

RESEARCH INSIGHT

AI Medical Devices Carry 2.5 Times the Recall Risk of Non-AI Devices

The added risk sits in software and design rather than manufacturing, and validation budgets for AI devices should move with it.

Anqi Wu (College of Business, Florida International University, anwu@fiu.edu) and Jong Youl Lee (College of Business, Florida International University, jonglee@fiu.edu)

The U.S. Food and Drug Administration had authorized about 60 AI-enabled medical devices by the end of 2016. It now clears more than 300 a year, and the cumulative total passed 1,600 in mid-2026. Nearly all reach the market through the 510(k) pathway, which requires no prospective clinical testing. Once an AI device is on the market, is it any more likely than a comparable conventional device to be recalled for a safety problem?

We linked the FDA's AI-Enabled Medical Devices List with registration, clearance, recall, and adverse event records for 26,535 devices cleared between 2004 and 2024, and used survival analysis to compare AI-enabled devices with conventional devices serving the same clinical function under the same regulatory standard. AI-enabled devices face roughly 2.5 times the hazard of a first recall. The result holds within product categories and years, after accounting for adverse event volume and manufacturer portfolio size, and in a stricter comparison limited to categories containing at least one AI device.

The nature of the risk matters as much as its size. The excess hazard is concentrated almost entirely in design- and administrative-related failures, including software execution errors, flawed algorithmic assumptions, inadequate validation, data handling problems, and user-interaction issues, with no comparable increase in manufacturing defects. About half of AI device recalls trace directly or indirectly to the AI functionality itself, and the hazard is highest for those recalls.

Two factors shape the excess recall hazard associated with AI. It is substantially larger for devices reviewed by FDA-accredited third-party organizations than for devices reviewed directly by FDA staff, despite the use of the same formal criteria. It is also smaller for manufacturers with stronger research capability. Together, both external review capacity and internal research capacity appear to buffer the risk that AI adds.

We then develop a decision support model that converts a device's failure risk profile into a split of the validation budget between design and software validation and manufacturing quality validation. A manufacturer that applies conventional benchmarks to an AI device systematically underinvests in design validation. In our calibration, the optimal design validation share rises from about 55 percent for a conventional device to about 78 percent for an AI device reviewed directly by the FDA, and higher still under third-party review. Stronger research capability reduces the required adjustment.

For manufacturers, the message is not simply to spend more on validation, but to rebalance validation resources toward algorithms, software, and broader design-related risks. For regulators, third-party review of AI-enabled devices stands out as a place where review capacity may lag the complexity of what is reviewed. Because the study is observational, these estimates should be interpreted as associations rather than causal effects. Even so, they give both manufacturers and regulators a concrete starting point for balancing speed to market against post-market safety.

Supporting Material

Table 1. Key findings on post-market safety of AI-enabled medical devices (AIMDs), FDA 510(k) devices cleared 2004 to 2024

Finding	Estimate	Evidence
Hazard of first recall, AIMDs vs. comparable non-AIMDs	2.52 times (hazard ratio)	Cox model with product code and year fixed effects, 26,535 devices, 3,932 recalls
Hazard under third-party review vs. direct FDA review	About 3.0 times under third-party review; 1.4 times and not statistically significant under FDA review	Interaction of AI indicator with review pathway
Where the excess risk sits	Concentrated in design and administrative failures (software, algorithmic assumptions, validation, data handling, human factors); no significant increase in manufacturing recalls	Interaction of AI indicator with failure type from FDA recall reasons
Role of manufacturer research capability	Excess hazard attenuates as research share of CMS Open Payments rises; robust to Compustat R&D measure	2018 to 2024 subsample, 1,579 devices from firms with payment activity
Share of AIMD recalls tied to AI functionality	About one half, directly or indirectly; hazard is highest for directly AI-related recalls	Manual classification of FDA recall narratives
Optimal design validation share (calibrated, $\beta V = 4$)	Non-AI device 55%; AI device under FDA review 78% (low research capability) to 59% (high); AI device under third-party review 80% (manufacturing floor binds) to 69%	Decision support model calibrated on the estimates above

Notes. Hazard ratios are exponentiated Cox coefficients. Standard errors are clustered at the firm level. Calibrated shares are conditional decision benchmarks and inherit the uncertainty of the empirical inputs.

References

Wu, A., & Lee, J. Y. (2026). AI-enabled medical devices and post-market safety: Empirical evidence and prescriptions for validation resource allocation. Working paper, Florida International University. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6818152