Deterministic Radiographic Scoring in Inflammatory Arthritis:

A Validator-Gated Architecture for Reproducible Endpoint Generation

Olga Goodman, MD
contact_us@rheumaview.com

Disclosure: Patent pending; related U.S. patent applications have been filed. No external funding was received for this work.

Abstract

Background: Radiographic scoring remains central to structural damage assessment in rheumatoid arthritis (RA), psoriatic arthritis (PsA), axial spondyloarthritis (axSpA), and related inflammatory conditions. Conventional visual scoring workflows are sensitive to inter-reader variability, calibration drift, and thresholding inconsistencies, limiting reproducibility in longitudinal monitoring and clinical trial endpoint generation—challenges that may be further exacerbated when AI-assisted tools operate as probabilistic 'black boxes' without explicit validation lineage.

Methods: We describe a deterministic, validator-gated radiographic analysis platform organized into two architectural layers: (1) a structured clinical rendering layer that produces standardized, anatomy-organized reports from validated descriptors, enforcing clinical purity through strict isolation from research-tier analytics; and (2) a research analytics layer deriving version-controlled endpoint measures from the same validated descriptor lineage. The platform enforces sequential validation gates addressing data integrity, descriptor completeness, projection adequacy, laterality verification and correction, symmetry normalization, and longitudinal eligibility when comparative scoring is requested. Schema-level safeguards prevent 'orphaned' quantitative statements, ensuring that longitudinal comparisons and composite integrations maintain structural and semantic completeness. All outputs include explicit operator and threshold version identifiers and audit-traceable provenance metadata.

Deterministic properties (by design): For identical validated descriptor inputs processed under the same operator and threshold versions, the system produces identical clinical renderings and endpoint fields. This architectural guarantee enables reproducible longitudinal scoring and version-stable exports suitable for multi-center research workflows and trial-oriented endpoint generation, subject to context-of-use validation for regulated submissions.

Conclusions: A validator-gated deterministic architecture offers a principled approach to reducing reader-dependent variability in radiographic scoring. By constraining downstream rendering and analytics to a controlled descriptor lineage with explicit versioning and audit traceability, the platform addresses key reproducibility challenges in current scoring methodologies and may provide a foundation for future regulatory-aligned endpoint generation in multi-center clinical trials, contingent on appropriate validation and qualification. Validation studies appropriate to the intended context of use remain required for any regulated endpoint application or clinical deployment.

Keywords: *radiographic scoring; inflammatory arthritis; deterministic analytics; structured reporting; modified Total Sharp Score (mTSS); mSASSS; reproducibility; structural damage assessment; clinical trial endpoints; validator-gated architecture*

1. Introduction

Radiographic assessment of structural damage is widely used for diagnosis support, prognosis, and monitoring in inflammatory arthritis. Scoring systems such as Sharp/van der Heijde and modified Total Sharp Score (mTSS) are commonly applied in clinical trials and research. However, conventional workflows rely on human visual judgment and manual aggregation, which can introduce inter-reader variability, thresholding disagreements (e.g., grade boundaries), and longitudinal inconsistencies.

Published reproducibility analyses report substantial variability across readers and timepoints in routine settings, even with training and calibration programs. These limitations can reduce sensitivity to change and complicate comparisons across sites and studies (see, e.g., multi-reader studies and reproducibility reviews in RA radiographic scoring).

2. Limitations of Current Radiographic Scoring Workflows

Key sources of variability include: (i) thresholding differences for small or subtle changes; (ii) inconsistent application of projection and laterality assumptions; (iii) variability in descriptor completeness and terminology; (iv) reader-state dependence and calibration drift; and (v) inconsistent longitudinal alignment (e.g., side mismatches, projection non-comparability, or inadequate baseline anchoring).

In clinical trials, these issues can produce discordant progression calls and increase the need for adjudication. In clinical practice, they can obscure true change versus measurement noise. A deterministic approach aims to address these failure modes by shifting the core scoring logic from subjective human selection to validated, fixed-order operators applied to a controlled descriptor lineage.

3. Methods: Deterministic Dual-Layer Architecture

The platform is organized into a validator-gated pipeline that enforces a fixed execution order. The pipeline is designed to prevent bypass or reordering of gates that establish eligibility for downstream interpretation and analytics.

3.1 Multi-Stage Validation Pipeline

Validation gates include, at minimum: data integrity checks; descriptor-completeness enforcement; projection adequacy verification; laterality and symmetry normalization; and longitudinal eligibility checks when comparative scoring is requested. Gate failures halt downstream analytics or constrain outputs to eligible subsets, preventing invalid or incomplete inputs from contaminating endpoint measures.

3.2 Structured Clinical Rendering Layer

The clinical rendering layer generates structured radiographic reports with deterministic section ordering and anatomy-organized findings. The output is intended to be consistent across users and reading environments, and to support downstream consumption by clinical systems and quality programs.

3.3 Research Analytics Layer and Endpoint Operators

The research layer applies deterministic operators to validated descriptors to produce reproducible endpoint measures. Operators are versioned (operator_version) to preserve traceability, and categorical bands/classes use versioned threshold sets (threshold_version) so that longitudinal comparisons remain interpretable across updates.

Example deterministic indices include: a Radiographic Stability Index (RSI) that summarizes stability/progression signals on a continuous scale, and a Composite Damage Trajectory Index (CDTI) that summarizes longitudinal trajectory behavior. These indices are intended as standardized summary measures; their use in regulated submissions requires appropriate validation for the target context of use.

4. Harmonized Trial Export and Audit Traceability

For multi-center research and trial workflows, the platform supports exporting deterministic endpoint fields with explicit versioning and audit traceability. A provenance hash can be used to reconstruct the audit trail linking outputs to validated input descriptors, operator versions, and threshold sets.

Table 1. Harmonized Export Fields

Field	**Description**
RSI	Radiographic Stability Index (0–1 continuous scale)
RSI_band	Categorical band derived from versioned thresholds (e.g., Stable / Borderline / Progressive)
CDTI	Composite Damage Trajectory Index (0–1 continuous scale)
CDTI_class	Trajectory class for stratification (versioned thresholds)
drift_category	Structural drift category label (categorical; versioned thresholds)
threshold_version	Version identifier for threshold set / bands / cutpoints
phenotype_profile	Structural phenotype profile label (e.g., patterns typically seen in RA, PsA, axSpA, OA, CPPD; non-diagnostic)
operator_version	Version identifier for composite operator
provenance_hash	Hash enabling audit-trail reconstruction

5. Regulatory and Deployment Considerations

Regulated research workflows commonly expect documented traceability, audit trails, and reproducible endpoint derivation. Deterministic operators and explicit versioning can support these expectations by enabling repeatable re-execution and reconstruction of outputs. However, any deployment intended for clinical decision support or regulated endpoint generation should be preceded by validation studies appropriate to the intended context of use, and may require applicable regulatory review/clearance depending on jurisdiction and claims.

6. Discussion and Future Work

Deterministic scoring is complementary to training and calibration: rather than attempting to eliminate reader variability solely through education, it reduces variability by constraining interpretation and analytics to validated descriptor lineages and fixed-order operators.

Future work includes expanded validation across disease subsets and imaging protocols, refinement of deterministic indices for specific trial endpoints, and integration with multimodal research workflows where appropriate. The platform is designed to allow modular extensions without compromising deterministic execution, provided that extensions remain subject to the same validation gates and versioning controls.

7. Conclusion

A validator-gated deterministic pipeline can reduce reader-dependent variability in radiographic scoring and support auditable, traceable endpoint generation. By separating structured clinical rendering from research analytics and enforcing versioned operators, the platform is intended to provide reproducible outputs suited to multi-center research workflows and standardized trial exports.

References

1. Guillemin F, Billot L, Boini S, Gerard N, Ødegaard S, Kvien TK. Reproducibility and sensitivity to change of 5 methods for scoring hand radiographic damage in patients with rheumatoid arthritis. J Rheumatol. 2005;32(5):778–786. PMID: 15868609.
2. Swinkels HL, Laan RF, van 't Hof MA, van der Heijde DM, de Vries N, van Riel PL. Modified Sharp method: factors influencing reproducibility and variability. Semin Arthritis Rheum. 2001;31(3):176–190. PMID: 11740798.
3. Lukas C, Landewé R, Fatenejad S, van der Heijde D. Subtle changes in individual joints result in both positive and negative change scores in a patient: results from a clinical trial in patients with rheumatoid arthritis. Ann Rheum Dis. 2009;68(11):1691–1695. doi:10.1136/ard.2008.097816. PMID: 19029169.
4. Gandjbakhch F, Granger B, Freund R, et al. Multireader assessment as an alternative to reference assessment to improve the detection of radiographic progression in a large longitudinal cohort of rheumatoid arthritis (ESPOIR). RMD Open. 2017;3(1):e000343. doi:10.1136/rmdopen-2016-000343. PMID: 28123779.
5. U.S. Food and Drug Administration (FDA). Clinical Trial Imaging Endpoint Process Standards: Guidance for Industry. April 2018.
6. European Medicines Agency (EMA). Guideline on computerised systems and electronic data in clinical trials. EMA/INS/GCP/112288/2023.
7. Bird A, Oakden-Rayner L, McMaster C, et al. Artificial intelligence and the future of radiographic scoring in rheumatoid arthritis: a viewpoint. Arthritis Res Ther. 2022;24(1):268. doi:10.1186/s13075-022-02972-x.
8. Sun D, Nguyen TM, Allaway RJ, et al. A crowdsourcing approach to develop machine learning models to quantify radiographic joint damage in rheumatoid arthritis. JAMA Netw Open. 2022;5(8):e2227423. doi:10.1001/jamanetworkopen.2022.27423. PMID: 36036935.
9. Honda S, Yano K, Tanaka E, Ikari K, Harigai M. Development of a scoring model for the Sharp/van der Heijde score using convolutional neural networks and its clinical application. Rheumatology (Oxford). 2023;62(6):2272–2283. doi:10.1093/rheumatology/keac586. PMID: 36218391
10. Venäläinen MS, Biehl A, Holstila M, Kuusalo L, Elo LL. Deep learning enables automatic detection of joint damage progression in rheumatoid arthritis—model development and external validation. Rheumatology (Oxford). 2025;64(3):1068–1076. doi:10.1093/rheumatology/keae215.