Sağlık Kuruluşlarında Karar Zekâsı: İnsan Uzmanlığı ile Yapay Zekâyı Bütünleştiren Bir Yönetişim Çerçevesi

Özgür ÖZMEN
Nişantaşı Üniversitesi
dr.ozgurozmen@gmail.com
https://orcid.org/0009-0001-5244-6106

ÖZET

Bu çalışmanın amacı, sağlık kuruluşlarının yapay zekâ destekli kararları yönetme kapasitesini kavramsal olarak tanımlamak ve bu kapasiteyi kamuya açık verilerle gözlemlemektir. Sağlık kuruluşları, yapay zekâyı klinik karar süreçlerine, bu kararları yönetecek kapasiteyi geliştirdiklerinden daha hızlı dâhil etmektedir. Çalışma, karar zekâsını bir örgütsel yönetişim kavramı olarak geliştirmektedir: kararların insan uzmanları ile algoritmik sistemler arasında bilinçli biçimde dağıtılması (epistemik tahsis), algoritma performansının kuruluşun kendi bağlamında doğrulanması (yerel doğrulama) ve kararlar melez hâle geldiğinde hesap verebilirliğin izlenebilir tutulması (hesap verebilirlik bütünleştirmesi). Kavram iki tür kanıtla incelenmiştir. Birincisi, Amerika Birleşik Devletleri Gıda ve İlaç Dairesi'nin yetkilendirme listesindeki 1.524 yapay zekâ ve makine öğrenmesi temelli tıbbi cihazın tamamının (1995-2026 kararları) sayımıdır; cihazların hangi klinik alanlara girdiği, hangi düzenleyici yol üzerinden girdiği ve bu dağılımların yoğunlaşması betimsel olarak çözümlenmiştir. İkincisi, çerçeve ışığında okunan üç belgelenmiş örgütsel vakanın yapılandırılmış karşılaştırmasıdır. Yetkilendirmeler 2015 yılında 6 iken 2025 yılında 333'e yükselmiş; cihazların yüzde 96,2'si önceden onaylı bir eşdeğere dayanan 510(k) yolundan girmiş; yazında bildirildiğine göre bu yol, premarket onay ve De Novo yollarının gerektirdiği bağımsız yeni klinik kanıtı genellikle gerektirmemektedir, bu pay benimseme ölçeklendikçe artmıştır. Yetkilendirmeler tıbbın geneline yayılmak yerine radyolojide yoğunlaşmış, tedarik tarafı ise dağınık kalmıştır. Üç vaka, yönetişim başarısızlığının genel değil mekanizmaya özgü olabileceğine işaret etmektedir: her vaka tahsis, doğrulama ve hesap verebilirlik boyutlarında farklı bir güç ve zayıflık örüntüsü sergilemekte, yanlış mekanizmaya yönelen çözümler asıl kırılganlığı gidermemektedir. Sonuç olarak çalışma, yönetişimi teknolojiden kararın kendisine taşıyan tanısal bir kavram sunmakta ve sağlık kuruluşlarına karar riskiyle orantılı gözetim tasarımı için bir yapı kazandırmaktadır.

Anahtar Kelimeler: *yapay zekâ, klinik karar destek sistemleri, cihaz onayı, sağlık bakımının sunumu, sağlık bakım kalitesi.*

Makalenin Geliş Tarihi: 24/07/2026 - Makale Kabul Tarihi: 29/09/2026
Sorumlu Yazar: Özgür Özmen
DOİ:10.17932/IAU.ASD.2015.007/asd_v012i3003

Decision Intelligence In Healthcare Organizations: A Governance Framework Integrating Human Expertise And Artificial Intelligence

ABSTRACT

This study specifies the capability healthcare organizations need to govern decisions made with artificial intelligence, and observes it in publicly available data. Healthcare organizations are admitting artificial intelligence into clinical decisions faster than they build the capability to govern them. This paper develops decision intelligence as an organizational governance construct: the capability to allocate decisions between human professionals and algorithmic systems deliberately (epistemic allocation), to validate algorithmic performance in the organization's own context (local validation), and to keep accountability traceable once decisions become hybrid (accountability integration). The construct is examined through two forms of evidence. The first is a census of all 1,524 artificial intelligence and machine learning enabled medical devices in the United States Food and Drug Administration authorization list, covering decisions from 1995 to 2026, analysed for the clinical domains algorithms enter, the regulatory pathway they take, and the concentration of both. The second is a structured comparison of three documented organizational cases seen through the framework. The authorization record increased from six devices in 2015 to 333 in 2025. Across the record, 96.2 per cent entered through the 510(k) predicate-equivalence route, a pathway generally reported not to require the independent new clinical evidence involved in premarket approval and De Novo review. The predominance of this route grew as adoption increased. Radiology accounted for most authorizations instead of an even spread across specialties. At the same time, suppliers remained widely dispersed. The three cases reveal different kinds of governance weakness. Problems in allocation, validation and accountability do not necessarily occur together, so an intervention aimed at one mechanism may do little to correct a weakness in another. The study contributes a diagnostic construct that relocates governance from the technology to the decision, and a structure for designing oversight proportionate to decision risk.

Keywords: *artificial intelligence, decision support systems, clinical, device approval, delivery of health care, quality of health care.*

INTRODUCTION

Healthcare organizations are admitting algorithmic systems into consequential clinical decisions faster than they are building the capability to govern those decisions, and the published record now contains enough documented episodes to study that asymmetry rather than merely deplore it. This paper contributes a decision-level governance construct, decision intelligence, by defining it as three separable organizational mechanisms and examining it against two forms of real evidence: a complete census of the artificial intelligence and machine learning enabled medical devices authorized for clinical use in the United States, and a structured comparison of three documented organizational cases. The construct relocates the analytical unit of healthcare artificial intelligence governance from the technology to the decision, and the evidence indicates that governance failure in clinical artificial intelligence is mechanism-specific rather than generic.

Three encounters with clinical artificial intelligence illustrate the problem. In 2021, investigators at Michigan Medicine published an external validation of a sepsis prediction model embedded in an electronic health record platform used by hundreds of hospitals, and found that a tool already alerting clinicians in daily practice discriminated poorly, was badly calibrated, and generated a volume of alerts likely to fatigue the clinicians it was meant to assist (Wong et al., 2021). Some years earlier, the University of Texas System Audit Office documented how one of the world's most respected cancer centres spent at least sixty-two million United States dollars on a Watson-powered advisory system procured largely outside the institution's standard oversight processes, a system that never entered clinical use (University of Texas System Audit Office [UTSAO], 2016). In the same period a large academic health system was building something different: a standing oversight structure that subjects every algorithm intended for patient care to defined checkpoints across its whole lifecycle (Bedoya et al., 2022). Three organizations, three encounters, three very different governance outcomes.

What separates these outcomes is not the sophistication of the underlying technology, and it is not the presence or absence of governance in the generic sense, since all three organizations possessed boards, compliance functions and clinical quality structures of recognized standing. The difference lies in something more specific: the organizational capability to govern the decision itself, that is, to determine deliberately which decisions or decision components should be delegated to an algorithm, to verify in the organization's own patient population that the algorithm performs as its developer claims, and to keep accountability traceable once human judgement and machine output are jointly implicated in a clinical choice. This study names that capability decision intelligence and treats it as the proper object of governance.

A substantial literature now addresses artificial intelligence governance in healthcare, and this study builds on it rather than beside it. Proposed governance models specify principles, fairness, transparency, trustworthiness and accountability, and translational pathways for moving models from development into practice (Bedoya et al., 2022; Reddy et al., 2020). What this literature has not yet supplied is a construct that explains why organizations with comparable governance principles produce divergent governance outcomes, and that locates the source of divergence at the level of the decision rather than the level of the technology or the policy document. The absence matters in practice as well as in theory: an organization that cannot identify which specific governance capability it lacks will tend to respond to failures with generic remedies, another committee or another policy, that leave the operative weakness untouched.

The study is guided by one primary question and two subsidiary questions. Its primary question asks how the capability to govern algorithmic decisions can be specified and observed in healthcare organizations. The first subsidiary question asks which clinical decisions the sector is admitting algorithms into, and through which regulatory route those algorithms arrive. The second subsidiary question considers how the documented organizations performed across the mechanisms identified by the framework. The two evidence strands address different parts of the inquiry and sit at different levels of analysis: the device census, whose unit is the authorized device, informs the first subsidiary question, while the comparison of organizational cases, whose unit is the organization, addresses the second. Together they provide the empirical basis for considering the primary question, with the census supplying the setting and the cases supplying the organizational observation.

The remainder of the article is organized as follows. Section 2 sets out the decision intelligence framework, explains its three mechanisms and develops four propositions. Section 3 describes the materials and methods used for the device census and the comparison of the three organizational cases. The findings are presented in Section 4. Section 5 considers their theoretical and managerial implications, the boundaries of the study and directions for further research, before Section 6 concludes. Across these sections, the central argument is that healthcare AI governance is better understood at the level of the decision than at the level of the technology. The census describes the environment in which the three mechanisms have to operate, while the case evidence is used to consider how weaknesses in decision allocation, local validation and accountability can produce different forms of governance failure.

THE DECISION INTELLIGENCE FRAMEWORK

1. Bounded rationality, professional judgement and an unoccupied position

Organizational decision-making has long been understood as taking place under limits to human rationality. Decision-makers rarely have complete information, unlimited time or unlimited cognitive capacity; instead, they work within these constraints and settle for workable courses of action. Organizations partly exist to make this constrained form of decision-making manageable (Simon, 1955). Clinical practice presents the same problem in a particularly demanding setting. Seen in this way, algorithmic decision support extends an established organizational effort to support judgement under conditions of bounded rationality. Yet machine learning introduces a governance problem that earlier decision aids did not create to the same degree. Guidelines, checklists and formularies could generally be inspected and understood by the professionals using them. Many contemporary models are far less transparent. As a result, the division of decision work between clinician and algorithm cannot simply be allowed to develop by default; it requires an explicit governance choice.

Healthcare also differs from many organizational settings because clinical work is professional work. It depends on trained discretion in situations that cannot be completely standardized and is governed to a substantial extent by the profession itself (Freidson, 2001). A hospital decision is therefore more than a task that can be optimized for efficiency. Clinical decisions remain within professional jurisdiction, with the duties that follow from licensure, liability and accountability to professional peers. The presence of an algorithm does not transfer or remove those duties. Even when an algorithm materially influences the choice made, the clinician remains legally and morally accountable for the clinical decision. Governance is therefore inadequate if the use of an algorithm makes it difficult to identify where human responsibility ultimately rests, irrespective of the model's technical accuracy. The healthcare artificial intelligence governance literature acknowledges this in principle, listing accountability among its requirements (Reddy et al., 2020), but a principle is not yet a mechanism. At the intersection of bounded rationality, the sociology of professions and the governance literature lies an unoccupied position: a construct specifying the capability by which a healthcare organization governs the division and integration of decision labour between accountable professionals and algorithmic systems. Decision intelligence is proposed for that position.

2. Definition and three mechanisms

Decision intelligence is defined here as the organizational capability to allocate decision authority between human professionals and algorithmic systems deliberately, to validate algorithmic performance in the organization's own operating context before and during reliance on it, and to maintain traceable

accountability for decisions jointly produced by human judgement and algorithmic output. The unit of analysis is the decision, not the technology. An organization does not, on this account, govern an algorithm; it governs the class of decisions into which the algorithm has been admitted, and the same algorithm may be well governed in one decision class and ungoverned in another within a single institution. The framework is configurational and process-based, and it specifies three mechanisms whose joint operation constitutes the capability.

Epistemic allocation is the mechanism by which an organization decides, explicitly and in advance, which decisions or components of decisions are delegated to algorithmic systems, which are reserved to professional judgement, and which are shared under defined conditions. It is weak when the delegation decision is made implicitly, by procurement, by default configuration or by vendor design, rather than by a deliberate institutional judgement about which decision components the algorithm may legitimately occupy.

Local validation is the mechanism by which an organization verifies that an algorithm's claimed performance holds in its own patient population, case mix and workflow, before clinical reliance begins and continuously thereafter. It rests on a fact about machine learning systems that distinguishes them from most prior clinical technologies: their performance is conditional on the statistical properties of the data environment in which they operate, so developer-reported performance is a hypothesis about the adopting site, not a property the tool carries with it (Habib et al., 2021; Wong et al., 2021). Local validation is weak when an organization treats vendor evidence, regulatory clearance or peer adoption as substitutes for verification in its own context.

Accountability integration is the mechanism by which responsibility for a hybrid decision remains traceable to identifiable, accountable actors across the decision's whole lifecycle, from the choice to deploy the algorithm through each clinical use to eventual retirement. Its warrant is the professional logic already discussed: since the accountable professional cannot transfer liability to a model, governance must ensure that ownership of the model, ownership of the clinical decision and ownership of the monitoring obligation are explicitly assigned. It is weak when a tool operates with no named clinical owner, when monitoring obligations are unassigned, or when the organization cannot reconstruct, after the fact, who decided that the algorithm should participate in a given class of decisions.

What this decomposition adds is best seen by contrast with what the existing literature already provides. Principle-based models tell an organization that accountability matters; the present framework tells it where accountability breaks,

at allocation, at validation or at integration, and the evidence assembled below indicates that these are empirically distinct failure sites rather than restatements of one another. Table 1 summarizes the three mechanisms.

Table 1. The three mechanisms of decision intelligence

Mechanism	Governance question it answers	Characteristic weakness	Theoretical grounding
Epistemic allocation	Which decisions, or decision components, may the algorithm occupy, and on whose authority?	Delegation occurs by procurement, default configuration or vendor design rather than institutional judgement	Simon (1955); Freidson (2001)
Local validation	Does the algorithm perform as claimed in this organization's patients, case mix and workflow?	Vendor evidence, regulatory clearance or peer adoption treated as substitutes for verification in context	Wong et al. (2021); Habib et al. (2021)
Accountability integration	Who owns the model, the clinical decision and the monitoring obligation at every lifecycle stage?	Tools operate with no named clinical owner; monitoring unassigned; deployment decisions unreconstructable	Freidson (2001); Reddy et al. (2020); Bedoya et al. (2022)

Source: author's own elaboration.

3. Propositions

Four propositions follow from the framework. They are offered as theoretically grounded expectations to structure future empirical work, and they are examined below, to the extent the present evidence permits, against a census and three cases rather than tested as pre-specified hypotheses.

Proposition 1. Healthcare organizations with equivalent formal governance policies will differ in realized governance outcomes according to the strength of epistemic allocation, local validation and accountability integration, and these three mechanisms will vary at least partly independently of one another.

Proposition 2. Algorithmic tools admitted to clinical use through channels that bypass epistemic allocation, procurement-led adoption, default activation within a larger platform, or executive sponsorship outside standing governance, will exhibit higher rates of post-deployment failure and abandonment than tools admitted through deliberate allocation, independent of the tools' technical quality.

Proposition 3. The gap between developer-reported and locally observed algorithmic performance will be larger where the adopting organization's patient population, case mix or clinical workflow differs more from the development environment, making local validation most consequential precisely where organizational incentives to rely on vendor evidence are strongest.

Proposition 4. Sustained decision intelligence requires all three mechanisms jointly: organizations strong in two mechanisms but weak in the third will exhibit governance failures characteristic of the weak mechanism, so that assessment should report a three-mechanism profile rather than a composite score.

MATERIALS AND METHODS

This study uses a mixed descriptive design with two evidence strands unified by the framework. The first strand is a quantitative census of the population of artificial intelligence and machine learning enabled medical devices authorized for clinical use. Its unit of analysis is the authorized device, so it characterizes the environment in which decision intelligence has to be exercised, that is, which clinical decisions algorithms are arriving in and through which evidentiary route, and not the capability of any organization. The second strand is a structured comparison of three documented organizational cases. Its unit of analysis is the organization, and it is therefore the only strand in which the three mechanisms are observed directly. The two strands are kept analytically separate because they sit at different levels, and no inference is drawn from the device level to the organizational level: the census bounds the problem the framework addresses, while the cases carry whatever observation of the mechanisms the study offers.

1. Device census: data source, variables and procedure

For the quantitative analysis, the study uses the full public list of artificial intelligence and machine learning enabled medical devices authorized by the United States Food and Drug Administration. The record was retrieved on 19 August 2026 and includes final decisions through 30 March 2026 (U.S. Food and Drug Administration [FDA], 2026). The unit of analysis is the authorized device, and the analysed population is the authorization list itself, a census of all 1,524 devices it records between 1995 and 2026 rather than a sample; the list is maintained by the regulator for administrative purposes and is not an exhaustive inventory of every AI-enabled device in clinical use, a point returned to in the limitations. Analysing this list is an established empirical approach to the artificial intelligence device landscape (Benjamens et al., 2020; Muehlematter et al., 2021).

Three variables were used directly from the record, the year of final decision, the lead clinical panel (specialty), and the company, and two were derived by transparent rules. The clearance pathway was parsed from the submission-number prefix into 510(k), De Novo and premarket approval (PMA), the distinction that matters for the local-validation argument because the 510(k) route clears a device by demonstrating substantial equivalence to an existing predicate and generally does not require the independent new clinical evidence that the De Novo and premarket approval routes involve (Muehlematter et al., 2023; Wu et

al., 2021). Decision years were grouped into four periods (1995 to 2015, 2016 to 2020, 2021 to 2025, and the partial year 2026) to describe change over time. The dataset was built with a reproducible pipeline that emits its own transformation log, and a three per cent verification sample drawn against the source record returned complete agreement. The analytical file regenerates from the raw record by a single command, and the record contains no personal data. Because the study uses publicly available, de-identified secondary data, no ethics-committee approval was required.

2. Case strand: selection and coding

The qualitative strand develops a structured comparison of three cases, each documented in the peer-reviewed literature, a formal audit record, or both, and each selected because it exposes a different configuration of the framework's mechanisms under real institutional conditions. The cases are the Algorithm-Based Clinical Decision Support oversight programme at Duke Health (Bedoya et al., 2022); the Oncology Expert Advisor collaboration between the University of Texas MD Anderson Cancer Center and IBM Watson, as documented by the University of Texas System Audit Office (UTSAO, 2016); and the multi-site deployment of the Epic Sepsis Model read through its external validation at Michigan Medicine (Habib et al., 2021; Wong et al., 2021). Selection followed documentary richness and configurational contrast rather than sampling from a defined population. Each case was examined against all three mechanisms, and the cross-case pattern was then read against the propositions of Section 2.3.

3. Statistical analysis

Analyses are descriptive of an enumerated population and were conducted in Python (version 3) with the numpy and scipy libraries. Reported quantities are counts, percentages and period-specific distributions; concentration is summarized with the Herfindahl-Hirschman index (HHI), computed as the sum of squared shares, for clinical panels and for companies. A chi-square test of the pathway-by-period contingency table is reported to describe the non-uniformity of the pathway composition across periods within the population; as the data are a census, the test describes population structure rather than supporting inference to a superpopulation. In keeping with the descriptive design, all statements are held to a descriptive and associational reading, and no causal language is used. Percentages after the arithmetic distributions are shares of the relevant total.

RESULTS

1. The authorized device population

Authorizations of artificial intelligence and machine learning enabled medical devices accelerated sharply over the study period. The count rises from six in 2015 to 333 in 2025, a compound annual increase of about 49 per cent across those

ten years, and 71.3 per cent of the entire 1995 to 2026 record falls within 2021 to 2025 alone. The 92 devices recorded in the partial year 2026 are consistent with a continuing high rate. Figure 1 shows the annual pattern.

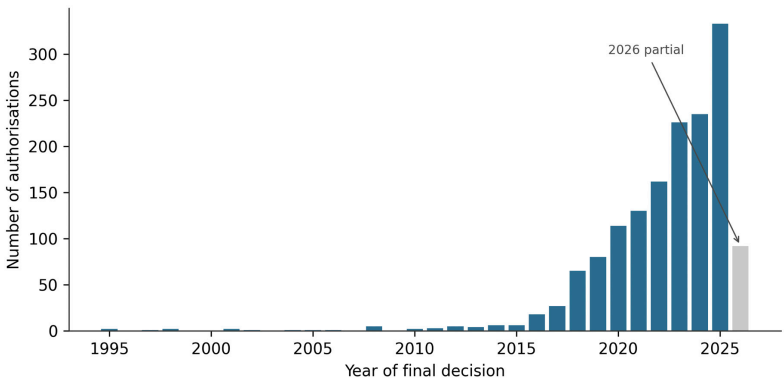


Figure 1. Annual authorizations of AI/ML-enabled medical devices, 1995 to 2026 (2026 partial).

The regulatory route through which these devices entered clinical use is heavily skewed. Across the whole population, 96.2 per cent (1,466 devices) were cleared through the 510(k) predicate-equivalence route, against 2.6 per cent (39) through De Novo and 1.2 per cent (19) through premarket approval. The pattern intensified as adoption scaled. The 510(k) share rose from 73.8 per cent in 1995 to 2015 to 97.1 per cent in 2021 to 2025, while the combined De Novo and premarket approval share, the routes generally reported to involve a higher clinical-evidence burden (Muehlematter et al., 2023; Wu et al., 2021), fell from 26.2 per cent to 2.9 per cent over the same span. Within the enumerated population the pathway composition is not uniform across periods (chi-square = 144.4, degrees of freedom = 4, $p < 0.001$, reported descriptively for a census). Figure 2 displays the shift.

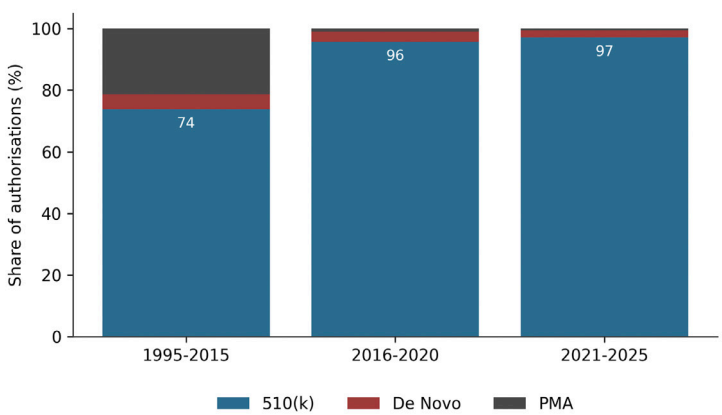


Figure 2. Clearance-pathway composition of authorizations by period.

The clinical domains into which algorithms are admitted are highly concentrated, and the concentration deepened rather than broadened over time. Authorizations span 18 clinical panels, but a single panel, radiology, accounts for 76.4 per cent of the whole population, and the panel-level concentration index for the full record is 0.596. Read by period, radiology's share rose from 35.7 per cent in 1995 to 2015 to 79.5 per cent in 2021 to 2025, and panel concentration rose from an index of 0.181 to 0.641. The population therefore does not show artificial intelligence spreading evenly across medicine; it shows authorizations concentrating in image-interpretation decisions as the field scaled. Figure 3 presents radiology's share alongside the concentration index by period, and Table 2 sets out the combined descriptive profile.

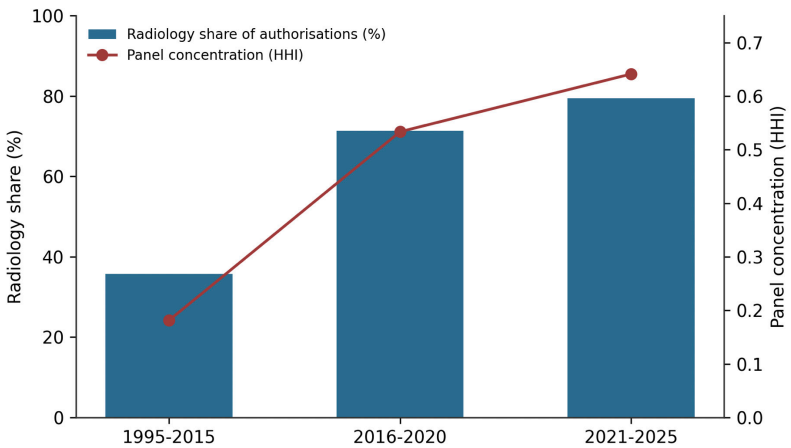


Figure 3. Radiology share of authorizations and panel concentration (HHI) by period.

Table 2. Descriptive profile of authorized AI/ML-enabled medical devices by period

Period	Devices	510(k) %	De Novo %	PMA %	Radiology %	Panel HHI
1995-2015	42	73.8	4.8	21.4	35.7	0.181
2016-2020	304	95.7	3.3	1.0	71.4	0.534
2021-2025	1,086	97.1	2.3	0.6	79.5	0.641
2026 (partial)	92	96.7	2.2	1.1	75.0	0.578
Total	1,524	96.2	2.6	1.2	76.4	0.596

Source: author's analysis of the FDA AI/ML-enabled medical devices list (FDA, 2026). HHI denotes the Herfindahl-Hirschman index.

The supply side presents a contrasting structure. The 1,524 authorizations are distributed across 781 distinct companies, the company-level concentration index is only 0.005, the five largest firms account for 10.8 per cent of authorizations, and 67.6 per cent of firms hold a single authorized device. A clinically concentrated target is thus being served by a dispersed, entrant-heavy supplier base. Read through the framework, this population describes the conditions under which epistemic allocation and local validation have to be exercised: which decisions algorithms are arriving in, and the evidentiary route by which they arrive. It does not measure how any organization allocated decision authority or whether any organization validated performance locally, since neither is recorded at the device level. Those observations belong to the case strand examined next, and the census is read here as a description of the environment that strand operates in.

2 Three organizational cases read through the framework

Duke Health's Algorithm-Based Clinical Decision Support oversight programme is the most fully documented attempt by a single institution to construct decision intelligence as a standing capability. The published framework superimposes the developmental logic of medical devices onto the software lifecycle, defining four phases through which every algorithmic solution intended for patient care must pass, model development, silent evaluation, effectiveness evaluation and general deployment, each separated by governance checkpoints, with an oversight committee spanning clinical, statistical, nursing, legal and informatics expertise maintaining a registry of every solution under consideration (Bedoya et al., 2022). At Duke, the three mechanisms can be identified in specific governance practices. The registry and risk-classification process require an explicit decision about whether an algorithm should be allowed to influence patient care. Purchase or technical activation alone is therefore insufficient. The model is then evaluated silently on Duke's own patients before it is allowed to inform clinical decisions, providing the local validation mechanism. Responsibility is also attached to named clinical and technical owners for every registered solution, and that ownership continues after deployment through monitoring. Taken together, these arrangements show how allocation, validation and accountability can become part of ordinary institutional governance rather than separate controls added after implementation.

The MD Anderson collaboration with IBM on the Oncology Expert Advisor is particularly informative because the University of Texas System Audit Office examined the initiative in a special review (UTSAO, 2016). That review records expenditure of at least sixty-two million United States dollars over four years, twelve extensions of the original contract, departures from standard procurement procedures and limited involvement of institutional information-technology governance. A substantial part of the expenditure proceeded without

board approval. The system was piloted in leukaemia and lung cancer using an earlier records system, but it was not integrated with the institution's Epic electronic health record and never entered clinical use. The audit was concerned with procurement and governance, not with the scientific performance of the technology. That distinction matters here because it places the failure at the level of organizational governance rather than algorithmic performance. It was not a failure of local validation in the strict sense, since the system never advanced far enough for validation in clinical use to become the operative question. It lay upstream, in epistemic allocation at the strategic level: the decision that an algorithmic advisory system should be built and should eventually participate in oncological decision-making was itself never processed through the institution's allocation machinery. The case is consistent with Proposition 2: a tool admitted through channels that bypassed allocation failed at extraordinary cost in an institution whose clinical resources were among the best in the world.

The third case differs in kind, since it concerns not one institution's programme but the diffusion of a single algorithm across a sector. The Epic Sepsis Model, a proprietary prediction tool embedded in a widely used electronic health record platform, was implemented at hundreds of hospitals. In 2021 investigators at Michigan Medicine published an external validation covering 27,697 patients and more than 38,000 hospitalizations, and found discrimination substantially below the developer-reported range, inadequate calibration, and, at the recommended alerting threshold, a high volume of alerts set against a low yield, with most sepsis cases not already evident to clinicians left undetected, a combination the authors linked to alert fatigue (Wong et al., 2021). The accompanying editorial treated these findings as a governance issue extending beyond the performance of one model. Its concern was that a proprietary model had been deployed widely before independent external validation, while responsibility still remained with adopting hospitals to determine whether the model performed adequately in their own patient populations (Habib et al., 2021). Read through the framework, the case is a sector-scale illustration of weak local validation coexisting with formally unobjectionable adoption. The tool arrived inside an already-procured platform, so its activation was closer to a configuration choice than to an institutional decision about decision authority, and where allocation is compressed into a platform default, local validation becomes the only remaining checkpoint before clinical reliance, which is exactly the checkpoint that adoption-by-default is structurally likely to skip. The device census reinforces the point: with 96.2 per cent of authorized devices entering through the predicate route, the sector-scale conditions under which platform-embedded models reach the bedside are precisely those in which fresh clinical validation is least likely to have been demanded before adoption.

The three cases vary along the lines the framework distinguishes rather than along a single axis of governance quality. Duke exhibits deliberate construction of all three mechanisms and functions as an existence proof; MD Anderson exhibits catastrophic weakness in strategic-level epistemic allocation, with the consequence that the remaining mechanisms never acquired jurisdiction; and the Epic Sepsis Model episode exhibits, across many adopting institutions at once, the substitution of platform default for allocation and of vendor evidence for local validation, with accountability integration left indeterminate. Three features bear on the propositions. First, mechanism strength does not travel together: the three cases occupy different points on each mechanism rather than lining up on a single gradient, which is consistent with Proposition 1 and is what a composite notion of governance maturity would obscure. Second, each failing case failed through its own weakest mechanism regardless of its strength elsewhere, MD Anderson through allocation despite clinical resources among the best in the world, and the Epic deployments through validation despite formally unobjectionable adoption. This is the configuration Proposition 4 describes when it holds that the three mechanisms are jointly required and that assessment should report a three-mechanism profile rather than a composite score, although three purposively selected cases cannot establish that the requirement holds generally. Third, the point at which an algorithm enters an organization, deliberate registration, executive sponsorship, or platform default, largely determines which mechanism is placed under stress, which suggests that the framework may serve as a diagnostic instrument and not only as a taxonomy. The present evidence does not bear on Proposition 3, since neither the device census nor three purposively selected cases supplies the paired developer-reported and locally observed performance measures, across sites differing in population and workflow, that the proposition concerns; it is reserved for the replication design set out in the research agenda. Table 3 sets out the cross-case configuration. Taken together, the three cases are read here as consistent with Propositions 1, 2 and 4 and as silent on Proposition 3; consistency with a small, purposively selected set of cases is not a test, and none of the propositions is treated as confirmed by this evidence.

Table 3. Cross-case configuration of the three mechanisms

Mechanism	Duke Health (ABCDS)	MD Anderson (Oncology Expert Advisor)	Epic Sepsis Model (multi-site)
Epistemic allocation	Performed as a distinct institutional act via registry and risk-tiered review before any clinical influence	Bypassed at strategic level; adoption decision made and renewed outside standard oversight	Compressed into a platform default; activation closer to configuration than institutional decision
Local validation	Institutionalized as mandatory silent evaluation on the organization's own patients before deployment	Never reached; the system did not advance to clinical use, so validation never acquired jurisdiction	Widely absent at adoption; performed post hoc by one institution, revealing performance below developer claims

Mechanism	Duke Health (ABCDS)	MD Anderson (Oncology Expert Advisor)	Epic Sepsis Model (multi-site)
Accountability integration	Named clinical and technical owners required for every registered solution through monitoring	Diffuse; audit documented oversight responsibilities unengaged and spending largely without board approval	Indeterminate at many sites; ownership of the decision to rely on embedded alerts frequently unassigned

Source: author's elaboration based on the published record of each case (Bedoya et al., 2022; Habib et al., 2021; UTSAO, 2016; Wong et al., 2021).

DISCUSSION

1 Theoretical implications

The evidence bears on organization theory beyond the immediate territory of healthcare artificial intelligence. For the bounded-rationality tradition, the material indicates that algorithmic extension of organizational decision capacity does not relax the governance burden but relocates it: the organization that once economized on bounded rationality by structuring who decides must now also structure what decides. Epistemic allocation, in this light, is the contemporary form of a very old organizational problem, and its neglect in current governance debates reflects the technological framing of those debates rather than any novelty in the underlying question. The Epic Sepsis Model brings another issue into the discussion of professional jurisdiction. The difficulty in this case did not arise because the algorithm replaced the clinician. Rather, its performance and alert burden altered the environment in which clinical judgement had to be exercised, making useful signals harder to distinguish while leaving the clinician's formal responsibility unchanged (Wong et al., 2021). From this perspective, accountability integration helps preserve professional responsibility when clinical judgement is increasingly exercised alongside algorithmic outputs.

These findings also qualify an important assumption in the healthcare AI governance literature. Organizations may endorse similar principles while differing substantially in the mechanisms through which those principles are put into practice. The three cases existed within a policy environment in which accountability, transparency and related governance principles were already well established (Reddy et al., 2020), yet allocation was bypassed in one case and local validation was absent across many adopting hospitals in another. The framework therefore extends, rather than replaces, existing work. Lifecycle approaches such as the Duke model show how governance can be organized in practice (Bedoya et al., 2022), while evidence on model diffusion and external validation shows why organizations need to establish performance in their own settings (Habib et al., 2021; Topol, 2019). This study brings allocation, validation and accountability together at the level of the decision, but does not collapse them into a single mechanism. The device census places that argument in a wider sectoral setting: concentrated clinical adoption and the predominance of the predicate route in the

regulatory record extend beyond the individual organizational cases examined here.

2 Practical and managerial implications

For boards and executive teams, the framework is most useful before an algorithm is allowed to influence a consequential decision. Its three mechanisms translate into a short set of practical checks. Who authorized this algorithm to shape this class of decisions, and could that authorization be reconstructed if it were audited? What evidence shows that the tool works as claimed for our patients, our case mix and our workflow, and how recently was that evidence reassessed? The third check is whether, should an algorithm-influenced decision be challenged a year from now, the organization can still identify the clinical owner, the technical owner and the monitoring record responsible for it. An organization that can answer all three has the principal elements of decision intelligence in place. Where an answer is weak or absent, it also indicates where governance needs attention first.

The device census also raises a procurement issue. Most authorized tools entered through a predicate-equivalence pathway, which on published accounts of the pathway is generally not held to the same requirement for new clinical evidence as the other routes considered here (Wu et al., 2021). The pathway is a characteristic of the submission rather than a verdict on any individual device, and some 510(k) submissions do include clinical data. Some algorithms, however, are embedded within larger platforms and may reach clinicians simply through a configuration choice rather than a distinct purchasing decision. In either situation, the adopting organization carries much of the responsibility for establishing local performance. An embedded algorithm therefore warrants the same scrutiny over allocation as a standalone product: enabling a feature can itself determine what is permitted to influence clinical judgement. Contracts should also secure access to the data required for local validation and preserve the organization's ability to disable an embedded model. These are governance conditions, not merely technical details. The issue may be particularly relevant to family-owned and founder-led hospital groups, which are common in the Turkish private healthcare sector. When major technology decisions are concentrated in a founder or family executive, the MD Anderson case shows what can happen when senior sponsorship moves outside established oversight. The answer is not to restrain innovation. Leadership-sponsored AI initiatives should instead pass through the same allocation process required of other consequential clinical technologies.

3. Limitations and boundary conditions

These findings should be interpreted in relation to the two evidence sources on which the study rests. The quantitative analysis concerns devices recorded by one regulator in the United States market. The FDA list is an administrative

authorization record, not a complete inventory of every AI-enabled device in clinical use. The results therefore describe the record available at the frozen retrieval date rather than the full population of deployed systems. The pathway and concentration measures should be read on the same basis, and the partial 2026 observations are excluded from year-on-year comparisons. Regulatory pathway provides an indication of the pre-market evidence route, but it cannot establish whether an adopting organization subsequently carried out local validation, and it is a property of the submission rather than a judgement on the evidence behind any individual device. The two strands also sit at different levels of analysis, the device and the organization, and the study draws no inference from one to the other: the census characterizes the environment the mechanisms must govern, while only the cases observe the mechanisms themselves. The case analysis has a different boundary. The three cases were selected because their documentary records were sufficiently detailed and because they represented contrasting configurations; they were not intended to form a representative sample. They indicate that allocation, validation and accountability are distinguishable mechanisms that can be developed independently, but they cannot establish how frequently any one configuration occurs, and the reading of the cases against the propositions is a check for consistency rather than a test of them. The framework also assumes an institutional setting in which professional accountability retains legal and normative force. Were regulators to grant algorithmic systems independent authority over decisions, the accountability-integration mechanism would require reconsideration.

4. Future research

The framework points to three empirical directions, beginning with measurement. A future instrument could examine each mechanism separately. For allocation, documentary evidence would show how a tool entered the organization. Validation could compare local performance evidence with developers' claims across deployed models. Accountability could be examined through the presence of named owners and monitoring records that are demonstrably in use. Applied across a stratified sample of hospitals, such a design would allow Propositions 1 and 4 to be examined directly. The second is a comparative study of entry channels, following cohorts of algorithmic tools admitted through deliberate registration, executive sponsorship and platform default within the same institutions and comparing post-deployment abandonment, which would examine Proposition 2. The third is a multi-site replication programme in the spirit of the Michigan Medicine validation, quantifying the developer-to-local performance gap across sites that differ measurably in population and workflow, which would examine Proposition 3. Each design is feasible with existing methods.

CONCLUSION

This study began from an asymmetry: healthcare organizations are admitting algorithmic systems into consequential decisions faster than they are building the capability to govern those decisions. Its response has been to conceptualize the missing capability as decision intelligence, the organizational capacity to allocate decision authority between professionals and algorithms deliberately, to validate algorithmic performance in the organization's own context, and to keep accountability traceable across hybrid decisions, and to examine that capability against real evidence. The census of 1,524 authorized devices shows a sector admitting artificial intelligence rapidly, overwhelmingly through a predicate route that, on published accounts of the pathway, does not generally require fresh independent clinical evidence, and concentrating that admission in image-interpretation decisions rather than spreading it across medicine. The comparison of Duke Health, MD Anderson and the Epic Sepsis Model is consistent with the three mechanisms being separable and independently fallible, shows in the Duke case that they can be deliberately built, and suggests that the channel through which an algorithm enters an organization bears on which mechanism its arrival stresses. Healthcare artificial intelligence governance has been written, overwhelmingly, at the address of the technology. This study argues, on a documented record, that governance lives at the address of the decision, and that the field's persistent puzzle, why institutions with identical principles and comparable resources diverge so sharply in outcomes, becomes tractable once the mechanisms operating at that address are separated and examined.

DECLARATIONS

Ethics approval: This study uses publicly available, de-identified secondary data (a regulatory device register and the published record of documented cases), and therefore did not require ethics-committee approval.

Data availability: The device data analysed are openly available from the United States Food and Drug Administration (FDA, 2026). The reproducible build pipeline, the transformation log, the analysis code and the derived analytical dataset are provided as supplementary material.

Funding: This research received no external funding.

Conflict of interest: The author declares no conflict of interest.

Use of artificial intelligence tools: An artificial intelligence assistant was used to support data processing and language editing. The author reviewed and verified all analyses, figures and references, and takes full responsibility for the content of this work.

REFERENCES

Bedoya, A. D., Economou-Zavlanos, N. J., Goldstein, B. A., Young, A., Jelovsek, J. E., O'Brien, C., Parrish, A. B., Elengold, S., Lytle, K., Balu, S., Huang, E., Poon, E. G., & Pencina, M. J. (2022). A framework for the oversight and local deployment of safe and high-quality prediction models. *Journal of the American Medical Informatics Association*, 29(9), 1631–1636. https://doi.org/10.1093/jamia/ocac078

Benjamens, S., Dhunnoo, P., & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: An online database. *npj Digital Medicine*, 3, 118. https://doi.org/10.1038/s41746-020-00324-0

Freidson, E. (2001). *Professionalism: The third logic*. Polity Press.

Habib, A. R., Lin, A. L., & Grant, R. W. (2021). The Epic Sepsis Model falls short: The importance of external validation. *JAMA Internal Medicine*, 181(8), 1040–1041. https://doi.org/10.1001/jamainternmed.2021.3333

Muehlematter, U. J., Daniore, P., & Vokinger, K. N. (2021). Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): A comparative analysis. *The Lancet Digital Health*, 3(3), e195–e203. https://doi.org/10.1016/S2589-7500(20)30292-2

Muehlematter, U. J., Bluethgen, C., & Vokinger, K. N. (2023). FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. *The Lancet Digital Health*, 5(9), e618–e626. https://doi.org/10.1016/S2589-7500(23)00126-7

Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, 27(3), 491–497. https://doi.org/10.1093/jamia/ocz192

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99–118. https://doi.org/10.2307/1884852

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

U.S. Food and Drug Administration. (2026). *Artificial intelligence and machine learning (AI/ML)-enabled medical devices*. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices

University of Texas System Audit Office. (2016). *Special review of procurement procedures related to the M.D. Anderson Cancer Center Oncology Expert Advisor project*. https://www.utsystem.edu

Wong, A., Otles, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Cooley, O., Pestrue, J., Phillips, M., Konye, J., Penoza, C., Ghous, M., & Singh, K. (2021). External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine*, 181(8), 1065–1070. https://doi.org/10.1001/jamainternmed.2021.2626

Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D. E., & Zou, J. (2021). How medical AI devices are evaluated: Limitations and recommendations from an analysis of FDA approvals. *Nature Medicine*, 27(4), 582–584. https://doi.org/10.1038/s41591-021-01312-x