Journal of Medical Radiation Sciences

Open Access

ORIGINAL ARTICLE

Understanding reasons for image rejection by radiologists and radiographers

Robin Decoster, MSc[1], Rachel Toomey, PhD[1], Dirk Smits, PhD[2], Tamara Miner Haygood, MD PhD[3], & Marie-Louise Ryan, PhD[1]

[1]School of Medicine, University College Dublin, Dublin, Ireland
[2]Health Care Science, Odisee University College, Brussel, Belgium
[3]MD Anderson Cancer Center, The University of Texas, Houston, Texas, USA

Keywords
diagnostic imaging, image quality, interprofessional, radiographer, radiology

Correspondence
Robin Decoster, School of Medicine, University College Dublin, Belfield, Dublin, Dublin 4, Ireland; Tel/Fax: +32 476 96 65 49; E-mail: robin.decoster@ucdconnect.ie

Received: 29 March 2022; Accepted: 28 November 2022

J Med Radiat Sci **0** (2022) 1–10

doi: 10.1002/jmrs.641

Abstract

Introduction: A core element of a radiographer's role is the decision on whether a radiograph is sufficient for diagnosis, or a repeat examination is needed. Studies illustrate the disagreement on the diagnostic value of radiographs between radiographers and radiologists, which may influence repeat examinations. This study investigates if parameters contributing to image quality are possible determinants to explain the difference between professions. **Methods:** A total of 74 radiographers and radiologists from three different countries assessed three data sets (chest PA, hip HBL, c-spine lateral), each containing 25 radiographs. All observers scored image quality in terms of anatomical visualisation, positioning, collimation, detector exposure and judged the diagnostic value using the ACR RadLex classification. All assessments were performed on a clinically relevant display. Visual grading characteristics were used to compare image quality evaluations between groups. **Results:** Radiographers scored the visualisation of anatomical structures lower than radiologists though the difference was not statistically significant. A difference in classification using the RadLex categories – with radiographers rejecting more radiographs – was demonstrated. Only the subjective evaluation of the detector exposure correlated statistically with RadLex ratings. There was no difference between radiographers and radiologists when reviewing patient positioning and collimation. **Conclusion:** Radiographers and radiologists agree on the visualisation of anatomical structures, but radiographers are more critical towards the diagnostic value. Within the criteria studied, the evaluation of anatomical structures does not explain the difference. Radiographs have a higher change of being rejected if the observer (subjectively) assessed the detector exposure as inappropriate. This correlation is stronger for radiographers.

Introduction

The production of a radiograph of sufficient quality to enable accurate diagnosis is of common interest to radiologists and radiographers. The radiographer visually and subjectively estimates whether the acquired image quality matches the quality required to answer the clinical question.[1,2] In this assessment, the radiographer should consider that the benefit of a retake must surpass the negative effect of the additional radiation dose,[1,3] as balancing patient radiation dose and image quality is a fundamental aspect of good practice in radiography.[4,5] Radiographers, when satisfied with the quality of a radiographic image, end the examination and make the radiographs available for reporting by the radiologist.[1,6]

The image quality requirements of a radiograph may vary depending on factors such as the clinical question to be answered. Therefore, Uffmann et al.[7] presented categories of image quality which might be required in different cases. For example, a high image quality was expected for primary bone tumours and non-displaced fractures. In contrast, follow-up examinations for pneumonia or scoliosis only needed a low image quality.[7] If radiographers reject and retake radiographs that would have been sufficient to answer the clinical question, this results in a needless radiation exposure. It is difficult, as described by Hofmann,[8] Waaler and Hofmann,[5] to determine the rate at which radiographers reject images which would be considered sufficient for diagnosis. In many cases, radiographs are deleted even before they reach the radiologist.[2,5,8]

Mount[1] investigated the rejection rate of lateral knee radiographs in a monocentric study. In conclusion, Mount[1] hypothesised a different approach to image quality assessment by different professions may be the primary cause for the different acceptance rate. Radiographers were found 'to consistently rate image quality more harshly than radiologists'.[1] Their research indicated a more technical evaluation (defined as exposure, visualisation of certain structures, collimation and positioning) by radiographers, compared to a more clinical assessment (diagnostic value) by radiologists.[1] The discrepancy between both professions is in line with the review by Waaler and Hofmann[5] who explain the difference based on the absence of the image quality discussion between radiologist and radiographers.

However, Mount's[1] methodology focused on the initial perception of a lateral knee radiograph and did not request individual scores for the different factors, such as anatomical criteria, noise and so on, allowing only a statistical difference between radiographers and radiologists on the general assessment of the radiograph. Therefore, the explanation for this difference, and how radiographers' overall judgements of image quality relate to their perception of the visibility of anatomical structures, remains unclear.

Other researchers have also investigated how radiographers judge image quality and whether it is a structure-based or more global evaluation. Work by Larsson et al.[9] and Lundvall et al.[10] found that radiographers judge the quality of the radiograph based on patient conditions during the examination and the visibility of the pathology. Their findings are supported by the findings of Prime and Le Masurier,[11] who point out that radiographers use 'illness scripts' or a mental models combining information collected by the different sources (observation, request card) and knowledge (anatomy, pathology, radiographic technique). These models are used to predict and assess the overall examination in an overall judgement by comparing the current patient with the mental model.[9–15]

Regardless of the reasoning by which radiographers accept or reject a radiograph, the fact remains that experiments by Mount,[1] Kjelle et al.[16] and Dunn et al.[17] demonstrated a difference between radiographers and radiologists in terms of what was considered acceptable image quality, with radiographers demonstrating higher reject rates. The current study set out to investigate the (quantitative) difference between radiologists and radiographers based on an overall judgement ('technical' factors and general assessment of clinical usability) in comparison to the scrutiny of anatomical structures. Additionally, this study aimed to assess if the technical (such as positioning and noise) or anatomical factors can explain the possible differences.

Methods

To investigate the factors influencing overall image quality judgements by radiographers and radiologists, three sets of radiographs (DICOM format) with different characteristics were assembled and presented to members of each profession (radiographers and radiologists). These participants were then asked to rate various image criteria, but also to rate the overall holistic image quality using the RadLex scale.[18] This allowed the authors to determine which factors best correlated with overall quality judgements in each profession.

Images

The images used in the study were selected out of a larger data set (sequence of 100 consecutive clinically accepted examinations with random start date) from three clinical PACSs (Picture Archiving and Communication System). While it is not possible to establish an absolute 'truth' for ratings of image quality, the selected images were considered by three radiographers to represent a range of image qualities. The first set contained 23 PA (posterior–anterior) chest radiographs to investigate the relationship in a typically low-contrast radiograph. The second set contained 23 lateral cervical spine radiographs with higher contrast in the bony structures but still with an important soft tissue component in the pre-vertebral region. The third set contained 23 lateral hip radiographs (horizontal beam) where bony structures are the primary concern, but the projection is generally considered by radiographers to be harder to position. To assess inter-observer variability, one radiograph from each set was presented two additional times at random points; therefore, each observer viewed a total of 25 images per data set. The repeated radiographs were only used for inter-observer variability and discarded for other statistics. In the hip data set, no results were summarised for the United States of America, as the observers did not

Table 1. Overview of the anatomical criteria to be evaluated, based on Carmichael et al.[19]

	Cervical spine lateral	Chest PA	Hip Horizontal Beam Lateral (HBL)
Q1	The tubercle posterior of C1 is easily identified	Visually sharp reproduction of the diaphragm	Visually sharp reproduction of the acetabulum
Q2	C7-T1 intervertebral joint space is clearly seen	Reproduction of the trachea	Visually sharp reproduction of the femoral head
Q3	The visually sharp reproduction of the facet joint	Reproduction of the costophrenic angles	Visually sharp reproduction of the femoral neck
Q4	The visually sharp reproduction of spinous processes	Visually sharp reproduction of the heart, aorta and mediastinum	Clear visualisation of the trochanters
Q5	Sufficient contrast between bone, air (trachea) and tissue	Visually sharp reproduction of thoracic spine	Sufficient contrast between bone and tissue

volunteer for this data set as they declared to be unfamiliarity with the positioning technique.

To correlate the diagnostic quality to the assessment of anatomical structures, observers rated the representation of five relevant anatomical criteria (Table 1) selected from European guidelines[19] for each data set on a Likert scale ranging from 'not fulfilled'[1] to 'fulfilled'. To the best of the authors' knowledge, there are no alternative guidelines on image quality for radiographs. The guidelines merely describe general accepted criteria such as the sharp reproduction of the diaphragm. To minimise the risk of a bias due to the unfamiliarity of the American observers, the guidelines were approved by a US radiologist as a part of the ethical review.

To assess the influence of non-anatomical factors, the observers also rated the collimation, radiographic positioning and the detector dose (i.e. detector exposure, appearance of noise) on a Likert scale ranging from 'unacceptable'[1] to 'excellent'.[5] To determine the diagnostic quality, the observers categorised each radiograph using the RadLex categories (Radiological Lexicon, Radiological Society of North America). The description for each of the four categories (Table 2) was provided for their reference. Observers also rated their confidence in their RadLex ratings on a Likert scale, ranging from 'guess'[1] to 'very confident'.[5]

In a pilot study, three observers assessed the data sets to evaluate the data collection tool. Based on the findings, the logic of the data collection tool was improved. Additionally, the observers used all categories of image quality, indicating that the data set contained the full spectrum of image quality.

Observers

The grading of the radiographs for the different anatomical criteria, positioning, noise and so on is a time-intensive and focussed task. Therefore, a convenience sample of observers were recruited through volunteering in Ireland, Belgium and the United States of America. The countries were selected based on access to the radiographer and radiologist workforce by the researchers following previous collaborations, and to include countries which had differences in educational programmes of radiographers.[20–22] In Ireland and Belgium, the study was advertised through the professional bodies and through academic or teaching hospitals. In the United States of America, the study was

Table 2. The definition of the RadLex categories as presented in the RadLex online lexicon.[18]

Category	Description
Non-diagnostic	Little or no clinically usable diagnostic information (e.g. system failure or extensive motion artefact). Insufficient information to answer the primary clinical question (e.g. area of interest does not project, superposition of structures or artefacts limits visibility in area of interest).
Limited	Not as much diagnostic information as is typical for an examination of this type but likely sufficient to answer the primary clinical question. For example, motion artefact, body habitus or patient positioning might limit visualisation of some body regions, but the area of interest is sufficiently visualised. Such imaging might need to be repeated, depending on the clinical circumstances.
Diagnostic	Image quality that would be expected routinely when imaging cooperative patients.
Exemplary	Image quality that can serve as an example that should be emulated.

conducted in close collaboration with a teaching hospital where one author is based, and at an event hosted by the American Board of Radiology national representative bodies (e.g. radiologists presenting for 10-year recertification examinations with the American Board of Radiology volunteering in their free time). In total, 74 radiologists and radiographers participated in the study. A strength of this recruitment is the independence from one centre or clinical routine. At the same time, the fact that the radiographers and radiologists were recruited separately and do not work together on a daily basis might be considered a limitation of the study, because it might result in the absence of mutual criteria for acceptance of a radiograph that would exist within a given institution and bias the data collection. As presented in the paper of Waaler and Hofmann,[5] routine discussions on image quality might align the two professions in the definition of acceptable and minimum necessary image quality; this could not be investigated in this case.

Power calculations were performed prior to participant recruitment to determine the minimum numbers of participants required in each country and profession. With a power set at 0.80 (two-tailed), the sample size needed to be at least seven observers in each group to detect a significant difference of 0.25 at the 0.05 level. To exclude a lack of statistical power, post hoc power calculations were also conducted. The required numbers were achieved or exceeded in all cases. Demographic information (e.g. years of experience) of the observers was collected at the start of the study. The research was submitted to the home institution's Human Research Ethics Committee (University College Dublin) and declared exempt from full ethical review. One study site further requested a full ethical review, which was approved by that institution's ethics committee (The University of Texas MD Anderson Cancer Center). The data were collected between 2014 and 2017. During this time, the protocol was not amended.

Image presentation

All observers used ViewDex to review the images and record their responses, allowing them to use clinically applicable tools, such as window width/level adjustments.[23–27] A DICOM calibrated primary display (6MP; L_{max} 1050 cd/m^2; L_{min} 0.7 cd/m^2) was used for radiologists and a secondary display (2MP; L_{max} 400 cd/m^2; L_{min} 0.47 cd/m^2) for radiographers[28,29] in line with the display types typically used by each profession in clinical practice. Average lighting conditions of 28.29 ± 7.99 lx were recorded across all sites. To give a common clinical context to the observers, they were informed that the clinical indication was 'trauma evaluation'.

Statistical analysis

To explore differences in the rating of the anatomical structures, a Visual Grading Characteristics curve (AUC_{VGC}) was calculated for each structure.[9,11,19,24,30,31] The differences in the categorisation by diagnostic value (RadLex) between radiologists and radiographers were tested with a Mann–Whitney U test. Spearman correlations were used to explore correlations between RadLex categories, the rating of anatomical structures, collimation, positioning and delivered dose. Significant correlations were further investigated with a multinomial logistic model to determine which factors increased the chance of a radiograph being non-diagnostic. For example, this model calculates the chance of a radiograph being classified in a lower category in relation to its score for positioning. Inter-observer and intra-observer variability were evaluated using an interclass correlation coefficient or ICC.[32,33] All statistical calculations were performed using SPSS (Statistical Package for the Social Sciences. version 23).

Results

Reliability of observers

In total, 74 radiologists and radiographers volunteered and participated in the study (Table 3). The reliability of observers was analysed by comparing their RadLex scores for the radiographs repeated in each image set, with every observer in the study demonstrating good intra-observer consistency (ICC > 0.8).

Variations in mean years of experience were noted between radiographers and radiologists in all countries. The radiologists were all above 5 years of experience. To check if a difference in radiographers' experience significantly influenced RadLex ratings, a Mann–Whitney

Table 3. Overview of the number (*N*) of participants and their experience (years) in the field of radiology.

County	Observer	Number of Participants	Experience (years)	
			Mean	Std. Dev.
Belgium	Radiographer	13	5.67	6.04
	Radiologist	9	10.33	3.81
Ireland	Radiographer	17	5.21	4.29
	Radiologist	8	15.69	14.92
USA	Radiographer	10	9.6	5.5
	Radiologist	17	24.29	11

U test was applied to compare the RadLex classification groups of greater than or less than 5 years of experience for each country. This showed that the more experienced radiographers' ratings were not significantly different in some cases but were lower for several image types (USA radiographers in both c-spine and chest, ($P < 0.01$); Belgian radiographers for hip ($P < 0.05$) and chest ($P < 0.05$) and Irish radiographers ($P < 0.01$) for the hip). The impact of any experience-related effects is discussed later in this section.

Assessment of anatomical structures

Visual grading characteristics (VGC) were used to investigate a difference between the two professions in the assessment of anatomical structures. As summarised in Table 4, only a few significant (*) differences were found for certain structures.

As presented in Table 4, differences between radiographers and radiologists mostly concerned one or two structures from the list of criteria. For instance, the USA radiographers graded significantly lower compared with the USA radiologists the visualisation of diaphragm (Q1, $P = 0.03$), or the Belgium radiographers graded the contrast between bone and soft tissue (Q5) in the hip images significantly ($P < 0.05$) lower in comparison with the Belgian radiologists. In contrast, the c-spine is notable as the only examination where multiple anatomical criteria were assessed differently by radiographers and radiologists for any examination or in any country. This is explored further in the discussion.

Factors influencing RadLex rating

The purpose of the RadLex categories was to compare the opinions of observers on the overall diagnostic value of the radiographs, rather than the quality of individual features. The Mann–Whitney U test demonstrated a significant difference ($P < 0.01$) in ranking between the radiographers and the radiologists for every country and radiograph type.

Table 5 shows the frequency of each overall rating given by each profession in each country for all examinations. It also shows the total percentage of radiographs achieving a rating at a certain level or lower (cumulative percent), which allows the reader to compare the total percentage of cases at or below a given quality rating more easily – for example, allowing easy comparison of how many images were rated 'limited' or below.

Based on the cumulative percent (Table 5), it might be observed that radiographers tend to classify radiographs into lower categories than radiologists, especially for images of lower quality; for example, the percentage of images rated ether 'non-diagnostic' or 'limited' is consistently higher for radiographers than radiologists (Belgium: 49.2% vs. 32.6%; Ireland 49.8% vs. 30.1%; US 32.6% vs. 15.9%), while the percentage categorised as 'exemplary' is more similar. Although radiographers were stricter in their overall RadLex classifications, they reported less confidence ($P < 0.05$) than radiologists.

It was noted above that the radiographer groups in each country had lower mean years of experience than

Table 4. Overview of the VGC analyser results; $AUC_{VGC} = 0.5$ signifies equal ratings by radiographers and radiologists; > 0.5 indicates a lower rating by radiographers; <0.5 a lower rating by radiologists.

			Belgium		Ireland		USA	
			AUC_{VGC}	*P*-Value	AUC_{VGC}	*P*-Value	AUC_{VGC}	*P*-Value
Chest	Q1	Visually sharp reproduction of the diaphragm	0.52	0.703	0.58	0.291	0.668	0.031*
	Q2	Reproduction of the trachea	0.584	0.17	0.481	0.804	0.663	0.057
	Q3	Reproduction of the costophrenic angles	0.529	0.477	0.552	0.486	0.652	0.057
	Q4	Visually sharp reproduction of the heart, aorta and mediastinum	0.599	0.915	0.592	0.238	0.692	0.02*
	Q5	Visually sharp reproduction of thoracic spine	0.429	0.334	0.557	0.421	0.615	0.21
C-Spine	Q1	The tubercle posterior of C1 is easily identified	0.696	0.03*	0.624	0.114	0.761	0.002*
	Q2	C7-T1 intervertebral joint space is clearly seen	0.57	0.29	0.512	0.798	0.578	0.245
	Q3	The visually sharp reproduction of the facet joint	0.55	0.37	0.567	0.411	0.718	0.012*
	Q4	The visually sharp reproduction of spinous processes	0.56	0.2	0.587	0.279	0.748	0.005*
	Q5	Sufficient contrast between bone, air (trachea) and tissue	0.64	0.06	0.6	0.213	0.745	0.006*
Hip	Q1	Visually sharp reproduction of the acetabulum	0.59	0.15	0.576	0.247		
	Q2	Visually sharp reproduction of the femoral head	0.55	0.37	0.575	0.329		
	Q3	Visually sharp reproduction of the femoral neck	0.58	0.19	0.61	0.1		
	Q4	Clear visualisation of the trochanters	0.51	0.84	0.612	0.134		
	Q5	Sufficient contrast between bone and tissue	0.65	0.02*	0.616	0.14		

Table 5. Frequencies for the classification of the radiographs in RadLex categories, sorted for the countries and types of observers.

Country	Observer Type		Frequency	Percent (%)	Cumulative Percent (%)
Belgium	Radiographer	Non-diagnostic	108	11.1	11.1
		Limited	372	38.2	49.2
		Diagnostic	435	44.6	93.8
		Exemplary	60	6.2	100
		Total	975	100	
	Radiologist	Non-diagnostic	49	7.5	7.5
		Limited	163	25.1	32.6
		Diagnostic	393	60.5	93.1
		Exemplary	45	6.9	100
		Total	650	100	
Ireland	Radiographer	Non-diagnostic	158	12.4	12.4
		Limited	477	37.4	49.8
		Diagnostic	560	43.9	93.7
		Exemplary	80	6.3	100
		Total	1275	100	
	Radiologist	Non-diagnostic	40	7	7
		Limited	133	23.1	30.1
		Diagnostic	349	60.7	90.8
		Exemplary	53	9.2	100
		Total	575	100	
USA	Radiographer	Non-diagnostic	40	8.0	8.0
		Limited	123	24.6	32.6
		Diagnostic	244	48.8	81.4
		Exemplary	93	18.6	100
		Total	500	100	
	Radiologist	Non-diagnostic	23	3.3	3.3
		Limited	88	12.6	15.9
		Diagnostic	438	62.6	78.4
		Exemplary	151	21.6	100
		Total	700	100	

their radiologist counterparts and categorised radiographs in a lower category. In addition, less experienced radiographers tended to rate some image types (e.g. Chest PA) less harshly than those who were more experienced. While the difference in experience between radiologist and radiographer groups in this study is acknowledged, it is thought based on the above that experience-matched groups would yield similar results or even greater differences between radiologist and radiographer RadLex ratings.

Correlation between overall RadLex ratings and individual image features

No significant correlation between the RadLex scores and the collimation, the positioning or anatomical criteria was found. However, across all radiograph types, a strong overall correlation (Table 6) was found between the RadLex classification and the evaluation of the delivered dose for the radiographers. In comparison, a similar but weaker correlation is present within the group of the radiologists. Therefore, it appears that the judgement of the 'delivered dose' has a correlation with the overall judgement of clinical usability.

Factors leading to increased likelihood of image rejection

To further investigate the correlation between the RadLex ratings and evaluation of the delivered dose, an ordinal regression model (pooled across countries) was used. In contrast to the Spearman correlations, which allows identification of overall correlations, an ordinal regression estimates the chance of a radiograph being classified in a lower Radlex category based on the evaluation of the delivered dose (e.g. whether perception of delivered dose makes a radiograph more likely to be considered 'limited' than 'diagnostic', or 'diagnostic' than 'exemplary').

For both the radiographers and the radiologists (Table 7), a lower rating of the delivered dose increased the chance of a radiograph being assigned a lower RadLex category, implying that observer's impressions of delivered dose influence the overall judgement of acceptability of a radiograph. This effect was greater for

Table 6. Correlations between RadLex and the visual judgement of the detector dose with the confidence interval based on a bootstrap (1000 samples, 95% confidence). All correlations are significant at the 0.01 level (two-tailed).

Country	Observer	Overall	Chest	C-Spine	Hip
Belgium	Radiographer	0.648 [0.605–0.688]	0.606 [0.531–0.674]	0.53 [0.349–0.611]	0.808 [0.762–0.845]
	Radiologist	0.505 [0.441–0.571]	0.375 [0.231–0.506]	0.503 [0.384–0.61]	0.646 [0.599–0.724]
Ireland	Radiographer	0.587 [0.548–0.625]	0.427 [0.335–0.513]	0.608 [0.537–0.675]	0.649 [0.581–0.706]
	Radiologist	0.55 [0.48–0.609]	0.441 [0.318–0.549]	0.624 [0.514–0.719]	0.584 [0.475–0.669]
USA	Radiographer	0.757 [0.717 - 0.79]	0.707 [0.631 -0.768]	0.761 [0.695 - 0.816]	
	Radiologist	0.53 [0.469 - 0.587]	0.559 [0.49 -0.621]	0.519 [0.41 - 0.61]	

Table 7. Output of the ordinal logistic model for the visual dose estimation (QC3). The model calculates the chance (estimate) of a radiograph being classified in a lower (negative estimate) or higher (positive estimate) RadLex category. For example, a radiograph with a satisfactory dose for a radiographer has a higher change of being classified in a lower category (estimate −0.335) in comparison to the classification by a radiologist (estimate −0.286).

	Radiographer			Radiologist		
	Estimate	Std. Error	Sig.	Estimate	Std. Error	Sig.
Delivered dose – Unacceptable	−0.139	0.135	0.303	0.031	0.211	0.881
Delivered dose – Very poor	−0.421	0.085	0.000	−0.183	0.109	0.093
Delivered dose – Satisfactory	−0.335	0.077	0.000	−0.286	0.085	0.001
Delivered dose – Very good	−0.137	0.073	0.060	−0.311	0.072	0.000
Delivered dose – Excellent	0.117	0.096	0.223	0a		

radiographers than radiologists, with a higher change for a radiograph to be classified in a lower category when a radiographer estimates a similar delivered dose compared to a radiologist.

Discussion

Previous studies have considered various potential influences on the decision-making process by radiographers when assessing image quality. Some outline the importance of a global assessment of the radiograph (incl. patient), such as Prime and Le Masurier,[11] Lundvall et al[10] and Larsson et al.[6,9] were others focus on the use of technical elements.[1] The aim of this study was to compare, between radiographers and radiologists, the use of a general image quality assessment in comparison to the approach with anatomical structures. Although data were collected in different countries, this paper will focus on the difference between professions and not between countries. Nevertheless, the similar findings in each country make the findings generalisable and the cultural difference an interesting area for further research.

Differences in RadLex rating

The RadLex categories allowed the observers to classify each radiograph into a certain category of clinical acceptability. This study demonstrated a significant difference between radiologists and radiographers when allocating a radiograph to a certain category. The Mann–Whitney U test indicated the significant differences ($P < 0.01$) in ranking for all three image types. Even more important, the radiographers were stricter in their assessment of radiographs compared to radiologists, which echoes the findings of multiple previous studies.[1,9,34] This entails radiographers potentially discarding radiographs with sufficient diagnostic value for a radiologist and performing an unnecessary repeat of the examination.

Anatomical structures

The difference in assessment based on the ratings of overall clinical acceptability, as measured by the RadLex categories, was more pronounced than the differentiation based on the visual grading analyses (VGA) of the visibility of individual anatomical structures, which only demonstrated significant differences for certain very specific structures. In contrast to the RadLex categories, no overall difference was found between radiologists and radiographers for all structures of a radiograph. The visual grading did not reflect the difference in the assessment of clinical usability between radiographers and radiologists. These findings imply that the decision to clinical accept a radiograph for diagnosis is not entirely or even primarily judged based on anatomical structures.

This might be problematic with, firstly, anatomical structures often being used in parameter optimisation research studies to assess the image quality. The main argument for this approach or visual grading is the simplification of the problem based on the ability of a radiologist to assess the reproduction of the anatomy required for making a diagnosis.[24,35] Secondly, the findings have implications on the approaches for auditing a radiographer's work. Approaches, as presented by Tesselaar,[36] recommend the use of visual grading. Based on the results of the current study, it might be questioned if the use of anatomical criteria sufficiently reflects the clinical usability of a radiograph. Thirdly, frameworks, such as the EU criteria that describe image quality based on structures,[19] may be questionable as basis for the assessment of clinical acceptability.[37]

Non-anatomical or technical factors

The judgement of the detector dose (i.e. detector exposure, appearance of noise) has the strongest Spearman correlation with the RadLex classification for both professions. The ordinal regression demonstrated that the estimation of the detector exposure predicts the classification into the RadLex categories. Meaning that, compared to factors such as positioning, a radiograph has a higher change of being classified as limited or non-diagnostic if the observer finds the detector exposure inappropriate. Additionally, the effect is greater for radiographers than radiologists. These findings entail that, based on the general assessment of the detector exposure, a radiograph will be discarded sooner by a radiographer than a radiologist. The effect of the detector dose, or the quantum noise, is in strong contrast with the review of Waaler and Hofmann.[5] Waaler and Hofmann conclude that the rejects in digital systems are mainly related to the radiographers' skills (positioning) and not exposure related (while they noted that, in the past with film screen systems, exposure-related rejects were the primary cause for a reject). The difference between radiographers and radiologists raises questions around radiographers unnecessarily discarding images, challenges to align both professions in their assessment and the emphasis on the evaluation of the detector exposure in training.

One might question the low impact of positioning due to the data set containing an insufficient amount of positioning errors. On the one hand, and despite being collected from clinical PACS systems, the data demonstrated classification by participants into all categories (exemplary – non-diagnostic), illustrating the use of the full scale. On the other hand, the data set is rather small; hence, there is a possible variance for a certain criterion, which may be considered a limitation of the study.

The decision on the clinical usability was made in the context of a single clinical question for consistency, and it could be argued that the decision on the clinical usability of the radiograph was influenced by the question. The chosen clinical question (trauma) was however very broad, such that images might include subtle (i.e. degeneration, pneumothorax) and massive pathologies (i.e. collapse, fractures, dislocations). The difference between radiographers and radiologist might be caused by how confident they are in answering the clinical question based on the information in the radiograph. In line with the review by Waaler and Hofmann,[5] the radiographer might be doubtful if the radiograph is sufficiently answering the clinical question for the radiologist due to the lack of discussion between them. Therefore, an interesting approach would be to evaluate the decision-making of reporting radiographers while they progress in their training. The effect of reporting findings in addition to judging the clinical usability might enlighten the gap between both professions in the judgement of clinical acceptability.

The selection of display technologies can be perceived as both a strength and limitation of this research. The use of different displays adds a factor which, on the one hand, might complicate the comparison of the two professions by providing radiologists with higher specification displays and therefore potentially altering the presentation of pertinent image features. On the other hand, the same bias is present in the daily clinical practice, where radiographers usually do work with a lower class of display than radiologists. The replication of clinical practice strengthens the clinical relevance of the study. A further potential limitation is the absence of varied clinical indications, which might have allowed more realism in the scenario.

Conclusion

The study demonstrates differences between radiologists and radiographers in the judgement of diagnostic usability, indicating that radiographers are more critical and classify more radiographs as non-diagnostic or limited for a specific question. Radiographers and radiologists both demonstrated consistency and agreement with their professions. In contrast to prevailing views, the evaluation to anatomical structures does not explain the difference between both professions when assessing the clinical acceptability. Radiographs have a higher chance of being considered less clinically acceptable if the observer (of either profession) deems the delivered dose (i.e. detector exposure) to be inappropriate. This correlation is stronger for radiographers than radiologists. This may be an important but previously unrecognised factor in understanding how radiographers and radiologists judge diagnostic image quality for which attention in training is needed.

Acknowledgement

Open access funding provided by IReL.

References

1. Mount J. Reject analysis: A comparison of radiographer and radiologist perceptions of image quality. *Radiography [Internet]* 2016; **22**: e112–7.
2. Kjelle E, Chilanga C. The assessment of image quality and diagnostic value in X-ray images: a survey on radiographers' reasons for rejecting images. Insights. *Imaging* 2022; **13**(1): 13–36.
3. Shet N, Chen J, Siegel EL. Continuing challenges in defining image quality. *Pediatr Radiol* 2011; **41**: 582–7.
4. Nodine CF, Mello-Thoms C. The nature of expertise in radiology. In: Beutel J, Kundel HL, Van Metter R (eds). Handbook of Medical Imaging. Bellingham. SPIE Press, Washington, 2000 (Handbook of Medical Imaging; vol. 1).
5. Waaler D, Hofmann B. Image rejects/retakes - Radiographic challenges. *Radiat Prot Dosimetry* 2010; **139**: 375–9.
6. Larsson W, Aspelin P, Bergquist M, et al. The effects of PACS on radiographer's work practice. *Radiography* 2007; **13**: 235–40.
7. Uffmann M, Schaefer-Prokop C. Digital radiography: The balance between image quality and required radiation dose. *Eur J Radiol* 2009; **72**: 202–8.
8. Hofmann BB, Rosanowsky TB, Jensen C, Wah KHC. Image rejects in general direct digital radiography. *Acta Radiol Open* 2015; **4**: 205846011560433.
9. Larsson W, Aspelin P, Lundberg N. Learning strategies in the planning and evaluation phase of image production. *Radiography* 2013; **19**: 347–52.
10. Lundvall LLL, Dahlgren MA, Wirell S. Professionals' experiences of imaging in the radiography process A phenomenological approach. *Radiography* 2014; **20**: 48–52.
11. Prime NJJ, le Masurier SBB. Defining how we think: an investigation of decision making processes in diagnostic radiographers using the think aloud technique. *Radiography* 2000; **6**: 169–78.
12. Hayes BK, Adams RD, Higgs J. Parallels between the proces of clinical reasoning and categorization. Clinical reasoning in the health professions. Butterworth Heinemann, Oxford, 1995; 35–47.
13. Higgs J, Jones M, Higgs J. Clinical reasoning. Clinical reasoning in the health professions. Butterwort Heinemann, Oxford, 1995; 3–23.
14. Patel VL, Arocha JF, Higgs J. Methods in the study of clinical reasoning. Clinical reasoning in the health professions. Butterworth Heinemann, Oxford, 1995; 35–47.
15. Sedlmeier PRH. From associations to intuitive judgement and decision making: Implicit learning from experience. In: Betsch T, Haberstroh SHS (eds). The Routines of Decision Making. Psychology Press, New York, 2005; 67–83.
16. Kjelle E, Schanche AK, Hafskjold L. To keep or reject, that is the question - A survey on radiologists and radiographers' assessments of plain radiography images. *Radiography* 2021; **27**(1): 115–9.
17. Dunn MA, Rogers AT. X-ray film reject analysis as a quality indicator. *Radiography* 1998; **4**: 29–31.
18. RSNA. RadLex Radiology Lexicon - Playbook [Internet]. RSNA, Oak Brook, IL, 2018 Available from: https://www.rsna.org/en/practice-tools/data-tools-and-standards/radlex-radiology-lexicon.
19. Carmichael JHE, Maccia C, Moores BMM, et al. European Guidelines on Quality Criteria for Diagnostic Radiographic Images. ECSC-EC-EAEC; 1996.
20. McNulty JP, Rainford L, Bezzina P, et al. A picture of radiography education across Europe. *Radiography* 2016; **22**(1): 5–11.
21. England A, Geers-van Gemeren S, Henner A, et al. Clinical radiography education across Europe. *Radiography* 2017; **1**: S7–15.
22. EFRS (ed). European Qualifications Framework (EQF) Level 6 Benchmarking Document: Radiographers, 2nd edn. EFRS, Utrecht, 2018.
23. Börjesson S, Håkansson M, Båth M, et al. A software tool for increased efficiency in observer performance studies in radiology. *Radiat Prot Dosimetry* 2005; **114**: 45–52.
24. Båth M, Månsson LG. Visual grading characteristics (VGC) analysis: A non-parametric rank-invariant statistical method for image quality evaluation. *Brit J Radiol* 2007; **80**: 169–76.
25. Håkansson M, Svensson S, Zachrisson S, Svalkvist A, Båth M, Månsson LG. ViewDEX: an efficient and easy-to-use software for observer performance studies. *Radiat Prot Dosimetry* 2010; **139**: 42–51.
26. Båth M, Svalkvist A, Svensson S, Håkansson M, Båth M, Månsson LG. Viewdex: A status report. *Radiat Prot Dosimetry* 2016; **169**: 38–45.
27. Ivarsson J, Rystedt H, Asplund S, Johnsson ÅA, Båth M. The application of improved, structured and interactive group learning methods in diagnostic radiology. *Radiat Prot Dosimetry* 2016; **169**: 416–21.

28. Samei E, Badano A, Chakraborty D, et al. Assessment of display performance for medical imaging systems. *Med Phys* 2005; **32**(4): 1205–25.
29. Norweck JT, Seibert JA, Andriole KP, et al. ACR-AAPM-SIIM technical standard for electronic practice of medical imaging. *J Digit Imaging* 2013; **26**: 38–52.
30. Krupinski EA, Williams MB, Andriole KP, et al. Digital radiography image quality: Image processing and display. *J Am Coll Radiol* 2007; **4**: 389–400.
31. Williams MB, Krupinski EA, Strauss KJ, et al. Digital radiography image quality: Image acquisition. *J Am Coll Radiol* 2007; **4**: 371–88.
32. Kirkwood BR, Sterne JAC. Essential Medical Statistics. Blackwell Science, Malden, MA, 2003.
33. Field A. Discovering Statistics Using IBM SPSS Statistics, 4th edn. SAGE, London, UK, 2012.
34. Atkinson S, Neep M, Starkey D. Reject rate analysis in digital radiography: an Australian emergency imaging department case study. *J Med Radiat Sci [Internet]* 2019; **67**: 72–9.
35. Smedby Ö, Fredrikson M, de Geer J, et al. Quantifying the potential for dose reduction with visual grading regression. *Br J Radiol* 2013; **86**: 20110784.
36. Tesselaar E, Dahlström N, Sandborg M. Clinical audit of image quality in radiology using visual grading characteristics analysis. *Radiat Prot Dosimetry* 2015; **169**: 340–6.
37. Commission of the European Communities (CEC), Commission U. European Guidelines on Quality Criteria for Diagnostic Radiographic Images:(EUR 16260 EN) 1996. 71 p. Available from: https://publications.europa.eu/en/publication-detail/-/publication/d59ccc60-97ed-4ce8-b396-3d2d42b284be/language-en.