Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

FDA regulatory lifecycles of robotic platform generations under clinician control

jia Zeng
West China Hospital of Sichuan University
Jiao Luo
West China Hospital of Sichuan University
hui Xu
West China Hospital of Sichuan University
Linjun Yun
244144417@qq.com

West China Hospital of Sichuan University

Research Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11601193/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose

To examine how reported patient-based clinical evidence relates to clinical-use expansion across US Food and Drug Administration (FDA) authorisations of clinician-controlled robotic platform generations.

Methods

We assembled a source-supported inception cohort first authorised between 1 January 2015 and 20 August 2026. Platform-level (CORE) decisions were linked within technical generations. Post-entry expansions and other changes were compared using risk differences and whole-generation bootstrap intervals. Kaplan–Meier analysis described first expansion; post-hoc analyses examined evidence definitions, generation boundaries and unresolved candidates.

Results

The cohort comprised 26 generations and 109 CORE decisions. Among 78 complete post-entry decisions, clinical evidence was reported or specifically cited in 21/26 expansions versus 2/52 non-expansions: risk difference 76.9 percentage points (generation-bootstrap 95% interval 46.9–93.1). Restricting the endpoint to device-specific descriptive or stronger evidence reduced the difference to 53.8 points. Twelve generations expanded and 14 were censored; median time was 1,499 days (95% interval 657 days to not estimable). Specified candidate and missing-label scenarios gave risk differences of 59.3–78.8 points and median times of 1,314–2,395 days; these finite ranges did not bound unidentified platforms.

Conclusion

Clinical-use expansions more often reported patient-based clinical evidence, but the magnitude depended on what counted as evidence. Decision-linked histories distinguish authorised-use changes from technical modification; they do not establish comparative patient benefit, clinical adoption or complete ascertainment of eligible FDA platforms.

Introduction

A robotic platform can acquire new authorised clinical uses while its instruments, imaging and control functions continue to change. A single authorisation or product name therefore provides only a partial account of its development. Reviews have mapped the autonomy and regulatory routes of FDA-cleared surgical robots [1] and the evaluation maturity of minimally invasive platforms [2]. For teams considering

a new use or configuration, however, the relevant question is which clinical evidence supports that particular change.

The IDEAL framework treats surgical robotics evaluation as a process extending from development to comparative assessment and long-term monitoring [3]. Public FDA records also permit examination of the evidence supporting device authorisation, as demonstrated in research on novel therapeutic devices entering through De Novo [4]. These perspectives motivate a longitudinal question: within a technical generation, do decisions that extend clinical use report patient-based evidence more often than decisions addressing other platform changes? Linking decisions is necessary because the presence of evidence at entry does not establish its applicability to later uses or configurations.

We therefore reconstructed generation-linked FDA histories and compared reported patient-based clinical evidence between subsequent clinical-use expansions and other platform changes. Secondary objectives were time from entry to first expansion and documented predicate ancestry after De Novo entry. The aim was to describe evidence reporting through regulatory development, rather than compare the clinical effectiveness of robotic systems.

Methods

Design and ascertainment

This retrospective inception cohort analysed public FDA 510(k) and De Novo metadata, decision documents, indications for use and technical sources. The 510(k) pathway assesses substantial equivalence to a legally marketed predicate [5]; De Novo provides a classification route for certain devices without such a predicate [6]. Each generation entered at its first authorisation between 1 January 2015 and 20 August 2026 and was followed to the common cutoff. Earlier generations were ancestry context unless sources supported substantive redesign. The inception boundary was clarified post hoc; no patient-level dataset was assembled.

Official bulk metadata retrieved on 25 October 2026 contained 36,133 in-window identifiers (35,772 K; 361 DEN). A technical-term, 22-product-code and all-De-Novo query identified 1,645 records, all initially screened. Reconciliation with a 354-record historical register, including 96 outside query hits, yielded 1,741 identifiers; targeted authorisation-chain tracing added three, giving 1,744. Metadata screening, source-description review and technical adjudication were recorded separately. Query-negative metadata and inaccessible documents were not technical exclusions. Exact queries, dispositions and source locators are in Online Resources 2 and 3.

Study unit and generation assignment

The retained unit was a reusable integrated platform with source-supported software-mediated, continuous direct clinician control of powered patient-side procedural movement. Input had to be

operable spatially apart from the execution body, including detachable controllers or controllers on another instrument. Planning, approval or start/stop of an autonomous task alone did not qualify. The scope covered invasive soft-tissue, endoluminal, natural-orifice and microsurgical procedures and robotic bronchoscopy. Powered endoscope positioning and directly controlled needle guides could qualify despite manual insertion.

Exclusions comprised bone/dental systems, stereotactic neural navigation, radiotherapy/HIFU, electrophysiology robots, hand-guided cooperative supports, task-autonomous procedures, isolated single-purpose actuators, self-contained handheld instruments and entirely single-use systems. Disposable patient-contact parts did not exclude reusable control platforms. These post-hoc boundaries were study choices, not a universal FDA definition. Cross-source inferences concerning invendoscopy software/control correspondence and the SIRIUS reusable-control configuration were flagged and tested by stricter-source omission.

A generation required an identifiable authorised integrated platform with substantive architecture, core-control or instrument-interface redesign relative to its own predecessor. Rebranding, accessories and external-manufacturer predicates alone were insufficient. Commercial supply was not an eligibility gate; removal of an inconsistent supply condition was a post-hoc amendment. ARES and MONARCH were retained separately and also analysed jointly.

Decision roles and clinical evidence

CORE decisions concerned the integrated platform, authorised clinical use or platform-level software, control, interface or architecture. Isolated accessory, instrument, reprocessing and use-life decisions were SATELLITE unless platform-level involvement was established. Each CORE decision received one primary role (Table 1). Independently, expansion was YES, NO or UNCLEAR for an added specialty, procedure, patient population, anatomical site or approach; entry established baseline use. Technical change was graded 0 (none), 1 (one domain), 2 (multiple domains/system-level) or 3 (index/core architecture).

Table 1
Characteristics of index generations and post-entry CORE decisions Panel A Index generations N = 26

Characteristic	Category	n (%)
FDA entry route	510(k)	19 (73.1%)
	De Novo	7 (26.9%)
Post-entry CORE per generation	0	9 (34.6%)
	1–2	6 (23.1%)
	3–5	6 (23.1%)
	≥ 6	5 (19.2%)
Panel B Post-entry CORE decisions N = 83		

Patient-based clinical evidence was YES, NO or UNCLEAR for identifiable clinical procedural-performance or safety results reported or specifically cited in that decision's public dossier. Clinical technical accuracy qualified; comparative benefit, a unique-patient denominator and newly generated data were not required. Cadaver, animal, simulation and human-factors-only evaluations were excluded. Explicit patient-safety wording within human-factors sections was flagged for permissive sensitivity analysis. Generic predicate/usability citations and external publications did not automatically transfer clinical results.

Descriptive evidence intensity was 0 (none), 1 (indirect/reused patient evidence), 2 (device-specific descriptive results), 3 (prospective comparative evidence) or 4 (randomised evidence). This was not a validated quality grade or a measure of newly generated data. Online Resource 2 maps source units and known reuse; the analysis unit remained an FDA decision.

Statistical and sensitivity analyses

The complete-case comparison excluded UNCLEAR expansion/evidence labels without recoding them NO. We calculated Wilson proportions, Newcombe–Wilson risk differences, log-Wald risk ratios and two-sided Fisher probabilities. Interpretation emphasised 10,000 whole-generation bootstrap resamples of contributing complete-case generations (sorted index identifier; NumPy default_rng/PCG64; seed 20260820; linear percentiles). Non-estimable draws were counted; replicate-specific 0.5 correction applied only to zero-cell bootstrap risk ratios. Conventional decision-level estimates did not adjust for clustering.

Follow-up ran from index authorisation to first subsequent confirmed CORE expansion; event-free generations were administratively censored on 20 August 2026. Kaplan–Meier (KM) estimates used Greenwood log–log limits, complemented/reversed for cumulative expansion; median intervals inverted

those limits. Risk counts were immediately before labelled times, with years=days/365.25. Withdrawal and succession were not measured competing events. Formal same-generation De Novo descent required documented predicate paths, not names, references or chronology.

Post-hoc analyses examined evidence intensity and labels, generation omission, scope boundaries, documented commercial evidence and ARES/MONARCH merging. Four endpoint definitions addressed technical accuracy and human-factors safety wording. A separate 27-scenario set examined stricter E200/E210 and SIRIUS omission, broader Spiro-VISTA/iotaSOFT units, candidate-specific generation/CORE states and four entry-date-only timing tests. The latter did not estimate candidate-specific risk differences without supported later CORE mapping.

A finite joint set retained the original 26 generations and examined 720 candidate configurations; completing unknown post-entry binary fields yielded 41,472 combinations. It was not crossed with every source omission or broader-unit scenario. Conditional intervals and logical extrema were distinguished: extrema were not confidence intervals or bounds on unidentified platforms. Definitions, missingness, selectors and exploratory comparisons are in Online Resources 1 and 3.

Source verification and adjudication

Jia Zeng and Jiao Luo independently reviewed source documents and coding for all included decisions, covering eligibility, generation, CORE classification, expansion and clinical evidence. Disagreements were discussed, with Linjun Yun adjudicating if consensus was not reached. Classifications and statistical results were checked against sources and the final dataset; unresolved limitations were retained.

Results

Cohort and clinical evidence

The cohort contained 26 generations and 109 unique CORE decisions: 26 index and 83 post-entry; 19 generations entered through 510(k) and seven through De Novo (Table 1; Fig. 1). Nine remained entry-only. Expansion was YES/NO/UNCLEAR in 27/55/1 post-entry decisions, independently of primary regulatory role. Clinical evidence was YES/NO/UNCLEAR in 35/67/7 decisions overall and 23/56/4 post-entry. Table 1 gives the distribution of primary roles.

Five post-entry decisions had unknown binary labels: clinical evidence for K230694, K232773, K251761 and K252528, and expansion for K260695. Among 78 complete decisions from 17 generations, clinical evidence appeared in 21/26 expansions (80.8%) versus 2/52 non-expansions (3.8%): risk difference 76.9 points (Newcombe 95% CI 56.2–88.0; generation-bootstrap interval 46.9–93.1; Table 2; Fig. 2). No primary bootstrap draw was non-estimable. Conventional risk-ratio and Fisher estimates are reported in Table 2.

Table 2
Clinical-use expansion and reported patient-based clinical evidence Panel A
Complete-case post-entry decisions n = 78

Expansion	Clinical evidence YES n (%)	Clinical evidence NO n (%)	Total
Yes	21 (80.8%)	5 (19.2%)	26
No	2 (3.8%)	50 (96.2%)	52
Total	23 (29.5%)	55 (70.5%)	78
Panel B Effect estimates and uncertainty			

Restricting clinical evidence to device-specific descriptive or stronger results yielded a 53.8-point difference. Excluding technical-accuracy-only evidence gave 78.8 points; including explicit human-factors safety wording gave 75.0 points; both changes together gave 76.9 points. Thirty-two completions of the five within-cohort unknowns gave 65.9–77.9 points without altering first-expansion timing, because K260695 followed its generation's already documented first expansion.

Expansion timing and boundary scenarios

Twelve generations expanded and 14 were administratively censored. Median time was 1,499 days (95% CI 657 days to not estimable; Fig. 3); only one generation remained at risk at ten years. Merging ARES/MONARCH gave 25 generations and a 1,314-day median. Omitting both E200/E210 and SIRIUS gave 24 generations/105 CORE decisions, a 76.8-point risk difference and a 1,314-day median.

Broader Spiro-VISTA inclusion added one INDEX and no subsequent comparison record. Broader iotaSOFT inclusion required both DEN190055 and its actual same-generation paediatric expansion K252339 (expansion/evidence YES); together with Spiro this yielded 28 generations/112 CORE decisions, a 77.6-point difference and 13 expansion events. Across 720 conditional candidate configurations and 41,472 binary completions, denominators were 26–35 generations/109–119 CORE, risk differences 59.3–78.8 points and medians 1,314–2,395 days. These were finite modelling ranges, not observed membership estimates (Online Resource 1).

Regulatory ancestry and source limitations

Six of seven De Novo generations had 19 chronological same-generation 510(k) follow-ups; 18 had documented formal predicate paths to their own De Novo entry. Anovo K260804 formally cited Versius DEN230078 and identified an Anovo decision as reference; references were not counted as predicates.

The 1,744-record register included 109 CORE and 59 SATELLITE decisions alongside exclusions, context and unresolved candidates. Five original eligibility questions, two additional source-access referrals and one conditional Vathin follow-up remained distinct. Targeted tracing added iotaSOFT K252339 and two Pristine records outside the original query/register; none established new retained-cohort membership.

Earlier Pristine/Vathin authorisations prevented treating later unresolved records as proven first entries. Dispositions and review depth are in Online Resources 2 and 3.

Discussion

Clinical-use expansions were more often accompanied by identifiable patient-based evidence than other post-entry platform changes. Linking decisions within technical generations shows where evidence appears during regulatory development, complementing platform inventories and evaluation reviews [1, 2]. The principal finding is the contrast by type of change, rather than an overall evidence count. Its magnitude decreased when evidence had to be device-specific, while time to expansion was more sensitive to generation boundaries. These distinctions matter when interpreting how a platform's authorised capabilities evolve.

A plausible explanation is that extending use to another procedure, anatomical site or patient population raises clinical questions not fully addressed by the original evidence. FDA guidance describes circumstances in which modified indications or technological differences may require clinical data, particularly when nonclinical testing cannot adequately address performance [5]. Conversely, some software, imaging or interface changes may be evaluated chiefly through engineering and nonclinical testing. However, we did not observe reviewers' requests or manufacturers' study decisions. Differences in public-summary detail could also contribute. The association therefore cannot establish that expansion caused additional research, or that non-expansion decisions were inadequately validated.

The reduction from 76.9 to 53.8 points under the device-specific endpoint shows that the measured contrast depends on evidence attribution and definition. Reuse can support a regulatory argument, but the number of evidence-positive decisions does not count independent studies. Device-specific evidence likewise need not have been generated for that particular submission. Ficarra and colleagues found that most studies of novel multiport urological platforms were noncomparative developmental or exploratory series [7]. Our findings add a different perspective: evidence may accompany a new authorised use without resolving the comparative effectiveness questions emphasised by IDEAL [3].

Clinical procedural performance and comparative patient benefit also answer different questions. The retained Micromate/Navi+ records illustrate this distinction: clinical guidance accuracy qualified for our endpoint, yet 54 Navi+ interventions did not establish 54 unique patients or comparative benefit (Online Resource 2). By contrast, Kim and colleagues examined recovery and patient-reported scar outcomes after SP versus Single-Site colectomy [8]; Obayashi and colleagues examined perioperative, functional, oncological and economic outcomes after hinotori versus da Vinci prostatectomy [9]. These retrospective clinical comparisons address outcomes that a binary dossier label cannot capture. They were not used to recode FDA evidence labels, and their findings cannot be quantitatively compared with our regulatory risk difference.

The 1,499-day median describes time to an additional authorised clinical use, not time to clinical adoption, regulatory review duration or a universal innovation cycle. Platforms entering with broad

indications may have fewer opportunities for a coded expansion; later entrants have less follow-up. Changing generation assignments can alter entry dates and risk sets; omission changes the composition of the risk set. An INDEX-only addition can therefore affect KM estimates while contributing nothing to the post-entry evidence comparison. The conditional 1,314–2,395-day range makes that dependence visible, rather than defining a population development benchmark. Moreover, retirement and succession were unmeasured and only one generation remained at risk at ten years. Event-free follow-up should not be interpreted as technical failure, and late cumulative-expansion estimates require particular caution.

Predicate ancestry and technical continuity also answer different questions. De Novo classification can establish a predicate for later 510(k) submissions [6, 10]. Eighteen of the 19 chronological follow-up decisions had documented same-generation predicate paths to their own De Novo entry. The Anovo decision citing Versius illustrates why a brand-based history cannot substitute for reading the formal predicate and reference sections. Conversely, a predicate link does not establish an unchanged control architecture or make an earlier clinical cohort applicable to every subsequent configuration. Technical-generation assignment and the relevance of supporting evidence therefore require separate judgments.

For surgical and perioperative teams, the practical implication is to review an update at the level of its intended use and configuration. Relevant questions include whether the evidence concerns the proposed procedure and population, whether it evaluates the current device or a predecessor, and whether its denominator represents patients, procedures or interventions. Changes to control, instruments or imaging should also inform local discussions about training and monitoring. This approach is consistent with IDEAL's emphasis on iterative evaluation and with governance recommendations concerning credentialing and quality assurance [3, 11]. Our study supports organising that review; it did not test an implementation strategy or measure improvements in safety or cost.

Several limitations constrain these interpretations. First, ascertainment was source-supported rather than exhaustive. Directed tracing found three records outside the original query/register, and technical eligibility was not established for all query-negative records. Exact-model gaps and the cross-source inferences for E200/E210 and SIRIUS remain relevant even after verification. Second, study-unit, inception, vascular, commercial-gate and generation boundaries were clarified post hoc. The finite candidate analyses quantify consequences of specified alternatives; they neither assign membership probabilities nor bound undiscovered eligible systems. Retrospective source retrieval also does not reconstruct exactly what was publicly accessible at each decision date.

Third, the pooled comparison is not a within-platform paired or confounder-adjusted effect. Whole-generation resampling addresses observed within-generation dependence, but does not remove differences in specialty, regulatory pathway, calendar period, platform maturity or initial indication breadth. Only 17 generations contributed complete cases, with two evidence-positive non-expansions; repeated clinical material and variable disclosure further limit inference. Evidence intensity was descriptive, not a validated quality assessment. Future work should combine prospectively specified

generation rules, audits of query-negative records and explicit links to clinical study identifiers, while recording adoption, withdrawal and succession. Such data would address uncertainty that additional hypothetical configurations alone cannot resolve.

Conclusion

Within this source-supported cohort, clinical-use expansions more often reported patient-based clinical evidence than other platform changes. The contrast depended on evidence definitions, while expansion timing depended on generation and entry assumptions. Assessing a robotic update therefore requires separating its authorised use, technical configuration and supporting clinical evidence; these histories do not establish comparative patient benefit.

Declarations

Competing interests

The authors declare no relevant financial or non-financial competing interests.

Ethics approval and consent to participate

Not applicable. Public regulatory documents were analysed without collecting patient-level or identifiable participant data; no institutional approval, exemption or waiver is claimed.

Consent for publication

Not applicable. No identifiable participant materials are included.

Supplementary information captions

Online Resource 1

Statistical methods and sensitivity analyses, including conventional estimates, source and study-unit scenarios, finite candidate configurations, missing-label analyses, regulatory ancestry and expansion timing

Online Resource 2

Ascertainment and source-review appendix, including screening dispositions, eligibility and generation boundaries, source verification, retained-cohort clinical-evidence mapping and separate conditional-candidate mapping

Online Resource 3

Data and executable code for the retained-cohort analyses, metadata-query replay, source register, clinical-evidence mapping, endpoint definitions and conditional scenario analyses

Funding

No funding, grants or other financial support were received.

Author Contribution

Jia Zeng contributed to conceptualisation, data curation, source verification, and drafting the original manuscript. Jiao Luo contributed to data collection, independent source review, and verification of coding records. Hui Xu contributed to data collection and verification. Linjun Yun was responsible for study design, supervision, adjudication of unresolved disagreements, and critical revision of the manuscript. All authors reviewed and approved the final manuscript.

Data Availability

The data analysed in this study were derived from publicly available US Food and Drug Administration (FDA) 510(k) and De Novo regulatory records. The curated decision-level datasets, source register, classification records, supporting source documentation, and analysis scripts are provided in Online Resource 3 and its accompanying data and code archive. Detailed statistical methods, sensitivity analyses, and source verification information are available in Online Resources 1 and 2. The original regulatory records are publicly accessible through FDA databases, with source identifiers and references provided in the supplementary materials. No patient-level data were collected.

References

1. Lee A, Baker TS, Bederson JB, Rapoport BI (2024) Levels of autonomy in FDA-cleared surgical robots: a systematic review. npj Digit Med 7:103. https://doi.org/10.1038/s41746-024-01102-y
2. Boal M, Di Giovene C, Tesfai F et al (2024) Evaluation status of current and emerging minimally invasive robotic surgical platforms. Surg Endosc 38:554–585. https://doi.org/10.1007/s00464-023-10554-4
3. Marcus HJ, Ramirez PT, Khan DZ, IDEAL Robotics Colloquium et al (2024) The IDEAL framework for surgical robotics: development, comparative evaluation and long-term monitoring. Nat Med 30:61–75. https://doi.org/10.1038/s41591-023-02732-7
4. Johnston JL, Dhruva SS, Ross JS, Rathi VK (2020) Clinical evidence supporting US Food and Drug Administration clearance of novel therapeutic devices via the De Novo pathway between 2011 and

2019. JAMA Intern Med 180:1701–1703. https://doi.org/10.1001/jamainternmed.2020.3214

5. US Food and Drug Administration (2014) The 510(k) Program: Evaluating Substantial Equivalence in Premarket Notifications [510(k)]. Guidance for Industry and Food and Drug Administration Staff. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/510k-program-evaluating-substantial-equivalence-premarket-notifications-510k. Accessed 3 November 2026
6. US Food and Drug Administration (2021) De Novo Classification Process (Evaluation of Automatic Class III Designation). Guidance for Industry and Food and Drug Administration Staff. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/de-novo-classification-process-evaluation-automatic-class-iii-designation. Accessed 3 November 2026
7. Ficarra V, Rossanese M, Giannarini G et al (2024) Evaluation of clinical research on novel multiport robotic platforms for urological surgery according to the IDEAL framework: a systematic review of the literature. Eur Urol Open Sci 67:7–25. https://doi.org/10.1016/j.euros.2024.06.014
8. Kim KE, Jeong WK, Baek SK, Bae SU (2025) Robot-assisted colectomy for left-sided colon cancer: comparison of da Vinci SP and Single-Site platforms. Int J Med Robot 21:e70079. https://doi.org/10.1002/rcs.70079
9. Obayashi K, Akatsuka J, Nishino T et al (2025) Clinical and economic assessment of robot-assisted radical prostatectomy with the hinotori surgical system: insights from a Japanese single-centre study. Int J Med Robot 21:e70084. https://doi.org/10.1002/rcs.70084
10. Aboy M, Crespo C, Stern A (2024) Beyond the 510(k): the regulation of novel moderate-risk medical devices, intellectual property considerations, and innovation incentives in the FDA's De Novo pathway. npj Digit Med 7:29. https://doi.org/10.1038/s41746-024-01021-y
11. Burke E, Devine M, Harkins P, Fenn S, Khan MF, McGuire BB (2025) Robotic surgery governance structures: a systematic review. J Robot Surg 19:218. https://doi.org/10.1007/s11701-025-02356-8

Figures

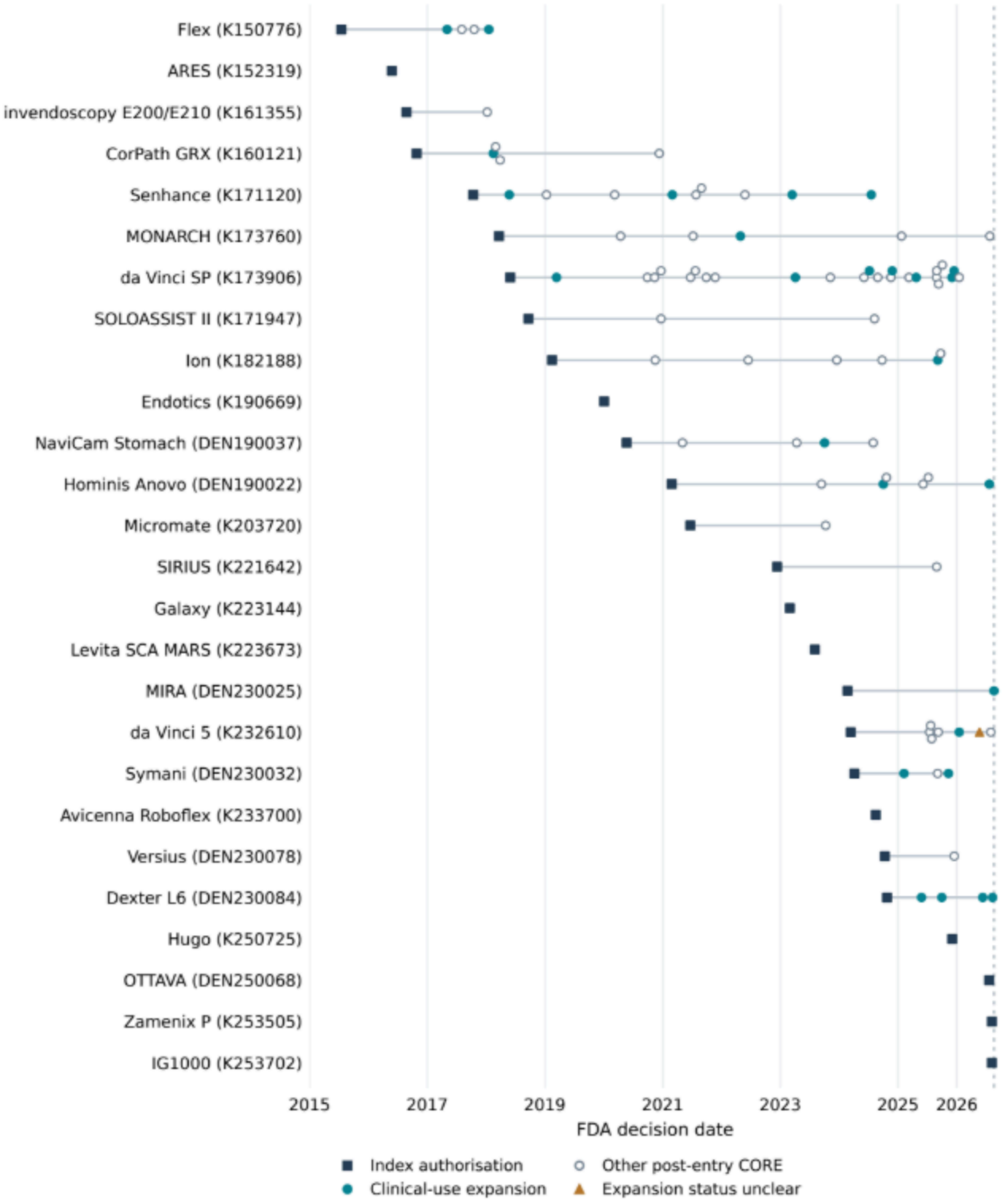

Figure 1

FDA timelines of 109 CORE decisions across 26 generations. Squares indicate index authorisation; filled circles, clinical-use expansion; open circles, other post-entry CORE; triangles, unclear expansion. Vertical offsets separate nearby records without changing dates. Connecting lines do not imply continuous commercial use; the dashed line marks 20 August 2026

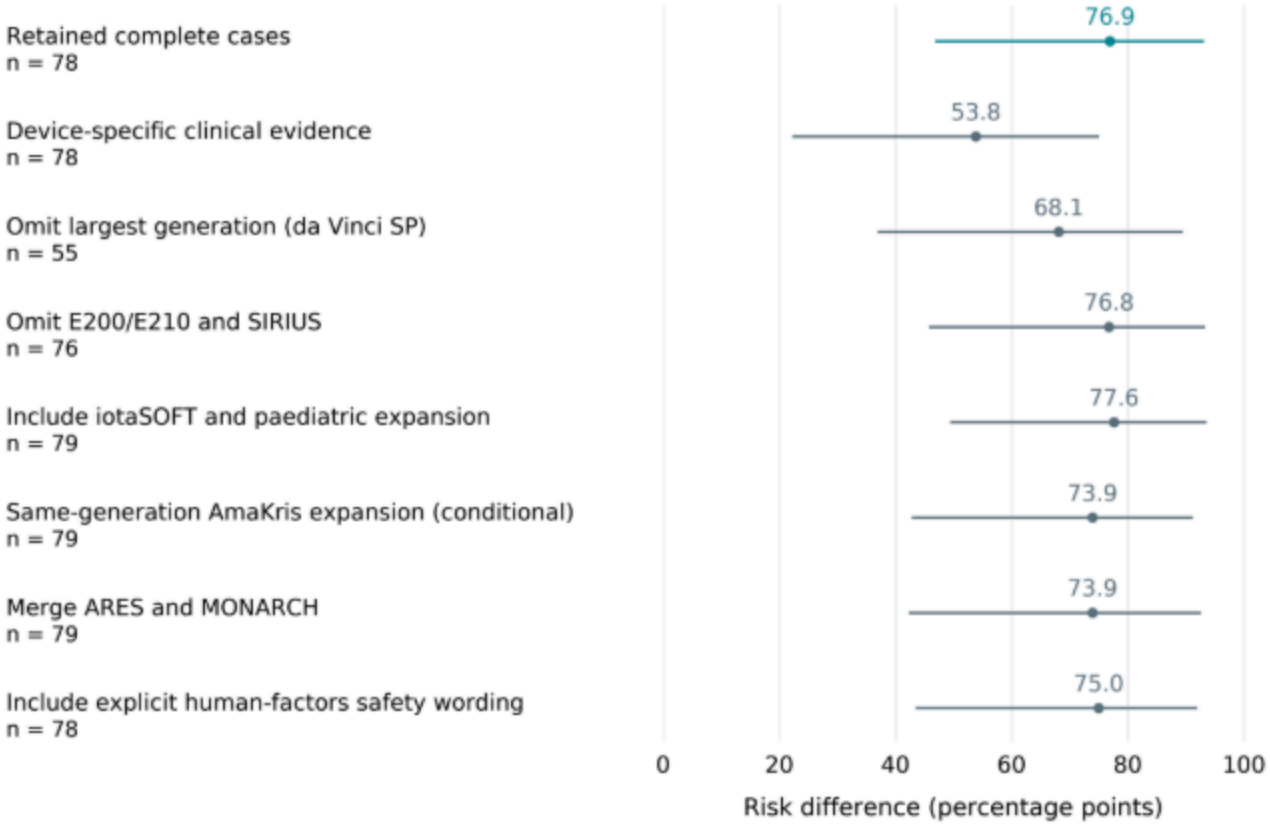

Figure 2

Risk differences under selected evidence, source, generation and study-unit interpretations. Bars are 95% percentile intervals from 10,000 whole-generation bootstrap resamples; n is the complete post-entry comparison size. Candidate AmaKris inclusion is conditional on technical eligibility and same-generation assignment. Intervals condition on each stated scenario and do not bound unidentified eligible platforms

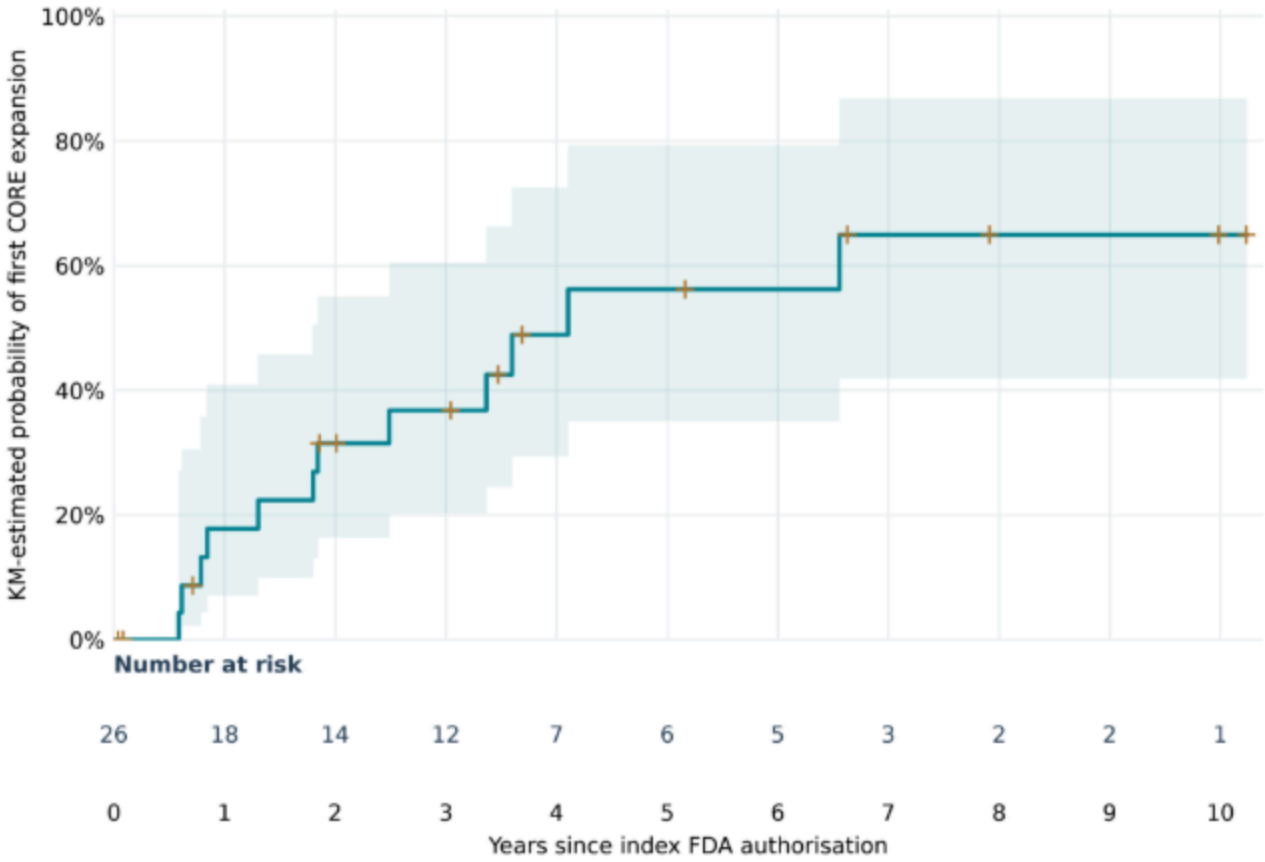

Figure 3

Kaplan–Meier-estimated probability of first subsequent CORE clinical-use expansion in 26 generations. Shading complements/reverses Greenwood log–log 95% survival limits; crosses indicate administrative censoring at 20 August 2026. Risk counts are immediately before labelled times. One generation remains at ten years; late estimates are unstable

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.docx
- ESM2.docx
- ESM3.docx