Radiography 21 (2015) e1–e4

Contents lists available at ScienceDirect

Radiography

journal homepage: www.elsevier.com/locate/radi

Post-processing, is it a burden or a blessing? Part 1 evaluation of clinical image quality

Robin Decoster [a,*], Harrie Mol [b], Dirk Smits [a,b]

[a] *Hogeschool-Universiteit Brussel, Brussels, Belgium*
[b] *KU Leuven, Leuven, Belgium*

ARTICLE INFO

Article history:
Received 1 November 2013
Received in revised form
21 May 2014
Accepted 3 June 2014
Available online 21 June 2014

Keywords:
Radiography
Digital detectors
Post-processing
Image perception
VGA

ABSTRACT

Background: The dynamic range of modern detectors tolerates a higher detector dose or Detector Air Kerma (DAK) without negative effect on image quality. Necessary image quality is closely related to the clinical question: in order to keep the patient dose as low as possible, image quality criteria are needed for each type of radiography. The purpose of this study was to investigate the relation between image quality, described as an overall judgement versus the visibility of well-defined structures, and DAK in clinically accepted radiographs.

Methods and materials: 168 AP radiographs of the knee and 152 radiographs of the pelvis were collected randomly in 19 radiologic centres. Six radiologists with at least five-years experience scored the overall image quality and the visibility of seven different anatomic structures, in a controlled environment on a five-point scale, using a Visual Grading Analysis (VGA). The relation between the DAK and the VGA Score (VGAS) was evaluated.

Results: The VGAS was 3.92 for the knee and 3.71 for the pelvis. The VGAS for CR and DR were significantly different ($p < 0.01$). Intra-observer variability was not significant and inter-observer correlations were high and significant. Only for the pelvis radiographs produced with computed radiography, a rather weak but significant correlation was found between DAK and VGAS.

Conclusion: The VGA revealed an image quality higher than diagnostically necessary in both datasets, and high inter-observer correlation. Based on the DAK-range, it could be hypothesized that below a certain noise level no further visible improvement of the image quality was reached.

Introduction

The transition to digital detectors was a major innovation for medical imaging departments.[1,10–14] It changed existing procedures, due to the larger dynamic range, offering the possibility to 'tailor' the patient dose and image quality for a specific task.[1,10,15] This paradigm shift demands for new guidelines and it illustrates the current challenge for practitioners to regain control over the relation between dose and image quality.[1,8]

The large dynamic range of the modern detectors tolerates a wider range in detector dose, here defined as detector air kerma (DAK), without observing an adverse effect on the image quality. Higher DAKs deliver a crisp and sharp appearance due to a lower noise level, which is favoured by the radiologist.[1,2] The preference of the radiologist for such images can lead to higher DAK and also a higher image quality than needed to answer the diagnostic question. Moreover, the necessary image quality is closely related to the clinical question, as not all imaging tasks demand the same level of diagnostic information and detail. Both aspects interfere in the relation between the diagnostic task ahead and possible dose reductions.[1,5–7]

First, image quality is not easy to define. A single, straightforward definition is lacking. As previously mentioned, it even depends on the specific diagnostic task ahead and on a personal preference of the radiologist. If clear image quality criteria are lacking for specific examinations, it is plausible that different observers will rely on different aspects of an image when evaluating the overall image quality and can use different weighting procedures in order to combine these aspects to one general judgement on image quality. Such an approach, we assume, will result in at most a moderate inter-observer reliability when evaluating the image quality of a fixed set of radiographs, due to interpersonal differences between observers. A second approach to image quality

* Corresponding author.
E-mail address: robin.decoster@hubkaho.be (R. Decoster).

http://dx.doi.org/10.1016/j.radi.2014.06.003
1078-8174/

consists of a predefined set of criteria in terms of the visibility of certain anatomical structures or the degree of detectability of those structures, which will be used as a measure of image quality. A systematic evaluation of image quality based on, what will be called Minimal Detectable Structures (MDS) to diagnose, should deliver a higher inter-observer agreement resulting in a high correlation between observers.[26]

Using MDS resembles the methodology of a Visual Grading Analysis (VGA) but also the more general approach of image evaluation during the installation of radiological equipment. It can be defended as the visualization of normal anatomy correlates with the detection of pathological structures and agrees with ROC based methods.[3,4] If well-defined clinical image quality criteria are present for a given type of examination, the radiographer can use their detectability in combination with an as low as possible patient dose as a proxy for good image quality.[3,4,6,8,9] A good image combines both aspects: the image comprises all necessary information to answer the clinical question and is produced with an as low as possible patient-dose. Reliable image quality analysis, which can be used in the clinical routine, is an important first step in the determination of the necessary dose and in regaining control over the dose – image quality relationship.

The purpose of this paper is therefore twofold[1]: to investigate the relation between 2 measures of image quality; overall perceived image quality and image quality based on MDS, and[2] to investigate the relation between image quality and DAK based on clinically accepted radiographs.

Method & materials

To assess image quality in orthopaedic radiography two incidences were selected: the anterior–posterior (AP) radiograph of the knee and pelvis. Radiographs of the pelvis with lead protection were excluded due to potential superposing with structures up for evaluation. Radiographs of the pelvis with the presence of two hip prostheses and radiographs of the knee AP with prosthetic material were excluded, because of the potential destruction of MDS during placement. In total 320 radiographs, 168 for the knee AP and 152 for the pelvis, were randomly collected by including the first 20 patients who met the criteria in each 17 radiology centres across Flanders. Next to the generator settings, the age and BMI were collected. The radiographs were anonymized and stored on optical carriers, such as CD or DVD. Additional information on positioning, patient's morphology and technique were collected anonymously. Subgroups were defined based on detector technology: CR and DR, including indirect and direct digital radiography.

First, to measure image quality, 2 measures were used[1]: the general appreciation of the radiography by the radiologist on a scale ranging from 1 to 5; and[2] an absolute Visual Grading Analysis (VGA) to evaluate and quantify the visibility of different anatomical structures, the MDS.[4,16,17] Those anatomical structures to evaluate the pelvis AP radiograph were collected from the guidelines published by the European Commission and from radiography textbooks.[3,4,18–20] As these guidelines did not include any criteria for the knee AP, criteria were collected from different publications with a similar purpose.[18–22] Moreover, five radiologists reviewed the 2 sets of criteria during structured interviews in order to verify the completeness of the list (Table 1). None of these radiologists were involved any further in the assessment of the radiographs or other aspects of this study.

Six radiologists, with at least five year of experience, who did not participate in reviewing the list of criteria but agreed with that list, assessed the radiographs. All assessments were carried out with ViewDex® (Version 2.0) on a five-point scale (from 1 'bad' to 5 'excellent'). The mid-point was equalized to diagnostic image quality i.e. the image quality that would be expected routinely when imaging cooperative patients.[4,5,16,17] The first five questions focused on the assessment of an anatomical structure and the last question inquired the general impression of the radiography.

All observations took place in a controlled environment on a standard workstation (Windows7-64bit) equipped with a Barco's® Coronis display (6 MP) and Barco® QAWeb for quality control. The display operated within the boundaries of the AAPM TG18. Options such as window/level, pan and zoom were available.[1,23] Prior to the VGA, each observer received a training dataset of twenty radiographs. Results from the training dataset were discarded. Additionally, twenty radiographs were repeated during the VGA to determine intra-observer variability.

For each image, a VGA score (VGAS) was calculated using the following equation:

$$\text{VGAS} = \frac{\sum_{i=1}^{I}\sum_{s=1}^{S}\sum_{o=1}^{O} G_{i,s,o}}{I \times S \times O}$$

where $G_{i,s,o}$ is the grading for image i, structure s and observer o. De denominator is formed by I, the total number of images, S for the number of evaluated structures and O the number of observers in the study.[24] This numeric expression, the Visual Grading Analysis Score (VGAS), defines the mean score over all observations.[4] The VGAS will serve as an indicator for the image quality as appreciated by the radiologists.

Statistical comparisons of the VGAS were performed with a one-way ANOVA followed by post hoc tests with Bonferroni correction using SPSS 19.0. A p-value <0.05 denoted a significant difference between two data points. To analyse intra-observer variability a

Table 1
Criteria for the evaluation of the radiographs (MDS).

Knee AP	Pelvis AP
1. Visualization of the patella	1. Visualization of sacroiliac joints
2. Visually sharp reproduction of the tibial plateau	2. Visualization of the middle third of the iliac crest
3. Visually sharp reproduction of the intercondylar eminence and fossa	3. Visualization of the pubic and ischial rami
4. Visually sharp reproduction of the femoral condyles	4. Visually sharp reproduction of the femoral necks
5. Visually sharp reproduction of the fibula head	5. Visually sharp reproduction of the trochanters
6. Transition of cortical to trabecular bone	6. Visually sharp reproduction of the Shenton's line

Table 2
Results for the knee AP.

	N	Sex		BMI	Tube voltage [kV]	Tube load [mAs]	FRD [cm]	DAK [μGy]	VGAS	General appreciation
		Male	Female							
CR	96	48	48	25.65 ± 4.8	69.63 ± 4.8	30.39 ± 62	103.44 ± 5.95	4.08 ± 1.47	3.82 ± 0.38	3.84 ± 0.54
DR	72	38	34	25.73 ± 3.45	64.28 ± 2.58	8.82 ± 3.97	115.16 ± 4.5	2.99 ± 2.14	4.07 ± 0.32	3.96 ± 0.51
Total	168	86	82	25.68 ± 4.23	67.33 ± 4.79	20.56 ± 47	108.46 ± 7.92	3.63 ± 1.85	3.91 ± 0.39	3.9 ± 0.53