CLINICAL AUDIT OF IMAGE QUALITY IN RADIOLOGY USING VISUAL GRADING CHARACTERISTICS ANALYSIS

Erik Tesselaar, Nils Dahlström and Michael Sandborg

Linköping University Post Print

Tweet

N.B.: When citing this work, cite the original article.

Clinical audit of image quality in radiology using visual grading characteristics analysis

Erik Tesselaar[1], Nils Dahlström[2,3] and Michael Sandborg[1,3]

1. Department of Radiation Physics, and Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

2. Department of Radiology and Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

3. Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden

Running title

Image quality audit using VGC analysis

Corresponding address

Erik Tesselaar

Radiation Physics

Department of Medical and Health Sciences

Linköping University

SE-58183 Linköping, Sweden

tel. +46 10 1038059

email: erik.tesselaar@liu.se

Clinical audit of image quality in radiology using visual grading characteristics analysis

Erik Tesselaar, Nils Dahlström and Michael Sandborg

ABSTRACT

The aim of this work was to assess whether an audit of clinical image quality could be efficiently implementedwithin in a limited time frame using visual grading characteristics (VGC) analysis.

Lumbar spine radiography, bedside chest radiography and abdominal CT were selected. For each examination, images were acquired or reconstructed in two ways. Twenty images per examination were assessed by 40 radiology residents using visual grading of image criteria. The results were analysed using VGC. Inter-observer reliability was assessed.

The results of the visual grading analysis were consistent with expected outcomes. The inter-observer reliability was moderate to good and correlated with perceived image quality ($r^2=0.47$). The median observation time per image or image series was within 2 minutes.

These results suggest that the use of visual grading of image criteria to assess the quality of radiographs provides a rapid method for performing an image quality audit in a clinical environment.

O12-5 Image quality audit using VGC analysis

INTRODUCTION

A clinical audit in radiology is a systematic evaluation of radiological procedures. The purpose of an audit is to improve the quality and the outcome of patient care through structured review (EU 2013/59/Euratom). It is a multidisciplinary activity and an integral part of the quality management system [1]. According to the Swedish Radiation Safety Authority enactment [2], an audit should be conducted at regular intervals.

Audits in radiology can have various purposes. A patient dose audit is commonly initiated after the introduction of a new imaging system, and should then be carried out at regular intervals [3]. Also, the justification for a specific radiographic procedure can be audited. A fundamental aspect of diagnostic imaging is the quality of the images. Receiver operating characteristics (ROC) analysis [4] is regarded as the gold standard for evaluation of image quality in the detection of a known pathology. However, a limitation of ROC-analysis is that knowledge of the ground truth is required, but is often not readily available. In addition, the displayed pathology must be close to the limit of being detectable by the observer for the ROC method to work properly.

To be of practical interest, a clinical audit of image quality should not only be accurate and precise, but also reasonably fast and cost-effective to carry out in a clinical environment. An alternative approach to ROC-analysis is visual grading of the clarity of important anatomical structures. These structures may form the basis of image criteria, as defined by the European Commission [5,6]. One important advantage of visual grading is that observers can assess almost any radiograph, provided that the image criteria are properly selected, and have validity and relevance to the observer. Visual grading characteristics [7] and visual grading regression [8] are two practical implementations of visual grading techniques in clinical settings. They both rely on the hypothesis that the possibility to detect pathology correlates with the clarity of the reproduction of important anatomy. Their main advantage lies in the fact that they are less time consuming and less difficult to set up than ROC-type

3

evaluations [9]. It is therefore anticipated that, even in a busy clinical environment, a large proportion of radiologists would be able to take time to participate in visual grading studies. This increased participation rate would increase the statistical power of the result.

Computer image analysis software is sometimes used to increase the precision of the analysis and remove the subjective nature of visual evaluation of test-phantom images or image series [10, 11]. It has been shown that physical image quality measurements or calculations of signal-to-noise ratio are positively correlated with human observer studies [12-15]. However, radiology still relies exclusively on subjective visual assessment by radiologists for a positive or negative diagnosis. Therefore, including human observers in the clinical audit is essential. Furthermore, the use of images of real patients instead of phantoms incorporates the expertise of the radiographers required when producing high quality radiographs.

The aim of this work was to assess whether an audit of clinical image quality could be efficiently implemented within a limited time frame by using visual grading characteristics analysis of image criteria.

MATERIALS AND METHODS

Imaging systems

Three types of x-ray examinations for adult patients were evaluated by using two different image acquisition settings (A or B) per examination type. The first examination type was an anteroposterior (AP) projection of the lumbar spine acquired with a digital system (Siemens Axiom Aristos FX Plus, Siemens Healthcare, Upplands Väsby, Sweden). The acquisition settings included either sensitivity class 400 (Setting A) or sensitivity class 560 (Setting B), resulting in two different air kerma at the image detector. The two sensitivity classes correspond to an air kerma at the image detector plane of 2.5 and 1.8 μGy, respectively. The second examination type was bedside chest radiography AP,

using either a mobile computed radiography system (Siemens Mobilett 2, Siemens Healthcare, Upplands Väsby, Sweden, with image plates, (CR ST-V), Philips Healthcare, Stockholm, Sweden, Setting A) or a mobile digital imaging system with a portable flat panel image detector (Movix Dream, Mediel AB, Mölndal, Sweden,, Setting B). The third examination type was abdominal computed tomography (CT) with intravenous contrast (image series of approximately 80 images each), using either iterative image reconstruction (Setting A) or traditional filtered backprojection (Setting B). Setting A had a 50% reduction in $CTDI_{vol}$ compared with Setting B. The chest and lumbar spine examinations were carried out at Linköping University Hospital, Linköping, and the CT examination at Vrinnevi hospital, Norrköping, Sweden. Table 1 summarises the characteristics of the imaging systems and patient dose indicators relevant for the study.

Observers and cases

Forty radiology residents (observers) participated in the assessment of image quality of one or more of the three types of examinations (Table 1), as part of a national radiological protection training course in Linköping, Sweden. They had undergone between 1 and 4 years of specialist training. Twelve observers evaluated the lumbar spine radiography, 23 evaluated bedside chest radiography and 14 evaluated the abdominal CT examinations, respectively. Some observers evaluated more than one type of examination. No consensus meeting was held prior to the evaluation of the images, and only a single-page instruction was provided about how to use the image grading software.

Twenty patients (10 men, range 44-76 years) that had undergone chest examinations between January 2013 and June 2013, and 20 patients (10 men, range 36-73 years) that had undergone lumbar spine examinations between December 2011 and July 2012, were randomly selected, 10 for each A and B setting. The 10 patients subject to setting A were different from those in setting B. Patients were matched in terms of age and gender but not in terms of body dimensions. For the CT

examinations, 10 patients, examined between December 2010 and September 2011, were selected (5 men, range 45-87 years). These patients were scheduled for abdominal CT examination on two separate occasions (interval 1-310 days), for which the imaging reconstruction technique changed from filtered backprojection (setting B) to iterative reconstruction with the software SAFIRE level 1 (setting A), using the same type of x-ray equipment.

Assessment of image quality

The image assessment was made using image criteria suggested by the European guidelines on quality criteria for radiography [6] and CT [5]. The image criteria are listed in Table 2. For the CT examinations, the criteria are similar to those used by Borgen et al [16]. All images or image series (cases) were independently evaluated by the observers on diagnostic viewing stations. The displays of the viewing stations were calibrated according to AAPM TG18 [17]. No constraints were applied for viewing distance or duration. The images were presented to the observers using DICOM-compatible image viewing software (ViewDEX 2.0, Viewer for Digital Evaluation of X-ray Images) [18]. The ViewDEX program randomised the cases and logged the graded response and the image assessment time of the observers in a separate log file. The graded assessment for all criteria was made according to a five point ordinal scale, where grade 5 is expressed as 'I am sure that the criterion is fulfilled'; 4 'I am almost sure that the criterion is fulfilled'; 3 'I am not sure if the criterion is fulfilled or not'; 2 'I am almost sure that the criterion is not fulfilled'; 1 'I am sure that the criterion is not fulfilled'.

O12-5 Image quality audit using VGC analysis

Data Analysis

Visual grading analysis (VGA) was used for all cases. For each type of examination and for each criterion, the number of cases that scored each of 1 to 5 was computed for setting A and B separately. The software ROCfit [19] was used for visual grading characteristics (VGC) analysis. For each setting A and B, the ROCfit software computed the cumulative fractions of cases with score=5, score $\geq$4, score $\geq$3, score $\geq$2 and score $\geq$1, respectively. The software fitted a curve to the six data points, including the origin. The area under the curve, A_{VGC} was computed, as well as an estimated standard deviation of A_{VGC}, s_{VGC} (see Figure 1). If $A_{VGC} < 0.5$, setting A has a preferred image quality to setting B. If $A_{VGC} > 0.5$, the opposite is true. If $|A_{VGC}-0.5|>2s_{VGC}$ then the difference in perceived image quality was considered significant (p=0.05) (Table 2). If $|A_{VGC}-0.5|>3s_{VGC}$ or if $|A_{VGC}-0.5|>4s_{VGC}$ then the significance level was set to p=0.01 and p=0.001, respectively. If $|A_{VGC}-0.5|<s_{VGC}$ then the difference was considered not significant (n.s.). Inter-observer reliability was assessed for each setting, criterion and type of examination using Gwet's AC_1 coefficient [20]. The AC_1 coefficient is an alternative to the Kappa coefficient, which is similar in formulation but does not suffer from some of the statistical problems of the Kappa coefficient [21]. The AC_1 coefficient reflects the strength of agreement, ranging from poor (0 – 0.20), fair (0.21 – 0.40), moderate (0.41 – 0.60), good (0.61 – 0.80) to very good (0.81 – 1) [22].

RESULTS

Visual grading and inter-observer reliability

A typical VGC-curve for each examination type is shown in Figure 1 (criterion 3, criterion 4 and criterion 2 for the lumbar spine AP, chest AP and abdominal CT examinations, respectively). The results of the visual grading analysis for each criterion for lumbar spine AP are presented in Table 3.

The parameters presented are the area under the VGC-curve, A_{VGC}, the standard deviation s_{VGC}, and the inter-observer reliability (AC_1). Table 4 and 5 show the corresponding results for chest AP and for abdominal CT, respectively.

For all criteria for lumbar spine AP, the quality of the images was graded significantly higher for images acquired using a sensitivity class of 400 ($|A_{VGC}-0.5|>3s_{VGC}$, $p < 0.01$). This difference was particularly seen for the reproduction of the pedicles. Also, the inter-observer reliability (AC_1) was higher for the images acquired at sensitivity class 400 (0.69 ± 0.04) than for those acquired at sensitivity class 560 (0.59 ± 0.06). This indicates that observers agreed more on the quality of these images compared to those acquired at a lower dose level (sensitivity class 560). For chest AP, image quality was graded significantly higher for the DR system (flat panel detector) for all criteria ($|A_{VGC}-0.5|>2s_{VGC}$, $p < 0.05$) and the inter-observer reliability was similar between the two settings (CR: 0.58 ± 0.06 and DR: 0.61 ± 0.08). Finally, there were no differences in perceived image quality between abdominal CT images reconstructed by filtered backprojection and iterative reconstruction ($|A_{VGC}-0.5|<2s_{VGC}$), despite an average 50% reduction in $CTDI_{vol}$ with the iterative reconstruction setting. An exception was the criterion regarding the pancreatic contours (C1), for which the image quality was graded significantly higher for filtered backprojection ($|A_{VGC}-0.5|>2s_{VGC}$, $p < 0.05$). The overall inter-observer reliability for the abdominal CT examination was lower (iterative reconstruction: 0.41 ± 0.10 and filtered backprojection: 0.42 ± 0.08) and the AC_1 coefficients were more variable across criteria than for the other examinations.

Table 6 shows the proportion of observers that grade the images as either 4 or 5, i.e. are sure or almost sure that the criterion is fulfilled for each examination and imaging system configuration. The ratios are typically larger for the lumbar spine system with lower sensitivity (400 instead of 560) and the chest DR system (instead of CR). The ratios were similar for the abdominal CT examinations reconstructed by iterative reconstruction and by filtered backprojection, with the exception of the

criterion regarding the pancreatic contours (C1), for which a larger proportion of observers graded the filtered backprojection-reconstructed images as 4 or 5.

Administration and image assessment time

The time needed to extract the patient images from the image archive was approximately one day per examination type including finding appropriate cases and anonymising the images. Setting up the image viewing software took an additional two days for an experienced study coordinator. This included deciding on the image criteria with an experienced radiologist. The time required to analyse the data as shown in Table 3-6, was one working day.

The median (95-percentile) observation time was 28 (106) seconds per image for lumbar spine examinations, 41 (189) seconds per image for the chest examinations and 84 (276) seconds per image series for CT examinations. On average the observers spent between 18-38 minutes on their tasks, assessing the 20 images or image series. In total, the 40 observers spent approximately 25 hours or 3 full working days on assessing the images.

DISCUSSION

For a clinical audit to be successful, it needs to be carried out reasonably fast and cost-effective in a clinical environment. The main finding of this study was that by using visual grading techniques and image criteria, a clinical audit of image quality can be performed effectively and in a reasonably short time. In our experience, one of the challenges with image quality audits is to recruit observers, as many radiologists may consider the image grading tasks difficult to combine with their clinical workload. However, by including the image grading tasks as part of a radiologist training course, not

only were we able to introduce the concepts of auditing and visual grading techniques to the course participants, but at the same time it was fairly easy to recruit a large number of observers. In a radiology department, a similar number of observers would be able to do the audit in a distributed fashion over a period of a few weeks, i.e. not requiring any particular individual scheduling.

By using visual grading techniques, audits can be made of almost any radiological examination where a number of image criteria can be formulated. In a previous study, Gonzalez et al. [23] performed an image quality audit at a large number of private Spanish hospitals using EU image criteria [6]. They allowed both external and internal radiologists to assess four types of x-ray procedures (chest, lumbar spine, abdomen and mammography). Similar to the findings of the current study, a larger fraction of the lumbar spine images were evaluated as being of good or very good quality, compared to the chest images, which were judged as fair or defective. Fabiszewska et al. compared 248 mammographic screening facilities regarding physical image quality, clinical image quality and breast absorbed dose [24]. Given the large number of sites, however, only two mammograms per site were evaluated by three experts using the EU guidelines. This underlines the challenge of recruiting observers for clinical audits, in particular when large numbers of images or images from different imaging facilities are to be evaluated. Martin et al. reported on the optimisation of paediatric chest examinations and found visual grading analysis to be very sensitive to subtle differences in perceived image quality [25].

Our results partially contrast with the results of a study by Hardie et al., who studied the preferred strength of the sinogram-affirmed iterative reconstruction algorithm (SAFIRE) setting for abdominal CT examinations [26]. They found that images reconstructed using the lowest SAFIRE level, i.e. the level used in our study, were preferred over images reconstructed using filtered backprojection. However, the perceived image quality in our study was the same for both filtered backprojection and the lowest SAFIRE level. This difference in result may be explained by the difference in criteria. In

the study by Hardie et al., the observers were asked to rank the images based on overall appearance required for clinical interpretation, whereas in this study, the image criteria were related to specific anatomical structures and, thereby, diagnostic tasks. Interestingly, only a fair agreement between observers for abdominal CT images was found, which suggests that the preference for the reconstruction method (filtered backprojection or SAFIRE) varied substantially between observers. The least agreement was found for the criterion related to the perceived image noise. The experience with iterative reconstruction methods varied among the observers, and this may have influenced the results of the study. The inter-observer reliability, as assessed using Gwet's AC_1 measure, seems to provide a useful measure to further assess and explain the results of the visual grading analysis.

It should be noted that while patient absorbed doses were substantially reduced in both lumbar spine (setting B) and abdominal CT (setting A), image quality was only maintained in CT using iterative reconstruction. As expected, it was not maintained in projection radiography, where the effect of increased quantum noise associated with increasing sensitivity class (400 to 560) resulted in a significant reduction of the proportion of observers that graded the images as either score 4 or 5 (Table 6). For bedside chest examinations with the older CR-system (setting A), the proportion of observers that graded the images as either score 4 or 5 was low (<0.50), but significantly higher with the modern DR-system (setting B) in particular regarding the visibility of the retrocardiac vessels and the hilar region.

There was some notable correlation between the proportion of observers that graded the images as either score 4 or 5 and the inter-observer reliability AC_1 ($r^2=0.47$). A higher perceived image quality resulted in higher AC_1 (i.e. more observer agreement). It is likely that in situations with perceived poor image quality, assessment is more challenging which might lead to an increase in observers' score variation.

This study is not without its limitations. As the observers in this study were based in different parts of Sweden, there was no practical opportunity to have a consensus meeting on how to interpret the image criteria on the ordinal grading scale. However, the distributed reading of the images in this study was successfully completed using only a single A4 sheet of written instructions, emphasising the practicality of the visual grading method for use in clinical audit. All observers were radiology residents with varying but typically less than 5 years of clinical experience, which may have affected the outcome of the VGC analysis. However, the level of observer experience has previously been shown to have very little impact on the images assessment outcome in the case of CT images reconstructed by iterative methods [26]. Intra-observer variability has not been studied. The patients that underwent chest and lumbar spine examinations were not matched in terms of body dimensions. Further studies could focus on investigating the effect of observer experience on the results in VGC analysis.

CONCLUSION

Visual grading of image criteria to assess the quality of images in radiology can serve as an efficient method for performing an image quality audit in a clinical environment. The possibility of using the results as a baseline for future audits needs to be investigated.

ACKNOWLEDGEMENT

O12-5 Image quality audit using VGC analysis

REFERENCES

1. Sandborg, M., Båth, M., Järvinen, H., Falkner, K. Justification and optimization in clinical practice. In: Diagnostic Radiology Physics - A handbook to teachers and students. Dance D R et al. (editors), IAEA, Vienna (2014). ISBN 978-92-131010-1.

2. Swedish Radiation Safety Authority. Regulations on general obligations in medical and dental practices using ionising radiation. SSMFS 2008:35 (2008)

3. International Commission on Radiological Protection. Managing patient dose in digital radiology. A report of the international commission on radiological protection. Ann ICRP 34, 1-73 (2004).

4. Metz, C.E. ROC methodology in radiologic imaging. Invest. Radiol. 21:720-733 (1986).

5. Menzel, H., Shibilla, H., Teunen, D. European guidelines on quality criteria for computed tomography. EUR 16262 EN (2000).

6. Carmichael, J.H.E., Maccia, C., Moores, B.M. European guidelines on quality criteria for diagnostic radiographic images EUR 16260 EN (1996).

7. Båth, M., Månsson, L.G. Visual grading characteristics (VGC) analysis: A non-parametric rank-invariant statistical method for image quality evaluation. Br J Radiol 80, 169-176 (2007).

8. Smedby, O., Fredriksson, M. Visual grading regression: Analysing data from visual grading experiments with regression models. Br. J. Radiol. 83, 767-775 (2010).

9. Tingberg, A., Herrmann, C., Lanhede, B., Almén, A., Besjakov, J., Mattsson., S., Sund, P., Kheddache, S. and Månsson, L.G.. Comparison of two methods for evaluation of the image quality of lumbar spine radiographs. Radiat. Prot. Dosim. 90, 165-168 (2000).

10. Pascoal, A., Lawinski, C.P., Honey, I., Blake, P. Evaluation of a software package for automated quality assessment of contrast detail images-comparison with subjective visual assessment. Phys. Med. Biol. 50, 5743-5757 (2005).

11. Tapiovaara, M.J., Sandborg, M. How should low-contrast detail detectability be measured in fluoroscopy? Med. Phys. 31, 2564-2576 (2004).

12. De Crop, A., Bacher, K., Van Hoof, T., Smeets, PV., Smet, B.S., Vergauwen, M., Kiendys, U., Duyck, P., Verstraete, K., D'Herde, K. et al. Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study. Radiology 262, 298-303 (2012).

13. Sandborg, M., McVey, G., Dance, D.R., Alm Carlsson, G. Comparison of model predictions of image quality with results of clinical trials in chest and lumbar spine screen-film imaging. Radiat. Prot. Dosim. 90, 173-176 (2000).

14. Sandborg, M., Tingberg, A., Dance, D.R., Lanhede, B., Almén, A., McVey, G., Sund, P., Kheddache, S., Besjakov, J., Mattsson, S. et al. Demonstration of correlations between clinical and physical image quality measures in chest and lumbar spine screen-film radiography. Br. J. Radiol. 74, 520-528 (2001).

15. Sandborg, M., Tingberg, A., Ullman, G., Dance, D.R., Alm Carlsson, G. Comparison of clinical and physical measures of image quality in chest and pelvis computed radiography at different tube voltages. Med. Phys. 33, 4169-4175 (2006).

16. Borgen, L., Kalra, M.K., Laerum, F., Hachette, I.W., Fredriksson, C.H., Sandborg, M., Smedby, O. Application of adaptive non-linear 2D and 3D postprocessing filters for reduced dose abdominal CT. Acta Radiol. 53, 335-342 (2012).

17. Samei, E., Badano, A., Chakraborty, D., Compton, K., Cornelius, C., Corrigan, K., Flynn, M.J., Hemminger, B., Hangiandreou, N., Johnson, J., et al. Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report. Med. Phys. 32, 1205-1225 (2005).

18. Hakansson, M., Svensson, S., Zachrisson, S., Svalkvist, A., Båth, M., Månsson, L.G. VIEWDEX: An efficient and easy-to-use software for observer performance studies. Radiat. Prot. Dosim. 139, 42-51 (2010).

19. Eng, J. ROC analysis: Web-based calculator for ROC curves. Available via http://www.jrocfit.org. Accessed 26 Aug 2014 (2014).

20. Gwet, K.L. Computing inter-rater reliability and its variance in the presence of high agreement. Br. J. Math. Stat. Psychol. 61, 29-48 (2008).

21. Wongpakaran, N., Wongpakaran, T., Wedding, D., Gwet, K.L. A comparison of cohen's kappa and gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. BMC Med. Res. Methodol. 13, 61-2288-13-61 (2013).

22. Altman, D. Practical statistics for medical research. Chapman and Hall, London (1991).

23. Gonzalez, L., Vano, E., Oliete, S., Manrique, J., Hernáez, J.M., Lahuerta, J., Ruiz, J. Report of an image quality and dose audit according to directive 97/43/Euratom at Spanish private radiodiagnostics facilities. Br. J. Radiol. 72, 186-192 (1999).

24. Fabiszewska, E., Grabska, I., Jankowska, K., Wesolowska, E., Bulski, W. Comparison of results from quality control of physical parameters and results from clinical evaluation of mammographic images for the mammography screening facilities in poland. Radiat. Prot. Dosim. 147, 206-209 (2011).

25. Martin, L., Ruddlesden, R., Makepeace, C., Robinson, L., Mistry, T., Starritt, H. Paediatric x-ray radiation dose reduction and image quality analysis. J. Radiol. Prot. 33:621-633 (2013).

26. Hardie, A.D., Nelson, R.M., Egbert, R., Rieter, W.J., Tipnis, S.V. What is the preferred strength setting of the sinogram-affirmed iterative reconstruction algorithm in abdominal CT imaging? Radiol. Phys. Technol. 8, 60-63 (2015).

O12-5 Image quality audit using VGC analysis

LEGENDS TO TABLES

Table 1. Characteristics of the imaging systems for settings A and B. Abreviations: AP: anteroposterior, CT: computed tomography, SAFIRE: Sinogram Affirmed Iterative Reconstruction. P_{KA} : air kerma area product.

Table 2. Image criteria for the three examinations (lumbar spine AP, bedside chest AP and abdominal CT with intravenous contrast).

Table 3. Area under the VGC-curve, A_{VGC}, its standard deviation s_{VGC}, and the inter-observer reliability (AC_1) for each criterion for lumbar spine AP. Setting A represents sensitivity class 400, Setting B sensitivity class 560.

Table 4. Area under the VGC-curve, A_{VGC}, its standard deviation s_{VGC}, and the inter-observer reliability (AC_1) for each criterion for bedside chest AP. Setting A represents a CR system, Setting B a DR system with a flat panel detector.

Table 5. Area under the VGC-curve, A_{VGC}, its standard deviation s_{VGC}, and the inter-observer reliability (AC_1) for each criterion for abdominal CT. With Setting A, post-processing was done using iterative reconstruction. With Setting B, filtered backprojection was used.

Table 6. Proportion of observers that graded the images as either 'I am sure that the criterion is fulfilled' (5) or 'I am almost sure that the criterion is fulfilled' (4) for each criterion, examination and imaging system configuration.

O12-5 Image quality audit using VGC analysis

TABLES

Table 1.

Examination	Setting A	Setting B
Lumbar spine AP	Siemens Aristos FX, 81 kV, 0.1mm Cu, sensitivity class 400, mean P_{KA} = 0.64 Gycm^2	Siemens Aristos FX, 81 kV, 0.1mm Cu, sensitivity class 560, mean P_{KA} = 0.45 Gycm^2
Chest bedside AP	Siemens Mobilett 2 with Philips CR ST(V) image plates, P_{KA} = 0.10-0.15 Gycm^2 †	Mediel Movix Dream with Canon DR system, mean P_{KA} =0.15 Gycm^2
Abdominal CT with intravenous contrast	Siemens Somatom Definition AS Plus, Iterative reconstruction, SAFIRE level 1, 5 mm slice thickness, mean CTDI$_{vol}$ = 5,8 mGy	Siemens Somatom Definition AS Plus, Filtered back projection, 5 mm slice thickness, mean CTDI$_{vol}$ = 12,1 mGy

†: Since no patient dosemeter was mounted on the imaging system in clinical use, the P_{KA}-value was estimated by exposing an anthropomorphic chest phantom (Multipurpose Chest phantom N1 "Lungman", Kyoto Kagaku Co, Japan) using the exposure parameters used clinically (133 kV, 1.25 mAs) and by measuring P_{KA} with an external kerma area product meter.

Table 2.

Image criteria for Lumbar spine AP	
C1	Visually sharp reproduction of the upper and lower-plate surfaces
C2	Visually sharp reproduction of the pedicles
C3	Visually sharp reproduction of the cortex and trabecular structures
C4	Image noise does not interfere with my clinical assessment
C5	Overall image quality is sufficient for diagnosis

Image criteria for Chest AP	
C1	Vessels seen 3 cm from the pleural margin are sharply visualized
C2	Retrocardiac vessels are visualized
C3	Hilar region is sharply visualized
C4	Carina with main bronchi is visualized
C5	Overall image quality is sufficient for diagnosis

Image criteria for Abdominal CT with intravenous contrast	
C1	Visually sharp reproduction of the pancreatic contours
C2	Visually sharp reproduction of the kidneys and proximal ureters
C3	Visually sharp reproduction of the portal and liver veins
C4	Visually sharp reproduction of the gallbladder wall
C5	Visually sharp reproduction of the extrahepatic biliary ducts
C6	Image noise does not interfere with my clinical assessment

Table 3.

	$A_{VGC} \pm s_{VGC}$	p	Inter-observer reliability (AC_1)	
			Setting A	Setting B
C1 (upper and lower-plate surfaces)	0.37 ± 0.04	<0.01	0.73 ± 0.08	0.62 ± 0.05
C2 (pedicles)	0.32 ± 0.04	<0.001	0.70 ± 0.06	0.60 ± 0.08
C3 (cortex and trabecular structures)	0.37 ± 0.04	<0.01	0.67 ± 0.06	0.64 ± 0.06
C4 (image noise)	0.36 ± 0.04	<0.01	0.64 ± 0.07	0.50 ± 0.08
C5 (overall image quality)	0.35 ± 0.04	<0.01	0.72 ± 0.06	0.57 ± 0.08
All criteria	**0.36 ± 0.02**	**<0.001**	**0.69 ± 0.04**	**0.59 ± 0.06**

Table 4.

	$A_{VGC} \pm s_{VGC}$	p	Inter-observer reliability (AC_1)	
			Setting A	Setting B
C1 (vessels 3 cm from pleural margin)	0.62 ± 0.03	<0.001	0.52 ± 0.04	0.58 ± 0.03
C2 (retrocardiac vessels)	0.56 ± 0.03	<0.05	0.54 ± 0.04	0.50 ± 0.06
C3 (hilar region)	0.65 ± 0.03	<0.001	0.55 ± 0.05	0.62 ± 0.02
C4 (carina with main bronchi)	0.59 ± 0.03	<0.01	0.61 ± 0.06	0.66 ± 0.06
C5 (overall image quality)	0.62 ± 0.03	<0.001	0.67 ± 0.04	0.70 ± 0.04
All criteria	**0.60 ± 0.01**	**<0.001**	**0.58 ± 0.06**	**0.61 ± 0.08**

Table 5.

	$A_{VGC} \pm s_{VGC}$	p	Inter-observer reliability (AC_1)	
			Setting A	Setting B
C1 (pancreatic contours)	0.57 ± 0.04	<0.05	0.51 ± 0.11	0.68 ± 0.09
C2 (kidneys and proximal ureters)	0.48 ± 0.03	n.s.	0.37 ± 0.06	0.43 ± 0.09
C3 (portal and liver veins)	0.48 ± 0.04	n.s.	0.65 ± 0.08	0.55 ± 0.10
C4 (gallbladder wall)	0.50 ± 0.04	n.s.	0.39 ± 0.10	0.38 ± 0.08
C5 (extrahepatic biliary ducts)	0.52 ± 0.04	n.s.	0.52 ± 0.09	0.41 ± 0.08
C6 (image noise)	0.52 ± 0.04	n.s.	0.32 ± 0.07	0.46 ± 0.07
All criteria	**0.51 ± 0.01**	**n.s.**	**0.41 ± 0.10**	**0.42 ± 0.08**

Table 6.

	Image quality criterion					
	C1	C2	C3	C4	C5	C6
Lumbar Spine AP						
sensitivity class 400	0.85	0.81	0.78	0.79	0.83	-
sensitivity class 560	0.73	0.59	0.63	0.67	0.67	-
Chest AP						
CR	0.51	0.41	0.37	0.59	0.52	-
DR	0.68	0.51	0.58	0.71	0.71	-
Abdominal CT						
iterative reconstruction	0.57	0.57	0.65	0.47	0.40	0.58
filtered backprojection	0.71	0.51	0.59	0.43	0.33	0.59

O12-5 Image quality audit using VGC analysis

LEGENDS TO FIGURES

Figure 1. Cumulative image criteria score (●) for setting B as function of cumulative image criteria score for setting A. Left panel: lumbar spine AP (C3: cortex and trabecular structures); Middle panel: bedside chest AP (C4: carina with main bronchi); Right panel: Abdominal CT (C2: kidneys and proximal ureters). The solid line represents the fitted data and the dashed lines ± 2 standard deviations of the fitted data. The error bars represent the 95% confidence interval for the four data points. The dotted gray line (….) indicates the diagonal with equal performance.

FIGURES

Figure 1