Evidence-Steered Medicine: A Safety-First Control Logic for Clinical Decision-Making

By **Konstantin Gurbanov**, MD, PhD[1]

Keywords: evidence-based medicine, patient safety, human-in-the-loop, uncertainty, de-implementation, dose optimization, clinical AI governance

[1]Independent Researcher, Berlin, Germany

Disclosure: The author declares that they have no relevant or material financial interests.

Submitted
February 8, 2026
Accepted
March 3, 2026
Published
April 15, 2026

Gurbanov K. Evidence-Steered Medicine: A Safety-First Control Logic for Clinical Decision-Making. *Patient Safety*. 2026;8(2):159210. doi:10.33940/001c.159210.

Abstract

Background: Evidence-based medicine strengthens decision-making, but contemporary care—including clinical artificial intelligence (AI)—often operates under uncertainty, heterogeneous patient contexts, and shifting performance. A common failure mode is committing too early to actions that are difficult to reverse, monitor, or repair.

Methods: We developed a pragmatic, safety-first control logic by synthesizing concepts from patient safety (including Safety-II), implementation science, and deployment risks in clinical decision support. We operationalized these concepts as a repeatable decision-episode discipline and derived testable hypotheses with pragmatic evaluation designs.

Results: Evidence-steered medicine (ESM) structures decisions as controlled micro-steps: (1) a brief support check, (2) uncertainty banding that constrains action strength, (3) a low-dose action grammar prioritizing reversible micro-interventions paired with short-horizon readouts, and (4) reason-coded governance that enables auditability, learning, and rapid de-escalation/repair. The model yields measurable predictions on severe safety events, recoverability (checkpointing and de-escalation pathways), time to detection of unsafe trajectories, learning efficiency from reason-code distributions, and (in AI workflows) automation-bias and drift-trigger events.

Conclusion: ESM complements evidence-based medicine by making uncertainty operational: It specifies how to act safely when evidence is incomplete. The hypotheses can be evaluated using retrospective replay, prospective pilots, and stepped-wedge rollouts without replacing standard of care.

Introduction

Evidence-based medicine (EBM) is often summarized as the conscientious integration of best research evidence with clinical expertise and patient values.[1] Its success has been amplified by systematic reviews and structured appraisal methods that reduce arbitrary practice variation.[2] At the same time, the epistemic and operational conditions of contemporary care have shifted. Clinical decisions increasingly occur under three simultaneous pressures:

- Evidence is abundant but unevenly applicable to a given patient
- Interventions interact across comorbidities, polypharmacy, and complex care pathways
- Clinical decision support, including clinical artificial intelligence (AI), introduces new forms of uncertainty such as model miscalibration, data shift, and automation bias[3]

In these conditions, the problem is not merely that a clinician lacks evidence; it is that the decision is treated as a one-time choice rather than a controlled sequence. Many harms arise not because the initial direction was unreasonable, but because the chosen action was too forceful, too difficult to monitor, or too difficult to undo once new information emerged. Patient safety scholarship has emphasized that safety is not only the absence of adverse events (Safety-I) but also the capacity to succeed under varying conditions through monitoring, adaptation, and recovery (Safety-II).[4] Implementation science similarly underscores that interventions should be designed for uptake, feedback, and local adaptation, rather than assuming idealized compliance.[5]

This paper proposes evidence-steered medicine (ESM), a safety-first control logic that structures everyday decisions as controlled microsteps. The central claim is that evidence should not only justify a chosen intervention, it should also steer the form of action—its reversibility, monitoring plan, and governance—so that uncertainty is managed through controlled microsteps. ESM is not meant to replace EBM. Rather, it supplements EBM with a discipline that emphasizes:

- Explicit uncertainty bands that determine admissible action strength
- Low-dose action grammars that prioritize reversible micro-interventions and short-horizon readouts
- Reason-coded logs that make decisions auditable and learnable

When evidence is indirect or trial populations differ from real-world patients, transportability methods formalize what assumptions are needed to generalize effects across settings and populations.[6]

De-implementation initiatives, such as the American Board of Internal Medicine Foundation's Choosing Wisely campaign, illustrate how safety and value can improve when low-value practices are explicitly identified, monitored, and reduced over time.[7]

Methods

We developed ESM as a pragmatic theory of action by synthesizing recurring safety and implementation challenges reported across patient safety, clinical decision-making, and clinical AI deployment literature, then translating these into an operational four-move control loop. We further derived testable hypotheses and example study designs to enable empirical evaluation. To illustrate applicability at the point of care, we include a brief worked example showing how uncertainty banding and the low-dose action grammar constrain admissible actions and specify monitoring.

Results

The Evidence-Steered Medicine (ESM) Model

ESM is defined as a four-move discipline that structures a decision episode from appraisal to governance. Each move is deliberately simple so it can be executed under routine time constraints, yet each move has a distinct safety function. **Figure 1** summarizes the loop and its feedback pathways.

Move 1: The Support Check

The support check is a brief, structured appraisal of what is known and what is merely assumed. It asks four questions:

- What is the strongest relevant evidence for the proposed action (guideline, trial, meta-analysis, mechanistic rationale)?
- For whom does that evidence apply (eligibility, comorbidity exclusions, age, frailty, baseline risk)?
- What is the expected time to benefit and time to harm?
- What competing actions are plausible (including inaction and de-escalation)?

Move 2: Uncertainty Bands As Action Constraints

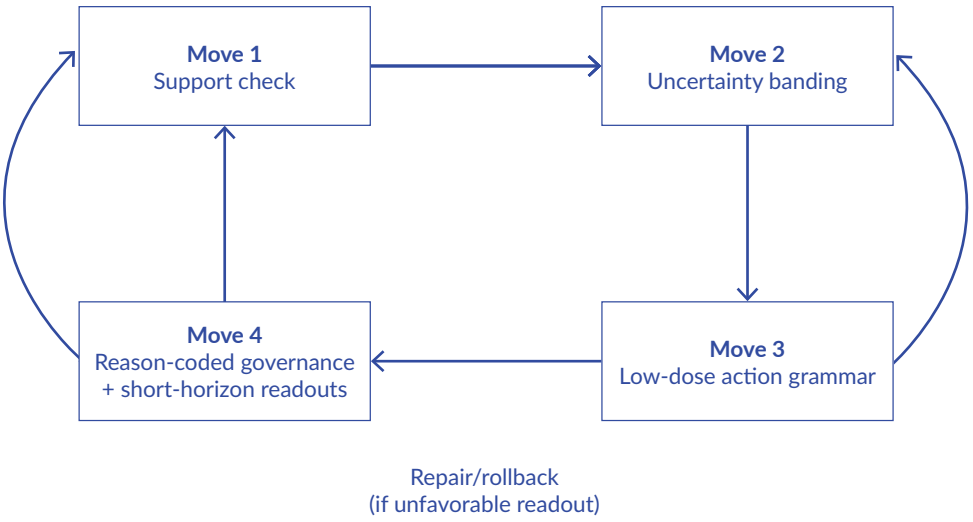
From the support check, the episode is assigned to an uncertainty band that constrains the admissible force of action. ESM proposes three default bands:

- Green (supported): evidence is strong and applicable; routine application is reasonable, with standard monitoring
- Amber (contestable): evidence is mixed, indirect, or weakly applicable; actions should be reversible-by-design and paired with short-horizon checkpoints
- Red (unsafe-to-commit): evidence is insufficient, applicability is poor, or potential harms are high relative to monitoring capacity; default is pause, escalation, or a conservative micro-intervention with immediate review

Move 3: The Low-Dose Action Grammar

Once uncertainty is explicit, ESM applies a low-dose action grammar: choose the smallest action that (a) is plausibly beneficial, (b) generates interpretable information within a short horizon, and (c) remains easy to stop or reverse if the trajectory is unfavorable.

Figure 1. The ESM Four-Move Loop



The grammar is not anti-treatment; it is pro-recoverability. In oncology, dose and schedule often matter as much as drug identity, and combination therapy can confer benefit via patient-to-patient variability even without drug synergy.[8-10] ESM generalizes this logic beyond oncology: It favors microsteps, rapid readouts, and default reversibility.

Move 4: Reason-Coded Governance and Repair

The fourth move turns the episode into an auditable unit of learning. ESM requires that each action be accompanied by a short, standardized reason code (e.g., "guideline-supported," "trial-ineligible patient," "toxicity risk dominates," "AI disagreement," "patient preference," "resource constraint"). These codes populate a local log that supports accountability, learning, and repair; unfavorable trajectories can be linked to the action and its rationale, enabling targeted correction.

Worked Example (Conceptual)

Example episode: A patient meets partial criteria for a guideline-supported intervention, but has comorbidities excluded from the pivotal trial. Support check identifies indirect applicability and uncertain time to harm → band assignment: Amber. Admissible action is constrained to a reversible microstep (e.g., lower-intensity initiation or a monitoring-first action) coupled to a short-horizon readout (e.g., safety labs between 24 and 72 hours, symptom tracking, or early physiologic signal). Prespecified thresholds determine escalation vs de-escalation. The action and rationale are logged using a reason code (e.g., "trial-ineligible patient") so patterns can be audited and protocols improved.

Testable Hypotheses Derived From ESM

ESM is proposed as a testable theory of safe action under uncertainty. The hypotheses in **Table 1** can be assessed in settings with or without clinical AI, using routinely collected outcome data and safety monitoring.

These hypotheses are empirically testable. If ESM does not measurably improve recoverability, monitoring performance, and auditability, then the proposed mechanism for harm reduction is not supported.

Discussion

ESM is offered as a theory of safe action under uncertainty: a compact control logic that complements EBM by ensuring that uncertainty is translated into operational constraints, monitoring, and repair pathways. Its plausibility rests on a well-established safety insight: When the world is variable, safety improves when systems are designed for observation, adaptation, and recoverability.[4]

Limitations deserve emphasis. First, ESM does not eliminate the need for clinical judgment; it structures judgment into a sequence that can be audited and improved. Second, ESM may be less applicable where actions are inherently irreversible and time-critical; in such cases, the main value may lie in reason-coded governance and post hoc learning rather than in low-dose steps. Third, ESM depends on feasible capture of short-horizon readouts; workflows without reliable outcome capture may require investment before ESM can be meaningfully evaluated.

ESM also suggests a constructive way to integrate clinical AI into care. Rather than asking clinicians to trust or distrust models globally, ESM asks institutions to specify what kinds of actions are permitted under what uncertainty conditions, and to govern decisions through reason-coded logs. In this framing, "human in the loop" becomes a concrete control policy rather than a slogan.[3]

Table 1. Testable Hypotheses Derived From the ESM Theory and Example Operationalizations

Hypothesis	Primary operational measure	Example evaluation design
H1 Safety events	High-severity adverse decision events per 1,000 eligible episodes	Prospective pilot; stepped-wedge rollout
H2 Recoverability	Share of episodes with documented checkpoint + deescalation pathway	Prospective pilot with process measures
H3 Early detection	Time from action to review/repair trigger (median, IQR)	Prospective pilot; retrospective replay
H4 Learning efficiency	Concentration of reason codes (e.g., top-k codes explaining escalations) and time-to-Qi cycle	Implementation evaluation with audit logs
H5 AI robustness	Rate of harmful automation bias events; audit triggers from clinician–model discordance	Deployment study in AI-supported workflow
H6 De-implementation	Change in inappropriate use rate for targeted low-value practice	Interrupted time series; stepped-wedge decision support

Abbreviations: IQR=interquartile range, Qi=quality index

Conclusion

Evidence-steered medicine (ESM) reframes patient-safety practice for clinical AI: Evidence should guide not only whether an intervention is justified, but also how it is executed under uncertainty—through reversible micro-actions; predefined monitoring windows; and explicit governance for escalation, auditability, and repair. ESM yields testable predictions and can be evaluated using pragmatic designs (retrospective replay, prospective pilots, stepped-wedge rollouts) without replacing standard of care.

Disclosures & Acknowledgments

Ethics Review is not applicable. This article reports a conceptual framework and did not involve studies with human participants, human data, or animals. There was no external funding.

The author used OpenAI ChatGPT for language editing and formatting support; all conceptual content, interpretations, and final editorial decisions are the author's own.

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence Based Medicine: What It Is and What It Isn't. *BMJ*. 1996;312(7023):71–72. doi:10.1136/bmj.312.7023.71

2. Mulrow CD. Rationale for systematic reviews. *BMJ*. 1994;309:597–599.

3. Bakken S. AI in Health: Keeping the Human in the Loop. *J Am Med Inform Assoc*. 2023;30(7):1225–1226. doi:10.1093/jamia/ocad091

4. Hollnagel E. *Safety-I and Safety-II: The Past and Future of Safety Management*. CRC Press; 2014. doi:10.1201/9781315607511.

5. Nilsen P, Ingvarsson S, Hasson H, von Thiele Schwartz U, Augustsson H. Theories, Models, and Frameworks for De-Implementation of Low-Value Care: A Scoping Review of the Literature. *Implement Res Pract*. 2020;1:2633489520953762. doi:10.1177/2633489520953762

6. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of Trial Results Using Inverse Odds of Sampling Weights. *Am J Epidemiol*. 2017;186(8):1010–1014. doi:10.1093/aje/kwx164

7. Rosenberg A, Agiro A, Gottlieb M, et al. Early Trends Among Seven Recommendations From the Choosing Wisely Campaign. *JAMA Intern Med*. 2015;175(12):1913–1920. doi:10.1001/jamainternmed.2015.5441

8. Palmer AC, Sorger PK. Combination Cancer Therapy Can Confer Benefit via Patient-to-Patient Variability Without Drug Additivity or Synergy. *Cell*. 2017(7);171:1678–1691.e13. doi:10.1016/j.cell.2017.11.009

9. Plana D, Palmer AC, Sorger PK. Independent Drug Action in Combination Therapy: Implications for Precision Oncology. *Cancer Discov*. 2022;12(3):606–624. doi:10.1158/2159-8290.CD-21-0212

10. Zirkelbach JF, Shah M, Vallejo J, et al. Improving Dose-Optimization Processes Used in Oncology Drug Development to Minimize Toxicity and Maximize Benefit to Patients. *J Clin Oncol*. 2022;40(30):3489–3500. doi:10.1200/JCO.22.00371

About the Author

Konstantin Gurbanov (konstantin.gurbanov@gmail.com) is a physician and pharmacovigilance and medical consultant. His work spans clinical development and post-marketing safety, including signal detection, benefit–risk evaluation, risk-management planning, labeling updates, and cross-functional safety governance, with experience across oncology/hematology and broader therapeutic areas.