AI in Clinical Trials: *From Pilot to Production*

Clinical trials have never been more important or expensive. A single new drug now takes over a decade and billions of dollars to bring to market and trials account for much of that cost and time. They are also growing more complex: more data, more sites, more decentralised elements and rising regulatory expectations. This is precisely why AI matters now. Applied well, it can compress the slowest, most repetitive parts of the process, screening records, cleaning data, interpreting protocols, drafting documentation and return skilled people to the work that needs them. The difficulty lies in moving from a promising pilot to something that works safely, inside a regulated environment. AI has already reshaped early discovery, protein structure prediction, target identification, molecule design, virtual screening but the dominant cost and uncertainty lie downstream, in demonstrating that a candidate works in humans.[11]

Two recent industry reports examine the same shift from different angles.[1,2] One asks what a modern, AI-enabled clinical trial requires in terms of regulation and system design. The other asks a more pointed question: why do so many AI projects succeed in a pilot, then stall before reaching everyday use? Read together, they reach the same conclusion. The obstacle is rarely the AI itself. It is the infrastructure around it, the rules, the data and the way people work which is not yet built to carry AI into regulated, inspection-ready use.

Clinical research is a very data-disciplined field and it has been slow to give its investigators and coordinators true support again, because of the infrastructure the technology must operate within, not the technology itself. This paper sets out what that infrastructure needs to be, why partial fixes fail and what sponsors, technology providers and sites should each do differently. It pays particular attention to the site, the principal investigators (PIs) and study coordinators (SCs) because that is where the gap between what AI can do and what teams can actually use is widest.

The Regulatory Direction Is Already Set

Trials have grown more complex while the pressure to run them faster and more transparently keeps rising. Remote visits, wearables and electronic consent have widened access since the pandemic; the Apple Heart Study enrolled more than 400,000 people entirely remotely.[3] AI now supports protocol design, patient matching and

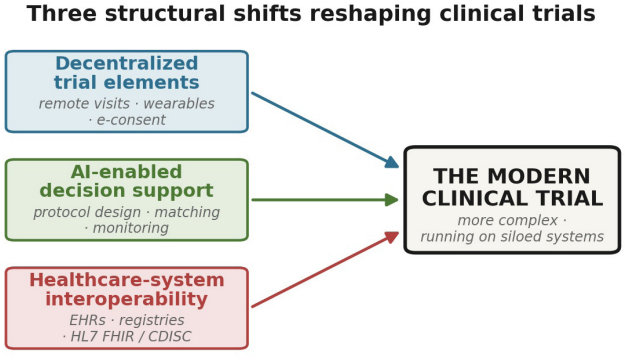


Figure 1. Three structural shifts reshaping clinical trials

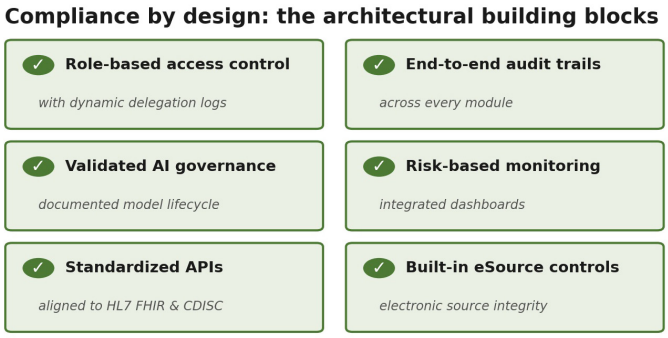


Figure 2. Compliance by design: the architectural building blocks

monitoring and interoperable data networks have made very large trials possible, the ADAPTABLE trial randomised over 15,000 patients through a network linked to electronic health records.[4] Yet most trials still run on separate, disconnected systems, one each for trial management, data capture, outcomes, consent and safety. The same information is entered several times, audit trails fail to align, safety signals surface late and inspection risk rises. The regulations demand accountability and traceability; the systems quietly work against both.

Regulators have been explicit about where this is heading. The FDA expects remote data tools to be fit for purpose and to transmit data securely.[7] The EMA has stated plainly that 'black-box,' models, those whose reasoning cannot be explained, are unlikely to be acceptable for high-stakes decisions without human oversight [8]. ICH E6(R3) keeps the sponsor accountable for vendor-run systems, including their validation and audit trails, even when the work is outsourced,[9] and the risk-based quality principles of ICH E8(R1) and Q9(R1) push oversight toward the issues that matter most. Digital endpoints are welcome, but only with rigorous, pre-planned validation: the EMA has, for example, accepted a wearable-based walking measure as a primary endpoint in a muscular dystrophy trial.[5]

Europe now adds a further layer. The EU AI Act, the first broad AI regulation, operates alongside GDPR, with its obligations phasing in through 2026 and 2027. AI that acts as part of a medical device, or that materially shapes decisions about a person's health, is classed as 'high-risk,' and must meet requirements for risk management, transparency, human oversight and ongoing monitoring. The Act also obliges organisations to ensure staff have adequate AI literacy, turning a desirable skill into a legal duty. Notably, its risk-based logic mirrors a distinction this paper returns to below: tools that handle patient data carry a far heavier burden than tools that merely help coordinate the work.

The conclusion is straightforward. Compliance must be designed into a trial from the outset, access controls, complete audit trails, clear governance for AI and standard data connections rather than retrofitted once a platform is already in use.

Why Good System Design Is Not Enough on Its Own

A second body of research interviews with fifteen leaders across clinical development, operations, quality and IT security, together with findings from a 2026 industry conference[2] asked why AI stalls

even when the system design is sound. The answer is that three obstacles must be addressed at once. Solve one and neglect the others and the project stops for reasons no one anticipated.

Trust means two different things. To operations staff, trust means audit trails and provenance, being able to state exactly who changed what, when and why. To scientific and medical leads, it means something quite different: a model validated for their disease, whose strengths and limits a clinician can judge without needing an AI specialist. A system built only for audit trails will not satisfy the scientists and one built only for predictive accuracy will not survive an inspection. Both forms of trust must be earned simultaneously.

Scattered data defeats AI before governance even arises. Trial data typically lives in dozens of systems, each with its own labels and permissions. One large company consolidated more than 60 terabytes from over 50 sources to build an AI assistant. It reached roughly 95 percent accuracy on plain-language questions,[2] but only after the team collapsed hundreds of fragmented tables into a small, consistent set with agreed definitions. Before that clean-up, even a capable model produced confident wrong answers, not because the AI was faulty but because the data beneath it was ambiguous.

AI scales only if it fits how people already work. Small teams often one technical lead paired with one clinical operations lead moved faster than large, centralised AI programmes. Broad 'chat with everything,' tools disappointed, while narrow, high-value uses succeeded. Adoption spread from the bottom up: a tool became standard only after early users demonstrated its value and brought colleagues along. Mandates issued from above, before the tool had proven itself, tended to stall.

One finding stands out. In a tool that reviewed draft protocols for design risks, more than half its suggestions told teams nothing they did not already suspect. Their value lay elsewhere: they turned a shared hunch into a written, source-linked rationale that could withstand an inspection. AI's most useful role here was not generating new ideas, but giving people documented grounds to act on judgments they already trusted.

It is worth asking, too, whether the manual workflow AI would replace deserves the trust it already receives. In one multisite cancer study, copying data from the hospital record to the trial database by hand produced errors in 5.8 percent of fields, which an AI-assisted tool with human oversight cut to 1.2 percent,[2] yet a 2025 survey found 72 percent of sites still transfer data by hand.

Two Kinds of AI Tool for Site Staff

One practical rule runs through both reports: the AI projects that succeed layer on top of existing systems rather than replacing them, since sponsors will not and realistically cannot tear out a global trial

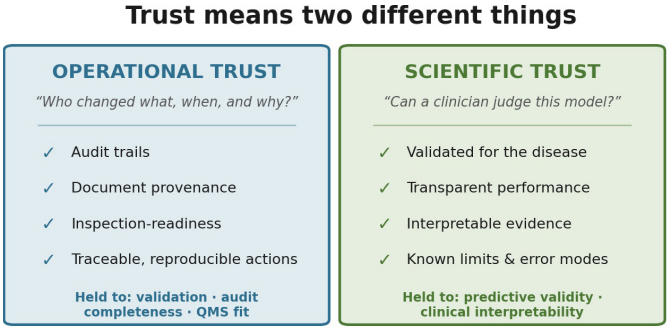


Figure 3. Operational versus scientific trust

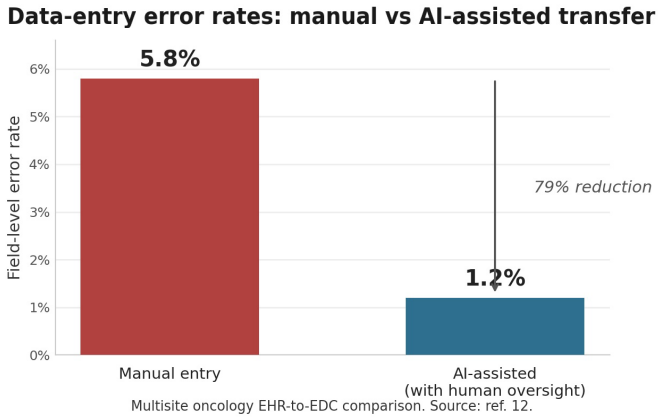


Figure 4. Data-entry error rates: manual versus AI-assisted transfer

infrastructure that works well enough. Against that backdrop, two kinds of tool are emerging: those that operate inside the regulated world of patient data and those that stay above it, helping people run and coordinate the trial. Knowing which is which matters, both for planning and for how a site evaluates a new tool.

The first kind handles patient data directly. Some tools help a PI review outcome data, assembling summaries, flagging what needs attention and signaling when data is ready for sign-off. Others manage admin and compliance: patient communication, visit scheduling and consent management, including monitoring for updated consent forms. Some even ease staff turnover by guiding new coordinators through a study's steps in real time. Because all of these touch patient data, they operate fully inside the regulated environment ICH E6(R3), 21 CFR Part 11, audit trails, so the validation burden is high and they must be formally qualified within the site's quality system.

The second kind never touches patient data. What it operates on is the trial's operational structure, the tasks, responsibilities, timelines and dependencies of running a study across many roles and sites, against a protocol that keeps changing. Today much of that lives in email, spreadsheets and people's heads, with no shared picture anyone can fully trust; a coordination tool makes that structure explicit and current. Because these tools stay above the patient-data line, they can be layered onto existing systems without entering the regulated data world. That also explains why they are adopted so readily: they fit how teams already work rather than asking them to replace validated systems and, echoing the bottom-up pattern in the SCOPE interviews,[2] spread as users see value rather than by mandate.

The boundary is not always perfectly clean. Coordinators inevitably ask questions connected to patient information even when the tool never operates on patient records. The distinction holds because the coordination layer acts on the trial's operational structure, who does what, by when and in what order, not on the source data: it is the operational plan that is shared and kept current, not the clinical record.

The two kinds are complementary rather than competing: a site might run a patient-data tool for PI oversight and a coordination tool for daily work at once, with no conflict. The distinction clarifies how to judge a tool patient-data tools on validation, audit trails and quality-system fit; coordination tools on workflow fit, adoption speed and friction removed.

The risk lies in conflating the two – subjecting a coordination tool to the scrutiny meant for patient data, or assuming it needs none at all. Knowing what a tool does and which side of the patient-data line it falls on, is the starting point for any realistic plan.

AI in Patient Recruitment: Where the Bottleneck Meets the Tool

Recruitment is the single most common reason trials fall behind, and most of the weight falls on site staff, coordinators reading through dozens of records to find one eligible patient, checking dense criteria against scattered notes under time pressure. The realistic prize is not that AI recruits patients. It is that AI clears away the least clinical, most time-consuming part of the task, returning coordinator hours to the work that genuinely needs a person: talking with patients, building trust and taking consent. High-volume, rule-based searching is exactly what AI does well; the harder question is how a site gets there from where it stands today.

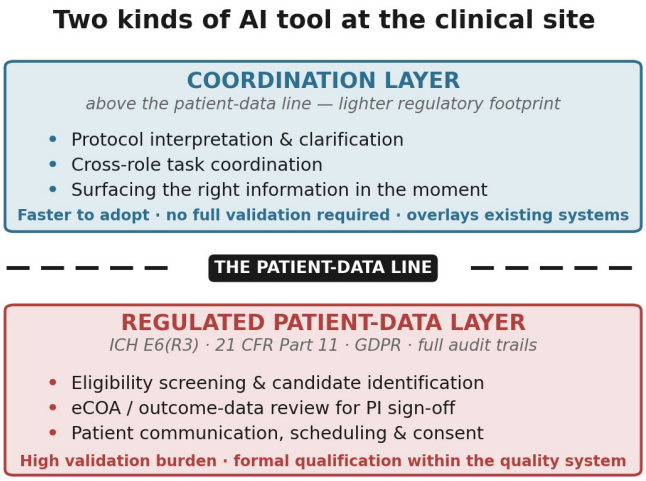


Figure 5. Two kinds of AI tool at the clinical site

From unregulated to regulated: a staged path. Moving from the unregulated side into the regulated one is demanding validation, audit trails, quality-system integration and lawful data access all take real effort but for anyone serious about AI in recruitment it is the journey that has to be made and it need not be a single leap. Much of recruitment today already operates in the unregulated engagement layer, where specialists work closely with sites to coach them past their recruitment barriers. AI fits naturally into that way of working, spotting patterns across sites, flagging when one falls behind its screening targets, preparing local-language materials faster supporting the people rather than replacing them. Introduced here first, it proves its value and builds the governance habits regulated use will later require. The unregulated side is not the destination; it is the on-ramp. Where a given application falls on that journey depends on the data it touches and the three main uses divide as follows.

Finding eligible patients – regulated. The most mature use is AI-assisted eligibility screening. The tool reads inclusion and exclusion criteria and matches them against health records, pathology reports and notes, cutting a chart review that took weeks down to a fraction of the time. Searching records this way has also been shown to enroll more patients from under-represented groups than reliance on the usual referral patterns. But because these tools read identifiable patient data and feed directly into an eligibility decision, they belong firmly in the regulated world: GDPR and national health-data law govern the access, GCP governs the decision. The AI does not decide who is eligible it shortens the list so a clinician can judge the ones that matter, with the physician's review and the investigator's sign-off still required. Full validation and audit trails apply.

Testing a protocol before it is locked – conditionally unregulated. A different use comes earlier, before any site opens. A model compares the draft eligibility criteria against de-identified records to estimate how many patients a given rule would exclude and how heavy the visit schedule would be, so a sponsor can correct a problem before the protocol is finalised. The condition attached to this category carries real weight and belongs up front: only genuinely anonymised data falls outside GDPR. Pseudonymised records where a key still exists that could re-identify individuals remain personal data and an application built on them stays inside the regulated environment however aggregate its outputs appear. Where that threshold is genuinely met, the analysis shapes design rather than any individual's care and falls outside the regulated world. Given that a substantial amendment costs $141,000 in Phase 2 and $535,000 in Phase 3,[6] catching the problem early is well worth the effort. Recent work shows how far this can be pushed. EmulatRx, a multi-agent framework published in 2026, coordinates five specialised agents that refine protocols against real-world electronic health record data; in one emulated cohort it identified 33 additional eligible patients from clinical notes that structured data alone would have missed.[12] Its authors are explicit that broader validation is required before deployment, capability demonstrated, deployment gated by validation. Reading identifiable notes to surface named candidates, it also falls inside the regulated environment.

Communicating with patients - regulated and consent-sensitive. Automated outreach, chat-style pre-screening, visit reminders, follow-up on patient-reported outcomes and predicting who might drop out all extend the same idea to keeping patients in the trial, often a greater challenge than enrolling them. Because these interact directly with participants and their data, they occupy the most demanding part of the regulated world: consent, GDPR and where they influence a person's health decisions, potentially the high-risk provisions of the EU AI Act.[10]

One caution applies throughout: the evidence is strongest that AI makes screening faster, while whether it improves final enrolment or retention is less well established and sites should keep the two claims apart when evaluating a tool.

Conclusion

Neither the regulations nor the technology is what holds AI back in clinical trials. What is missing is the quieter work of building trust, standardising the data and reshaping how people work so that AI becomes a governed member of the team rather than a bolt-on experiment. The teams making real progress share one habit: they treat design, data and workflow as a single problem, not three to be solved in sequence. That places a different demand on each group. Sponsors should fold AI oversight into the quality system they already operate rather than run it as a separate track; technology providers should design compliance in from the start and favour explainable models over opaque ones; sites should treat digital literacy as a core competency; and regulators must keep advancing shared guidance while using pilots and sandboxes to test new tools under controlled conditions. The same thinking extends to areas set aside here, from





safety monitoring to medical writing. The organisations that take on trust, data and workflow together, rather than waiting for one to be perfect before beginning the next, will move AI from a conference demonstration to real, inspection-ready use and, for the patients these trials exist to serve, close the gap between what is possible and what reaches the clinic.

This paper synthesizes two review articles published in DIA Global Forum, drawing on the primary sources they cite. Where a figure is reported by those reviews rather than independently verified against the primary study, the DIA source is indicated.

REFERENCES

1. Rajasimha HK, Rodriguez-Chavez IR. Building AI-Ready Clinical Trials Through Unified, Compliant Systems. DIA Global Forum, 2026.
2. Pirhaji L. Operationalizing GenAI in Clinical Development: The Interlocking Barriers to Regulated AI Deployment. DIA Global Forum, 2026. (Source of the EHR-to-EDC error-rate comparison, the 2025 eSource survey, and the SCOPE 2026 enterprise deployment data cited here.)
3. Perez MV, Mahaffey KW, Hedlin H, et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. N Engl J Med. 2019;381(20):1909–1917.
4. Jones WS, Mulder H, Wruck LM, et al. Comparative Effectiveness of Aspirin Dosing in Cardiovascular Disease. N Engl J Med. 2021;384(21):1981–1990.
5. European Medicines Agency. Qualification Opinion for Stride Velocity 95th Centile (SV95C) in Ambulatory Duchenne Muscular Dystrophy Studies. EMA, 2023.
6. Smith Z, DiMasi JA, Getz KA. The Impact of Protocol Amendments on Clinical Trial Performance and Cost. Ther Innov Regul Sci. 2016;50(4).
7. US Food and Drug Administration. Digital Health Technologies for Remote Data Acquisition in Clinical Investigations. Guidance for Industry, 2023.
8. European Medicines Agency. Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle. EMA, 2024.
9. International Council for Harmonisation. ICH E6(R3) Good Clinical Practice, 2025; ICH E8(R1) General Considerations for Clinical Studies; ICH Q9(R1) Quality Risk Management.
10. Regulation (EU) 2024/1689 (Artificial Intelligence Act). Official Journal of the European Union, 2024.
11. Can AI Make Better Drugs? Not on Wall Street's Timeline. The Wall Street Journal, 12 July 2026.
12. Li H, Pan W, Zang C, Rajendran S, Wang F. Empowering clinical trial design with agentic intelligence and real-world data. Nat Commun. 2026. doi:10.1038/s41467-026-74501-2.

Stefan Åström



Stefan Åström holds a PhD in molecular biology and began his career at Pfizer and AstraZeneca. A serial CRO entrepreneur, he founded Hylae Clinical Research, Astrom Research International and AdDoc (an early EDC company). He is now an independent scientific advisor to biotech and CRO companies and serves on the Editorial Advisory Board of the Journal for Clinical Studies.