Attention Reallocation: A Randomized Experiment on AI Decision Support and Clinician Information Processing

Sean Sylvia[a] Lirui Jiao[a] Sophia Fort[a] Junier Oliva[b] Tara Templin[a]

[a]*Department of Health Policy and Management, University of North Carolina at Chapel Hill*
[b]*Department of Computer Science, University of North Carolina at Chapel Hill*

February 2026

The global market for clinical decision support AI is estimated to exceed $2 billion annually, with systems deployed across radiology, emergency triage, oncology, and chronic disease management (DataM Intelligence 2025; Abbas, Jeong, and Lee 2025). Hospitals, insurers, and policy makers are investing heavily on the assumption that these tools improve clinical decisions. The evidence is far less encouraging. Vaccaro, Almaatouq, and Malone (2024) meta-analyze over 100 studies of human-AI teams and find that AI augmentation does not, on average, improve performance — and in some cases makes it measurably worse (g = -0.23). The technology is being deployed at scale, yet we do not reliably understand how it changes the way clinicians think.

The standard evaluation question, *does AI help?*, is incomplete. Agrawal, Gans, and Goldfarb (2018) characterize AI as a prediction cost reduction, implying that prediction and judgment are complements: as prediction cheapens, returns to judgment rise. Understanding the nature of this complementarity is essential for designing effective AI clinical decision support systems. The policy-relevant question is how AI restructures the decision process, and whether clinicians' behavioral response reflects this complementarity. A clinician confronted with an AI recommendation does not simply shift toward or away from the algorithm's output. She reorganizes which evidence she attends to, how she allocates effort across features, and how she weighs her own judgment against the system's. Understanding this reorganization, not just the net effect on accuracy, is what matters for designing systems and regulation grounded in evidence rather than assumption.

We provide experimental evidence on how providing clinical decision support in the form of AI risk predictions affects clinician information processing. Using a purpose-built experimental platform, we enrolled 56 healthcare professionals who provided screening recommendations for 2,015 colorectal cancer patient cases in a within-subjects randomized design. For each case review, clinicians were randomized into whether they received an AI-generated risk prediction alongside the

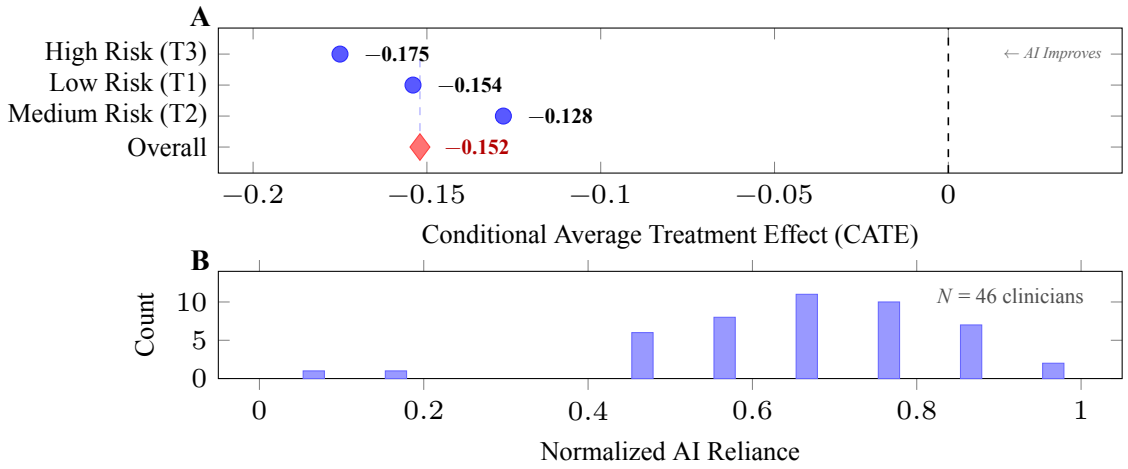
patient information. Approximately 19 percent of case reviews included AI predictions. Separately, we constructed two ground truth benchmarks to measure accuracy: consensus ratings from independent expert clinicians, and predictions from a machine learning model trained on real electronic health records from a major health system. The AI predictions shown to clinicians during the experiment are distinct from this ground truth benchmark: the former is the experimental treatment, while the latter provides an independent standard against which to evaluate clinician accuracy. The expert consensus benchmark captures clinical reasoning validity; the ML benchmark provides a population-level prediction standard with full case coverage.

The central innovation is the experimental design for identifying how AI changes information processing. Each case presented patient characteristics drawn from fields available in electronic health record systems (age, symptom burden, family history, comorbidities), and our platform independently randomized both whether an AI risk prediction was displayed and which patient characteristics were shown to the clinician. Because characteristic randomization is orthogonal to case content and clinicians are not told which characteristics are withheld, characteristic absence is uninformative conditional on case structure. This design lets us estimate the causal effect of any specific patient characteristic on a clinician's screening recommendation, and, critically, how the interactions between the display of AI risk predictions and the effects of each characteristic on clinical decision-making. Where Kleinberg et al. (2018, QJE) demonstrate the value of algorithmic prediction at the decision level, we extend this logic *within* a single decision, identifying which pieces of information clinicians delegate to the algorithm and which they continue to evaluate independently.

Turning to accuracy, we find that AI reduces the distance between clinician assessments and ground truth (Cohen's d = -0.14, p = 0.008 against ML predictions), increases exact agreement (OR = 1.36, 95% CI [1.06, 1.73], p = 0.015), and raises near-match accuracy (OR = 1.59, p = 0.020). A causal forest shows universal benefits across the case distribution, with the strongest effects for high-risk patients (CATE = -0.175) and complex cases with many observable characteristics. Case-order $\times$ AI interaction tests are jointly insignificant.

Beyond these average effects of AI clinical decision support on accuracy, our key finding is that AI fundamentally changes how clinicians weight patient information. The effects of displayed patient characteristics on clinician assessments interact significantly with the presence of AI decision support ($\chi^2(22) = 39.02$, p = 0.014). The pattern is asymmetric. With AI present, clinicians reduce their reliance on high-value diagnostic characteristics, those carrying the most predictive weight in the ML ground truth, while increasing attention to peripheral, low-value characteristics. The mean interaction coefficient for high-value characteristics is -0.129; for low-value characteristics, +0.037 (t = -2.86, p = 0.020). Total characteristic weights decline 39.5 percent under AI exposure, consistent with broad cognitive offloading.

The pattern of results is consistent with clinicians reallocating attention. One reading, rational delegation, holds that clinicians treat AI as a cognitive labor-saving device for prediction-intensive inferences, accepting algorithmic substitution on computationally demanding characteristics and redirecting freed attention toward characteristics requiring judgment the AI cannot supply. This mirrors rational inattention theory (Sims 2003; Maćkowiak and Wiederholt 2009): constrained agents allocate attention to maximize decision quality, and a tool that cheapens some signals shifts processing toward others. The competing reading, automation bias, holds that clinicians offload cognitively expensive characteristics regardless of the AI's actual comparative advantage on those



Note: Reliance normalized to [0,1] via min-max scaling of individual trust estimates (λ_i, range −0.27 to 0.34). Overall mixing parameter: $\lambda = 0.854$; confidence: $\beta = 0.18$, $p < 0.001$.

Figure 1: Treatment effect heterogeneity and clinician AI reliance. **(A)** CATEs from a causal forest by patient EHR risk level; all negative, with the strongest improvement for high-risk patients (diamond = overall mean). **(B)** Distribution of normalized individual AI reliance from a mixed-effects trust calibration model. Most clinicians cluster in the upper half of the reliance scale. $N = 46$ clinicians, 2,015 assessments.

dimensions. Distinguishing these accounts requires testing whether the magnitude of reallocation per characteristic tracks the AI's accuracy advantage.

Trust calibration estimates reinforce the delegation picture. Clinicians place approximately 85 percent weight on AI recommendations ($\lambda = 0.854$), and confidence rises significantly with AI exposure ($\beta = 0.18$, $p < 0.001$). The decision-maker who emerges from these data has substantially delegated information integration to the algorithm while retaining the role of evaluating individual cues. Whether this delegation is well-calibrated, that is, whether the AI truly outperforms on the characteristics clinicians offload most, determines whether the pattern reflects algorithm appreciation (Logg et al. 2019) rather than algorithm aversion (Dietvorst et al. 2015), and whether clinicians' reliance constitutes appropriate use or the kind of automation misuse documented by Parasuraman and Riley (1997).

This mechanism clarifies the puzzling heterogeneity in the broader literature. The meta-analytic null of Vaccaro, Almaatouq, and Malone (2024) is consistent with attention reallocation that sometimes improves outcomes — when clinicians offload characteristics the AI handles well — and sometimes does not — when they over-delegate to unreliable systems or redirect attention toward uninformative cues. The aggregate effect masks the behavioral shift driving it.

These findings have direct implications for the design and evaluation of clinical AI decision support. AI decision support changes not only the accuracy of clinical judgments but the process by which clinicians arrive at them, specifically how they allocate attention across available patient information. This implies that the design of AI clinical decision support systems should account for the nuances of how they affect the decision-making process of the humans using those predictions. In clinical settings and similar domains where AI predictions assist expert judgment, the complementarity between prediction and judgment, and the specific nature of that complementar-

ity, should be taken into account. Doing so may lead not only to improved outcomes but to more consistent decision-making more generally. Our findings are specific to the context studied: AI risk predictions for colorectal cancer screening recommendations. Whether similar patterns of attention reallocation arise with other forms of AI decision support, in other clinical domains, and under different institutional conditions is an important question for future research.

Funding: AI/ML for Advancing Health Equity and Researcher Diversity (AIM-AHEAD), National Institutes of Health; University of North Carolina at Chapel Hill. The experimental platform developed for this study (AugMed) is available at https://augmed.dhep.org.

References

Abbas, Q., Jeong, W., and Lee, S.W. (2025). Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges. *Healthcare,* 13(17), 2154. https://doi.org/10.3390/healthcare13172154

Agrawal, A., Gans, J., and Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence.* Harvard Business Review Press.

DataM Intelligence. (2025). Clinical Decision Support Systems Market Size, Growth and Industry Forecast 2025–2032. https://www.datamintelligence.com/research-report/clinical-decision-support-systems-market

Dietvorst, B.J., Simmons, J.P., and Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *Journal of Experimental Psychology: General,* 144(1), 114–126. https://doi.org/10.1037/xge0000033

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2018). Human Decisions and Machine Predictions. *Quarterly Journal of Economics,* 133(1), 237–293. https://doi.org/10.1093/qje/qjx032

Logg, J.M., Minson, J.A., and Moore, D.A. (2019). Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. *Organizational Behavior and Human Decision Processes,* 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Maćkowiak, B. and Wiederholt, M. (2009). Optimal Sticky Prices under Rational Inattention. *American Economic Review,* 99(3), 769–803. https://doi.org/10.1257/aer.99.3.769

Parasuraman, R. and Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors,* 39(2), 230–253. https://doi.org/10.1518/001872097778543886

Sims, C.A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics,* 50(3), 665–690. https://doi.org/10.1016/S0304-3932(03)00029-1

Vaccaro, M., Almaatouq, A., and Malone, T. (2024). When Combinations of Humans and AI Are Useful: A Systematic Review and Meta-Analysis. *Nature Human Behaviour,* 8(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1