RESEARCH BRIEFS · RB-026

# AI-Augmented Patient Recruitment: Network Performance Benchmarks

Clinitiative Health Research May 2026

CONTENTS

Executive Summary

1. Key Findings

2. Methodology and Phased Analysis

3. Performance Variation by Therapeutic Area

4. Bias, Equity, and the Credibility Lens

5. Strategic Implications

6. Conclusions

## Executive Summary

### From Tool Adoption to Operational Outcomes

Patient recruitment has long been the single largest source of timeline risk in clinical research. Sponsors and sites alike have invested heavily in technology platforms — patient registries, EHR queries, advertising automation — yet the industry-level enrollment shortfall has remained stubbornly consistent: an estimated 80% of trials miss their original enrollment timeline, and roughly 30% of sites enroll no patients at all. The arrival of mature natural language processing and machine learning systems has prompted broad industry optimism that AI can finally meaningfully accelerate the enrollment problem.

The promise is grounded in real capability. Modern NLP systems can extract relevant clinical concepts from unstructured physician notes with reported recall above 90%, and trial-matching systems have demonstrated eligibility-screening accuracy in the low-90s range against human assessment. Published efficiency claims include 30-50% trial timeline acceleration, 65% improvements in enrollment, and

recruitment timelines compressed from 18 months to 3-6 months. The challenge for sponsors and sites is to separate vendor-marketed claims from validated operational outcomes.

This research brief presents network-level performance benchmarks for AI-augmented patient recruitment based on 14 studies conducted at 47 Clinitiative network sites between October 2024 and March 2026. The studies span oncology, cardiometabolic, CNS, and rare disease therapeutic areas, and use a mix of AI screening platforms integrated with site EHR systems. The analysis isolates the marginal contribution of AI screening over traditional workflows by comparing AI-augmented sites against matched control sites running the same protocols without AI integration.

## 1 Key Findings

The benchmark analysis reveals consistent operational gains from AI-augmented recruitment, but also identifies the conditions under which AI delivers meaningful value versus the conditions under which traditional screening remains competitive.

**Table 1.** Key figures

| Figure | Measure | Detail |
|---|---|---|
| 38% | Reduction in Screening Cycle Time | AI-augmented sites reduced median screening-to-enrollment cycle time from 26 days to 16 days across the 14-study cohort. The savings concentrated in the eligibility verification phase, where NLP extraction of unstructured EHR data eliminated the manual chart review step that traditionally consumed 5-8 days per candidate. |
| 1.6x | Eligible Candidate Identification | AI-augmented sites identified 1.6 times more eligible candidates per 1,000 EHR records reviewed than traditional screening workflows (63% positive predictive value vs. 40%). The improvement was greatest in oncology (1.9x) and rare disease (2.1x), where eligibility criteria involve complex combinations of structured and unstructured clinical concepts. |
| 91% | Eligibility Determination Accuracy | AI-flagged candidates underwent investigator verification at a 91% confirmation rate, meaning AI eligibility determinations agreed with human investigator judgment on more than nine of every ten flagged candidates. The 9% disagreement rate concentrated in complex eligibility criteria involving prior treatment history and concurrent medications. |
| 34% | Coordinator Time Recovered | Site coordinators at AI-augmented sites recovered a median 34% of their pre-implementation screening time, reallocating those hours to candidate engagement, consent quality, and retention activities. The time savings were strongest at sites with mature EHR-AI integration and weaker at sites using AI as a stand-alone tool requiring manual data entry. |

# 2 Methodology and Phased Analysis

The benchmark analysis was structured in five sequential phases over an 18-month window, allowing the team to capture not only point-in-time performance but also the evolution of AI screening outcomes as sites matured in their use of the tools.

## Site and Study Selection (October 2024)

Studies were eligible for inclusion if they were multi-site protocols with at least one Clinitiative network site implementing AI-augmented screening and at least one Clinitiative network site continuing traditional screening on the same protocol. This matched-pair design controls for protocol-specific recruitment difficulty and allows the marginal contribution of AI to be isolated. 14 studies and 47 sites met the inclusion criteria, with 26 AI-augmented sites and 21 matched control sites.

## Baseline Period (Q4 2024)

A three-month baseline period captured traditional screening performance across all 47 sites prior to AI implementation at the augmented arm. Baseline metrics included screening cycle time, screen failure rate, candidate-to-consent conversion rate, and coordinator effort hours. The baseline phase established the statistical reference against which AI-augmented outcomes were subsequently measured.

## Implementation Ramp (Q1–Q2 2025)

AI screening platforms were deployed at the augmented arm sites with EHR integration, eligibility logic configuration, and coordinator workflow training. The implementation phase took a median of 8 weeks per site, with substantial variance based on EHR vendor (Epic-based sites averaged 6 weeks; sites using less common EHRs averaged 12 weeks). Performance during the implementation phase was excluded from the steady-state benchmark to avoid contamination from early learning effects.

## Steady-State Measurement (Q3 2025–Q1 2026)

Steady-state measurement captured 6 months of mature operational performance at the AI-augmented sites alongside parallel performance at the matched control sites. All comparative metrics presented in this brief derive from this steady-state window. The 6-month duration was selected to dampen the influence of any single recruitment surge or pause and to ensure that the comparison reflected operational reality rather than a launch-period spike.

### Outcome and Bias Analysis (Q1–Q2 2026)

The final analytical phase included a bias review: comparing the demographic composition of AI-flagged candidates against the broader site EHR population to identify whether AI screening systematically over- or under-represented specific demographic groups. The bias analysis was prompted by the FDA's draft AI guidance and the joint FDA/EMA Guiding Principles, both of which emphasize algorithmic bias as a credibility risk.

## 3 Performance Variation by Therapeutic Area

AI-augmented recruitment performance varied substantially across therapeutic areas, reflecting the underlying complexity of eligibility criteria and the structure of available clinical data.

**Figure 1.** Eligible candidate identification gain by therapeutic area (AI-augmented vs. traditional screening)




**Figure 2.** Screening cycle-time reduction by therapeutic area




### Oncology

Oncology trials showed the largest absolute improvement: 1.9x more candidates identified, 43% cycle-time reduction. NLP excelled at extracting cancer staging, biomarker status, and prior treatment history from unstructured pathology reports and oncology notes. AI was particularly effective at identifying

patients whose biomarker-defined eligibility was documented in narrative form rather than in discrete structured fields.

### Rare Disease

Rare disease studies showed the largest relative gain (2.1x candidate identification). When the eligible population is small, the cost of missing eligible candidates is high; AI screening systematically surfaced patients whose diagnostic information was buried in unstructured clinic notes and would have been missed by structured-data EHR queries. Three rare disease studies in the cohort identified candidates who had been documented for years but never matched to applicable trials.

### Cardiometabolic

Cardiometabolic trials showed moderate gains (1.4x candidate identification, 32% cycle-time reduction). Eligibility criteria in this area rely more heavily on structured laboratory values and discrete medication lists, which traditional EHR queries already handle relatively well. AI contributed at the margin by capturing patients whose lifestyle factors and comorbidity descriptions in clinic notes determined eligibility.

### CNS / Neurology

CNS trials showed the smallest gain in candidate identification (1.3x) but a meaningful cycle-time reduction (29%). Many CNS eligibility criteria depend on cognitive assessments, functional scores, and caregiver-reported observations that are not fully captured in EHR data. AI contributed primarily by accelerating the screening of medication history and comorbidity rule-outs rather than primary eligibility determination.

### Sponsor Variability

AI performance varied 2.4x in candidate-identification gain across sponsor protocols even within the same therapeutic area. The dominant driver of variability was the quality of the eligibility logic configuration: protocols where sponsors invested in detailed AI eligibility logic during study startup outperformed protocols using out-of-the-box default logic by a wide margin. AI is not plug-and-play; configuration quality matters substantially.

### EHR Vendor Effects

Sites on Epic-based EHR systems achieved faster AI implementation timelines and slightly better candidate-identification rates than sites on other EHR platforms, likely due to the breadth of Epic-AI

integration tooling and the maturity of Epic FHIR APIs. Site EHR maturity is a meaningful determinant of AI recruitment performance and should be considered in feasibility for AI-augmented trials.

## 4 Bias, Equity, and the Credibility Lens

The bias analysis surfaced an important nuance: AI-augmented recruitment did not eliminate demographic disparities in candidate identification, and in two studies it amplified them. AI models trained on historical EHR data inherit the access disparities encoded in those records — patients who appear infrequently in the EHR (for example, those with less consistent care contact) are under-represented in AI candidate pools relative to their share of the disease population.

In 12 of the 14 studies, AI-flagged candidate demographics tracked closely to overall site EHR demographics with no meaningful skew. In two studies — one in cardiology and one in rheumatology — the AI-flagged candidate pool over-represented patients with stronger care continuity (more frequent EHR encounters) and under-represented patients in the same disease category with sparser EHR documentation. Both sponsors responded by adjusting the AI eligibility logic to weight candidates with sparser records more heavily during the manual review step.

This finding is consistent with the FDA's emerging credibility framework and the joint FDA/EMA Guiding Principles, both of which emphasize that AI outputs in patient recruitment must be monitored for demographic bias and that human oversight remains essential. The 91% accuracy rate against human investigator judgment is necessary but not sufficient — credibility also requires evidence that the AI screening does not systematically exclude eligible patients from underrepresented groups.

## 5 Strategic Implications

The data support an evidence-based view of AI-augmented recruitment that sits between the most enthusiastic vendor claims and the most skeptical industry critiques. AI delivers measurable, repeatable improvements in screening efficiency and candidate identification — but only when implementation is mature, eligibility logic is well-configured, and bias is actively monitored. The 38% cycle-time reduction and 1.6x candidate-identification gain observed in this dataset are operationally meaningful, but they are not the order-of-magnitude transformations sometimes described in industry marketing.

For sponsors, the strategic question is not whether to use AI in recruitment but how to invest in the implementation conditions that determine whether the AI delivers. Detailed eligibility configuration, EHR integration depth, and site-level workflow redesign matter more than choice of platform. Sponsors who treat AI as a procurement decision rather than an operational program will under-realize the available value.

For sites, AI recruitment shifts the coordinator role from chart reviewer to candidate engagement specialist. The 34% time recovery is real, but only if site management actively reallocates the recovered time toward higher-value activities. Sites that simply absorb the time savings into reduced staffing miss the strategic opportunity to improve consent quality, retention, and patient experience. Sites that reinvest the recovered time consistently report stronger enrollment and retention outcomes in subsequent studies.

## 6 Conclusions

AI-augmented patient recruitment is no longer experimental. The 14-study network benchmark demonstrates consistent operational gains across therapeutic areas and confirms that the technology has matured from pilot to production. At the same time, the data make clear that AI is a tool whose value depends entirely on the surrounding implementation: eligibility configuration, EHR integration, workflow redesign, and bias monitoring all materially shape the realized outcome.

The convergence of operational evidence with the FDA's emerging credibility framework creates an important moment. Sites and sponsors who establish disciplined AI governance now — clear configuration standards, demographic monitoring, documented human verification — will be positioned to scale AI use across their portfolios while meeting the documentation expectations that are forming around AI-derived evidence. Those who treat AI as a stand-alone screening shortcut will face increasing friction as regulatory expectations sharpen.