



Quality of reporting of studies on artificial intelligence in radiology: what is the current state of the field?

Ahmed Maiter^{1,2,3}

¹Sheffield Teaching Hospitals NHS Foundation Trust,
Department of Radiology, Sheffield, United Kingdom

²NIHR Sheffield Biomedical Research Centre,
Sheffield, United Kingdom

³University of Sheffield, School of Medicine and
Population Health, Sheffield, United Kingdom

High-quality research reporting is of paramount importance. Transparent descriptions of study design, methods, and limitations are essential for reproducibility and underpin trust in results. Furthermore, consistent reporting of methods and outcomes is necessary for meaningful comparisons across similar studies and effective evidence synthesis. In contrast, incomplete reporting limits the interpretability of studies, increases the risk of misinterpreting findings, and raises the likelihood of research waste.

Promoting, implementing, and maintaining high-quality reporting practices in the field of artificial intelligence (AI) in radiology can be challenging. Numerous factors contribute to model performance and require detailed descriptions, particularly the data used for model training and testing. Seemingly minor variations in data, such as scanner vendor or study participant age distribution, may result in dramatic shifts in performance, making transparent reporting essential.

Additionally, the ability to explain how AI systems function and reach decisions is valuable, yet often not straightforward. This so-called “black box” problem is exacerbated by the development of increasingly complex models and the proliferation of proprietary commercial AI devices. Moreover, AI studies in radiology encompass both clinical and technical dimensions, each requiring transparent reporting. This duality can be a hurdle for research teams whose expertise is predominantly rooted in one of these domains.

Over the past two decades, a variety of reporting guidelines have been developed for different types of medical research.^{1,2} These guidelines specify and standardize the components of a study that should be reported by manuscript authors. Their overarching goals are to improve the transparency and reproducibility of research, facilitate peer review, and support the comparison of findings across publications. Several of these guidelines are now well established and are required by journals to be completed and submitted alongside manuscripts.

The Checklist for Artificial Intelligence in Medical Imaging (CLAIM) was the first reporting guideline specifically developed for studies on AI in medical imaging. Initially published in 2020 and updated in 2024, CLAIM takes a broad approach, covering general information applicable to most studies in the field.³⁻⁵ Since its introduction, a growing number of systematic and other literature reviews have used CLAIM to evaluate publications, yielding insights into deficiencies in reporting quality in specific areas of AI in radiology.⁶ In addition, several other reporting guidelines focused on specific study designs in AI-based radiology research have also been developed.⁷

In this issue of Diagnostic and Interventional Radiology, Koçak et al.⁸ present an umbrella review of AI studies in radiology, conducting a comprehensive two-level assessment of adherence to CLAIM. Review articles published before August 2024 that evaluated studies using the original version of CLAIM were eligible for inclusion. Thirty-three review articles were assessed at the review level, encompassing a total of 1,458 studies. Of these, 421 studies were assessed individually at the study level, identified from 15 reviews, and CLAIM adherence was extracted at both levels as score and/or compliance values. Univariate and multivariate logistic regression analyses were performed to identify predictors of CLAIM adherence and critiques of CLAIM within the included reviews were also appraised.

Corresponding author: Ahmed Maiter

E-mail: a.maiter@sheffield.ac.uk

Received 04 April 2025; revision requested 08 April 2025;
accepted 13 April 2025.



Epub: 05.05.2025

Publication date:

DOI: 10.4274/dir.2025.253315

The results convey several important messages for the field of AI in radiology. First, there has been a clear dichotomy in how adherence to CLAIM has been measured in the literature, which poses challenges for the direct comparison of studies. Twenty-six reviews summarised CLAIM adherence as a score, indicating the total number of items reported out of all 42 in the checklist. In contrast, 18 reviews summarised adherence as compliance, referring to the proportion of applicable items that were reported. As Koçak et al.⁸ highlight, CLAIM is a broad reporting guideline, and not all items will be applicable to every study—for example, details about model parameter initialization are unlikely to apply to studies evaluating commercial devices in clinical settings. By accounting for item applicability, compliance is likely to be a more meaningful metric than score, and this work supports its adoption as the standard for measuring CLAIM compliance moving forward.

Second, Koçak et al.⁸ identified variability in the adherence of publications to CLAIM, reflecting inconsistencies in the quality of reporting across the field. The median score was 26 at both the review and study levels, whereas median compliance was 63% at the review level and 68% at the study level. In other words, around one-third of the CLAIM items were unreported in the included studies. Furthermore, the score was below 21 in 11% of studies, and compliance was below 50% in 10%, indicating that approximately one in ten studies reported only a minority of checklist items. Three variables were associated with higher CLAIM adherence: more recent year of publication, journal impact factor quartile, and specific radiology subfields. The link between impact factor quartiles and adherence suggests variable enforcement of CLAIM and reporting standards by journal editors and peer reviewers. Overall, the findings support greater use and application of CLAIM by researchers, journals, and peer reviewers.

Third, some items in CLAIM have been reported more frequently than others, suggesting systemic issues in the quality of reporting within the field. Eleven checklist items were reported in fewer than 50% of studies despite covering information that is crucial

for understanding the performance, generalizability, and scope of use of AI models. These frequently underreported items mostly relate to details about the data used or the methods and metrics of model evaluation. The inadequate reporting of certain items raises particularly serious concerns: a clear description of case inclusion and exclusion is necessary to identify selection biases in training and testing datasets; demographic and clinical characteristics of cases are essential for understanding model generalizability; measures of significance and uncertainty reflect the internal validity of model performance; and failure analysis provides insights into model limitations. Peer reviewers and editors may wish to pay particular attention to these 11 frequently underreported items when appraising manuscripts.

The comprehensive umbrella review by Koçak et al.⁸ highlights the variation in how CLAIM has been applied and the general variability in CLAIM adherence across studies on AI in radiology. Multiple shortcomings were identified, offering actionable insights for authors, editors, peer reviewers, and readers of publications in the field. There are, of course, limitations to the work, particularly the reliance on the quality and consistency of CLAIM evaluations in previous reviews and the inclusion of studies published before the original version of CLAIM. However, the two-level analysis, the consideration of both checklist score and compliance, and the identification of predictors of adherence together constitute a rigorous methodological approach.

Looking ahead, there is clear potential for implementing automated approaches to evaluating studies using CLAIM, which could help improve consistency across the field. Similar analyses could also be extended to other reporting guidelines relevant to AI in radiology. Lastly, it would be valuable to explore the relationship between CLAIM adherence and the downstream impact of studies, such as citation rates or clinical translation.

Footnotes

Funding

The author is funded by the National Institute for Health and Care Research (NIHR) Sheffield Biomedical Research Centre (NIHR203321). The views expressed are those of the author and not necessarily those of the NIHR or the Department of Health and Social Care.

Conflict of interest disclosure

The author is a member of the trainee editorial board for Radiology: Artificial Intelligence.

References

1. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. *Lancet*. 2008;371(9619):1149-1150. [\[Crossref\]](#)
2. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med*. 2010;8:24. [\[Crossref\]](#)
3. Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. 2020;2(2):e200029. [\[Crossref\]](#)
4. Tejani AS, Klontzas ME, Gatti AA, et al. Updating the checklist for artificial intelligence in medical imaging (CLAIM) for reporting AI research. *Nat Mach Intell*. 2023;5(9):950-951. [\[Crossref\]](#)
5. Tejani AS, Klontzas ME, Gatti AA, et al. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. *Radiol Artif Intell*. 2024;6(4):e240300. [\[Crossref\]](#)
6. Maiter A, Salehi M, Swift AJ, Alabed S. How should studies using AI be reported? lessons from a systematic review in cardiac MRI. *Front Radiol*. 2023;3:1112841. [\[Crossref\]](#)
7. Klontzas ME, Gatti AA, Tejani AS, Kahn CE Jr. AI reporting guidelines: how to select the best one for your research. *Radiol Artif Intell*. 2023;5(3):e230055. [\[Crossref\]](#)
8. Koçak B, Köse F, Keleş A, Şendur A, Meşe İ, Karagülle M. Adherence to the checklist for artificial intelligence in medical imaging (CLAIM): an umbrella review with a comprehensive two-level analysis. *Diagn Interv Radiol*. 2025. [\[Crossref\]](#)