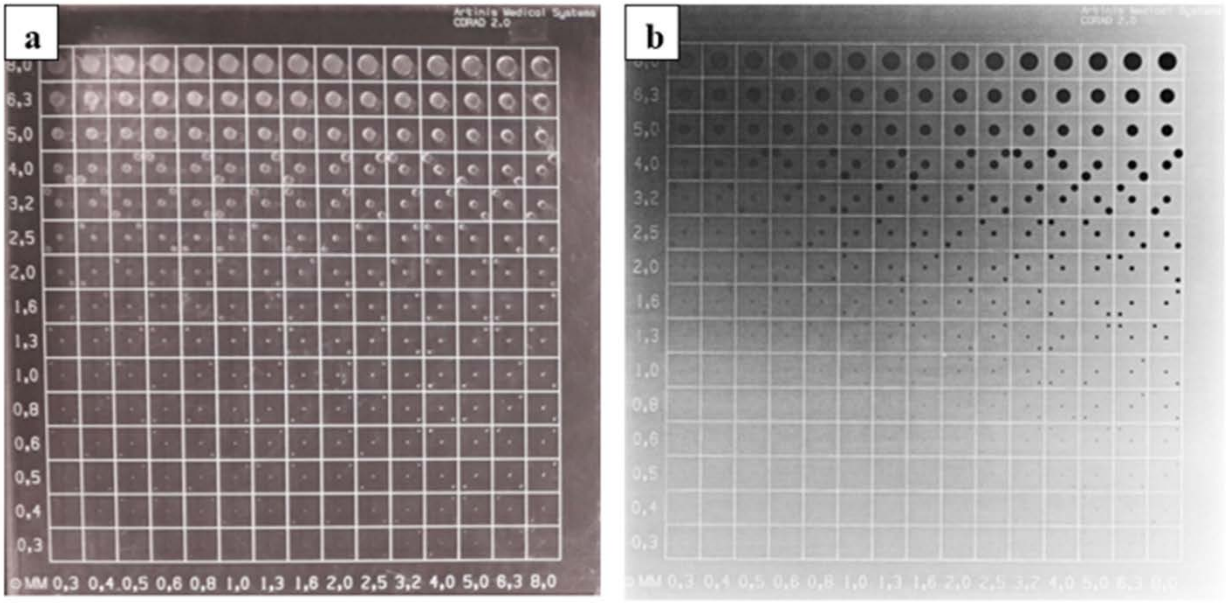

Table 1. Example of the IQ and PV criteria used[17,20].

Item	Criteria
IQ criteria	
1	Tracheal reproduction
2	Proximal bronchi reproduction
3	The small peripheral airways reproduction
4	Visually sharp diaphragm reproduction
5	Visually sharp costophrenic angles reproduction
6	Spinal reproduction
7	The amount of noise in the image
PV criteria	
1	The visibility of the pneumothorax
2	The visibility of the hyaline membrane disease

Statistical analysis

Data analysis was done with SPSS software (IBM Inc., Armonk, NY). The data distribution was examined via the Shapiro-Wilk test, and the findings indicated that IQF_{inv}, IQ, and PV were normally distributed. Pearson's correlation coefficient was used for measuring strength and direction of any relationship between the IQF_{inv} and each of IQ, PNV and HMV. The correlation (r) was interpreted as r=0.1–0.29 (small), r=0.30–0.49 (medium), and r=0.50–1.0 (large)[21]. An inter-class correlation coefficient (ICC) was performed to determine the inter-observer variation in the IQ, PNV and HMV evaluations by the three observers[22,23].

Results

The visual assessments of IQ and the physical measure of LCD detectability (IQF_{inv}) were shown to be strongly positively correlated (r=0.95; p0.001). IQF_{inv} and the visual assessments of PNV and HMV also showed a high positive connection (r=0.94; p0.001) and (r=0.92; p0.001), respectively. **Figures 3**, **4** and **5** demonstrate the respective linear regression curves between the IQF_{inv} and each of IQ, PNV and HMV, respectively.

Excellent agreement was achieved between the observers (ICC = 0.91; 95% CI 0.85–0. 95) for the IQ assessment, (ICC = 0.95; 95% CI 0.92–0. 97) for the PNV and (ICC = 0.91; 95% CI 0.86–0. 95) for the HMV.

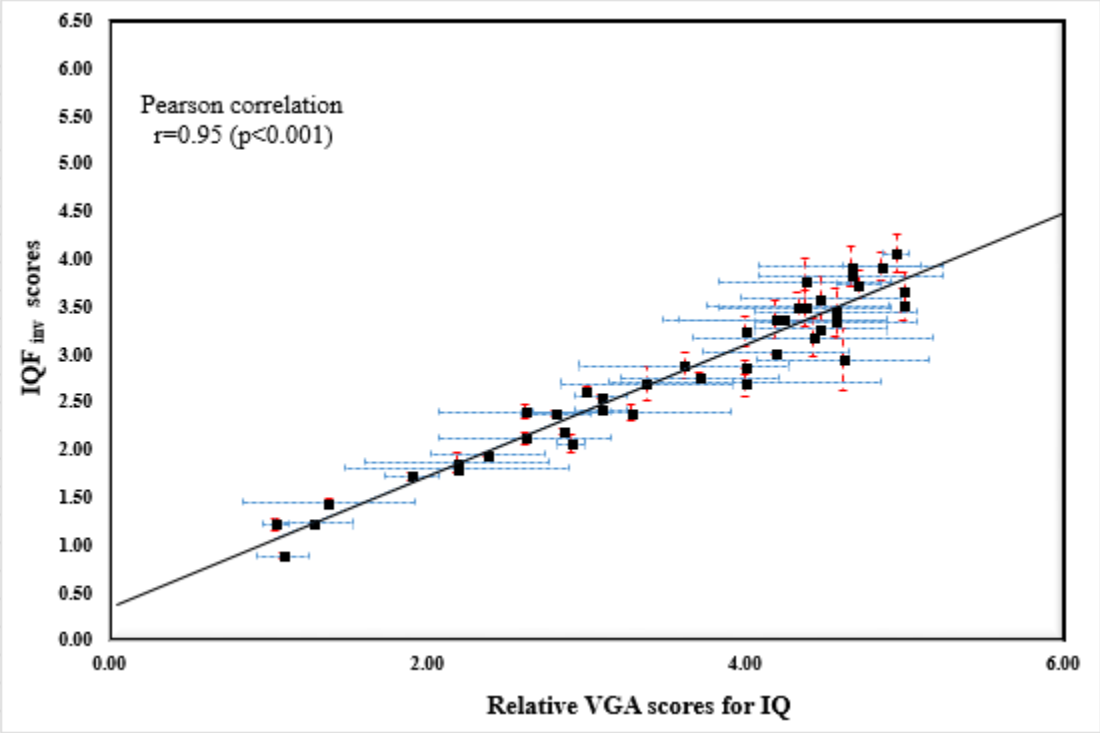

Figure 3. Demonstrates the linear regression curve between the mean scores of IQF_{inv} and the mean visual scores of IQ for the 42 images. Vertical error bars indicate the values of SD from the three CDRAD IQF_{inv} scores, while horizontal error bars across indicate the SD of the scores from the three observers.

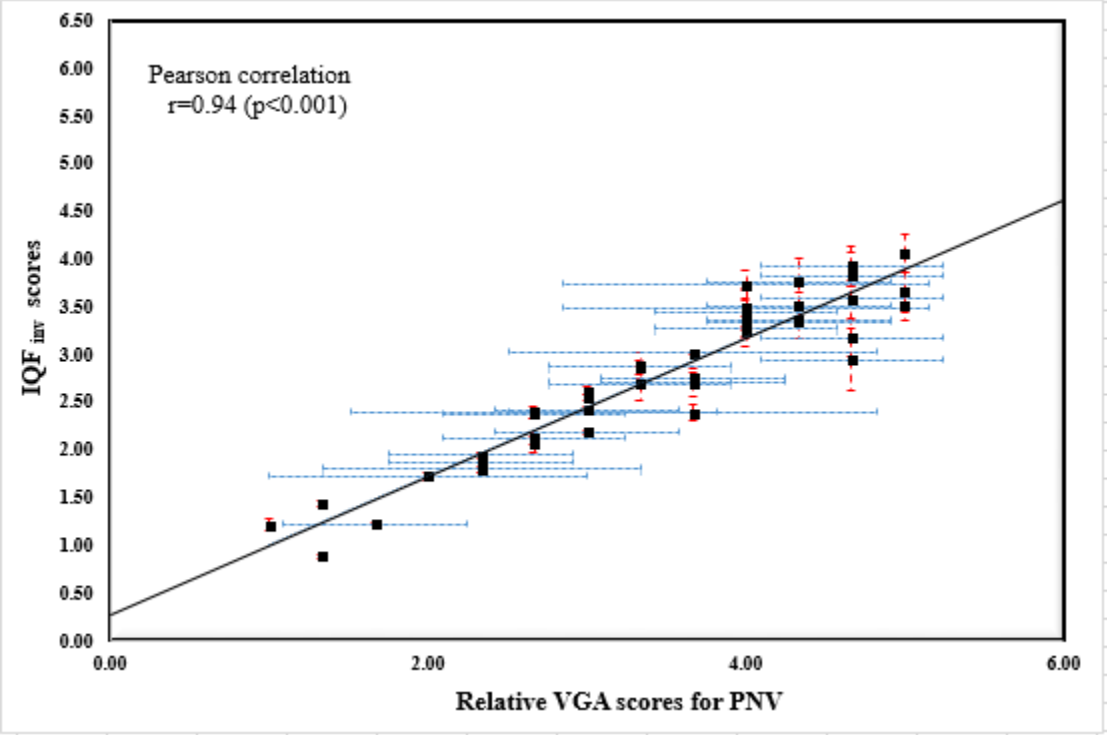

Figure 4. Demonstrates the linear regression curve between the mean scores of IQF_{inv} and the mean scores of PNV visual evaluation. Vertical error bars represent the values of SD from the three CDRAD IQF_{inv} scores, while the horizontal error bars across indicate the SD of the scores from the three observers.

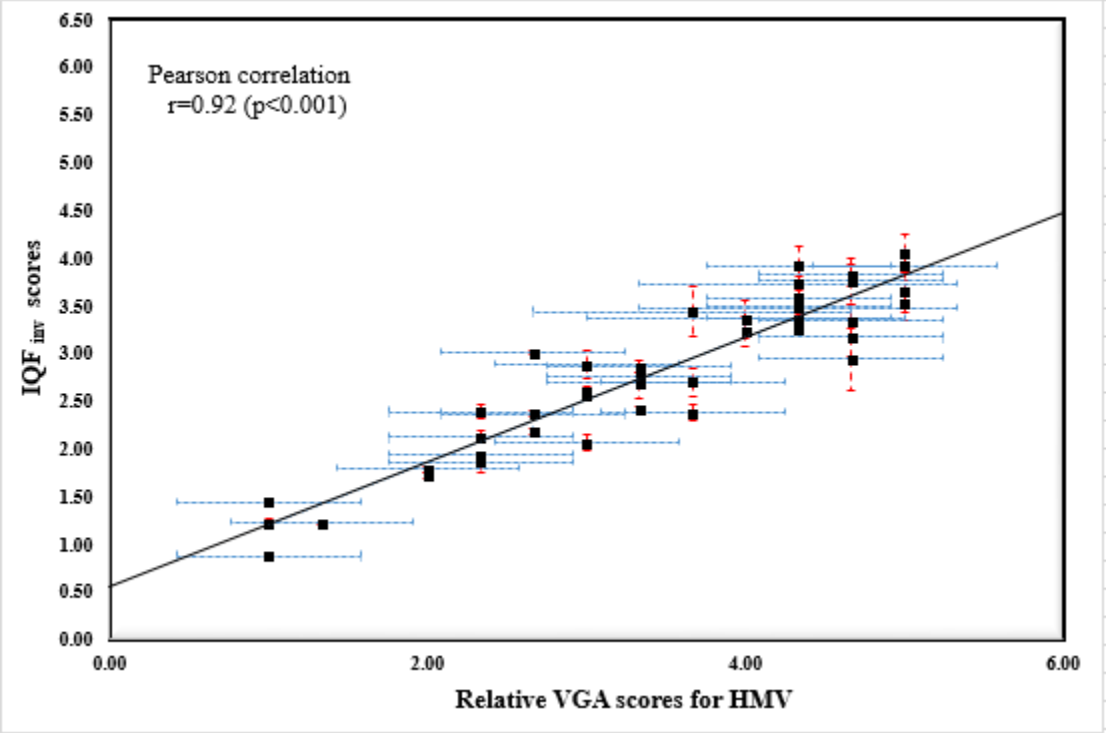

Figure 5. Demonstrates the linear regression curve between the mean scores of IQF_{inv} and the mean scores of HMV visual evaluation. Vertical error bars represent the values of SD from the three CDRAD IQF_{inv} scores, while the horizontal error bars across indicate the SD of the scores from the three observers.

Discussion

The results of this study revealed a strong positive correlation between the physical measure of the detectability of LCD (IQF_{inv}) generated via images from the CDRAD 2.0 phantom and the visual evaluations of IQ using relative VGA both general IQ ($r=0.95$; $p<0.001$) and PV. Specifically for the visibility of the pneumothorax ($r=0.95$; $p<0.001$) and hyaline membrance disease ($r=0.95$; $p<0.001$). This strong positive association between physical IQF_{inv} and the visual evaluation for each of IQ, PNV, and HMV shows that the CDRAD 2.0 phantom can be used in neonatal CXR optimisation research and routine quality assurance testing. However, more research is needed to investigate the clinical association between IQF_{inv} and the visual evaluation of IQ, PNV, and HMV in patients to prove it as a valid method.

In diagnostic imaging, there essentially two basic tasks 1) acquiring images and 2) interpreting the resultant images for the benefit of patient management[24]. The ability of clinicians to correctly interpret a medical image is largely governed by the quality of the image. Image quality can be evaluated using various approaches that fall into three categories: physical, psychophysical and observer performance[25].

In research studies, methods for measuring observer performance like receiver operating characteristic (ROC) and VGA are frequently used to assess clinical IQ. ROC and VGA approaches, however, can be difficult to practically implement for evaluating IQ as part of normal quality assurance programmes and when making comparisons within and between hospitals, since they are time consuming and need a significant number of images to be acquired[26–28].

The advantages of the CDRAD 2.0 phantom and the IQF_{inv} metric are based around the need not to acquire patient data and that images can be evaluated by a non-radiology trained observer or dedicated image analyser software. The physical IQ test based on the detectability of LCD via the CDRAD 2.0 phantom is therefore more practical than the VGA and ROC techniques. However, there has been a scarcity of evidence on their correlations with perceptual (visual) approaches in the past.

There are several restrictions when using the CDRAD 2.0 phantom to measure IQ and conduct optimisation studies. For instance, when imaging human anatomy, the CDRAD 2.0 phantom creates a homogeneous background that ignores the impact of anatomical noise [29]. According to

the Rose model, quantum noise represents the limiting element that can influence IQ (LCD detectability), and this phantom is based on that assumption[30]. However, it may be limited for a number of common diagnostic radiology clinical procedures. Rather than quantum noise, the anatomical noise is thought to be the most relevant component in CXR IQ evaluation within clinical practice[31–33].

From the literature, there are only three studies that have attempted to investigate the correlation between the IQF_{inv} physical measure using CDRAD 2.0 phantom and the visual evaluation of IQ when undertaking adult CXR examination[34–36]. However, to the best of the researchers' knowledge, previous studies have not examined the relationship between the physical measure of IQF_{inv} and the visual measure for each of IQ and diffuse PV represented by PNV and HMV in neonatal chest radiography.

Al-Murshedi and colleagues[34] investigated the relationship between IQF_{inv} from CDRAD 2.0 phantom and both visual IQ, and lesion visibility (LV) for images acquired using an adult anthropomorphic Lungman chest phantom (relative VGA technique). These authors discovered a high positive correlation ($r=0.91$; $p<0.001$) between the visual IQ and the IQF_{inv}, as well as a good positive correlation ($r=0.68$; $p<0.001$) between the IQF_{inv} and the LV. However, this study was based on adult CXR examinations, and the simulated lesion utilised was a small spherical ‘focal’ lesion, which is distinctly different from the current study, that was developed for neonates with various diseases that were appropriate for the respective age group.

A study by De Crop and colleagues investigated the association between the IQF_{inv} from CDRAD 2.0 phantom and visual IQ for cadaveric images, which were found to have a strong positive connection ($r=0.91$; $p=0.001$)[35]. However, this was again performed on adult CXR examinations and the connection between the visual PV and the physical measure of IQF_{inv} was not investigated. More recently, Yalcin and colleagues found a strong positive connection between the IQF_{inv} and the visual IQ for images from an adult Alderson anthropomorphic chest phantom[36]. However, within this work any correlations between the visual PV and the physical measure of IQF_{inv} was not investigated.

Based on this study's findings, future research could concentrate on developing a baseline IQF_{inv} value for neonatal CXR examinations that specifies the LCD detectability threshold necessary for establishing a suitable IQ for diagnosis. The accessibility of such baseline values is extremely

useful for routine quality assurance and quality control programmers and for ensuring that IQ remains within acceptable limits, as well as for optimisation studies to evaluate how different imaging techniques impact IQ and radiation dose in clinical practice.

It is necessary to acknowledge that using the Gammex phantom to simulate a neonate and produce CXR images has significant limitations due to the absence of movement and anatomical variation that would be evident in a human neonatal population. Also, the type of imaging machine utilised for obtaining the phantom images, which was a stationary X-ray machine, would not be fully reflective of the likelihood that neonatal imaging is commonly performed using mobile X-ray machines. Investigation of the relationship between IQF_{inv} and patient visual assessments of their IQ, PNV, and HMV is necessary and will be the subject of future research.

Conclusion

According to our findings, there is a strong correlation between the physical IQF_{inv} measure using CDRAD 2.0 phantom and the visual evaluation of both IQ and PV which was represented by PNV and HMV in a Gammex phantom when undertaking neonatal chest radiography. This demonstrates the potential of the physical measure of IQF_{inv} being used to predict visual IQ and PV and then for neonatal CXR optimisation studies.

References

1. Bacher K, Smeets P, Bonnarens K, De Hauwere A, Verstraete K, Thierens H. Dose Reduction in Patients Undergoing Chest Imaging: Digital Amorphous Silicon Flat-Panel Detector Radiography Versus Conventional Film-Screen Radiography and Phosphor-Based Computed Radiography. *Am J Roentgenol*. 2003;181(4):923-929. doi:10.2214/ajr.181.4.1810923

2. Veldkamp WJH, Kroft LJM, Geleijns J. Dose and perceived image quality in chest radiography. *Eur J Radiol*. 2009;72(2):209-217. doi:10.1016/j.ejrad.2009.05.039

3. Schäfer SB, Papst S, Fiebich M, Rudolph C, de Laffolie J, Krombach GA. Modification of chest radiography exposure parameters using a neonatal chest phantom. *Pediatr Radiol*. 2020;50(1):28-37. doi:10.1007/s00247-019-04522-1

4. Willis CE. Optimizing digital radiography of children. *Eur J Radiol*. 2009;72(2):266-273. doi:10.1016/j.ejrad.2009.03.003

5. Sensakovic WF, O'Dell MC, Letter H, et al. Image quality and dose differences caused by vendor-specific image processing of neonatal radiographs. *Pediatr Radiol*. 2016;46(11):1606-1613. doi:10.1007/s00247-016-3663-2

6. Mathews JD, Forsythe A V., Brady Z, et al. Cancer risk in 680 000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. *BMJ*. 2013;346(may21 1):f2360-f2360. doi:10.1136/bmj.f2360

7. Raissaki MT. Pediatric radiation protection. *Eur Radiol Suppl*. 2004;14(S1):74-83. doi:10.1007/s10406-004-0011-7

8. Al-Murshedi S, Hogg P, Meijer A, Erenstein H, England A. Comparative analysis of radiation dose and low contrast detail detectability using routine paediatric chest radiography protocols. *Eur J Radiol*. 2019;113(September 2018):198-203. doi:10.1016/j.ejrad.2019.02.017

9. Ween B, Olstad M, Jakobsen JÅ, Olsen DR. Pediatric digital chest radiography, comparison of grid versus non-grid techniques. *Eur J Radiogr*. 2009;1(4):201-206. doi:10.1016/j.ejradi.2010.09.002

10. Brosi P, Stuessi A, Verdun FR, Vock P, Wolf R. Copper filtration in pediatric digital X-ray imaging: its impact on image quality and dose. *Radiol Phys Technol*. 2011;4(2):148-155. doi:10.1007/s12194-011-0115-4

11. Dewerd L.A KM. *The Phantoms of Medical and Health Physics*. (DeWerd LA, Kissick M, eds.). Springer; 2014.

12. Burght R van der, Floor M, Thijssen M, Bijkerk R. *Manual CDRAD 2.0 Phantom and Analyser Software Version 2.1*. Artinis Medical Systems; 2014.

13. Smans K, Struelens L, Smet M, Bosmans H, Vanhavere F. Cu filtration for dose reduction

in neonatal chest imaging. *Radiat Prot Dosimetry*. 2010;139(1-3):281-286. doi:10.1093/rpd/ncq061

14. Precht H, Tingberg A, Waaler D, Outzen CB. New Developed DR Detector Performs Radiographs of Hand, Pelvic and Premature Chest Anatomies at a Lower Radiation Dose and/or a Higher Image Quality. *J Digit Imaging*. 2014;27(1):68-76. doi:10.1007/s10278-013-9635-2

15. Singh VH, Pradhan H. Neonatal chest radiography - Comparing image quality and dose for contact-techniques vs. under-tray techniques. *Radiogr Open*. 2015;2(1):65. doi:10.7577/radopen.1530

16. Cohen MD, Markowitz R, Hill J, Huda W, Babyn P, Apgar B. Quality assurance: A comparison study of radiographic exposure for neonatal chest radiographs at 4 academic hospitals. *Pediatr Radiol*. 2012;42(6):668-673. doi:10.1007/s00247-011-2290-1

17. Smet MH, Breysem L, Mussen E, Bosmans H, Marshall NW, Cockmartin L. Visual grading analysis of digital neonatal chest phantom X-ray images: Impact of detector type, dose and image processing on image quality. *Eur Radiol*. 2018;28(7):2951-2959. doi:10.1007/s00330-017-5301-2

18. Al-Murshedi S, Peter Hogg, England A. Neonatal chest radiography: Influence of standard clinical protocols and radiographic equipment on pathology visibility and radiation dose using a neonatal chest phantom. *Radiography*. 2020;26(4):282-287. doi:10.1016/j.radi.2020.02.005

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

20. Commission of the European Communities (CEC). *European Guidelines on Quality Criteria for Diagnostic Radiographic Images in Paediatrics:(UR 16261EN).*; 1996. ftp://ftp.cordis.lu/pub/fp5-euratom/docs/eur16261.pdf

21. Filed A. *Discovering Statistics Using SPSS Statistics*. Third edit. Sage publications Ltd; 2009.

22. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

23. Portney L, Watkins M. *Foundations of Clinical Research: Applications to Practice.New Jersey: Prentice Hall*.; 2000.

24. Tapiovaara MJ. Review of relationships between physical measurements and user evaluation of image quality. *Radiat Prot Dosim*. 2008;129(1-3):244-248. doi:10.1093/rpd/ncn009

25. M nsson LG. Methods for the Evaluation of Image Quality: A Review. *Radiat Prot Dosimetry*. 2000;90(1):89-99. doi:10.1093/oxfordjournals.rpd.a033149

26. Lanhede B, Båth M, Kheddache S, et al. The influence of different technique factors on

image quality of chest radiographs as evaluated by modified CEC image quality criteria. *Br J Radiol*. 2002;75(889):38-49. doi:10.1259/bjr.75.889.750038

27. Im JG, Chung MJ, Kim HY, et al. Detection of simulated chest lesions by using soft-copy reading: comparison of an amorphous silicon flat-panel-detector system and a storage-phosphor system. *Radiology*. 2002;224(1):242—246. doi:10.1148/radiol.2241011441

28. Yano Y, Yabuuchi H, Tanaka N, et al. Detectability of simulated pulmonary nodules on chest radiographs: Comparison between irradiation side sampling indirect flat-panel detector and computed radiography. *Eur J Radiol*. 2013;82(11):2050-2054. doi:10.1016/j.ejrad.2013.05.036

29. Månsson LG, Båth M, Mattsson S. Priorities in optimisation of medical X-ray imaging—a contribution to the debate. *Radiat Prot Dosimetry*. 2005;114(1-3):298-302. doi:10.1093/rpd/nch578

30. Rose A. The sensitivity performance of the human eye on an absolute scale. *J Opt Soc Am*. 1948;38(2):196-208. doi:10.1364/JOSA.38.000196

31. Båth M, Håkansson M, Börjesson S, et al. Nodule detection in digital chest radiography: part of image background acting as pure noise. *Radiat Prot Dosimetry*. 2005;114(1-3):102-108. doi:10.1093/rpd/nch523

32. Samei E, Flynn MJ, Eyler WR. Detection of Subtle Lung Nodules: Relative Influence of Quantum and Anatomic Noise on Chest Radiographs. *Radiology*. 1999;213(3):727-734. doi:10.1148/radiology.213.3.r99dc19727

33. Håkansson M, Båth M, Börjesson S, Kheddache S, Allansdotter Johnsson Å, Månsson LG. Nodule detection in digital chest radiography: Effect of system noise. *Radiat Prot Dosimetry*. 2005;114(1-3):97-101. doi:10.1093/rpd/nch525

34. Al-Murshedi S, Hogg P, England A. An investigation into the validity of utilising the CDRAD 2.0 phantom for optimisation studies in digital radiography. *Br J Radiol*. 2018;91(April):20180317. doi:10.1259/bjr.20180317

35. De Crop A, Bacher K, Van Hoof T, et al. Correlation of Contrast-Detail Analysis and Clinical Image Quality Assessment in Chest Radiography with a Human Cadaver Study. *Radiology*. 2012;262(1):298-304. doi:10.1148/radiol.11110447

36. Yalcin A, Olgar T, Sancak T, Atac GK, Akyar S. Correlation between physical measurements and observer evaluations of image quality in digital chest radiography. *Med Phys*. 2020;47(9):3935-3944. doi:10.1002/mp.14244