Reproducibility = Safety: Establishing Transparent Validation as a Quality Metric for Radiology AI

Colin Wu[1], Ravi Patel[1], Edward Pettyjohn, MD[2], Shreyas Meruga[3], Zhuangwei Kang, PhD[4], Sarah Pettyjohn, MD[5]

[1] University of the Incarnate Word School of Osteopathic Medicine [2] University of Illinois College of Medicine Peoria (UICOMP) [3] UT Health San Antonio, Department of Molecular Medicine [4] Independent Researcher [5] Rutgers Robert Wood Johnson Medical School

Introduction

To define reproducibility in AI as a patient-safety requirement and to propose transparent validation and version traceability as quality metrics.

Methods

AI tools affect image interpretation, triage, and treatment, yet many receive little quality assurance after deployment. Undocumented updates, calibration drift, and unmonitored performance create safety risks similar to operating imaging equipment without quality control. Although professional groups call for transparency, reproducibility is still treated as a research preference rather than protection against preventable diagnostic error.

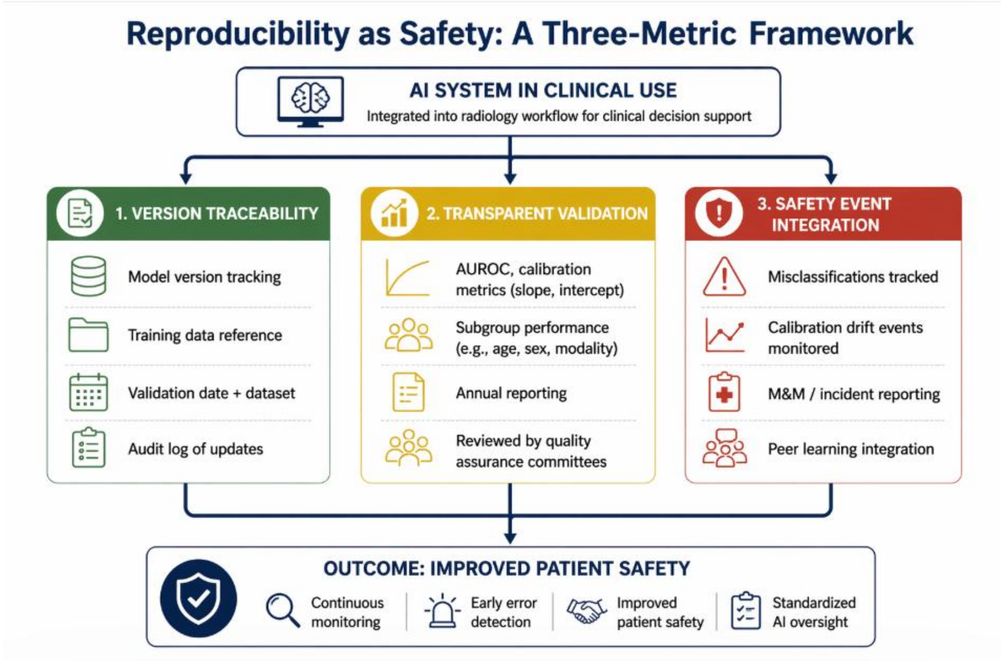
Results

We introduce a Reproducibility as Safety Framework that integrates AI oversight into existing safety structures through three auditable metrics:

1. Version Traceability: Each deployed algorithm records version identification, training data reference, and most recent validation date, functioning as an AI dose log that reveals silent drift.

2. Transparent Validation Reports: Annual summaries of AUROC, Brier score, calibration slope and intercept, and key subgroup performance are reviewed by quality committees similarly to radiation dose or physics reports.

3. Safety Event Integration: Algorithmic deviations, unexpected misclassifications, and calibration shifts are treated as reportable safety events within morbidity and mortality conferences, incident reporting, or peer learning.
These metrics can be supported by existing quality assurance infrastructure with minimal additional burden.



Conclusion

Embedding reproducibility within a safety framework reframes AI oversight as a patient-protection obligation and enables early detection of drift, fewer preventable discrepancies, and stronger clinician trust.
Reproducibility is a quality and safety standard.
Integrating transparent validation and version control into radiology programs helps ensure that AI tools evolve with the same rigor applied to imaging equipment and protocols.

References

1. Larson DB, Harvey H, Rubin DL, Irani N, Tse JR, Langlotz CP. Regulatory Frameworks for Development and Evaluation of Artificial Intelligence-Based Diagnostic Imaging Algorithms: Summary and Recommendations. J Am Coll Radiol. 2021 Mar;18(3 Pt A):413-424. doi: 10.1016/j.jacr.2020.09.060. Epub 2020 Oct 20. PMID: 33096088; PMCID: PMC7574690.