The British Journal of Radiology, 74 (2001), 520–528

Demonstration of correlations between clinical and physical image quality measures in chest and lumbar spine screen–film radiography

[1]M SANDBORG, PhD, [2]A TINGBERG, PhD, [3]D R DANCE, PhD, [4]B LANHEDE, MSc, [2]A ALMÉN, PhD, [3]G MCVEY, BSc, MSc, [4]P SUND, MSc, [5]S KHEDDACHE, PhD, MD, [6]J BESJAKOV, MD, PhD, [2]S MATTSSON, PhD, [4]L G MÅNSSON, PhD and [1]G ALM CARLSSON, PhD

[1]*Department of Radiation Physics, Faculty of Health Sciences, Linköping University, SE 581 85 Linköping, Sweden, Departments of* [2]*Radiation Physics and* [6]*Diagnostic Radiology at Malmö, Lund University, Malmö University Hospital, SE 205 02 Malmö, Sweden,* [3]*Joint Department of Physics, The Royal Marsden NHS Trust, Fulham Road, London SW3 6JJ, UK, and Departments of* [4]*Radiation Physics and* [5]*Radiology, Göteborg University, Sahlgrenska University Hospital, SE 413 45 Göteborg, Sweden*

Abstract. The ability to predict clinical image quality from physical measures is useful for optimization in diagnostic radiology. In this work, clinical and physical assessments of image quality are compared and correlations between the two are derived. Clinical assessment has been made by a group of expert radiologists who evaluated fulfilment of the European image criteria for chest and lumbar spine radiography using two scoring methods: image criteria score (ICS) and visual grading analysis score (VGAS). Physical image quality measures were calculated using a Monte Carlo simulation model of the complete imaging system. This model includes a voxelized male anatomy and was used to calculate contrast and signal-to-noise ratio of various important anatomical details and measures of dynamic range. Correlations between the physical image quality measures on the one hand and the ICS and VGAS on the other were sought. 16 chest and 4 lumbar spine imaging system configurations were compared in frontal projection. A statistically significant correlation with clinical image quality was found in chest posteroanterior radiography for the contrast of blood vessels in the retrocardiac area and a measure of useful dynamic range. In lumbar spine anteroposterior radiography, a similar significant correlation with clinical image quality was found between the contrast and signal-to-noise ratio of the trabecular structures in the L1–L5 vertebrae. The significant correlation shows that clinical image quality can, at least in some cases, be predicted from appropriate measures of physical image quality.

Both the International Commission on Radiological Protection [1] and the European medical exposure directive [2] recommend the optimization of image quality and patient dose in diagnostic radiology. Such optimization of the X-ray examination involves balancing clinical image quality against patient dose. Methods to measure individual patient absorbed doses are available [3], whereas methods to assess the quality of individual radiographic images are still under development. Methods to assess the quality of radiographic images often focus on the physical/technical aspects of the image [4], but methods that also include the radiologist in the assessment are required for a more realistic and complete treatment of the problem. Technical, physical, physiological and psychological elements are all involved in the transfer and interpretation of information by the radiologist. Consequently, the correlation between physical parameters of the imaging system and the relevant diagnostic information in the image is difficult to establish.

The image quality required will vary with the radiological task. For a selected number of routine radiographic projections, the European Commission has proposed sets of image criteria [5] that may be used for clinical image quality assessment. The image criteria are expressed as the

Received 30 June 2000 and in revised form 18 December 2000, accepted 24 January 2001.

This work has been supported by grants from the Commission of the European Communities (FI4P CT950005), the Swedish Radiation Protection Institute, SSI (P1892.95, P1018.97, P1083.98, P1158.99), the Swedish Medical Research Council, MFR (K98-17X-12652-01A) and Swedish Foundation for Strategic Research (R98:006).

visibility of characteristic features of imaged anatomical structures and are based on the normal anatomy. They apply to adult patients of standard size for the type of examination being considered. An underlying assumption and philosophy of these criteria is that if the normal anatomy is faithfully reproduced in the image, then the pathological lesions will also be visualized.

When the optimization of radiographic imaging systems is based on a study of physical parameters, it is important that the correlation between these physical parameters and clinical measures of image quality be established. However, in previous work on chest radiography using anthropomorphic test phantoms [6], the ability to predict clinical image quality based on physical parameters has been questioned. The authors studied 24 chest imaging systems and found no correlation between image quality assessed in a visual grading analysis study and system parameters such as the relative amount of scattered radiation in the image plane, beam quality (tube potential), sensitivity of the image receptor (speed class) and focal spot size. They did not evaluate the optical density nor the dynamic range of the image. They considered only single parameters at a time and not the combined effect of the parameters on the overall contrast and signal-to-noise ratio (SNR) of important details. However, a positive correlation was found between the number of low contrast details detected in the image of a contrast detail phantom and the best ranked systems. This is to be expected, since the detectability of small, low contrast details depends on how contrast, sharpness and noise combine to yield the SNR.

The inability to correlate individual system parameters with measures of image quality may be related to the multivariate nature of the problem and the difficulty of obtaining a controlled experimental situation when measurements are made with systems in several centres. To demonstrate a correlation, it is essential to look at the effects of system parameters in combination, to use appropriate physical measures of image quality and to obtain patient images in a controlled way, preferably at the same centre. The objective of the present work, therefore, was to search for correlations between physical image quality measures and the corresponding assessments of image quality of patient radiographs by expert radiologists.

The work has been performed as part of a European study of image quality in chest and lumbar spine imaging. It brings together separate work on the assessment of clinical image quality and the development of computer simulation models. 16 imaging alternatives for a postero-anterior (PA) chest examination and 4 imaging alternatives for an anteroposterior (AP) lumbar spine examination were considered. Assessment of patient images was based on the European image criteria [5] and the results for chest and lumbar spine imaging systems are reported elsewhere [7, 8]. The Monte Carlo computer simulation model [9] incorporated a voxel phantom to simulate the patient, with superimposed anatomical details for the calculation of contrast and SNR. The patient absorbed doses are needed for optimization and can be found in previous work [7–10].

Materials and methods

Imaging systems and patients

The chest imaging systems used in this study were obtained by varying the tube potential, screen–film system, antiscatter device and maximum optical density on the film. Parameters used were: tube potential (102 kV, 141 kV); screen–film system speed (Kodak Lanex 160, Lanex 320 screens); antiscatter device (grid, air gap); and maximum optical density (OD_{max}) on the film (1.3, 1.8), so that 16 chest imaging systems were formed. For lumbar spine imaging, two tube potentials (70 kV, 90 kV) and two screen–film system speeds (Kodak Lanex Regular Plus, Lanex Fast) were used to define four different imaging systems. Details of the systems are given in Table 1.

The chest examination was performed on volunteers (for which ethical approval was obtained), whereas the lumbar spine examination was performed on patients. The average height and mass of the two groups were 175.2 cm and 69.3 kg and 170.6 cm and 70.8 kg for the chest and lumbar spine subjects, respectively [10].

Measures of clinical image quality

The image criteria used to assess the quality of the patient images are listed in Table 2. Clinical trials were performed and images obtained with the different imaging techniques were assessed by seven European radiologists. In the analysis of lumbar spine radiographs, the seven original image criteria [5] were used. For chest radiography, the original image criteria were modified prior to the clinical trial [7]. Criteria devoted primarily to positioning of the patient were omitted as fulfilment of these criteria is likely to depend on the skill and training of the radiographer and not on the imaging system itself. In the revised criteria, the parenchyma, mediastinum and costopleural junction were separated and details to be visualized for each region were given ($C5_{CH}$–$C7_{CH}$). The criteria $C1_{CH}$–$C4_{CH}$ are the same as in the original criteria.

Table 1. Imaging system parameters used in the evaluation

System parameter	Chest	Lumbar spine
X-ray tube and generator		
Generator	Siemens Polydoros 50S	Medira 150/60
Waveform	High frequency	High frequency
X-ray tube	W-anode 15°, 0.6 mm focus	W-anode, 12°, 0.6 mm focus
Automatic exposure control	Yes	No
Tube potential (kV) and filtration (mm Al)	102 kV, 3.7 mmAl; 141 kV, 5.7 mmAl	70 kV, 3.5 mmAl; 90 kV, 3.5 mmAl
Bucky and scatter rejection		
Stand	Siemens Vertix E	CGR Penta X
Focus-to-film distance	150 cm with grid 390 cm with air gap	146 cm with grid —
Chest support plate/couch top	8 mm wood	6 mm carbon fibre
Grid ratio	12	10
Grid strip frequency (mm^{-1})	40	60
Lead strip width (μm)	40	36
Interspace and cover material	0.21 mm Al and 0.2 mm Al	0.155 mm carbon fibre and 0.5 mm Al
Cassette front	Kodak X-Omatic LW, 1.8 mm Mg	Kodak X-Omatic LW, 1.8 mm Mg
Air gap	29 cm (3 cm with grid)	12 cm
Image receptor		
Screen type	Kodak Lanex 160	Kodak Lanex Regular Plus
Speed class	160	400
Screen thickness and material	96 mg cm^{-2} Gd_2O_2S	133 mg cm^{-2} Gd_2O_2S
Type of screen	Kodak Lanex 320	Kodak Lanex Fast
Speed class	320	600
Screen thickness and material	126 mg cm^{-2} Gd_2O_2S	182 mg cm^{-2} Gd_2O_2S
Type of film	Kodak TmatL	Kodak TmatL
Film optical density (OD)	1.3 or 1.8 OD_{max}	1.02–1.36 OD_{med}

OD_{max}, maximum optical density; OD_{med}, median optical density.

Table 2. Image criteria used in the analysis of the imaging systems

Criteria	Description of the criteria
Chest [7]	
$C1_{CH}$	Reproduction[a] of the vascular pattern in the whole lung, particularly the peripheral vessels
$C2_{CH}$	Visually sharp reproduction[b] of the trachea and proximal bronchi, the boarders of the heart and aorta
$C3_{CH}$	Visually sharp reproduction[b] of the diaphragm and costophrenic angles
$C4_{CH}$	Visualization[c] of the retrocardiac lung and mediastinum
	Details to be sharply visualized[b] in the parenchyma:
$C5a_{CH}$	thin linear structures (0.5–2 mm): fissures, peripheral vessels
$C5b_{CH}$	rounded structures (2–6 mm): vessels seen *en face*
	Details to be reproduced[a] in the mediastinum:
$C6a_{CH}$	the carina with main bronchi
$C6b_{CH}$	the thoracic vertebra
$C6c_{CH}$	the mediastinum–lung interface
$C7_{CH}$	Details to be sharply visualized[b] in the costophrenic junction: the costopleural junction
Lumbar spine [5]	
$C1_{LS}$	Visually sharp reproduction[b] of the upper and lower plate surfaces, represented as lines in the central beam area
$C2_{LS}$	Visually sharp reproduction[b] of the pedicles
$C3_{LS}$	Reproduction[a] of the intervertebral joints
$C4_{LS}$	Reproduction[a] of the spinous and transverse processes
$C5_{LS}$	Visually sharp reproduction[a] of the cortex and trabecular structures
$C6_{LS}$	Reproduction[a] of the adjacent soft tissues, particularly the psoas muscle
$C7_{LS}$	Reproduction[a] of the sacroiliac joints

[a]Reproduction: details of anatomical structures are visible but not necessarily clearly defined; details emerging.
[b]Visually sharp reproduction: anatomical details are clearly defined; details clear.
[c]Visualization: characteristic features are detected but details are not fully reproduced; features just visible.

Two methods were used to score the images [7, 8]. The first was the image criteria score (ICS), defined as:

$$\mathrm{ICS} = \frac{\sum_{i=1}^{I}\sum_{c=1}^{C}\sum_{o=1}^{O} F_{i,c,o}}{ICO} \qquad (1)$$

where $F_{i,c,o}$ is the fulfilment of criterion (c) for image (i) and observer (o); I is the number of images assessed for each imaging system (I=15 for chest, I=10 for lumbar spine), C is the number of criteria (C=7) and O is the number of observers (O=7). If a criterion is fulfilled, $F_{i,c,o}$ is 1, and if it is not $F_{i,c,o}$ is 0. Since there were 16 chest and 4 lumbar spine imaging configurations, there were 240 chest and 40 lumbar spine images in total.

The second method of scoring was visual grading analysis. For this relative rating, each image was compared to a reference image. A patient image taken with 80 kV and 400 Lanex Regular Plus screen–film system was used as the reference image in the lumbar spine AP examination [8]. In the chest PA examination, the selection of reference image was more complicated. Prior to the collection of images, a statistical analysis of the necessary number of volunteers was performed [7]. The study was designed as an ''incomplete but balanced block trial''. 120 volunteers were required to test four technical factors (each under two conditions) and each volunteer was examined with 2 of the 16 techniques mentioned above. The observers viewed the radiographs in pairs. In the evaluation, the volunteers were then used as their own reference. If the structure in the image is reproduced much worse than in the reference image, it is given the score −2. If the structure is reproduced worse, equally, better or much better than in the reference image, it is given the score −1, 0, +1 or +2, respectively.

For a given system, a visual grading analysis score (VGAS) was calculated as:

$$\mathrm{VGAS} = \frac{\sum_{i=1}^{I}\sum_{s=1}^{S}\sum_{o=1}^{O} G_{i,s,o}}{ISO} \qquad (2)$$

where $G_{i,s,o}$ is the relative grading for a particular image (i), structure (s) and observer (o); S is the number of structures compared; and I and O are as described above.

In the clinical trial for chest [7], it was found that the modified criteria gave better discrimination between different techniques than the original criteria. The visual grading analysis was performed only on the modified criteria $C5_{CH}$–$C7_{CH}$, whereas the original criteria $C1_{CH}$–$C4_{CH}$ were assessed using both ICS and VGAS. It was therefore interesting to investigate whether the modified criteria show a more significant correlation with physical image quality than the original criteria. This was only possible for VGAS, since ICS values were not available for the modified criteria [7].

Measures of physical image quality

A Monte Carlo computer model of the complete imaging system was used to assess physical image quality. The model is an extension of previous work [11, 12]. It models the patient using an anthropomorphic 3-dimensional, segmented male anatomy (voxel phantom) originally developed elsewhere [13]. Appropriate anatomical details (Table 3) have been added to this phantom so that realistic estimates of the contrast and SNR of important details in the normal anatomy can be made. Estimates of the energy imparted per unit area to the image receptor at any point in the image plane were used to compute the optical density on the film by using the film's characteristic H&D curve. In this way it was possible to estimate the variations of the energy imparted to the screen–film system by scattered and primary photons and hence to assess the effects of the limited dynamic range of the screen–film system. The model takes specific account of the X-ray spectrum (anode material and angle, peak tube potential and ripple, and added filtration), antiscatter grid (strip frequency, lead strip width, grid ratio and material in interspaces and covers) or air gap, couch top or chest stand, and image receptor (cassette front, screen–film system and H&D curve). The computer program has been validated [9, 14] against measurements on phantoms and patients.

The properties of the anatomical details included in the phantoms are listed in Table 3. These details were selected on the basis of the image criteria and the list of important image details published by the European Commission [5].

To calculate the contrast or difference in optical density, ΔOD, beside and behind a particular detail, the effects of film gradient (γ) and imaging system unsharpness were considered. The film gradient was obtained from measurements of the H&D curve using the ISO standard [15]. The effect of unsharpness on ΔOD was calculated by considering the modulation transfer function (MTF) of receptor (screen), geometric (focal spot size and magnification) and motion unsharpness [16].

SNR was calculated in two steps. First, the $\mathrm{SNR_Q}$ due to quantum noise (index Q) only was calculated using the fluence of photons at the

Table 3. Properties of the anatomical details for which the difference in optical density (ΔOD) and signal-to-noise ratio were calculated in the two examinations. (The minimum and maximum calculated OD behind the detail for all the 16 simulated techniques are also listed for the chest examination)

Abbreviation	Description of the anatomical detail	Range of OD
Chest posteroanterior		
LLA	0.5 mm calcification in the left lung apex	0.9–1.5
RLA	0.5 mm calcification in the right lung apex	0.4–0.9
CRL	1.8 mm blood vessel in the central right lung	1.2–1.7
RCA	3.0 mm blood vessel in the retrocardiac area	0.5–1.0
CPA	1.8 mm blood vessel in the costophrenic angle area	1.0–1.6
Lumbar spine anteroposterior		
L1T	5.0 mm transverse process on the L1 vertebra	
L3T	3.5 mm transverse process on the L3 vertebra	
L5T	2.0 mm transverse process on the L5 vertebra	
L1D	1.0 mm bone marrow detail (trabecular structure) in the L1 vertebra	
L3D	1.0 mm bone marrow detail (trabecular structure) in the L3 vertebra	
L5D	1.0 mm bone marrow detail (trabecular structure) in the L5 vertebra	

screen and the single event size distribution of energy imparted to the screen [17]. The SNR_Q is based on the energy imparted per unit area to image elements beside and behind the detail and was calculated using the methodology in reference [11] and reference [18]. The SNR_Q overestimates the actual SNR. Multiplicative correction factors were applied to SNR_Q^2 to include the effects of additional noise from light emission from the screen and from film granularity. Methods from the literature [19] were used to derive these correction factors [16].

In addition to ΔOD and SNR, a measure of the dynamic range of the image data was computed. Dynamic range is important for the following reason: even though the object contrast may be large, the contrast on the film may be low owing to the low film contrast (gradient) in some parts of the image, and thus ΔOD will be reduced. Our measure of dynamic range was therefore defined as the percentage of pixels in the computed image having an OD such that the gradient γ(OD) exceeds a pre-set value, in our case 0.75 or 1.25 (chest) and 2.25 (lumbar spine). This physical image quality measure is thus an indication of how much of the image is properly exposed, *i.e.* with a ''reasonable'' film contrast, and is subsequently referred to as the PEF (properly exposed fraction). The pre-set values of the gradients were selected so that the PEF was sensitive to changes in imaging conditions (Table 1).

Statistical analysis

The software package Statistica® was used to compute the correlation coefficient r (the Pearson product-moment) between calculated values of ΔOD or SNR and ICS or VGAS. Calculated p-values were used to express the significance of the correlation (t-test). The correlation coefficient measures the magnitude, if any, of a linear causal relation. The null hypothesis is that there is no linear association between clinical and physical image quality. Correlations significant at $p<0.10$, $p<0.05$ and $p<0.01$ were identified. Correlations were sought between the physical image quality factors and the scores of the individual criteria as well as with the average scores when several or all criteria were used.

Results and discussion

Chest

Table 4a lists the correlation coefficients r between the ICS for the individual criteria $C1_{CH}$–$C4_{CH}$ as well as for the average ICS for all four criteria $C1_{CH}$–$C4_{CH}$ and the physical image quality measures. Tables 4b, c show the corresponding correlation coefficients with ICS and VGAS for the individual criteria $C5_{CH}$–$C7_{CH}$ and the average ICS and VGAS for the same criteria. A more significant correlation between clinical and physical image quality measures is found using the modified criteria $C5_{CH}$–$C7_{CH}$ (Table 4b) than using the original criteria $C1_{CH}$–$C4_{CH}$ (Table 4a) with ICS. Also, a more significant correlation with clinical image quality (using the modified criteria $C5_{CH}$–$C7_{CH}$) is found using the VGAS (Table 4c) than the ICS (Table 4b). Examples of the correlation between clinical and physical measures of image quality are given in Figures 1a–c at three levels of significance. In Figure 1a, the eight imaging systems with negative VGAS for the criteria $C5_{CH}$–$C7_{CH}$ (hence inferior clinical image quality) and low ΔOD for the detail in the retrocardiac area are all systems that use the lower maximum optical density in the lung region ($OD_{max}=1.3$). This shows the importance of not underexposing the chest film. Figures 1b and 1c show the correlation between $C3_{CH}$ (ICS) and ΔOD of

Table 4. Correlations between physical and clinical measures of image quality in the chest posteroanterior examination. The correlation between the difference in optical density (ΔOD) of the indicated details and the properly exposed fraction (PEF) and (a) the image criteria score (ICS) for criteria $C1_{CH}$–$C4_{CH}$, (b) the ICS for criteria $C5_{CH}$–$C7_{CH}$ and (c) the visual grading analysis score (VGAS) for criteria $C5_{CH}$–$C7_{CH}$ are presented. The last column shows the correlation between the sum of all the criteria scores and each of the physical measures

Physical image quality and anatomical detail	Correlation coefficient *r*						
	ICS $C1_{CH}$	ICS $C2_{CH}$	ICS $C3_{CH}$	ICS $C4_{CH}$	ICS $C1_{CH}$–$C4_{CH}$		
(a) ICS for criteria $C1_{CH}$–$C4_{CH}$							
PEF γ>0.75	0.33	0.78***	0.65***	0.79***	0.77***		
PEF γ>1.25	0.54**	0.84***	0.78***	0.88***	0.86***		
ΔOD LLA	0.03	0.11	0.01	0.23	0.15		
ΔOD RLA	0.41	0.69***	0.60**	0.77***	0.72***		
ΔOD CRL	0.05	0.08	0.05	0.22	0.14		
ΔOD RCA	0.47*	0.78***	0.72***	0.84***	0.81***		
ΔOD CPA	0.45*	0.60**	0.60**	0.69***	0.65***		
	ICS $C5a_{CH}$	ICS $C5b_{CH}$	ICS $C6a_{CH}$	ICS $C6b_{CH}$	ICS $C6c_{CH}$	ICS $C7_{CH}$	ICS $C5_{CH}$–$C7_{CH}$
(b) ICS for criteria $C5_{CH}$–$C7_{CH}$							
PEF γ>0.75	0.57**	0.64***	0.77***	0.86***	0.81***	0.76***	0.82***
PEF γ>1.25	0.72***	0.77***	0.84***	0.90***	0.90***	0.86***	0.91***
ΔOD LLA	0.45*	0.45*	0,03	0.32	0.25	0.25	0.26
ΔOD RLA	0.77***	0.78***	0.64***	0.82***	0.78***	0.78***	0.80***
ΔOD CRL	0.48*	0.48*	0.00	0.26	0.23	0.30	0.24
ΔOD RCA	0.78***	0.81***	0.74***	0.86***	0.85***	0.87***	0.87***
ΔOD CPA	0.78***	0.75***	0.54**	0.67***	0.69***	0.76***	0.71***
	VGAS $C5a_{CH}$	VGAS $C5b_{CH}$	VGAS $C6a_{CH}$	VGAS $C6b_{CH}$	VGAS $C6c_{CH}$	VGAS $C7_{CH}$	VGAS $C5_{CH}$–$C7_{CH}$
(c) VGAS for criteria $C5_{CH}$–$C7_{CH}$							
PEF γ>0.75	0.63***	0.65***	0.85***	0.88***	0.81***	0.77***	0.82***
PEF γ>1.25	0.70***	0.71***	0.92***	0.93***	0.89***	0.85***	0.90***
ΔOD LLA	0.75***	0.77***	0.21	0.39	0.41	0.41	0.46*
ΔOD RLA	0.92***	0.94***	0.79***	0.88***	0.89***	0.86***	0.91***
ΔOD CRL	0.76***	0.79***	0.21	0,39	0.44*	0.43*	0.47*
ΔOD RCA	0.88***	0.92***	0.87***	0.93***	0.95***	0.92***	0.95***
ΔOD CPA	0.93***	0.95***	0.70***	0.78***	0.86***	0.85***	0.86***

Significance of correlation: *p<0.10; **p<0.05; ***p<0.01.
LLA, left lung apex; RLA, right lung apex; CRL, central right lung; RCA, retrocardiac area; CPA, costophrenic angle area.

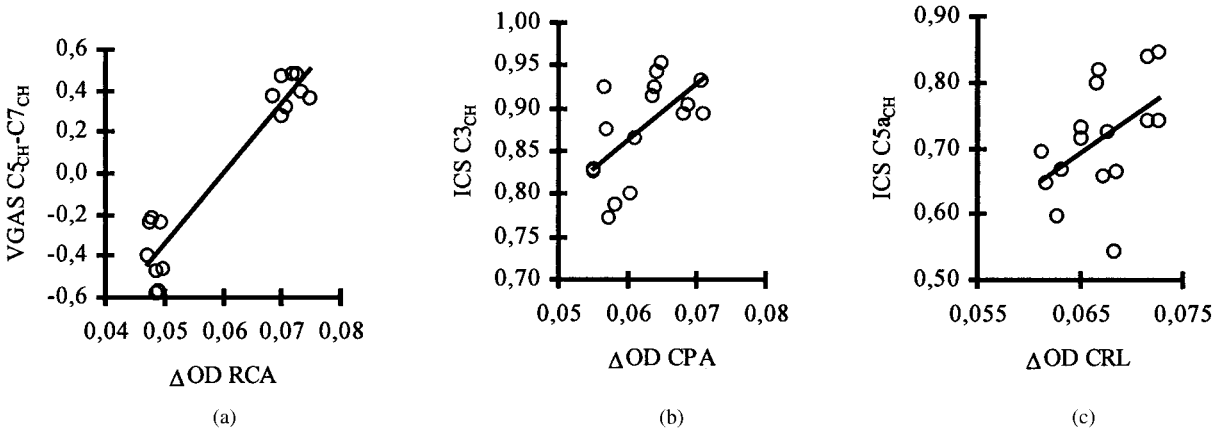

Figure 1. Correlation between clinical and physical measures of image quality in the chest examination at three levels of significance. Correlations between (a) visual grading analysis score (VGAS) for criteria $C5_{CH}$–$C7_{CH}$ and difference in optical density (ΔOD) for retrocardiac area (RCA) (r=0.95, p=0.00000001), (b) image criteria score (ICS) for criterion $C3_{CH}$ and ΔOD for costophrenic angle area (CPA) (r=0.60, p=0.013) and (c) ICS for criterion $C5a_{CH}$ and ΔOD central right lung (CRL) (r=0.48, p=0.060) are shown. The solid line is the linear regression line.

the costophrenic angle area, and between $C5a_{CH}$ (ICS) and ΔOD of the central right lung, respectively. These correlations are less significant.

The three physical image quality measures that show the most significant correlation with ICS and VGAS are the contrast of blood vessels in the retrocardiac area, our measure of dynamic range (PEF $\gamma>1.25$), and the contrast of calcifications in the right lung apex. This can be explained by considering the H&D curve. Both of these details are situated in areas where the OD is less than 1.0, hence on the toe of the H&D curve. The ΔOD is therefore much increased if the OD_{max} in the chest image is increased from 1.3 to 1.8. The ΔOD of details situated in areas where the OD is generally higher (OD>1.0), such as the central right lung and the left lung apex, does not show as significant a correlation with ICS and VGAS as details in regions with low optical density. A possible explanation for this is that the ΔOD of these details is already high enough and the criteria are therefore already fulfilled. For the same details, the ΔOD shows a more significant correlation with ICS and VGAS than the SNR. For example, there is no significant ($p<0.05$) correlation between ICS and SNR. However, there is a significant correlation between VGAS for criteria $C5a_{CH}$ and $C5b_{CH}$ and SNR, but this is not as significant as with ΔOD ($p<0.01$) for the same detail. This is an indication that clinical image quality is limited more by contrast than by noise in chest screen–film radiography.

The most significant correlation between clinical and physical image quality is found with criteria $C5a_{CH}$ and $C5b_{CH}$ (VGAS), whereas a poor correlation is found with $C1_{CH}$ (ICS). An explanation could be that the wording of $C1_{CH}$ is not as specific as the wording of $C5a_{CH}$ and $C5b_{CH}$, and that it may be difficult to find a single physical measure that correlates to such a general criterion as $C1_{CH}$.

No significant correlation was found between the physical parameters such as applied tube potential, screen–film speed and scatter–rejection technique on the one hand and ICS and VGAS on the other. However, a significant correlation was found between the maximum OD in the chest PA image (OD_{max}) and both ICS and VGAS. This indicates that the OD_{max} is the most important parameter of the four tested; the other three are of lesser importance. If the image is properly exposed (hence not underexposed, as with $OD_{max}=1.3$), the choice of screen speed, scatter–rejection technique and tube potential is not critical, or at least will not affect the image quality enough to generate significantly different ICS and VGAS in the clinical trial [7]. Similar conclusions have also been found on the basis of the computational model alone [14].

Contrary to earlier work [6], this work was able to demonstrate that clinical image quality can be predicted, provided that three conditions are satisfied. This may prove useful, as optimization based on clinical image quality alone can be difficult and time consuming. The conditions are as follows. First, it is important to characterize the imaging system in sufficient detail for the model calculations to agree with measurements on the imaging system on an absolute scale [18]. Second, the effect of the different radiographic technique factors (Table 1) must be acknowledged in combination and not used separately in attempts to correlate with clinical image quality. Finally, the effect of the different radiographic technique factors must be combined into appropriate measures of physical image quality (*i.e.* contrast and SNR) that correspond to the perception or visualization of relevant anatomical details, *i.e.* to specific diagnostic tasks.

Lumbar spine

Table 5 shows the correlation coefficients r between the clinical image quality measures ICS and VGAS for different combinations of image criteria and calculated physical image quality measures ΔOD and SNR of the anatomical details and our measure of dynamic range, PEF. Examples of the correlation between clinical and physical measures of image quality are given in Figures 2a–c at three levels of significance.

A positive correlation between clinical and physical measures of image quality was found for all tested comparisons. However, as expected, some correlations were more significant than others. Generally, a stronger correlation was found when all seven criteria were used in the ICS and VGAS evaluations than if only one ($C4_{LS}$ or $C5_{LS}$) or two ($C4_{LS}$ and $C5_{LS}$) criteria were used. Criteria $C4_{LS}$ and $C5_{LS}$ are of particular interest since they mention the anatomical details used in the model calculation of ΔOD and SNR (transverse processes (L1T–L5T) and trabecular details (L1D–L5D)). Typically, a stronger correlation was found with $C5_{LS}$ than with $C4_{LS}$. Also, a stronger correlation with clinical image quality was found for the physical image quality measures that use the trabecular structure detail than those that use the transverse processes. The visibility of the transverse processes is also influenced by the stomach content that may interfere with the perception of the processes.

The ΔOD and SNR of the L1D–L5D trabecular details were the best predictors of clinical image quality amongst those tested. The percentage of the calculated image with a film gradient larger than 2.25 (PEF) was not as good as the

Table 5. Correlations between physical and clinical measures of image quality in the lumbar spine anteroposterior examination. Correlation between the image criteria score (ICS) or visual grading analysis score (VGAS) for criteria $C1_{LS}$–$C7_{LS}$ [5] on the one hand and difference in optical density (ΔOD), signal-to-noise ratio (SNR) of indicated details and properly exposed fraction (PEF) on the other are presented. The first column shows the correlation between the sum of all the criteria scores and each of the physical measures

Physical image quality and anatomical detail	Correlation coefficient *r*							
	ICS $C1_{LS}$–$C7_{LS}$	ICS $C4_{LS}$–$C5_{LS}$	ICS $C4_{LS}$	ICS $C5_{LS}$	VGAS $C1_{LS}$–$C7_{LS}$	VGAS $C4_{LS}$–$C5_{LS}$	VGAS $C4_{LS}$	VGAS $C5_{LS}$
PEF γ>2.25	0.85	0.94	0.98**	0.90*	0.86	0.90*	0.96**	0.78
ΔOD L1D	0.99***	0.99***	0.96**	0.99***	0.99***	1.00***	0.97**	0.95**
ΔOD L3D	1.00***	0.99**	0.95*	0.99***	1.00***	1.00***	0.97**	0.96**
ΔOD L5D	1.00***	0.96**	0.90*	0.98**	1.00***	0.99**	0.93*	0.97**
ΔOD L1T	0.99***	0.96**	0.88	0.98**	0.99***	0.97**	0.93*	0.94*
ΔOD L3T	0.98**	0.93*	0.84	0.96**	0.98**	0.95**	0.90*	0.94*
ΔOD L5T	0.95*	0.87	0.76	0.92*	0.94*	0.90*	0.83	0.90*
SNR L1D	1.00***	0.97**	0.92*	0.98**	1.00***	0.99***	0.93*	0.98**
SNR L3D	1.00***	0.97**	0.92*	0.98**	1.00***	0.99***	0.94*	0.98**
SNR L5D	1.00***	0.96**	0.90*	0.97**	1.00***	0.99**	0.92*	0.98**
SNR L1T	1.00***	0.95**	0.88	0.97**	0.99***	0.98**	0.92*	0.96**
SNR L3T	0.99***	0.95*	0.88	0.97**	0.99***	0.98**	0.92*	0.97**
SNR L5T	0.99***	0.95*	0.87	0.97**	0.99***	0.97**	0.91*	0.96**

Significance of correlation: *p<0.10; **p<0.05; ***p<0.01.

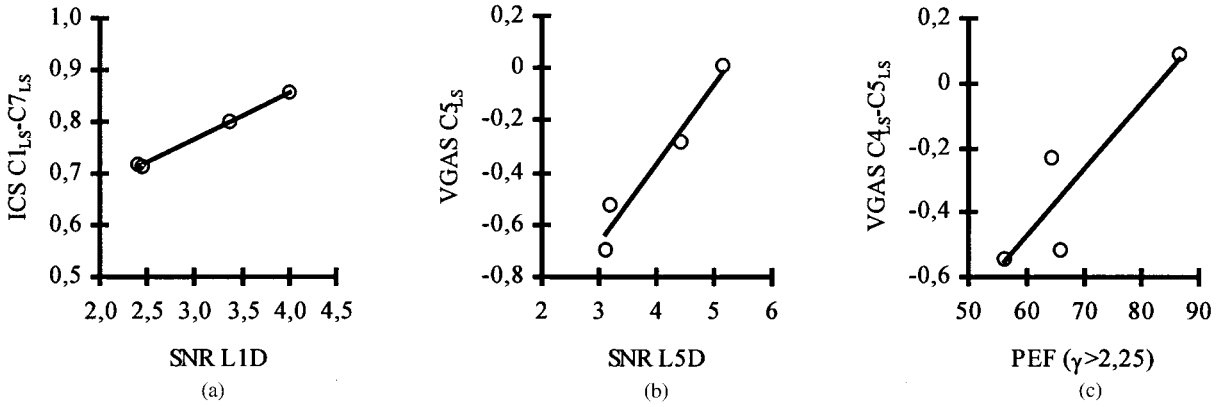

Figure 2. The correlation between clinical and physical measures of image quality in the lumbar spine examination at three levels of significance. Correlations between (a) image criteria score (ICS) for criteria $C1_{LS}$–$C7_{LS}$ and signal-to-noise ratio (SNR) L1D (r=1.00, p=0.0016), (b) visual grading analysis score (VGAS) for criterion $C5_{LS}$ and SNR L5D (r=0.98, p=0.022) and (c) VGAS for criteria $C4_{LS}$–$C5_{LS}$ and properly exposed fraction (PEF, γ>2.25) (r=0.90, p=0.099) are shown. The solid line is the linear regression line.

ΔOD and SNR of particular details. This may indicate that, provided the spine is properly exposed, the surrounding soft tissue with significantly higher optical density, possibly overexposed, is not a problem.

Conclusions

A statistically significant correlation exists between some physical image quality measures and clinical image quality as assessed by expert radiologists using the EU image criteria. For PA chest radiography, these physical measures are the contrast (ΔOD) of blood vessels in regions with comparatively low optical densities, such as the retrocardiac area. No significant correlation, however, was found between the SNR of details and clinical image quality. To quantify the effect of dynamic range on image quality, a new quantity, the properly exposed fraction, was introduced. The PEF shows a significant correlation with clinical image quality in chest imaging and demonstrates the importance of proper film exposure.

For AP lumbar spine radiography, the ΔOD and SNR of trabecular details in the L1–L5 vertebrae are the best predictors of clinical image quality, whereas the PEF is not as good.

The significant correlations found between clinical image quality and some physical image quality measures in this work are encouraging and show that, for the situations considered, the

clinical image quality can be predicted provided the imaging conditions are known in detail and relevant measures of physical image quality are used.

Acknowledgments

The authors would like to thank the following European expert radiologists, Dr C Gückel, Prof. M Laval-Jeantet, Prof M Maffessanti, Prof J-W Oestmann and Prof G Whitehouse, for many fruitful discussions and for evaluation of clinical image quality. Dr Francis R Verdun (Institute for Applied Radiophysics, Lausanne) is acknowledged for providing us with the measured film H&D curve.

References

1. International Commission on Radiological Protection. Radiological protection and safety in medicine, ICRP Publication 73. Annals ICRP 26 Oxford: Pergamon, 1996.
2. EU 1997. Council Directive 97/43/Euratom of 30 June 1997 on health protection of individuals against the dangers of ionizing radiation in relation to medical exposure, and repealing Directive 84/466/Euratom. Official Journal of the European Communities L180, 40, 22.
3. Kwan-Hong NG, Bradley DA, Warren-Forward HM, editors. Subject dose in radiological imaging. Amsterdam: Elsevier, 1998.
4. International Commission on Radiation Units. Medical imaging—the assessment of image quality, ICRU Report 54. Bethesda, MD: ICRU Publications, 1995.
5. European Commission. CEC quality criteria for diagnostic radiographic images and patient exposure trial, EUR 12952. Brussels: European Commission, 1990.
6. Leitz WK, Månsson LG, Hedberg-Vikström BRK, Kheddache S. In search of optimum chest radiography techniques. Br J Radiol 1993;66:314–21.
7. Lanhede B, Tingberg A, Månsson LG, Kheddache S, Widell M, Björneld L, et al. The influence of different technique factors on image quality for chest radiographs: application of the recent CEC image quality criteria. Radiat Prot Dosim 2000;90: 203–6.
8. Almén A, Tingberg A, Mattsson S, Besjakov J, Kheddache S, Lanhede L, et al. The influence of different technique factors on image quality of lumbar spine radiographs as evaluated by established CEC image criteria. Br J Radiol 2000;73: 1192–9.
9. Dance DR, McVey G, Sandborg M, Persliden J, Alm Carlsson G. Calibration and validation of a voxel phantom for use in the Monte Carlo modelling and optimisation of X-ray imaging systems. SPIE Proceedings 1999;3659:548–59.
10. Zankl M, Panzer W, Herrmann C. Calculation of patient doses using a human voxel phantom of variable diameter. Radiat Prot Dosim 2000;90: 155–8.
11. Sandborg M, Dance DR, Persliden J, Alm Carlsson G. Monte Carlo program for the calculation of contrast, noise and absorbed dose in diagnostic radiology. Comp Meth Prog Biomed 1994;42: 167–80.
12. Dance DR, Lester SA, Alm Carlsson G, Sandborg M, Persliden J. The use of carbon fibre material in radiographic cassettes: estimation of the dose and contrast advantages. Br J Radiol 1997;70:383–90.
13. Zubal IG, Harrell CR, Smith EO, Rattner Z, Gindi G, Hoffer PB. Computerised three dimensional segmented human anatomy. Med Phys 1994;21: 299–302.
14. Sandborg M, McVey G, Dance DR, Alm Carlsson G. Comparison of model predictions of image quality measures with results of clinical trials in chest and lumbar spine screen–film imaging. Radiat Prot Dosim 2000;90:173–6.
15. International Organization for Standardization. Photography—sensitometry of screen–film systems for medical radiography—Part 1: Determination of sensitometric curve shape, speed, and average gradient, ISO 9236. Geneva: International Organization for Standardization, 1996.
16. Sandborg M, Dance DR, Alm Carlsson G. Implementation of unsharpness and noise into a model of the imaging system: applications to chest and lumbar spine imaging. ISRN ULI-RAD-R-90-SE, http://huweb.hu.liu.se/inst/imv/radiofysik/publi/reports.html#99
17. Sandborg M, Alm Carlsson G. Influence of energy spectrum, contrasting detail and detector on the signal-to-noise ratio (SNR) and detective quantum efficiency (DQE) in projection radiography. Phys Med Biol 1992;37:1245–63.
18. Tapiovaara MJ, Sandborg M. Evaluation of image quality in fluoroscopy by measurements and Monte Carlo calculations. Phys Med Biol 1995;40:589–607.
19. Nishikawa RM, Yaffe MJ. Model of the spatial-frequency-dependent detective quantum efficiency of phosphor screens. Med Phys 1990;17:894–904.