

Sydney Medical School
University of Sydney
Sydney, NSW 2006
Australia
itod2305@uni.sydney.edu.au

December 10, 2025

Editor-in-Chief
Journal of the American Medical Informatics Association
Oxford University Press

Dear Editor-in-Chief,

Please find enclosed our manuscript, “**The Dimensional Validity Bound: Structural Limits of Clinical AI Evaluation in Multimorbidity**,” for consideration as a Research and Applications article in *JAMIA*. This manuscript provides a unifying theoretical framework that identifies conditions under which clinical decision support evaluation becomes structurally unreliable, independent of model architecture or training procedure.

The problem: While aggregate metrics like AUC are the gold standard for evaluating Clinical Decision Support Systems (CDSS), they implicitly assume that model validity is uniform across the patient population. Our study challenges this assumption. By applying information-theoretic principles to the MIMIC-IV database ($N=425,216$), we demonstrate the existence of a **Dimensional Validity Bound**—a critical complexity threshold below which Bayesian inference becomes structurally unstable.

What we show:

- A theoretical framework (the dimensional validity bound) derived from Fano’s inequality
- Three distinct failure mechanisms: projection error, noise amplification, hypothesis space explosion
- Simulation evidence for a critical threshold at dimensional ratio $r \approx 0.3$
- Real-world validation showing the predicted U-shaped performance curve in ICU mortality prediction
- A paradoxical “Zone of Maximum Entropy” in moderate-complexity patients where aggregate AUC remains high (0.835) despite structural posterior instability

Why JAMIA: This work addresses a fundamental methodological question about when CDSS evaluation metrics are valid—precisely the kind of cross-disciplinary, methods-focused research that *JAMIA* publishes. The dimensional validity bound offers a theoretical mechanism explaining why clinical AI tools often degrade unpredictably in complex clinical environments, distinct from issues of dataset shift or implementation quality. We propose that **effective dimensionality (D_{eff})** be reported as standard metadata for clinical datasets, alongside demographic parity and fairness metrics, providing a quantitative metric for algorithmic auditing and regulatory clearance of AI medical devices.

Significance: Current evaluation practice assumes that validated models will perform comparably on deployment data. We show this assumption fails when complexity distributions differ—even if marginal distributions match. The zone of maximum entropy represents a systematic blind spot in current practice: invisible to aggregate metrics, occurring in an intermediate regime that intuition misses, and containing instances where errors may be most consequential.

This manuscript is original, has not been published, and is not under consideration elsewhere. I am the sole author and approve submission.

Sincerely,

Ian Todd
Sydney Medical School
University of Sydney