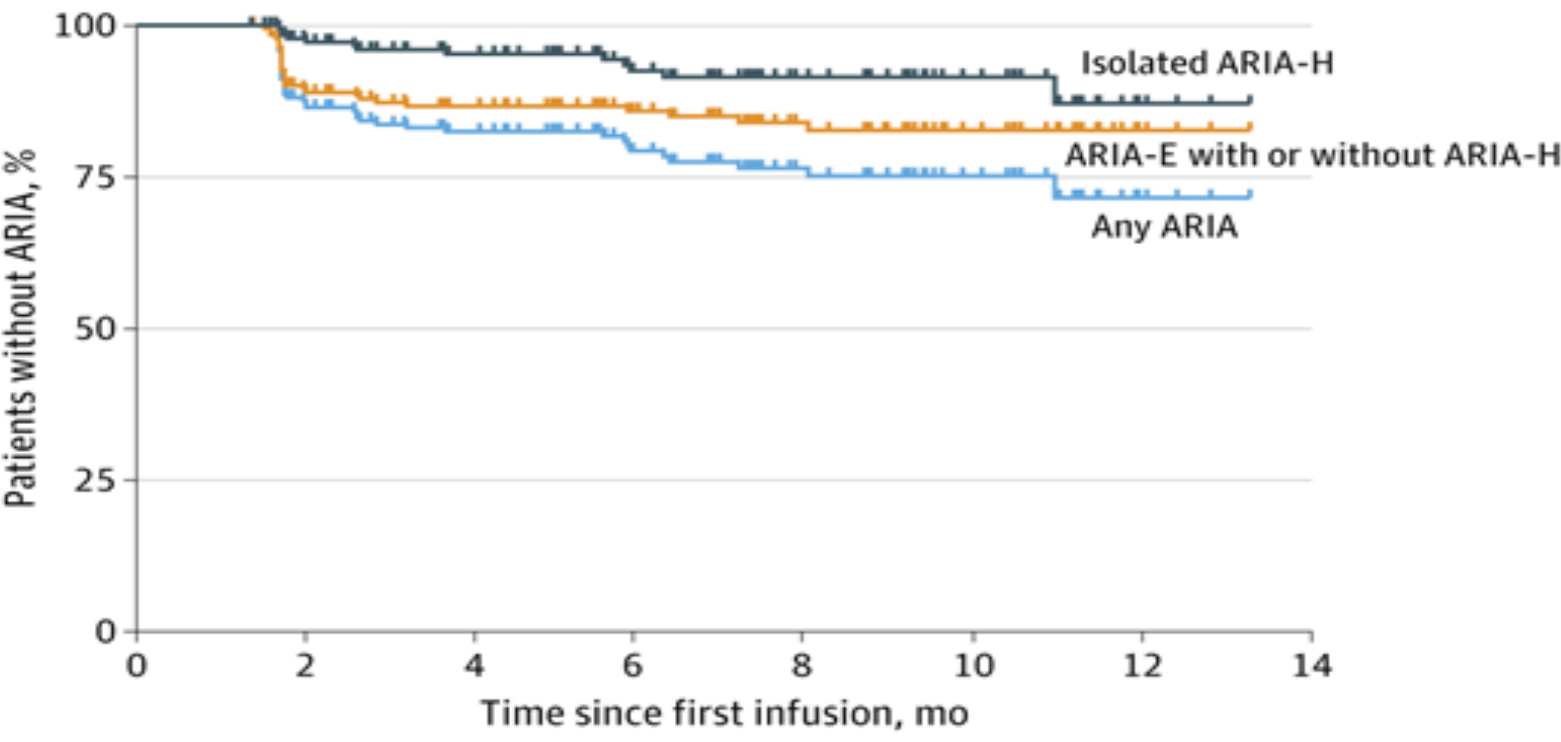


Standardizing ARIA Assessment: Development and Implementation of a Rater Study to Evaluate Grading Criteria Reliability

Background and Clinical Context

- The approval of anti-amyloid monoclonal antibodies has transformed Alzheimer's disease treatment, elevating radiologists to active gatekeepers of therapy due to Boxed Warnings for **Amyloid-Related Imaging Abnormalities (ARIA)**.
- The proportion of patients with an ARIA finding at 12 months of treatment is 29%: 17% with ARIA-E and 13% with ARIA-H. Clinical management, whether to continue or suspend dosing, hinges entirely on the precise grading of ARIA-E (edema/effusions) and ARIA-H (hemorrhages/siderosis) as "Mild," "Moderate," or "Severe".
- Current application of these quantitative criteria remains highly subjective, creating a critical gap in standardized reporting that threatens patient safety and necessitates a validated interpretation framework.



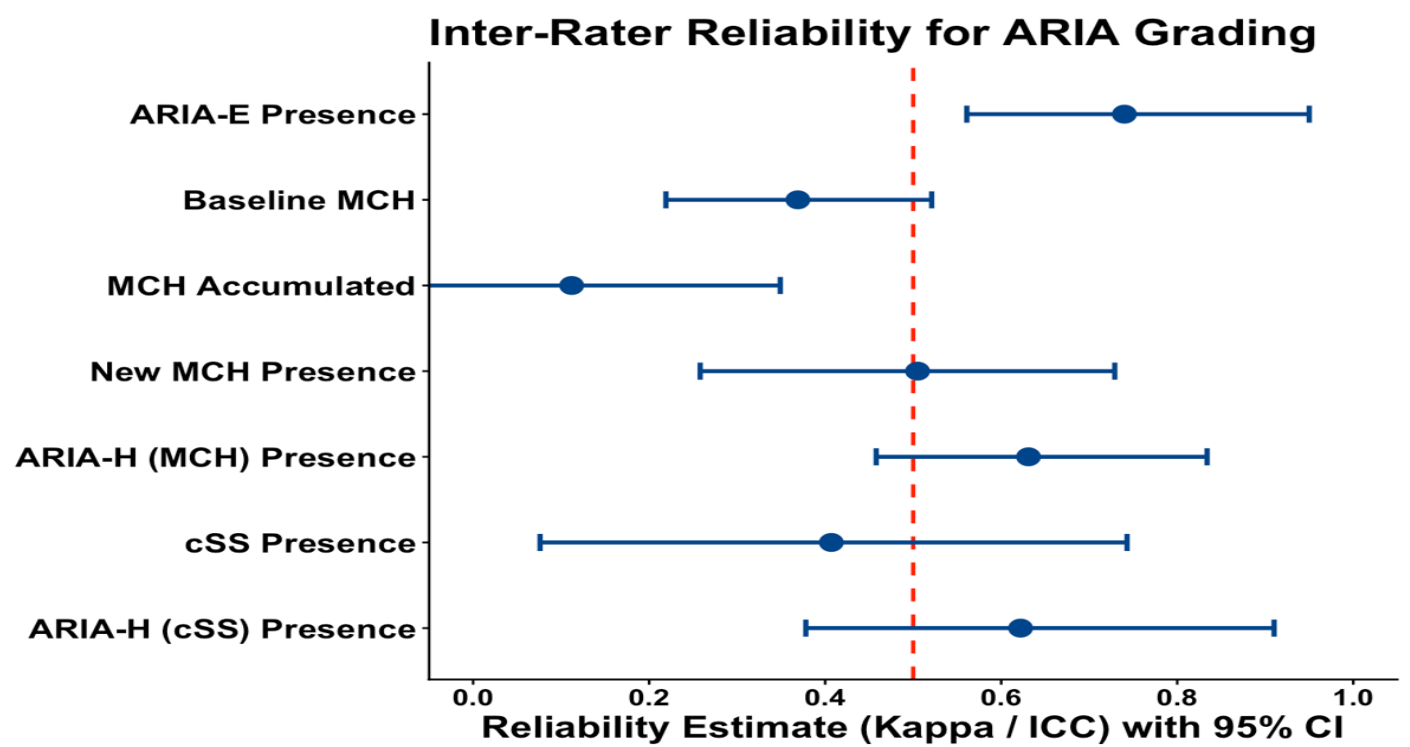
No. at risk						
Isolated ARIA-H	194	163	135	91	58	28
ARIA-E with or without ARIA-H	194	165	141	99	65	32
Any ARIA	194	163	135	91	58	28

Methods and Study Design

- Dataset:** A 13-case retrospective "ground truth" cohort representing a full severity spectrum of ARIA-E, ARIA-H, and negative controls.
- Curation:** Mined across a multi-hospital SECTRA PACS using natural language processing (NLP) queries: "amyloid," "lecanemab," "edema," and "siderosis."
- Imaging Protocols:** Baseline and follow-up MRIs utilized T2-FLAIR sequences for ARIA-E detection, alongside 2D T2* GRE and 3D Susceptibility-Weighted Imaging (SWI) for ARIA-H evaluation.
- Multi-Reader Setup:** After establishing reference standards via neuroradiologist consensus, three raters graded each case using strictly provided reference tables. To ensure standardized data capture and maintain reader blinding, each participant's diagnostic evaluations were independently recorded using a customized Microsoft Forms survey. This electronic platform presented the cases sequentially alongside the required reference grading tables, facilitating the efficient aggregation of categorical severity responses and microhemorrhage counts for subsequent analysis.

Results and Statistical Analysis

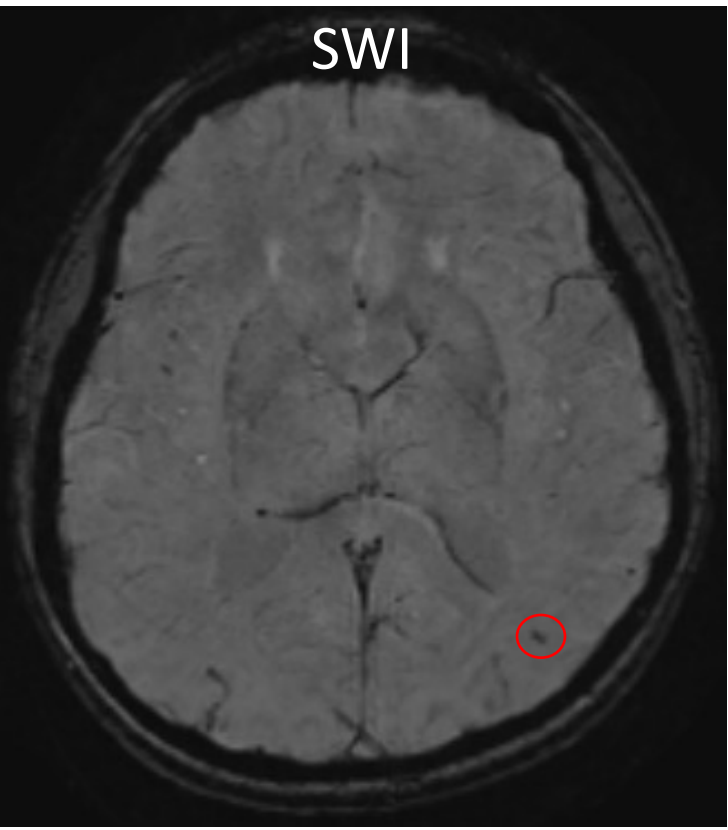
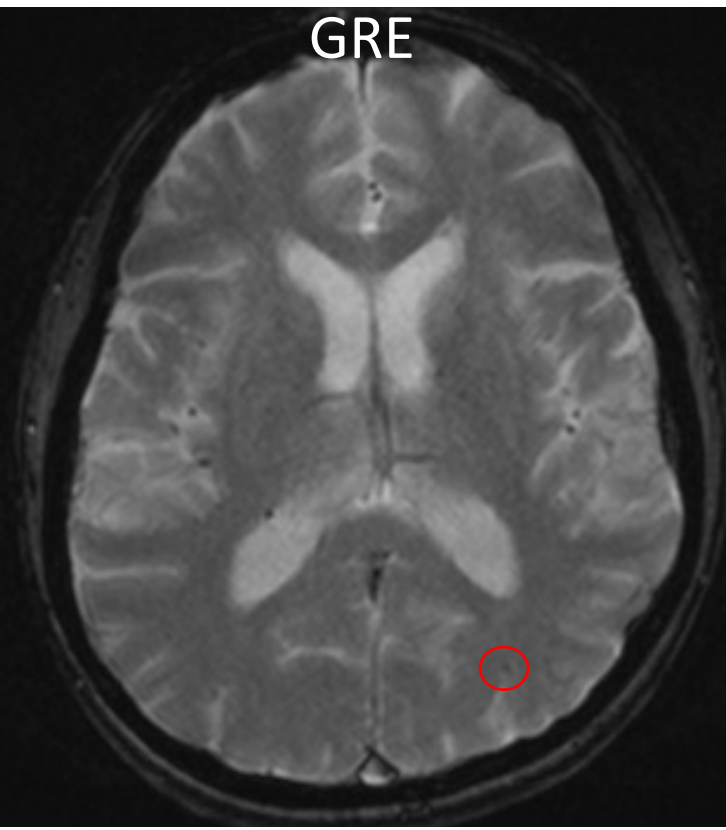
Variable Name	Statistical Test	Point Estimate	Lower CI	Upper CI	p-value	Interpretation
ARIA-E Presence	ICC 3,1	0.74	0.561	0.95	<0.001	Moderate reliability
Baseline MCH	ICC 3,1	0.369	0.219	0.521	0.007	Poor reliability
MCH Accumulated	ICC 3,1	0.112	-0.214	0.349	<0.001	Poor reliability
New MCH Presence	Cohen's Kappa	0.505	0.258	0.729	0.001	Moderate agreement
ARIA-H (MCH) Presence	Cohen's Kappa	0.631	0.458	0.834	0.028	Moderate Fair
cSS Presence	Cohen's Kappa	0.407	0.0761	0.743	0.075	agreement
ARIA-H (cSS) Presence	ICC 3,1	0.622	0.378	0.91	0.007	Moderate reliability



Methodology: Inter-rater reliability between the original clinical reports and the aggregated study reader was calculated using Cohen's Kappa for binary categorical variables and a Two-Way Mixed Effects Intraclass Correlation Coefficient (ICC 3,1) for absolute agreement on ordinal severity scales and microhemorrhage counts. Analyses were conducted in jamovi, with forest plots generated in R.

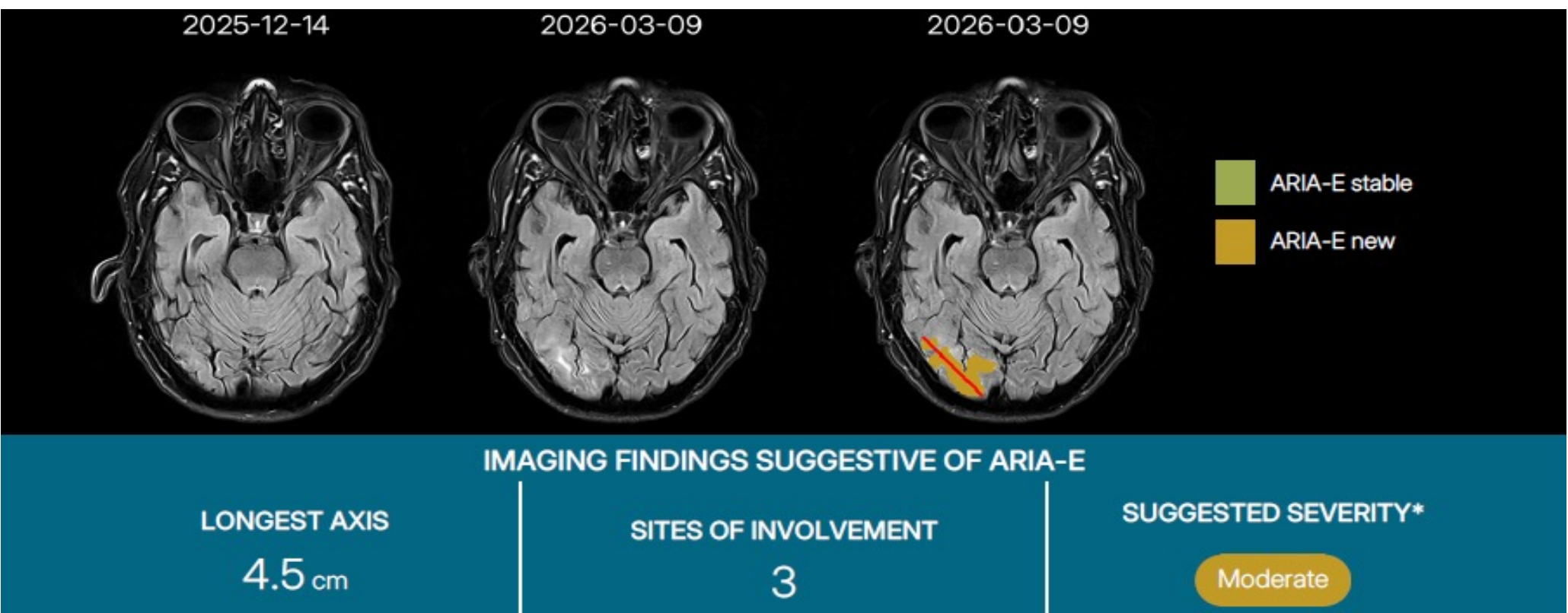
Imaging Pitfalls and Qualitative Discordance

- The Sulcal Spectrum (ARIA-E):** Preliminary reviews identify high discordance in differentiating "Mild" isolated sulcal hyperintensities from "Severe" presentations marked by gyral swelling and complete sulcal effacement.
- The Sequence Paradox (ARIA-H):** 3D SWI provides significantly higher sensitivity for detecting heme products compared to standard 2D T2* GRE. This technological gap directly alters total microhemorrhage (MCH) counts and shifts severity classification thresholds. The scans below show a case where 2 MCHs were counted on T2* GRE and 12 MCHs counted on SWI.



AI-Assisted ARIA Detection: *icobrain* Case Study

- Technology:** The FDA-cleared *icobrain aria* software by icometrix provides automated detection, volumetric quantification, and severity grading for both ARIA-E and ARIA-H.
- Mechanism:** The software utilizes a longitudinal neural network to directly register and compare a patient's baseline MRI with their post-dosing follow-up scan.
- Clinical Impact:** Application of one case study demonstrates that while AI assistance significantly boosts a radiologist's diagnostic sensitivity and classification accuracy for subtle ARIA lesions, literature indicates it may come at the cost of a slight reduction in overall specificity.



Conclusions and Future Directions

- Standardized reporting of ARIA is mandatory for patient safety in the era of disease-modifying Alzheimer's therapies.
- This study establishes a highly reproducible framework for assessing radiologist performance and characterizing interpretation errors.
- Our workflow is designed to be scalable for future deployment in larger cohorts and eventual validation against national registry data.

Acknowledgements

I am grateful for the support from the Benzinger Lab members, led by Tammie L.S. Benzinger MD, PhD. A special thanks to the radiologist readers, Dr. Thomas Kim and Dr. Omer Cakir. I would also like to acknowledge icometrix for their support in letting us use their software to test ARIA validation. Finally, this project is supported by the National Institutes of Health.