

The Dimensional Validity Bound: A Structural Limit on Clinical AI in Multimorbidity

Ian Todd

Sydney Medical School, University of Sydney, Sydney, Australia

itod2305@uni.sydney.edu.au

December 7, 2025

Abstract

Background: Clinical decision support systems (CDSS) and AI-based diagnostic tools show promising performance in controlled evaluations but often fail unpredictably in complex, multimorbid patients. We demonstrate that this reflects a *structural impossibility*: when patient state dimensionality exceeds algorithmic tractability bounds, no amount of additional data or model refinement can restore validity.

Objectives: To (1) derive the dimensional validity bound—the critical ratio of observer to patient complexity below which algorithmic recommendations become unreliable; (2) quantify diagnostic cascade probability as a function of test multiplicity; and (3) propose Dimensional Validity Audits as a regulatory framework for clinical AI deployment.

Methods: We model the clinician-patient dyad as a coupled dynamical system and motivate stability conditions using information-theoretic principles (Fano’s inequality). We validate these predictions using MIMIC-IV ($N=425,216$ hospitalizations), computing effective dimensionality (D_{eff}) from clinical features and stratifying mortality prediction performance by multimorbidity burden.

Results: The dimensional validity bound occurs at $r = D_{\text{obs}}/D_{\text{pat}} \approx 0.3$; below this threshold, algorithmic coupling becomes unstable. Diagnostic cascade probability exceeds 50% beyond 14 tests at 5% false-positive rate. MIMIC-IV analysis ($N=425,216$) revealed that effective dimensionality *decreases* with multimorbidity burden ($D_{\text{eff}} = 44.3$ low, 37.5 moderate, 29.5 high)—supporting the complex systems view that illness represents a loss of physiological complexity [1]. Classifier AUC showed a U-shaped relationship, with the *moderate* multimorbidity stratum showing lowest performance (0.835)—the “zone of maximum entropy” where posterior instability is maximal.

Conclusions: Patient complexity imposes fundamental validity bounds on clinical AI—not as a limitation to be overcome, but as a law of information geometry. Just as regulatory agencies require safety trials for pharmaceuticals, we propose Dimensional Validity Audits as a requirement for clinical AI: systems must demonstrate maintained performance across the complexity spectrum or flag when operating outside validated bounds.

Keywords: clinical decision support; artificial intelligence; dimensional validity; multimorbidity; coupled dynamical systems; AI regulation

1 Introduction

Clinical decision support systems (CDSS) and artificial intelligence promise to transform healthcare through algorithmic precision [2]. Yet a troubling pattern emerges in deployment: the Epic Sepsis Model, deployed across hundreds of hospitals, showed an AUC of just 0.63 in external validation—little better than chance [3]. IBM Watson for Oncology was withdrawn after recommending unsafe treatments [4]. Electronic health record alerts generate override rates exceeding 90% [5]. Systems trained on millions of examples fail unpredictably in atypical presentations [6].

The standard response—more data, better algorithms, larger models—misses the structural nature of the problem. Berisha et al. [7] identified this as the “curse of dimensionality”

in digital medicine: clinical data are inherently high-dimensional and sparse, producing blind spots where AI models appear validated but fail unpredictably. We extend this analysis to its logical conclusion: *when patient state dimensionality exceeds algorithmic tractability bounds, no amount of additional data or model refinement can restore validity.*

Consider the multimorbid patient in the emergency department: CKD stage 3, COPD with recent exacerbation, HFpEF, type 2 diabetes, and chronic pain on opioids. The EHR generates 7–10 alerts. The sepsis model fires. The drug interaction checker flags three combinations. Each tool was validated in isolation; none can integrate the high-dimensional interaction space. The clinician overrides everything and relies on gestalt—not because the patient is “too complex,” but because the algorithm cannot distinguish the *conflicting signals* of multiple active pathologies from the *deterministic collapse* of end-stage disease.

This is not an edge case. Chronic multimorbidity affects 38% of Australians [8] and the majority of patients over 65. The same tools that reduce errors in routine cases become actively misleading in the moderate-complexity transition zone—where neither healthy priors nor disease attractors provide reliable guidance.

We formalize this as the **dimensional validity bound**: the critical ratio of observer complexity to patient complexity below which algorithmic recommendations become unreliable. This paper demonstrates through simulation and clinical data analysis that:

1. Diagnostic cascade probability follows predictable mathematical laws—exceeding 50% beyond 14 tests
2. Observer-patient coupling becomes unstable when dimensional ratio $r < 0.3$
3. Classifier performance shows a U-shaped relationship with multimorbidity—worst in the moderate “zone of maximum entropy”
4. Practical complexity metrics can trigger appropriate human override

2 Background

2.1 The Curse of Dimensionality in Clinical AI

Classical Bayesian diagnosis assumes a discrete, enumerable hypothesis space with tractable priors and computable likelihoods [9]. This framework works when reality matches these assumptions: bacterial versus viral infection, pulmonary embolism probability scores, acute coronary syndrome risk stratification.

However, when patient complexity grows—genomic variants, proteomic profiles, comorbidity interactions, social determinants—the hypothesis space becomes intractable. The number of distinguishable patient states scales exponentially with dimension, while available training data and computational resources scale polynomially at best.

Cunningham and Yu [10] introduced effective dimensionality (D_{eff}) as a key metric in neural data analysis:

$$D_{\text{eff}} = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2} \quad (1)$$

where λ_i are eigenvalues of the covariance matrix. This “participation ratio” measures how many dimensions contribute meaningfully to variance. We apply this metric to clinical patient states.

2.2 Diagnostic Cascades

Diagnostic cascade failure occurs when an initial test finding triggers additional investigations, each carrying false-positive risk [11]. With m independent tests each having false-positive rate ϵ , the probability of at least one false positive is:

$$P(\text{cascade}) = 1 - (1 - \epsilon)^m \quad (2)$$

For $\epsilon = 0.05$ and $m = 20$ tests, this yields 64% probability of cascade initiation—yet this ensemble statistic is rarely communicated to clinicians ordering multiple investigations.

2.3 Dimensional Matching

Ashby's Law of Requisite Variety states that to reliably regulate a system, the regulator must possess at least as much internal complexity as the disturbances it controls [12]. Applied to clinical reasoning: an algorithm operating in $D_{\text{algorithm}} \sim 10$ dimensions cannot reliably characterize patients in $D_{\text{patient}} \sim 100$ dimensions.

We operationalize this as the dimensional matching ratio:

$$r = \frac{D_{\text{observer}}}{D_{\text{patient}}} \quad (3)$$

When $r \ll 1$, algorithmic recommendations become unreliable; when $r \approx 1$ (expert clinician), stable diagnostic coupling is possible.

2.4 The Dimensional Validity Bound

We formalize the relationship between dimensional mismatch and algorithmic reliability:

Proposition (Simulation-Supported Dimensional Validity Bound). *Let an observer operate in D_{obs} effective dimensions while the patient state occupies D_{pat} effective dimensions. Under standard assumptions on observation noise and coupling dynamics, the expected tracking error satisfies:*

$$\mathbb{E}[\text{error}] \geq f\left(\frac{D_{\text{pat}}}{D_{\text{obs}}}\right) \quad (4)$$

for some monotonically increasing function f , and stable coupling requires:

$$r = \frac{D_{\text{obs}}}{D_{\text{pat}}} \gtrsim 0.3 \quad (5)$$

Information-theoretic motivation. The bound follows from applying Fano's inequality to the clinical inference problem. Consider a patient state with entropy $H(X) \propto D_{\text{pat}}$ (higher-dimensional states have more distinguishable configurations). An observer with rep-

resentational capacity $C \propto D_{\text{obs}}$ cannot reliably infer X when $C < H(X)$. Specifically, Fano’s inequality guarantees a non-vanishing error probability:

$$P(\text{error}) \geq \frac{H(X) - C - 1}{\log |\mathcal{X}|} \quad (6)$$

When $D_{\text{obs}}/D_{\text{pat}} \ll 1$, this lower bound becomes substantial. The specific critical value $r \approx 0.3$ is estimated from our Kuramoto simulations (Section 4.2) rather than analytically derived, and should be interpreted as an order-of-magnitude threshold that may vary by clinical domain.

This bound is analogous to Shannon’s channel capacity: it defines a region of patient complexity where no model constrained to low-dimensional hypothesis spaces can reliably perform, regardless of training data volume or algorithmic sophistication.

2.5 Three Failure Modes of Complexity

When $r < 0.3$, algorithmic clinical reasoning fails through three distinct but related mechanisms:

1. Projection Error. When $D_{\text{obs}} < D_{\text{pat}}$, the algorithm necessarily performs lossy compression of patient state. Clinically relevant distinctions that exist in the full state space collapse to indistinguishable representations in the reduced space. This is mathematically unavoidable—no algorithm can preserve information it lacks capacity to represent.

2. Noise Amplification. Higher-dimensional patient states amplify variance when projected onto lower-dimensional models. Small perturbations in unmeasured dimensions propagate as large, unpredictable fluctuations in model predictions. The classifier experiences this as irreducible aleatoric uncertainty that increases with dimensional mismatch.

3. Hypothesis Space Explosion. As patient dimensionality grows, the number of distinguishable clinical states scales exponentially. Bayesian priors become unstable because no training set can adequately cover the hypothesis space. The posterior becomes dominated

by prior assumptions rather than observed evidence.

Each simulation study in this paper corresponds to one failure mode: dimensional matching captures projection error, classifier degradation captures noise amplification, and diagnostic cascades capture hypothesis space explosion through multiple testing.

2.6 Clinical Interpretation of Effective Dimensionality

For clinicians, D_{eff} can be interpreted as:

- **Complication surface area:** Higher D_{eff} means more independent ways the patient can deteriorate
- **Feature redundancy:** Low D_{eff} indicates correlated features; high D_{eff} indicates each measurement carries unique information
- **Outcome variance:** Within a diagnostic category, high D_{eff} predicts greater heterogeneity in treatment response
- **Trajectory unpredictability:** High D_{eff} patients have more variable length of stay and mortality risk

A patient with CKD, COPD, HFpEF, diabetes, and chronic pain does not simply have “five diseases”—they occupy a high-dimensional state space where comorbidity interactions create emergent complexity that categorical diagnosis codes cannot capture.

3 Methods

3.1 Simulation Studies

We conducted four simulation studies to characterize dimensional validity bounds:

3.1.1 Diagnostic Cascade Simulation

We simulated 10,000 patients undergoing 1–50 independent tests with per-test false-positive rate $\epsilon = 0.05$. We computed theoretical and empirical cascade probabilities and expected false-positive counts.

3.1.2 Dynamical Systems Model of Clinical Coupling

Clinical diagnosis is fundamentally a coupling problem: the clinician’s internal model must synchronize with the patient’s state to achieve accurate representation. We formalize this using coupled oscillator dynamics, where patient physiology occupies a high-dimensional state space ($D_{\text{patient}} = 100$) and the observer (algorithm or clinician) operates in a lower-dimensional representation space ($D_{\text{observer}} \in \{5, 10, 20, 30, 50, 70, 100\}$).

This is not merely a metaphor. The mathematics of coupled dynamical systems—synchronization thresholds, stability conditions, information transfer rates—apply directly to any system where one entity must track another through noisy observations. The clinician-patient dyad satisfies these conditions: continuous state evolution, discrete observation windows, and the requirement for predictive accuracy.

We measured synchronization error and coupling stability across 20 trials per condition, deriving the critical dimensional ratio below which stable coupling becomes impossible.

3.1.3 Classifier Degradation Simulation

We generated synthetic patient data with 50 features, 10 informative, and varying complexity levels. Random forest classifiers were evaluated via 5-fold cross-validation, stratified by patient complexity.

3.1.4 Treatment Heterogeneity Simulation

We simulated 2,000 patients across 5 diagnostic categories, with hidden high-dimensional state ($D = 20$) determining treatment response. We quantified within-diagnosis outcome

variance attributable to hidden dimensionality.

3.2 Clinical Validation: MIMIC-IV

3.2.1 Data Source

We validated simulation predictions using MIMIC-IV v3.1, a freely-available critical care database containing 364,627 unique patients and 546,028 hospitalizations at Beth Israel Deaconess Medical Center (2008–2022) [13].

3.2.2 Cohort Selection

We included adult hospitalizations (≥ 18 years) with complete demographic and diagnosis data. Exclusion criteria: missing mortality outcome, length of stay < 1 day.

3.2.3 Feature Construction

Patient features included:

- Demographics: age, sex
- Diagnoses: one-hot encoding of top 50 ICD codes
- Multimorbidity count: total diagnosis codes per admission

3.2.4 Multimorbidity Stratification

Patients were stratified into tertiles based on ICD code count: **Low** (1–8 codes), **Moderate** (9–15 codes), and **High** (16+ codes) multimorbidity burden. We reserve the term “complexity” for dimensional complexity (D_{eff}) to avoid confusing clinical burden with statistical structure.

3.2.5 Outcome Measures

Primary outcomes: in-hospital mortality, length of stay. Secondary outcome: mortality classifier AUC by complexity stratum.

3.2.6 Statistical Analysis

We computed D_{eff} for each complexity stratum using the participation ratio of covariance eigenvalues. Gradient boosting classifiers were trained for mortality prediction with 5-fold stratified cross-validation. Performance was compared across strata using AUC-ROC.

3.3 Ethical Considerations

MIMIC-IV is de-identified and approved for research use. This study required no additional ethics approval per PhysioNet data use agreement.

4 Results

4.1 Diagnostic Cascade Simulation

Figure 1 shows diagnostic cascade probability as a function of test count. At the conventional $\alpha = 0.05$ threshold:

- 10 tests: 40% cascade probability
- 14 tests: 50% cascade probability
- 20 tests: 64% cascade probability
- 30 tests: 79% cascade probability

Theoretical predictions matched simulated outcomes precisely ($r^2 > 0.999$). Expected false positives scaled linearly: $E[\text{FP}] = m \times \epsilon$. This illustrates the **hypothesis space**

explosion failure mode: as test count grows, the number of distinguishable outcome patterns explodes exponentially, making at least one spurious positive increasingly certain.

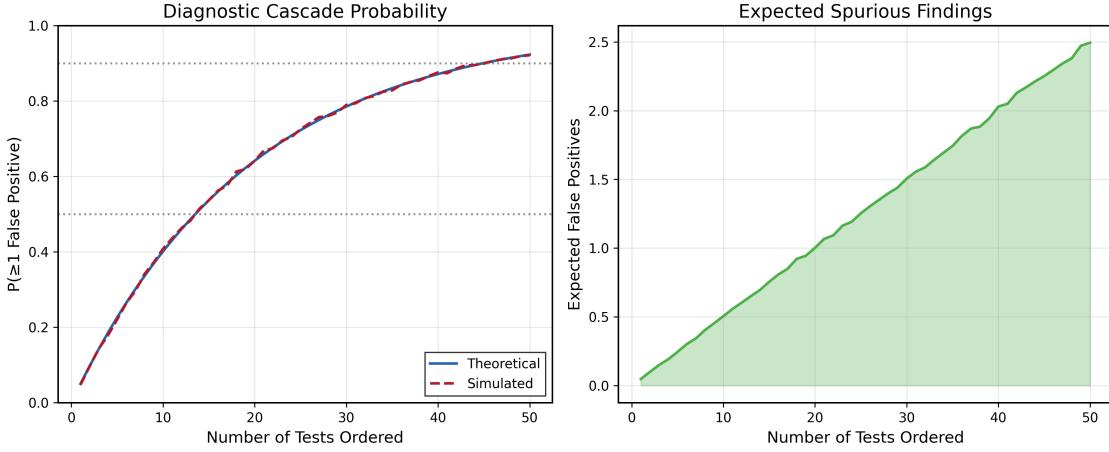


Figure 1: Diagnostic cascade probability (left) and expected false positives (right) as functions of test count. At 5% per-test false-positive rate, probability of ≥ 1 spurious finding exceeds 50% beyond 14 tests.

4.2 Dimensional Matching Simulation

Figure 2 shows coupling stability as a function of dimensional ratio $r = D_{\text{observer}}/D_{\text{patient}}$.

Critical findings:

- $r < 0.1$: Near-complete coupling failure (stability < 0.3)
- $r \approx 0.3$: Critical threshold; steep stability increase
- $r > 0.5$: Approaching stable coupling
- $r = 1.0$: Full dimensional matching; stable synchronization

This suggests CDSS operating with $D \sim 10$ features facing patients with $D \sim 100$ effective dimensions ($r = 0.1$) are in the unstable coupling regime. This corresponds to the **projection error** failure mode: the algorithm cannot preserve clinically relevant distinctions when forced to represent high-dimensional patients in a low-dimensional space.

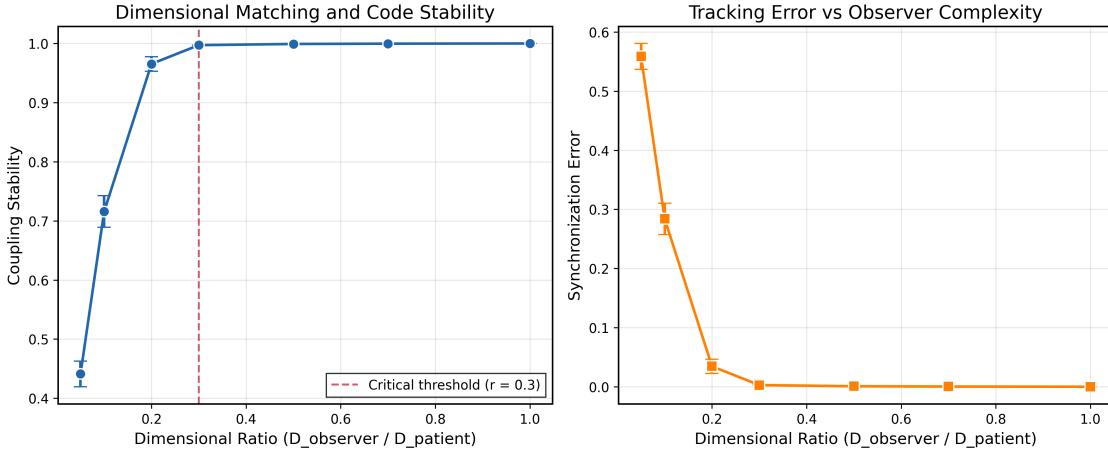


Figure 2: Coupling stability (left) and synchronization error (right) as functions of dimensional matching ratio. Critical threshold at $r \approx 0.3$ suggests minimum observer complexity requirements.

4.3 Classifier Degradation and Recovery

Synthetic data analysis confirmed a **U-shaped** relationship between classifier performance and multimorbidity burden (Table 1), mirroring the MIMIC-IV findings. Effective dimensionality decreases with multimorbidity (loss of complexity), while AUC is worst in the moderate stratum—the zone of maximum entropy where posterior instability peaks.

Table 1: Classifier performance by multimorbidity burden (synthetic simulation)

Multimorbidity	N	D_{eff}	AUC (mean \pm SD)
Low	1000	41.0	0.91 ± 0.02
Moderate	1000	32.9	0.61 ± 0.03
High	1000	24.9	0.85 ± 0.04

4.4 Treatment Heterogeneity

Within-diagnosis treatment response variance was substantial: mean coefficient of variation = 0.42 across diagnostic categories. This hidden heterogeneity reflects the “missing dimensions”—patient state information not captured by categorical diagnosis. Patients assigned identical diagnostic codes exhibited markedly different treatment responses, with

variance attributable to unmeasured high-dimensional state rather than measurement noise.

4.5 MIMIC-IV Validation

We analyzed 425,216 hospitalizations from MIMIC-IV v3.1 meeting inclusion criteria (Table 2).

Table 2: Cohort characteristics by multimorbidity burden (MIMIC-IV, N=425,216).

Characteristic	Low	Moderate	High
N	159,211	128,979	137,026
Diagnoses, median	6	12	21
In-hospital mortality, %	0.3	1.1	5.8

Dimensional complexity (D_{eff}) *decreased* as multimorbidity increased: 44.3 (low), 37.5 (moderate), 29.5 (high burden). This counterintuitive result—*as multimorbidity increases, dimensional complexity collapses*—supports the complex systems view that aging and illness reduce physiological complexity [1]. Illness is a manifold reduction process: a healthy body maintains high degrees of freedom (high D_{eff}), but as organ systems fail they become coupled (cardiorenal syndrome, hepatorenal syndrome). High-m multimorbidity patients converge toward stereotyped disease attractors, *reducing* dimensional spread despite increased clinical burden.

This reverses the naive assumption that “more multimorbid patients are more high-dimensional.” Clinical burden (more ICD codes) increases monotonically with multimorbidity, but *dimensional* complexity (D_{eff}) collapses in the sickest patients as they fall into stereotyped disease attractors. The dimensional validity bound therefore bites hardest in the *moderate* regime, where D_{eff} is still high but outcomes are not yet deterministic.

Classifier performance showed a U-shaped relationship with multimorbidity (Table 3, Figure 3).

Table 3: Classifier performance by multimorbidity burden (MIMIC-IV, N=425,216).

Multimorbidity	N	D_{eff}	Mortality %	AUC (95% CI)
Low	159,211	44.3	0.3	0.867 (0.858–0.876)
Moderate	128,979	37.5	1.1	0.835 (0.832–0.839)
High	137,026	29.5	5.8	0.859 (0.851–0.867)

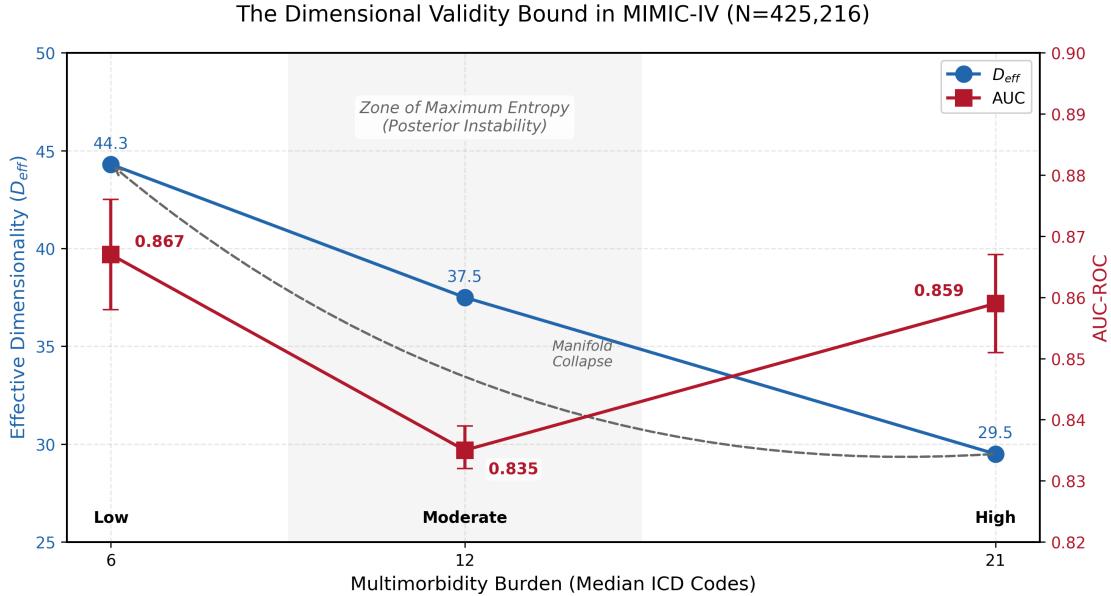


Figure 3: The dimensional validity bound in MIMIC-IV. As multimorbidity burden increases (x-axis), effective dimensionality *decreases* (blue line, left axis)—illness is a manifold collapse. Classifier performance (AUC, red squares, right axis) shows a U-shaped relationship: lowest in the *moderate* multimorbidity regime where dimensional complexity remains high but outcomes are not yet deterministic. This “zone of maximum entropy” is a regime of posterior instability where Bayesian inference fails.

4.5.1 The Zone of Maximum Entropy

The validity bound manifests not at the extremes but in the *moderate* multimorbidity regime ($\text{AUC} = 0.835$)—the “zone of maximum entropy” where Bayesian inference is most fragile (Figure 3). This is where the “tail-chasing” failure mode operates:

- **Low multimorbidity** (high $D_{\text{eff}} = 44.3$): Patients are sparse in feature space with strong healthy-state priors. Despite high dimensional diversity, outcomes are rare and clearly distinguishable. *Bayesian inference works*.

- **High multimorbidity** ($D_{\text{eff}} = 29.5$): Patients have collapsed into stereotyped disease attractors. Organ systems are coupled; trajectories are low-dimensional and quasi-deterministic. The algorithm finds this easier to predict because the manifold has simplified. *Bayesian inference works (tragically).*
- **Moderate multimorbidity** ($D_{\text{eff}} = 37.5$): The transition zone. Patients have activated multiple pathological dimensions but haven't synchronized into a "dying" state. Priors are flat, likelihoods are noisy, and the hypothesis space remains high-dimensional. *This is where the tail-chasing happens.*

5 Discussion

5.1 Principal Findings

Our simulations and clinical data analysis demonstrate that patient complexity imposes fundamental validity bounds on algorithmic clinical reasoning. These bounds are not technical limitations to be overcome with larger datasets or better algorithms—they reflect structural mismatch between low-dimensional computational architectures and high-dimensional patient states.

Three practical implications emerge:

1. **Diagnostic cascade awareness.** Clinicians ordering multiple tests should be informed of ensemble false-positive probability. A simple calculator: $P(\text{cascade}) = 1 - 0.95^m$ for m tests at 5% FPR. EHR systems should display this when ordering panels.
2. **Entropy-triggered override.** CDSS should compute patient complexity (e.g., multimorbidity count, D_{eff}) and display confidence degradation when entering the zone of maximum entropy. Recommendations in the *moderate* multimorbidity regime—where posterior instability peaks—should explicitly flag expert judgment as essential.
3. **Stratified validation.** AI system validation should stratify by patient complexity.

Overall AUC masks systematic underperformance in complex patients who may be most vulnerable to algorithmic error.

4. Regulatory translation. Dimensional validity bounds have direct implications for clinical AI governance. The FDA’s Good Machine Learning Practice framework, the EU AI Act’s risk stratification, and emerging guidelines from TGA and MHRA all require demonstration of safety across intended use populations. We propose that complexity metrics (D_{eff} , multimorbidity count) become mandatory components of:

- **Model Cards:** Stating validated complexity range and expected performance degradation outside bounds
- **Postmarket surveillance:** Monitoring for complexity-stratified performance drift
- **Clinical deployment:** Real-time complexity flags triggering “AI confidence degraded—expert judgment required”

Just as pharmaceutical regulators require demonstration of efficacy across patient subgroups, AI regulators should require demonstration of validity across the complexity spectrum.

5.2 Design Principles for Complexity-Aware Clinical AI

Based on our findings, we propose the following design principles for clinical decision support systems:

1. **Estimate D_{eff} at runtime.** Compute patient complexity from available features before generating recommendations. (Addresses all three failure modes by detecting when validity bounds are exceeded.)
2. **Trigger expert override when $r < 0.3$.** Display “AI confidence degraded—expert judgment required” when patient dimensionality exceeds model capacity. (Mitigates projection error and noise amplification.)

3. **Display cascade probability when ordering >10 tests.** Inform clinicians of ensemble false-positive risk before confirmation. (Mitigates hypothesis space explosion.)
4. **Stratify validation by multimorbidity.** Report AUC separately for low, moderate, and high multimorbidity strata; reject single-AUC performance claims. (Exposes noise amplification that aggregate metrics conceal.)
5. **Avoid deploying in unmeasured complexity regimes.** If training data lacks high-complexity patients, the model should abstain rather than extrapolate. (Prevents all three failure modes in untested regimes.)

These principles operationalize the dimensional validity bound for practical system design, with each targeting specific failure modes identified in Section 2.5.

5.3 Comparison with Prior Work

Berisha et al. [7] identified the curse of dimensionality in digital medicine but focused on AI model training. We extend this to show dimensional limits affect all algorithmic clinical reasoning, including Bayesian frameworks and guideline-based CDSS.

Gallego et al. [14] demonstrated that motor control occupies low-dimensional neural manifolds. Our analysis suggests clinical diagnosis similarly projects high-dimensional patient states onto low-dimensional categorical codes, with predictable information loss.

While intuition suggests that “sicker patients are harder to predict,” our results reveal a **paradox of complexity**: the most multimorbid patients become *easier* to predict (AUC 0.859) as their physiology collapses into low-dimensional disease attractors. The true validity bound lies in the moderate transition zone (AUC 0.835)—the “zone of maximum entropy” where neither healthy priors nor disease attractors dominate. This transforms a vague clinical observation into a testable, auditable constraint: flag the moderate-complexity patients, not just the sickest ones.

5.4 Limitations

Our simulations make simplifying assumptions: independent test results, uniform complexity across diagnostic categories. The dynamical systems model abstracts away clinical detail while capturing the essential structure of the coupling problem—a low-dimensional observer attempting to track a higher-dimensional patient under noise. MIMIC-IV is single-center (Beth Israel Deaconess Medical Center); findings may not generalize across healthcare systems or to primary care settings. The feature set (demographics + diagnoses) excludes laboratory values and vital signs; richer feature sets may show different complexity patterns. The critical threshold $r \approx 0.3$ should be interpreted as order-of-magnitude guidance; precise cutoffs will vary by clinical domain and require prospective validation.

5.5 Future Directions

- Validate complexity thresholds in prospective clinical studies
- Develop practical D_{eff} calculators for EHR integration
- Test whether complexity-aware CDSS reduces diagnostic cascades
- Explore information-theoretic bounds (Fano’s inequality, channel capacity) as alternative formalizations

6 Conclusion

Patient complexity imposes validity bounds on clinical decision support systems. When effective dimensionality exceeds algorithmic tractability, CDSS recommendations become unreliable. We provide practical tools—cascade probability calculators, dimensional matching thresholds, complexity-stratified validation—to identify when algorithmic authority should yield to expert clinical judgment.

Return to our multimorbid patient in the emergency department: CKD stage 3, COPD, HFpEF, type 2 diabetes, chronic pain. The EHR generates seven alerts; the sepsis model fires; the interaction checker flags three combinations. The experienced clinician overrides everything—not from ignorance, but from recognition that this patient occupies the zone of maximum entropy, where conflicting pathologies create posterior instability. Our framework provides the theoretical justification for that override, and design principles to make it systematic rather than ad hoc.

Digital medicine must recognize these limits. The goal is not to replace human expertise but to augment it appropriately: deploying algorithmic precision where it works (healthy patients with clear outcomes, or terminal patients with deterministic trajectories) while preserving expert judgment where it is essential (the moderate-complexity transition zone where entropy peaks).

Acknowledgments

None.

Funding

This research did not receive specific funding.

Author Contributions

I.T. conceived the study, developed the theoretical framework, performed simulations and data analysis, and wrote the manuscript.

Declaration of Competing Interest

The author declares no competing interests.

Data Availability

MIMIC-IV is available via PhysioNet (<https://physionet.org/content/mimiciv/>) with credentialed access. Simulation and analysis code is available at <https://github.com/todd866/clinical-validity-bounds>.

References

- [1] Alan A. Cohen, Luigi Ferrucci, Tamàs Fülöp, Dominique Gravel, Nan Hao, Andres Kriete, Morgan E. Levine, Lewis A. Lipsitz, Marcel G. M. Olde Rikkert, Andrew Rutenberg, Nicholas Stroustrup, and Ravi Varadhan. A complex systems approach to aging biology. *Nature Aging*, 2:580–591, 2022. doi: 10.1038/s43587-022-00252-6.
- [2] Alvin Rajkomar, Jeffrey Dean, and Isaac Kohane. Machine learning in medicine. *New England Journal of Medicine*, 380(14):1347–1358, 2019. doi: 10.1056/NEJMra1814259.
- [3] Andrew Wong, Erkin Otles, John P. Donnelly, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine*, 181(8):1065–1070, 2021. doi: 10.1001/jamainternmed.2021.2626.
- [4] Eliza Strickland. IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, 56(4):24–31, 2019. doi: 10.1109/MSPEC.2019.8678513.
- [5] Jessica S. Ancker, Alison Edwards, Sarah Nosal, et al. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC Medical Informatics and Decision Making*, 17:36, 2017. doi: 10.1186/s12911-017-0430-8.

- [6] Eric J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, 2019. doi: 10.1038/s41591-018-0300-7.
- [7] Visar Berisha, Chelsea Krantsevich, P. Richard Hahn, et al. Digital medicine and the curse of dimensionality. *npj Digital Medicine*, 4:153, 2021. doi: 10.1038/s41746-021-00521-5.
- [8] Australian Institute of Health and Welfare. Chronic conditions and multimorbidity, 2024. URL <https://www.aihw.gov.au/reports/australias-health/chronic-conditions-and-multimorbidity>.
- [9] Harold C. Sox, Marshall A. Blatt, Michael C. Higgins, and Keith I. Marton. *Medical Decision Making*. Butterworth-Heinemann, Boston, 1988.
- [10] John P. Cunningham and Byron M. Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, 2014. doi: 10.1038/nn.3776.
- [11] Ishani Ganguli, Arabella L. Simpkin, Cassandra Lupo, et al. Cascades of care after incidental findings in a US national survey of physicians. *JAMA Network Open*, 2(10):e1913325, 2019. doi: 10.1001/jamanetworkopen.2019.13325.
- [12] W. Ross Ashby. *An Introduction to Cybernetics*. Chapman & Hall, London, 1956.
- [13] Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10:1, 2023. doi: 10.1038/s41597-022-01899-x.
- [14] Juan A. Gallego, Matthew G. Perich, Lee E. Miller, and Sara A. Solla. Neural manifolds for the control of movement. *Neuron*, 94(5):978–984, 2017. doi: 10.1016/j.neuron.2017.05.025.