ORIGINAL ARTICLE

Interobserver and intra-observer agreement among radiographers in assessing the quality of chest radiograph

Dwi Rochmayanti, MEng[1,2], Kusworo Adi, Dr.[2], Catur Edi Widodo, Dr.[2], Akhmad Haris Sulistiyadi, MKes[1], Siti Daryati, MSc[1]

[1]Study Program of Information System, Postgraduate School, Diponegoro University, Semarang, Indonesia, [2]Department of Radiodiagnostic and Radiotherapy Technics, Poltekkes Kemenkes Semarang, Semarang, Indonesia

ABSTRACT

Introduction: Chest radiography is a commonly performed examination for screening, clinical assessment, and pre-intervention evaluation. Image quality plays an important role in supporting diagnostic interpretation. After image acquisition, radiographers decide whether the image should be accepted or repeated. Variation in these decisions may affect patient radiation exposure and radiology workflow. This study aimed to evaluate variation and agreement among radiographers in accepting or rejecting chest radiographs.

Materials and Methods: This cross-sectional study involved 91 certified radiographers in Indonesia. Thirty de-identified chest radiographs representing technical-quality categories were selected by an expert panel, and 12 duplicate images were added to assess intra-observer consistency. The resulting 42 images were presented in random order. The acceptance and rejection criteria were classified into 8 factors: accepted, artifact, blurring, mispositioning, incorrect body part, body part cut-off, exposure failure, and under/over exposure. Data were analysed using Cohen's kappa for duplicated image pairs and Fleiss' kappa for inter-observer agreement.

Results: There were radiographer's perception variations in accepting or rejecting X-ray chest images. Intra-observer agreement for duplicated image pairs ranged from κ=0.027 to κ=0.727, with a simple mean Cohen's kappa of κ=0.525, indicating moderate agreement. Overall inter-observer agreement across the 42 images was fair (Fleiss' κ=0.337). Category-specific Fleiss' kappa values were highest for artifact-related rejection (κ=0.474) and image acceptance (κ=0.436), and lowest for blurring (κ=0.068), exposure failure (κ=0.076), incorrect body part (κ=0.107), and incorrect positioning (κ=0.173).

Conclusion: Radiographers showed moderate intra-observer consistency but only fair inter-observer agreement when evaluating chest radiograph quality. The findings demonstrate variation in accepting or rejecting chest X-ray image. Collaboration between radiographers and radiologists is recommended to align evaluation standards.

KEYWORDS:
Chest radiography; observer variation; observer agreement; radiography quality assurance

INTRODUCTION

Chest radiography is commonly used to evaluate patient's condition in general, screening and pre-intervention assessments such as surgeries and other medical procedures.[1-2] High quality image that meets technical quality standards is required to provide accurate diagnostic value.[3] Patient movement, mispositioning, and image receptor-related artifacts can result in image rejection.[4] Repeated radiographic examinations lead to increased radiation exposure and additional economic costs.[2,5] To enhance image quality, radiographers have to adopt a proactive role, not only during image acquisition but throughout the entire imaging process.[6] Radiographers control the image acquisition process and the factors that determine final image quality. Therefore, radiographer-based strategies are essential to raise quality awareness, improve image outcomes, and reduce rejection rates.[7]

Diagnostic inaccuracies in radiology are estimated to occur in approximately 30% of image interpretations. Studies have shown that suboptimal image quality is correlated with false-positive and false-negative interpretations.[6] Quality assurance becomes a priority to improve client satisfaction and the sustainability of healthcare organizations8. An important aspect of radiology quality management is a systematic approach to ensuring high-quality imaging outcomes, emphasizing processes, protocols, and performance improvements. This approach includes minimizing image rejection and repetition. Radiologists and radiographers play a significant role in minimizing rejected image or repeated examination. However, subjectivity determines the justification for rejection or repetition. These decisions depend to the individual perception, leading to potential bias and inefficiencies due to inconsistent implementation.[2] As reported, image quality assessment is subjective and significantly varies among radiographers.[3] This study aimed to identify the perceptions profile for accepting or rejecting Chest X-ray images among radiographers.

MATERIALS AND METHODS

Study Design and Data Collection

This cross-sectional observer-agreement study involved 91 certified radiographers from 53 accredited hospitals and 2 clinics in Indonesia that applied radiology quality-control programme. Fifty-five radiographers (60.4%) were from

This article was accepted: 09 July 2026
Corresponding Author: Dwi Rochmayanti
Email: dwirochmayanti@poltekkes-smg.ac.id

Table I: Participant's characteristic

No	Characteristic	Specification	Number
1	Hospital Accreditation Level	Full	77
		Main	9
		Middle	4
		Basic	1
2	Working Experience	≤5 years	13
		> 5 years	78
3	Certified	National Certification	91
4	Involved in QA/QC	Radiography	91

QA= Quality Assurance; QC= Quality Control

Table II: The acceptance and rejection criteria

No	Decision	Rejection Factor	Definition	Number of Images
1	Accepted	None	Anatomy of interest is visualized with correct position, adequate brightness and contrast within collimation area. The image is free from artifact and un-sharpness.	5
2	Rejected	Incorrect body-part	The body part being visualized is different to the requested body part.	1
3	Rejected	Body-part cut off	Anatomy of interest outside the collimation.	9
4	Rejected	Incorrect positioning	Incorrect patient position, anatomical part position, and under/ over rotation.	13
5	Rejected	Blurring	Un-sharpness due to patient, object movement, and/ or equipment.	1
6	Rejected	Artifact	Visualization undesirable objects or body parts, grid lines, and/ or equipment related defects.	7
7	Rejected	Under/over exposure	Excessive noise due to underexposure or excessive density due to over exposure that reduced contrast.	5
8	Rejected	Exposure failure	Blank or double exposed image.	1

Table III: Agreement classification

Category	Range
Poor	<0.00
Slight	0.00–0.20
Fair0.21–0.40	
Moderate	0.41–0.60
Substantial	0.61–0.80
Almost Perfect	>0.80

Note: According to Landis and Koch.[20]

Table IV: Intra-observer consistency for duplicated image pairs

Image	Cohen's Kappa (κ)	Category
1 – 2	0.355	Fair
3 – 4	0.592	Moderate
7 – 8	0.582	Moderate
9 – 10	0.637	Substantial
12 – 13	0.027	Slight
17 – 18	0.494	Moderate
22 – 23	0.498	Moderate
24 – 25	0.653	Substantial
29 – 30	0.518	Moderate
31 – 32	0.489	Moderate
36 – 37	0.727	Substantial
38 – 39	0.727	Substantial
Average	0,525	Moderate

Note: The average is a simple unweighted descriptive mean of the 12 duplicated image-pair kappa values and should not be interpreted as a pooled inferential estimate.

Table V: Category-specific inter-observer agreement according to accept/reject decision

No	Decision	Factor	Fleiss' Kappa (κ)	Category
1	Accept	None	0.436	Moderate
2	Reject	Incorrect body-part	0.107	Slight
3	Reject	Body-part cut off	0.363	Fair
4	Reject	Incorrect positioning	0.173	Slight
5	Reject	Blurring	0.068	Slight
6	Reject	Artifact	0.474	Moderate
7	Reject	Under/over exposure	0.204	Fair
8	Reject	Exposure failure	0.076	Slight

Note: Category-specific Fleiss' kappa values were calculated as separate one-versus-all binary classifications for each response category.

Central Java Province and 36 (39.6%) were from other regions. Seventy-eight observers (85.7%) had more than five years of work experiences, whereas 13 (14.3%) had five years of work experiences or less. All participants were certified radiographers; however, the survey did not collect exact years of experience, hospital type or level, prior formal training in reject/repeat analysis, or vendor-specific digital radiography experience. The participant's characteristic is described in Table I.

Thirty chest radiographs were purposively selected from a de-identified institutional digital radiography archive. The images were selected by an expert panel comprising two consultant radiologists, three senior quality-control radiographers, and three medical physicists to represent the predefined acceptance and rejection categories listed in Table II. Patient identifiers were removed before the survey. The expert panel assigned a reference technical-quality label to each image for dataset construction, but these labels were not used as a gold standard for accuracy analysis. Disagreements were resolved through consensus discussion. Pathological findings, when present, were not used as selection criteria or outcome variables; observers were instructed to evaluate technical image quality and reject/repeat factors rather than diagnostic findings. Twelve images were duplicated and mixed with the 30 initial images, resulting in 42 images presented to observers in random order. The acceptance and rejection criteria were classified into eight factors, as described in Table II.

Data Analysis

The validity of the survey instrument was assessed by the expert panel, yielding a validity index of 0.98. Cohen's kappa was used to assess intra-observer consistency for each duplicated image pair. Fleiss' kappa was used to assess overall inter-observer agreement across all observers and images. Category-specific Fleiss' kappa values were calculated as separate one-versus-all binary classifications for each response category. Agreement strength was interpreted according to the Landis and Koch classification.[20] The average Cohen's kappa reported in Table III is a simple unweighted descriptive mean across duplicated image pairs and was not treated as a pooled inferential estimate.

RESULTS

Intra-observer agreement was assessed using Cohen's kappa for the 12 duplicated image pairs presented to the same observers. The kappa values ranged from 0.027 to 0.727, indicating varying levels of consistency across image pairs. Most duplicated image pairs showed moderate agreement, while four pairs showed substantial agreement, particularly image pairs 24-25, 9-10, 36-37, and 38-39. The lowest agreement was observed for image pair 12-13 (κ=0.027), indicating slight agreement. The simple unweighted mean Cohen's kappa across the 12 pairs was κ=0.525, which was interpreted descriptively as moderate agreement.

Overall inter-observer agreement across all 42 images was assessed using Fleiss' kappa. The overall value was κ=0.337, indicating fair agreement among the 91 observers. This result reflects agreement in image acceptance or rejection classifications, not the diagnostic accuracy or correctness of the decisions.

Category-specific inter-observer agreement varied according to the response category. The highest Fleiss' kappa values were found for artifact-related rejection (κ=0.474) and image acceptance (κ=0.436), both indicating moderate agreement. Lower agreement was observed for body-part cut-off (κ=0.363) and under/over exposure (κ=0.204), both categorized as fair agreement. The lowest agreement was observed for blurring (κ=0.068), exposure failure (κ=0.076), incorrect body part (κ=0.107), and incorrect positioning (κ=0.173), all categorized as slight agreement.

DISCUSSION

Chest radiography is a routine examination, but the decision to accept or repeat an image has direct implications for patient exposure, workflow, and quality assurance. Reject or repeat analysis is therefore an important component of radiology quality management because it identifies not only the number of rejected or repeated images but also the technical factors contributing to rejection.

Radiographers are key personnel in radiology quality assurance. After image acquisition, they visually evaluate whether the examination should be accepted or repeated. Accepting a suboptimal image may reduce diagnostic value, while unnecessarily rejecting an acceptable image may expose the patient to avoidable repeat radiation.

Interpretation of Consistency

The numerical findings show that radiographers were more consistent with their own repeated assessments than with one another. The simple mean Cohen's kappa for duplicated pairs was moderate (κ=0.525), whereas the overall Fleiss' kappa across all images was fair (κ=0.337). These findings indicate variability in technical-quality judgement and agreement; they do not establish whether individual decisions were correct or incorrect against a gold standard. The category-specific results provide further practical insight. Agreement was higher for artifact-related rejection (κ=0.474) and image acceptance (κ=0.436), but substantially lower for blurring (κ=0.068), exposure failure (κ=0.076), incorrect body part (κ=0.107), and incorrect positioning (κ=0.173). This pattern suggests that more visually obvious findings may be recognized more consistently, whereas positioning, motion, and exposure-related decisions may require clearer operational criteria.

From a clinical quality perspective, inconsistent acceptance or rejection decisions can affect both patient safety and operational efficiency. Unnecessary rejection may lead to avoidable repeat exposure, whereas acceptance of technically suboptimal images may reduce the value of the radiographic examination. Therefore, improving the reliability of radiographic image-quality evaluation remains a practical priority for radiology departments.

Link to Practice

The factors with the lowest agreement in this study--incorrect body part, incorrect positioning, blurring, and exposure failure--identify priority topics for radiographer training and quality assurance activities. Radiography education and continuing professional development programme may benefit from structured image-evaluation exercises, standardized reject/repeat criteria, peer-review sessions, and feedback mechanisms. These activities could reduce interpretive variability, particularly for positioning accuracy, motion blur, and exposure adequacy.

Hospital radiology departments should also consider structured reject/repeat analysis programme. Regular review of rejected images, group discussion, and consensus-building sessions may help align radiographers' perceptions and improve consistency in quality assessment. Standardized and field-relevant rejection categories are essential because users should be able to select clearly defined, actionable, and easily understood reasons for image rejection.[9] Department-wide rejection analysis can support process improvement and help reduce patient radiation doses.[10]

Joint discussions between radiographers and radiologists may also reduce subjectivity in borderline cases. Previous work has shown that radiologists and radiographers may differ in their decisions to keep or reject plain radiographic images, particularly for positioning errors.[11] Consensus discussion can therefore help harmonize acceptance parameters and improve confidence in reject/repeat decisions.[12-13] Rejected or repeated images should be systematically documented, analyzed for underlying causes, and followed by corrective action. The AAPM Task Group Report 305 provides guidance for standardizing vendor-neutral rejection categories, but each department should adapt these categories to its clinical workflow and equipment context.[9]

LIMITATIONS

Several limitations should be considered when interpreting the findings. First, the study did not employ a definitive gold standard for image acceptance or rejection. Although the images were selected and labelled by an expert panel consisting of radiologists, senior radiographers, and medical physicists, observer judgments were compared with each other rather than with an absolute reference standard. Consequently, the study evaluates observer agreement and variation, not decision accuracy. Second, the participant questionnaire recorded region and broad work-experience category but did not collect exact years of experience, hospital type or level, prior formal training in reject/repeat analysis, or experience with specific digital radiography systems. These unmeasured characteristics may have contributed to variation and should be captured in future studies.

Third, although the study involved radiographers from multiple regions of Indonesia, most participants were from Central Java Province. The findings may therefore not fully represent all clinical settings, equipment standards, or training backgrounds across Indonesia. Fourth, viewing conditions and display devices were not standardized across participants. Differences in monitor resolution, ambient lighting, or image-display settings may have influenced observers' perceptions of image quality. Despite these limitations, the study provides useful evidence of variation in radiographers' agreement when evaluating chest radiograph quality in the Indonesian clinical context.

Future Directions

The observed variability highlights an opportunity to develop automated image-quality assessment systems based on artificial intelligence (AI) and deep learning technologies. Such systems may support, rather than replace, radiographers' technical evaluation. AI-based image-quality assessment systems could assist radiographers by identifying common rejection factors such as positioning errors, exposure inadequacy, motion blur, and artifacts. These systems could provide real-time feedback during image acquisition and may help reduce repeat examinations and patient radiation exposure. AI-supported reject/repeat analysis has been reported to enhance technologist performance and image auditing.[12] Neural network approaches can also predict anatomical cut-off and motion-related errors in thoracic images, supporting radiographer and radiologist decision-making.[14-15]

Future AI research should use large annotated datasets that combine radiographers' perception profiles with expert-validated ground truth labels. This approach would help models learn from real-world clinical decisions while remaining anchored to objective technical-quality standards. Information systems can support radiology departments in maintaining quality standards and monitoring rejection patterns.[17] Imaging protocols should also be optimized during acquisition and post-processing to prevent repeat examinations and reduce unnecessary radiation exposure.[18]

Future studies should use multi-centre datasets with larger and more diverse samples to improve generalisability. Longitudinal studies are also needed to examine how radiographers' decisions change after structured training or quality-improvement interventions. A rigorous quality assurance programme remains fundamental to image quality. Routine acceptance testing, periodic inspections, maintenance evaluations, and annual assessments are needed to support patient safety and diagnostic value.[19] Combining standardized training, improved quality-assurance protocols, and AI-assisted image evaluation may lead to more consistent radiographic quality assessment and safer imaging practice.

CONCLUSION

This study found moderate intra-observer consistency and fair inter-observer agreement among radiographers when assessing chest radiograph image quality. Agreement was higher for artifact-related rejection and image acceptance, but lower for blurring, exposure failure, incorrect body part, and incorrect positioning. The findings show variation in radiographers' agreement rather than decision accuracy against a gold standard. Collaboration between radiographers and radiologists, standardized reject/repeat criteria, structured training, and AI-assisted decision-support tools may help improve consistency in image-quality assessment.

ACKNOWLEDGMENTS

We acknowledge the multi-years research fund, granted by Poltekkes Kemenkes Semarang in 2023-2024.

CONFLICTS OF INTEREST

There are no conflicts of interest to declare.

ETHICAL APPROVAL

This study was approved by Health Research Ethics Committee of Poltekkes Kemenkes Semarang. Approval Letter No. 0103/EA/KEPK/2024.

REFERENCES

1. Wambersie A, DeLuca PM, Seltzer SM, editors. Image quality in chest radiography. ICRU Report 70. Bethesda, MD: International Commission on Radiation Units and Measurements; 2003.
2. Rochmayanti D, Adi K, Widodo CE. Repeat analysis program as a quality assurance system for radiology management: causal repeat and challenges. E3S Web Conf. 2023; 448.
3. 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): 1-6.
4. Jones AK, Heintz P, Geiser W, Goldman L, Jerjian K, Martin M, et al. Ongoing quality control in digital radiography: Report of AAPM Imaging Physics Committee Task Group 151. Med Phys. 2015; 42(11): 6658-70.
5. Mesfin Z, Elias K, Melkamu B. Analysis and economic implication of X-ray film reject in Diagnostic Radiology Department of Jimma University Specialized Hospital, Southwest Ethiopia. Ethiop J Health Sci. 2017; 27(4): 421-6.
6. Reiner BI. Hidden costs of poor image quality: a radiologist's perspective. J Am Coll Radiol. 2014; 11(10): 974-8.
7. Tonks A, Jimenez Y, Gray F, Ekpo E. A stake in the game: can radiographer image interpretation improve X-ray quality? A scoping review. Radiography (Lond). 2024; 30(2): 641-50.
8. Papp J. Quality management in the imaging sciences. 6th ed. St Louis: Elsevier; 2019.
9. Little KJ, Reiser I, Apgar BK, Dalal P, Dave JK, Fisher R, et al. AAPM Task Group Report 305: guidance for standardization of vendor-neutral reject analysis in radiography. J Appl Clin Med Phys. 2023; 24(5): e13985.
10. DeMaio DN, Herrmann T, Noble LB, Orth D, Peterson P, Young J, et al. Best practices in digital radiography. Radiol Technol. 2019; 91(2): 198-201.
11. 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 (Lond). 2021; 27(1): 115-9.
12. Hardy M, Harvey H. Artificial intelligence in diagnostic imaging: impact on the radiography profession. Br J Radiol. 2020; 93(1108): 20190840.
13. Varghese AP, Naik S, Andrabi ASH, Luharia A, Tivaskar S. Enhancing radiological diagnosis: a comprehensive review of image quality assessment and optimization strategies. Cureus. 2024; 16(6); 1-10
14. Whaley JS, Pressman BD, Wilson JR, Bravo L, Sehnert WJ, Foos DH. Investigation of the variability in the assessment of digital chest X-ray image quality. J Digit Imaging. 2013; 26(2): 217-26.
15. Moradi M, Siegel E, Kashyap S, Karargyris A, Wu JT, Saboury B, et al. Artificial intelligence for point-of-care radiograph quality assessment. 2019; 128.
16. Maes F, Robben D, Vandermeulen D, Suetens P. The role of medical image computing and machine learning in healthcare. In: Ranschaert ER, Morozov S, Algra PR, editors. Artificial intelligence in medical imaging. Cham: Springer; 2019. p. 9-23.
17. Minnigh TR, Gallet J. Maintaining quality control using a radiological digital X-ray dashboard. J Digit Imaging. 2009; 22(1): 84-8.
18. Krupinski EA, Williams MB, Andriole K, Strauss KJ, Applegate K, Wyatt M, et al. Digital radiography image quality: image processing and display. J Am Coll Radiol. 2007; 4(6): 389-400.
19. Mahesh M. Essential role of a medical physicist in the radiology department. Radiographics. 2018; 38(5): 1665-71.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1): 159-74.