



Research Letter | AI in Health Policy

Early Recalls and Clinical Validation Gaps in Artificial Intelligence–Enabled Medical Devices

Branden Lee, BS; Patrick Kramer, BS; Sara Sandri, BS; Ritika Chanda, BS; Crystal Favorito, BS; Olivia Nasef, BS; Joseph S. Ross, MD, MHS; Joshua Sharfstein, MD; Tinglong Dai, PhD

Introduction

Artificial intelligence–enabled medical devices (AIMDs) are increasingly present in US clinical practice,¹ with nearly all cleared through the US Food and Drug Administration (FDA) 510(k) pathway.^{2–4} Because 510(k) clearance does not require prospective human testing, many AIMDs enter the market with limited or no clinical evaluation; meanwhile, recalls may undermine clinician and patient confidence in their performance. We examined whether lack of reported clinical validation and publicly traded manufacturer status were associated with AIMD recalls.

Methods

In this cross-sectional study, all FDA-cleared AIMDs¹ were matched to recall entries in the Center for Devices and Radiological Health database from November 15 to November 30, 2024. Each recall number–device pair was counted as 1 event; devices could contribute multiple events. Validation status was classified as none, retrospective dataset, or prospective trial based on FDA summaries. AIMD commercialization models were determined using S&P Capital IQ NetAdvantage and PrivCo databases using previously described schema.⁵ Manufacturer-attributed recall causes were categorized as diagnostic or measurement errors, functionality delay or loss, physical hazards, biochemical hazards, and postmarket changes. We followed the [STROBE](#) reporting guideline.

Kaplan-Meier analyses estimated recall-free survival; group differences were assessed using log-rank tests. Multivariable logistic regression (GraphPad Prism 10; 2-sided $\alpha = .05$) modeled the odds of any recall, adjusting for validation status, manufacturer type, clearance year (2020–2024 vs earlier), and clinical specialty. Because the study used publicly available data, it did not require institutional review board oversight or informed consent per the Common Rule (US 45 CFR 46).

Results

Among 950 AIMDs, 60 (6.3%) were associated with 182 recall events (mean [SD] recalls per device, 3.0 [4.3]) (**Figure 1A**). Diagnostic or measurement errors accounted for 109 recalls encompassing 935 063 units, followed by functionality delay (44 recalls, 755 647 units), physical hazards (14 recalls, 8192 units), and biochemical hazards (13 recalls, 76 257 units). Of these recalls, 79 (43.4%) occurred within the first 12 months of device clearance (**Figure 1B**), approximately double the rate reported for all 510(k) devices.⁴ Notably, 108 recalls (59.3%) remained unresolved at the study end date, with 20 recalls remaining unresolved for more than 3 years. Overall, recall-free survival was 96.6% at 1 year, 93.5% at 3 years, and 91.8% at 5 years (**Figure 1C**).

Devices without reported validation had more recalls per device (mean [SD], 3.4 [5.1]) than those with retrospective (1.9 [1.5]) or prospective (2.0 [1.4]) validation, and were associated with larger recalls (mean [SD] units recalled, 12 193 [82 405] vs 6523 [18 690]; $P < .001$). Public companies accounted for 53.2% of AIMDs but 91.8% of recalls and 98.7% of recalled units (**Figure 2A**). Among recalled devices from privately traded companies, 40% lacked clinical validation

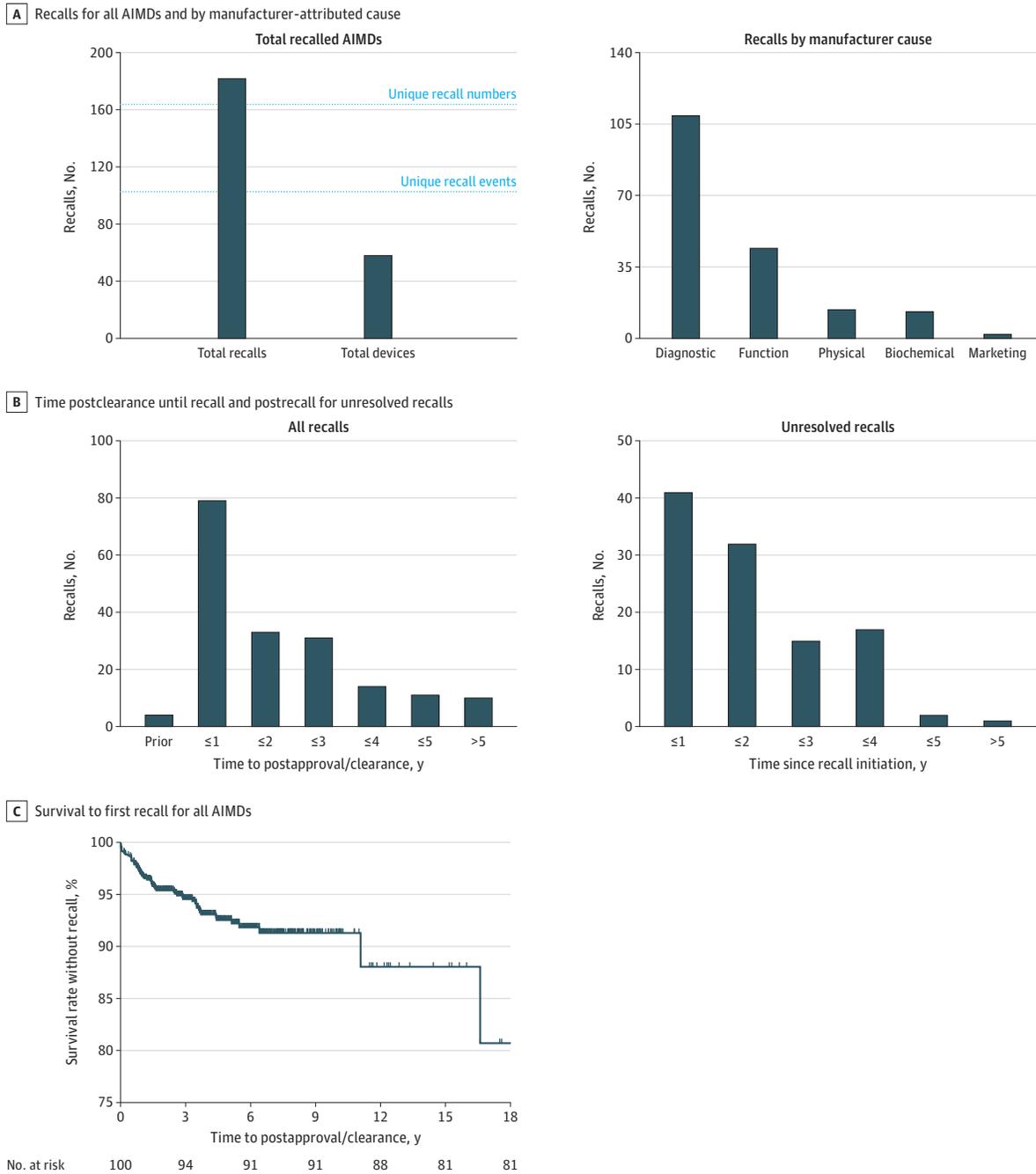
+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Open Access. This is an open access article distributed under the terms of the CC-BY License.

compared with 77.7% of those from established public companies and 96.9% from smaller public companies (Figure 2B). Multivariable analysis revealed that lack of clinical validation (odds ratio [OR], 2.8; 95% CI, 1.6-4.7) and public company status (OR, 5.9; 95% CI, 2.4-14.6) were independently associated with recall.

Figure 1. Overview of Artificial Intelligence-Enabled Medical Device (AIMD) Recalls



A, Total number of recalls (n = 182) and AIMDs recalled (n = 60) for all recalled AIMDs (left) and by manufacturer-attributed causes of recall (right). In the left panel, the total recalls represent each US Food and Drug Administration (FDA) recall number-device pairing, the unique recall numbers represent the distinct FDA-issued recall IDs, the unique recall events represent recall numbers arising from 1 issue, and the total devices

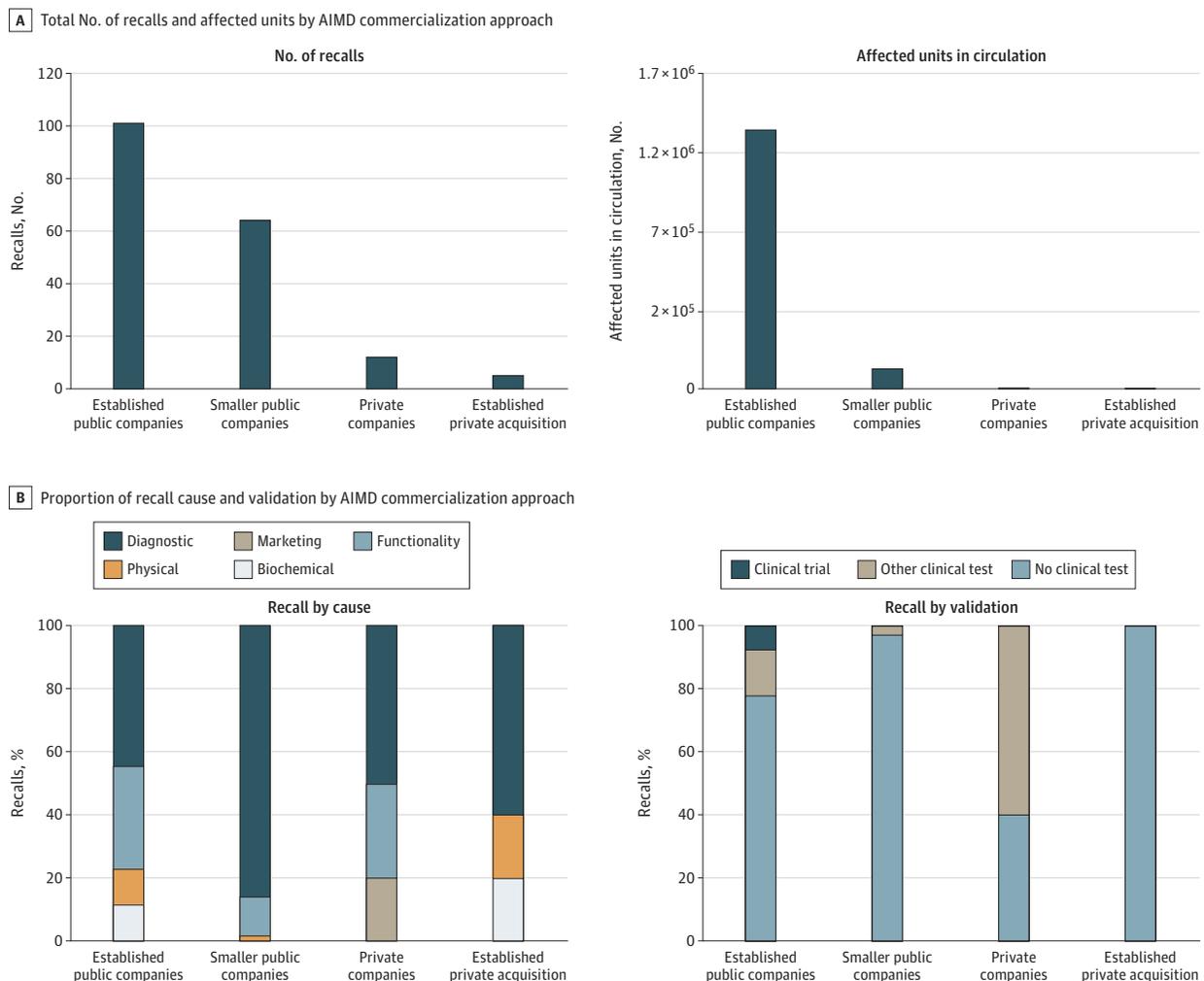
represent AIMDs with 1 or more recalls. B, Time elapsed between device clearance and recall initiation (left) and since recall initiation for nonterminated recalls (right). Each bin is a 1-year interval up to 5 years. C, Kaplan-Meier survival curve showing recall-free survival over time since FDA clearance.

Discussion

In this cross-sectional study, recalls of FDA-cleared AIMDs were uncommon but were concentrated early after clearance and predominantly involved products lacking clinical validation and manufactured by publicly traded companies, suggesting that the 510(k) process may overlook early performance failures of AI technologies. Requiring prospective evaluation or issuing time-limited clearances that lapse without confirmatory data may reduce these risks.²⁻⁴ Given that over 90% of recalled units were produced by public companies, heightened premarket clinical testing requirements and postmarket surveillance measures may improve identification and reduction of device errors, similar to risk-based strategies in pharmacovigilance.⁶ The association between public company status and higher recalls may reflect investor-driven pressure for faster launches, warranting further study.

Study limitations include reliance on publicly available validation reports, exclusion of software updates not labeled as recalls, and limited follow-up for devices cleared after 2022. Nevertheless, linking premarket evidence gaps and manufacturer type to postmarket actions offers practical guidance for regulators, clinicians, and health systems adopting AI-based tools.

Figure 2. Recall Counts, Cause, and Affected Units by Commercialization Model



A, Total number of recalled devices (left) and affected units (right) across commercialization models. For the segment break at 10 000 units, the scale below the segmentation is 5000 per tick mark, and the scale above the segmentation is 200 000

units per tick mark. B, Proportions of recall causes for recalled devices (left) and by differing degrees of clinical testing (right) across commercialization models.

ARTICLE INFORMATION**Accepted for Publication:** June 11, 2025.**Published:** August 22, 2025. doi:10.1001/jamahealthforum.2025.3172**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](#). © 2025 Lee B et al. *JAMA Health Forum*.**Corresponding Author:** Tinglong Dai, PhD, Carey Business School, Johns Hopkins University, 100 International Dr, Baltimore, MD 21202 (dai@jhu.edu).**Author Affiliations:** Department of Orthopaedic Surgery, Johns Hopkins Medicine, Baltimore, Maryland (Lee); School of Medicine, Johns Hopkins University, Baltimore, Maryland (Lee, Kramer, Sandri, Favorito, Nasef); Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Lee, Sharfstein); School of Medicine, Georgetown University, Washington, District of Columbia (Chanda); Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut (Ross); Department of Health Policy and Management, Yale School of Public Health, New Haven, Connecticut (Ross); Carey Business School, Johns Hopkins University, Baltimore, Maryland (Dai); Hopkins Business of Health Initiative, Johns Hopkins University, Washington, DC (Dai).**Author Contributions:** Mr Lee had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.*Concept and design:* Lee, Kramer, Sandri, Dai.*Acquisition, analysis, or interpretation of data:* All authors.*Drafting of the manuscript:* Lee, Kramer, Sandri, Chanda, Favorito, Dai.*Critical review of the manuscript for important intellectual content:* Lee, Kramer, Sandri, Nasef, Ross, Sharfstein, Dai.*Statistical analysis:* Lee, Kramer, Nasef.*Obtained funding:* Dai.*Administrative, technical, or material support:* Lee, Sandri, Dai.*Supervision:* Lee, Dai.**Conflict of Interest Disclosures:** Dr Ross reported receiving grants from the US Food and Drug Administration, Johnson & Johnson, the Medical Devices Innovation Consortium, the Agency for Healthcare Research and Quality, the National Heart, Lung, and Blood Institute, and Arnold Ventures outside the submitted work and reported serving as an expert witness in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc that was settled September 2022. Dr Dai reported collaborating on multiple studies with Digital Diagnostics (without receiving financial payments), serving as co-lead of the Bloomberg Distinguished Professorship Cluster on Global Advances in Medical AI at Johns Hopkins University, and serving as co-chair of the Johns Hopkins Workgroup on AI and Healthcare. No other disclosures were reported.**Funding/Support:** Access to financial databases was generously provided by the Welch Medical Library at Johns Hopkins University and Johns Hopkins Medicine. Dr Dai received funding from a Johns Hopkins University Discovery Award for the project Purposeful Design for AI-Augmented Healthcare: Harnessing Physician-in-the-Loop Systems to Improve the Patient Journey.**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.**Data Sharing Statement:** See the [Supplement](#).**REFERENCES**

1. Artificial intelligence–enabled medical devices. US Food and Drug Administration. Updated March 2024. Accessed November 15, 2024. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices>
2. Muehlematter UJ, Bluethgen C, Vokinger KN. FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. *Lancet Digit Health*. 2023;5(9):e618–e626. doi:10.1016/S2589-7500(23)00126-7
3. Mooghali M, Ross JS, Kadakia KT, Dhruva SS. Availability of unique device identifiers for class I medical device recalls from 2018 to 2022. *JAMA Intern Med*. 2023;183(7):735–737. doi:10.1001/jamainternmed.2023.0727

4. Dubin JR, Simon SD, Norrell K, Perera J, Gowen J, Cil A. Risk of recall among medical devices undergoing US Food and Drug Administration 510(k) clearance and premarket approval, 2008-2017. *JAMA Netw Open*. 2021;4(5):e217274.

doi:[10.1001/jamanetworkopen.2021.7274](https://doi.org/10.1001/jamanetworkopen.2021.7274)

5. Lee B, Patel S, Favorito C, et al. Development and commercialization pathways of AI medical devices in the United States: implications for safety and regulatory oversight. *NEJM AI*. 2025;2(7).

doi:[10.1056/AIra2500061](https://doi.org/10.1056/AIra2500061)

6. Handley JL, Krevat SA, Fong A, Ratwani RM. Artificial intelligence related safety issues associated with FDA medical device reports. *NPJ Digit Med*. 2024;7(1):351.

doi:[10.1038/s41746-024-01357-5](https://doi.org/10.1038/s41746-024-01357-5)

SUPPLEMENT.

Data Sharing Statement