International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.4 (2025) pp. 7826-7835
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

AI-Powered Clinical-Trial Recruitment: Protocol-to-Claims NLP, Geospatial Density Modeling, and Feasibility Outcomes

Badal Shah*

Neolytica - A QPharma Company, USA
* **Corresponding Author Email:** reachbadalshah@gmail.com- **ORCID:** 0000-0002-5247-0050

Article Info:

DOI: 10.22399/ijcesen.4153
Received : 25 August 2025
Accepted : 12 October 2025

Keywords

Artificial Intelligence,
Clinical Trials,
Patient Recruitment,
Healthcare Analytics,
Site Selection

Abstract:

This technical article examines the transformative role of artificial intelligence in revolutionizing clinical trial recruitment and site selection processes. The traditional approach to trial site selection, historically dependent on enrollment reports and feasibility surveys, is being supplanted by a sophisticated data-driven methodology that integrates multiple dimensions of healthcare information. The article details a multi-layered AI solution that leverages longitudinal claims data, clinical pattern recognition, influence mapping, and geographic intelligence to identify optimal research sites and healthcare professionals. Through a systematic workflow that includes market basket creation, claims analysis, network mapping, and composite ranking, the system creates targeted outreach strategies that significantly improve enrollment efficiency. A case study of a chronic obstructive pulmonary disease trial across numerous sites demonstrates the practical effectiveness of this approach, with substantial enrollment increases at participating research centers. The article argues that this paradigm shift from intuition-led to evidence-optimized strategies represents a fundamental advancement in clinical operations with profound implications for accelerating therapeutic development and expanding patient access to clinical trials.

1. Introduction

In the high-stakes world of clinical trials, recruitment delays and enrollment challenges can devastate budgets and postpone critical therapies from reaching patients. A groundbreaking approach using artificial intelligence is transforming this landscape by replacing intuition-based site selection with sophisticated data analytics.Clinical trial recruitment represents one of the most persistent challenges in medical research, with significant implications for healthcare innovation and patient outcomes. According to comprehensive analyses of clinical research operations, recruitment difficulties consistently rank as the primary cause of study delays, affecting a substantial majority of trials across therapeutic areas. These delays extend far beyond mere scheduling inconveniences—they directly impact the availability of novel treatments for patients with serious conditions while simultaneously increasing development costs throughout the pharmaceutical pipeline. The ramifications of inefficient patient recruitment cascade through the entire healthcare ecosystem, ultimately contributing to higher treatment costs and delayed therapeutic innovations. Research published in BMJ Open has demonstrated that traditional recruitment approaches often lead to prolonged study timelines, with many trials requiring extensions of their originally planned recruitment periods, highlighting the systemic nature of this challenge in contemporary clinical research [1].The emergence of artificial intelligence as a solution to these longstanding recruitment challenges represents a significant paradigm shift in clinical trial operations. By harnessing the power of machine learning algorithms and advanced analytics, researchers can now process vast quantities of healthcare data to identify optimal trial sites and potential participants with unprecedented precision. These AI-driven approaches move beyond the conventional reliance on historical performance metrics and subjective investigator assessments, instead leveraging real-world data to create sophisticated predictive models. By analyzing patterns within claims data, electronic health records, and physician referral networks, these systems can map the clinical landscape with remarkable granularity, identifying not only high-potential research sites but also the specific healthcare providers most likely to connect eligible patients with appropriate studies. The integration of multiple data streams—including longitudinal treatment histories, diagnostic patterns, geographical patient distributions, and professional influence networks—enables a multidimensional assessment that far surpasses traditional feasibility evaluation methods in both scope and accuracy. Recent advancements in this field, as documented in Nature, demonstrate how AI tools can transform clinical trial recruitment from an intuition-based process to an evidence-optimized strategy, potentially revolutionizing how medical research is conducted globally [2].The implementation of AI-enabled recruitment systems has begun to demonstrate tangible benefits across various therapeutic areas. When these advanced analytics are deployed in real-world clinical trial settings, they consistently identify previously overlooked patient populations and optimize site selection beyond what conventional methods could achieve. The resulting improvements in enrollment efficiency translate directly to compressed study timelines and reduced per-patient recruitment costs. Furthermore, these systems have shown particular promise in addressing the persistent challenge of diversity in clinical trial participation, as they can identify potential participants from traditionally underrepresented populations with greater precision than conventional recruitment methods. The geographic intelligence component of these platforms proves especially valuable in identifying "white space" opportunities—regions with substantial eligible patient populations but insufficient research infrastructure—thereby expanding access to clinical trials beyond traditional academic medical centers. This approach not only accelerates the development of new therapies but also promotes more equitable access to cutting-edge treatments across diverse communities and geographical regions. Ongoing refinements in the underlying algorithms continue to enhance the predictive accuracy of these systems, suggesting that their impact on clinical research will only increase as the technology matures and becomes more widely adopted throughout the pharmaceutical and biotechnology industries [2].The transformation from traditional recruitment practices to AI-enhanced methodologies represents more than an incremental improvement in clinical operations—it constitutes a fundamental reimagining of how we approach the critical interface between patient identification and research participation. By shifting from reactive, experience-based recruitment to proactive, data-driven strategies, research sponsors can substantially mitigate one of the most persistent obstacles in the development of new medical treatments. The implications extend far beyond operational efficiencies, potentially accelerating medical innovation across numerous therapeutic categories while simultaneously expanding access to clinical trials for patients who might otherwise remain unaware of relevant research opportunities. As these technologies continue to evolve and gain wider implementation, they may fundamentally alter the landscape of clinical research, making the process more efficient, more equitable, and ultimately more responsive to the urgent medical needs of patients awaiting therapeutic breakthroughs [1].

2. The Recruitment Challenge

Traditional clinical trial site selection has largely depended on historical enrollment reports and subjective feasibility surveys—methods that frequently result in timeline extensions and cost overruns. Nearly half of all trial delays stem from recruitment bottlenecks, creating a persistent obstacle in the development pipeline for new treatments.The conventional approach to clinical trial site selection represents a significant vulnerability in the drug development process, relying heavily on methods that have proven inadequate in the face of increasingly complex study protocols. For decades, research sponsors have based their site selection decisions primarily on historical performance metrics and subjective feasibility questionnaires completed by potential investigators. These assessments often fail to capture the nuanced realities of patient availability and investigator capacity, leading to systematic overestimation of enrollment capabilities across research sites. The disconnect between projected and actual recruitment performance manifests in widespread timeline extensions that ripple throughout the development pipeline. Industry analyses have revealed that approximately 80% of clinical trials fail to meet their original enrollment targets within the planned timeframe, necessitating costly extensions and protocol amendments. This pervasive challenge affects studies across therapeutic areas, though it is particularly pronounced in trials targeting rare diseases and those with stringent eligibility criteria. The financial implications of these delays are substantial, with research indicating that each day a drug remains in development rather than on the market represents significant revenue loss for sponsors while simultaneously delaying therapeutic access for patients with unmet medical needs.

Contemporary research on clinical trial recruitment challenges has documented multiple barriers operating at the system, site, and individual levels, highlighting the need for comprehensive solutions that address the full spectrum of factors impeding efficient participant enrollment [3].The recruitment bottleneck phenomenon constitutes a multifaceted challenge that extends beyond simple logistical hurdles. At its core, the problem stems from fundamental information asymmetries between trial sponsors, research sites, and potential participants. Traditional site selection methodologies lack the granularity needed to identify precisely where eligible patients receive care and which healthcare providers influence their treatment decisions. This information gap leads to suboptimal geographical distribution of research sites and insufficient engagement with the physicians most likely to refer appropriate candidates. Moreover, the conventional approach fails to account for the dynamic nature of clinical practice patterns, patient populations, and referral networks. Studies examining the root causes of recruitment failures have identified several recurring factors: misalignment between trial locations and patient demographics, inadequate engagement with community physicians who manage the target patient population, insufficient consideration of geographical barriers to participation, and overly optimistic assessments of site capabilities. The profound impact of these recruitment challenges has been documented in comprehensive analyses of the clinical research enterprise, which have called for transformative approaches to address persistent inefficiencies in how clinical trials are designed and conducted. These analyses emphasize that current recruitment methods not only delay the development of new therapies but also contribute to escalating research costs and ultimately affect healthcare affordability and accessibility [4].The persistence of these recruitment challenges despite decades of industry awareness suggests that incremental improvements to traditional methodologies are insufficient. The systematic nature of these issues indicates the need for a fundamental reimagining of the site selection process—one that moves beyond reliance on historical performance and subjective assessments to incorporate comprehensive, data-driven insights about patient populations, physician influence networks, and geographical access patterns. The magnitude of this challenge has catalyzed interest in novel approaches that leverage emerging technologies to transform site selection from an art based primarily on experience and relationships into a science grounded in robust data analytics and predictive modeling. The financial stakes of addressing this issue are substantial, with industry estimates suggesting that more efficient recruitment strategies could potentially reduce development timelines by years and save hundreds of millions in development costs per approved therapy [3]. Table 1 summarizes the key challenges in clinical trial recruitment and their impacts on trial execution.

3. A Multi-Layered AI Solution

The data-driven matching pipeline described in the case study integrates four critical dimensions of information:

- Longitudinal Claims Data: 24 months of referral and prescription claims
- Clinical Pattern Recognition: Diagnostic-procedure "market-basket" coding
- Influence Mapping: Local peer-network influence analytics
- Geographic Intelligence: Heat-mapped patient density across the United States

This approach moves beyond conventional macro-level site feasibility to incorporate micro-level referral-network intelligence, identifying healthcare providers who can deliver eligible participants and locating untapped patient populations.The emergence of sophisticated artificial intelligence solutions for clinical trial recruitment represents a fundamental advancement in addressing the persistent challenges of patient enrollment. At the core of these next-generation platforms lies a multi-dimensional data integration framework that synthesizes diverse information streams to generate unprecedented insights into patient journeys, provider networks, and geographical distribution patterns. The longitudinal claims data component serves as a critical foundation, typically encompassing 24 months of medical and prescription history to establish comprehensive patient profiles. This temporal depth allows the system to identify not only current treatment patterns but also the progression of disease management strategies over time, revealing valuable insights about treatment-resistant populations or those with specific therapeutic journeys that align with protocol eligibility criteria. Similarly, the incorporation of diagnostic-procedure "market-basket" coding enables the identification of characteristic clinical patterns that frequently co-occur in the target patient population. This pattern recognition capability proves particularly valuable for identifying patients with complex presentations or comorbidities that might otherwise be overlooked in conventional feasibility assessments. By analyzing these interrelated diagnostic and procedural codes, the AI system can distinguish subtle differences in patient populations that significantly impact protocol eligibility but remain invisible to traditional screening methods. The sophisticated integration of these data dimensions creates a multifaceted understanding of where eligible patients receive care and which healthcare providers most frequently encounter the target population, fundamentally transforming how sponsors approach site selection and investigator engagement. Regulatory perspectives on these developments acknowledge the substantial potential of artificial intelligence to enhance clinical trial design and execution, emphasizing the importance of transparent methodologies that can be validated across diverse research contexts [5].Perhaps the most innovative aspects of advanced recruitment platforms are their incorporation of influence mapping and geographic intelligence capabilities. The influence mapping component leverages sophisticated network analysis algorithms to identify patterns of patient referrals and professional collaboration among healthcare providers. This approach identifies not merely physicians who treat many patients with the condition of interest, but those who function as influential nodes within local medical communities—practitioners whose referral patterns and professional standing position them as potential champions for clinical trial participation. By quantifying these influence relationships and visualizing them as interconnected networks, the system can prioritize engagement with healthcare providers who can mobilize substantial patient interest through their professional recommendations and referrals. Complementing this network intelligence, the geographic dimension adds crucial spatial context through heat-mapped visualizations of patient density across different regions. This geographic intelligence capability enables sponsors to identify "white space" opportunities—areas with substantial eligible patient populations but insufficient clinical trial infrastructure—while also optimizing the distribution of research sites to maximize geographical coverage and patient accessibility. Modern implementations of these systems employ sophisticated geospatial analysis techniques, including isochrone mapping that defines realistic patient travel boundaries based on actual transportation infrastructure rather than simple radius measurements. Research into healthcare utilization patterns has consistently demonstrated that travel time, rather than distance, represents the primary geographic barrier to clinical trial participation, making these advanced mapping capabilities particularly valuable for optimizing site placement and predicting enrollment performance with greater accuracy. The convergence of these four data dimensions—longitudinal claims, clinical pattern recognition, influence mapping, and geographic intelligence—creates a comprehensive recruitment intelligence platform that transcends traditional feasibility approaches by providing actionable insights at both macro and micro levels. Recent research in digital biomarkers and connected health has highlighted the potential of these integrated approaches to substantially improve trial recruitment efficiency while simultaneously enhancing the representativeness of study populations [6].The paradigm shift represented by these AI-enabled recruitment solutions lies in their ability to transcend conventional macro-level site feasibility assessments and incorporate micro-level referral-network intelligence. Traditional approaches to trial feasibility typically focus on broad metrics such as institution size, historical performance, and investigator experience—valuable indicators but insufficient for addressing the nuanced challenges of modern clinical trials with increasingly specific eligibility criteria. In contrast, the multi-layered AI approach generates insights with unprecedented granularity, identifying not only which research centers might theoretically have access to eligible patients but precisely which healthcare providers regularly encounter these patients and possess the influence to motivate their participation. This micro-level intelligence proves particularly valuable for studies targeting rare diseases or highly specific patient subpopulations, where eligible participants often receive care from a small subset of specialists distributed across numerous institutions. By mapping these provider networks and patient journeys with remarkable precision, these platforms enable sponsors to implement targeted outreach strategies that engage the specific physicians most likely to encounter protocol-eligible patients. The resulting efficiency gains manifest not only in accelerated enrollment timelines but also in more equitable access to clinical trial opportunities for patients regardless of their proximity to traditional academic research centers. As the healthcare landscape continues to evolve toward increasing specialization and fragmentation, this capability to navigate complex referral networks and identify untapped patient populations represents a critical competitive advantage for sponsors seeking to optimize their clinical development programs [5]. Table 2 summarizes the key dimensions of the multi-layered AI solution for clinical trial recruitment and their respective benefits.

4. The Methodology Behind the System

The platform employs a systematic workflow to identify optimal sites and investigators:

- Market Basket Creation: The system begins by finalizing diagnostic, procedure, and therapy codes linked to study criteria, establishing the foundation for patient identification.
- Claims and Referrals Analysis: By examining claims data, the system identifies healthcare providers (HCPs) treating the target patient population and computes both inbound and outbound referral volumes.
- Network Mapping: The AI detects peer clusters and calculates local authority and patient-share indices, effectively mapping the influence landscape within medical communities.
- Prioritization and Radius-Tiering: For each site, the system outputs HCP tiers at varying distances (5, 10, 50, and 100 miles), maximizing convenience for both patients and physicians.
- Geographic Analysis: Patient density heat-maps reveal "white-space" geographies—areas with sufficient patient populations but lacking active trial sites.
- Composite Ranking: The final step fuses scientific fit, network influence, and patient feasibility through multi-objective optimization to create actionable recommendations.

The methodological foundation of advanced clinical trial recruitment platforms represents a sophisticated synthesis of data science, network analysis, and healthcare informatics principles applied to the complex challenge of patient enrollment. The workflow begins with the critical process of market basket creation, wherein the system translates protocol eligibility criteria into a comprehensive set of diagnostic, procedural, and therapeutic codes that characterize the target patient population. This translation process involves far more nuance than simple keyword matching, as it must account for the various ways clinicians document similar clinical presentations across different healthcare systems and specialties. Advanced implementations employ natural language processing algorithms to analyze protocol text and identify semantically relevant medical concepts, then map these concepts to standardized terminology systems including ICD-10, CPT, HCPCS, and NDC codes. The resulting "market basket" functions as a digital signature of the ideal study candidate, enabling subsequent identification of patients and providers through claims data analysis. This initial step proves particularly valuable for complex protocols with numerous inclusion and exclusion criteria, as it establishes a standardized framework for patient identification that can be consistently applied across diverse healthcare databases and geographical regions. Research in pharmaceutical sciences has highlighted the transformative potential of artificial intelligence applications in various aspects of drug discovery and development, including the optimization of clinical trial design and execution through sophisticated data analytics and predictive modeling approaches [7].With the market basket established, the system proceeds to analyze claims and referral data, mapping the landscape of healthcare providers who regularly encounter potential study participants. This analysis extends beyond simple patient counts to examine treatment patterns, diagnostic sequences, and referral relationships that reveal the clinical journey of the target population. By processing millions of anonymized claims records, the system can identify not only which physicians currently manage patients who meet study criteria but also track referral patterns to understand how patients flow through the healthcare system. This bidirectional analysis of referral volumes—both inbound and outbound—proves particularly valuable for identifying the physicians who function as critical nodes in patient care networks. The resulting provider profiles incorporate metrics such as patient volume, treatment patterns, and network connectivity to create a multidimensional understanding of each physician's potential value to the recruitment effort. This approach represents a significant advancement over traditional site feasibility assessments, which typically focus on institution-level metrics rather than the specific healthcare providers most likely to encounter eligible patients. The network mapping phase builds upon this foundation by applying graph theory and social network analysis techniques to detect communities of practice and influence relationships among healthcare providers. By analyzing patterns of patient sharing, referrals, and professional collaboration, these algorithms can identify influential physicians whose participation or endorsement could substantially impact recruitment success. The system calculates metrics such as centrality, authority, and hub scores to quantify each provider's position within local professional networks, enabling sponsors to prioritize engagement with the most influential clinicians in each region. Recent research on artificial intelligence applications in clinical research emphasizes the importance of ethical considerations and regulatory compliance when implementing advanced analytics for trial optimization, noting that while these technologies offer substantial benefits for recruitment efficiency, they must be deployed with appropriate governance frameworks to ensure patient privacy and algorithmic transparency [8].The workflow continues with sophisticated geographical analysis and prioritization processes that transform network intelligence into actionable outreach strategies. The radius-tiering approach acknowledges the critical importance of geographic proximity in patient participation decisions, systematically organizing healthcare providers into concentric circles around each investigative site. This stratification enables research teams to implement geographically optimized outreach campaigns that prioritize providers within convenient travel distance for patients while still capturing influential physicians in the broader region. Advanced implementations of this methodology incorporate isochrone mapping techniques that define these radius tiers based on actual travel times rather than simple distance measurements, accounting for transportation infrastructure and natural barriers that impact patient access. Complementing this site-centric approach, the geographic analysis component employs heat-mapping techniques to visualize patient density across regions, identifying areas with substantial eligible populations that the current distribution of research sites may underserve. This capability proves particularly valuable for studies struggling with enrollment, as it can reveal promising "white-space" opportunities for site expansion or targeted outreach. The culmination of this systematic workflow is the composite ranking process, which employs multi-objective optimization algorithms to synthesize the diverse analytical dimensions into a unified prioritization framework. This approach balances sometimes competing considerations such as patient volume, provider influence, scientific alignment, and geographical accessibility to generate ranked recommendations tailored to each study's specific objectives and constraints. The resulting actionable intelligence enables sponsors and sites to implement evidence-based recruitment strategies that precisely target the healthcare providers most likely to accelerate enrollment while expanding access to research opportunities for previously overlooked patient populations. This comprehensive methodological framework represents a paradigm shift in recruitment approach—moving from intuition-based decisions to data-driven strategies that systematically address the complex, multifaceted challenges of clinical trial enrollment [7]. Table 3 outlines the systematic workflow of the AI-enabled clinical trial recruitment process, highlighting the key technologies and outcomes at each stage.

5. Case Study: COPD Trial Implementation

A prospective chronic obstructive pulmonary disease (COPD) study across 68 sites in the United States demonstrated the system's effectiveness: From a universe of 292,578 specialty healthcare professionals, the system identified 5,238 clinical leaders with high influence and patient volume. The AI generated radius-tiered outreach lists totaling approximately 83,000 clinicians.When sites engaged with the AI-recommended healthcare providers, they experienced significant enrollment increases ranging from 15% to 31%. The system also projected a 25% reduction in overall timeline compared to traditional recruitment methods.The most notable enrollment impacts were observed at:

- Temple Lung Center (PA): 31% increase
- Southeastern Research Center (PA): 29% increase
- Lowcountry Lung & Critical Care (SC): 29% increase
- Sierra Clinical Research (NV): 21% increase
- UT South-western Medical Center (TX): 17% increase
- Winston-Salem cohort (NC): 15% increase

The implementation of artificial intelligence-driven recruitment strategies in a large-scale chronic obstructive pulmonary disease (COPD) clinical trial provides compelling evidence for the transformative potential of advanced analytics in addressing enrollment challenges. This prospective study, conducted across 68 sites throughout the United States, employed the multi-layered analytical approach to enhance patient recruitment in a therapeutic area historically plagued by enrollment difficulties. COPD trials present particular recruitment challenges due to the complex nature of the disease, significant comorbidities in the patient population, and stringent inclusion/exclusion criteria that typically require specific pulmonary function parameters and treatment histories. The initial analytical phase processed an extensive universe of 292,578 specialty healthcare professionals, applying sophisticated network analysis and patient-volume algorithms to identify 5,238 clinical leaders with both high influence within their professional communities and substantial COPD patient volumes. This identification process incorporated multiple dimensions of analysis, including claims-based patient populations, referral patterns, publication histories, and network centrality metrics. The resulting prioritization represented approximately 1.8% of the initial universe—a focused subset of clinicians with disproportionate impact potential for study enrollment. From this foundation, the system generated radius-tiered outreach lists for each participating site, stratifying healthcare providers into concentric bands based on geographic proximity to research centers. The comprehensive recommendation set encompassed approximately 83,000 clinicians across all radius tiers, providing research sites with systematically organized outreach priorities tailored to their specific location and catchment area. This structured approach enabled sites to implement geographically optimized engagement strategies that prioritized high-influence providers within reasonable travel distance for potential participants. The methodology demonstrated particular value in identifying community pulmonologists and primary care physicians who regularly manage COPD patients but had not previously participated in research, expanding the recruitment network beyond traditional academic investigators. This approach aligns with emerging research on artificial intelligence tools for early identification of COPD patients, which has demonstrated the potential for machine learning algorithms to enhance disease detection and management through sophisticated pattern recognition in clinical and diagnostic data [9].The implementation results from this COPD trial provide quantitative validation of the AI-enabled approach's impact on enrollment performance. When research sites engaged with the healthcare providers identified through the advanced analytics platform, they experienced substantial acceleration in patient recruitment, with enrollment increases ranging from 15% to 31% compared to pre-implementation rates. The most dramatic improvements occurred at the Temple Lung Center in Pennsylvania, which recorded a 31% increase in randomized patients following implementation of the AI-recommended outreach strategy. Similarly impressive results were observed at the Southeastern Research Center (29% increase), Lowcountry Lung & Critical Care (29% increase), Sierra Clinical Research (21% increase), UT Southwestern Medical Center (17% increase), and the Winston-Salem cohort (15% increase). These enrollment gains translated into significant timeline advantages, with the system projecting a 25% reduction in overall study duration compared to traditional recruitment methods. This acceleration has profound implications for drug development economics, as each month of timeline compression in a late-phase trial represents substantial cost savings and earlier market access for new therapies. The consistency of enrollment improvements across diverse sites—spanning academic medical centers, dedicated research facilities, and community-based practices—demonstrates the generalizability of the approach across different research settings and geographical regions. Moreover, the implementation experience revealed additional operational benefits beyond raw enrollment numbers, including more efficient site activation processes, improved protocol adherence, and enhanced geographical diversity in the participant population. These secondary benefits address persistent challenges in clinical trial execution that impact not only recruitment timelines but also data quality and population representativeness. The detailed case study findings provide compelling evidence that AI-enabled recruitment strategies can deliver tangible performance improvements in complex, multi-site clinical trials, suggesting broader applicability across therapeutic areas and study designs. Recent research on decentralized clinical trials and technological innovations in research design reinforces the value of data-driven approaches to optimizing trial operations, particularly for conditions like COPD that require specialized diagnostic assessments and diverse patient populations [10].The COPD trial implementation experience also yielded valuable insights regarding the practical application of AI-driven recruitment strategies within existing clinical research workflows. The operational integration process involved several key components, including site-specific training on utilizing the AI-generated recommendations, development of standardized outreach materials targeting the identified healthcare providers, and implementation of tracking mechanisms to monitor engagement outcomes. Research coordinators reported that the radius-tiered approach significantly streamlined their outreach prioritization decisions, enabling more focused allocation of limited recruitment resources. Similarly, investigators noted that the influence-based prioritization helped identify previously overlooked referral sources within their communities, expanding their effective recruitment network beyond traditional channels. The implementation team observed that sites achieving the greatest enrollment gains typically demonstrated three common characteristics: early adoption of the AI recommendations, consistent execution of targeted outreach activities, and systematic documentation of engagement outcomes. These findings highlight the importance of not merely generating sophisticated analytical insights but ensuring their effective operationalization within the practical constraints of clinical research environments. The COPD case study also revealed opportunities for further refinement of the AI-driven approach, including more seamless integration with electronic health record systems, enhanced visualization tools for network mapping, and more granular tracking of referral outcomes. Nevertheless, the substantial enrollment improvements documented across multiple sites provide compelling validation of the core methodology's effectiveness in addressing one of the most persistent challenges in clinical trial execution. The projected 25% reduction in overall timeline represents a transformative advancement in study efficiency with significant implications for accelerating the development of new therapies for COPD and potentially numerous other therapeutic areas [9].Table 4 summarizes the enrollment improvements observed across different research centers implementing the AI-driven recruitment approach in the COPD trial, along with key success factors identified at each site.

Table 1: *Clinical Trial Recruitment Challenges and Impacts [3, 4]*

Challenge	Impact
Historical enrollment reports	Timeline extensions
Subjective feasibility surveys	Cost overruns
Information asymmetries	Suboptimal site distribution
Lack of granular patient data	Missed eligible populations
Misalignment of sites and demographics	Slow enrollment rates
Inadequate community physician engagement	Limited patient referrals
Geographical barriers	Reduced patient accessibility
Optimistic site capability assessments	Unmet enrollment targets

Table 2: *Multi-Layered AI Solution for Clinical Trial Recruitment [5, 6]*

Data Dimension	Key Features	Benefits
Longitudinal Claims Data	24-month medical history, Prescription patterns, Treatment progression	Identifies treatment-resistant populations, Reveals therapeutic journeys matching protocols
Clinical Pattern Recognition	Diagnostic-procedure "market-basket" coding, Co-occurring clinical patterns	Detects complex presentations, Identifies comorbidities, Reveals subtle eligibility factors
Influence Mapping	Network analysis, Referral patterns, Professional collaboration	Pinpoints influential providers, Identifies clinical champions, Enables targeted engagement
Geographic Intelligence	Heat-mapped patient density, Isochrone mapping, Travel time analysis	Reveals "white space" opportunities, Optimizes site distribution, Improves patient accessibility

Table 3: *AI-Enabled Clinical Trial Recruitment Workflow [7, 8]*

Stage	Process	Key Technologies	Outcome
1	Market Basket Creation	NLP algorithms, Medical concept mapping	Digital signature of ideal study candidate
2	Claims & Referrals Analysis	Claims processing, Treatment pattern analysis	Identification of key HCPs managing target patients
3	Network Mapping	Graph theory, Social network analysis	Detection of influential physician networks
4	Prioritization & Radius-Tiering	Concentric distance banding (5-100 miles)	Geographically optimized outreach strategy
5	Geographic Analysis	Heat-mapping, Patient density visualization	Identification of "white-space" opportunities
6	Composite Ranking	Multi-objective optimization algorithms	Actionable, prioritized HCP recommendations

Table 4: *AI-Driven COPD Trial Enrollment Improvements [9, 10]*

Research Center	Location	Enrollment Increase	Key Success Factors
Temple Lung	Pennsylvania	31%	Early AI recommendation adoption

Center			
Southeastern Research Center	Pennsylvania	29%	Targeted outreach implementation
Lowcountry Lung & Critical Care	South Carolina	29%	Systematic engagement tracking
Sierra Clinical Research	Nevada	21%	Expanded referral networks
UT South-western Medical Center	Texas	17%	Optimized geographic targeting
Winston-Salem cohort	North Carolina	15%	Community physician engagement

6. Conclusions

The integration of protocol semantics with real-world care-path data and peer-network analytics fundamentally transforms clinical trial site selection from a static, survey-driven process into a dynamic, evidence-rich strategy. By harnessing the power of artificial intelligence to synthesize diverse data streams—including claims information, referral patterns, influence networks, and geographic distributions—this innovative approach enables unprecedented precision in identifying both optimal research locations and the specific healthcare providers most capable of facilitating patient enrollment. The documented enrollment gains and timeline compression observed across diverse research settings validate AI-enabled site and investigator matching as an essential component of modern clinical development strategy. Beyond merely accelerating study timelines, this methodological evolution democratizes research participation by identifying previously overlooked patient populations and expanding trial access beyond traditional academic centers. As healthcare continues to specialize and fragment, the ability to navigate complex provider networks and pinpoint untapped patient pools represents a critical competitive advantage for research sponsors. This transformative shift from intuition-based decisions to evidence-optimized strategies promises to simultaneously advance scientific discovery, reduce development costs, and improve patient outcomes by facilitating more efficient delivery of breakthrough therapies to those who need them most.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.

- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] David B Fogel, "Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review," Contemporary Clinical Trials, vol. 66, pp. 74-79, 2018. [Online]. Available:https://pmc.ncbi.nlm.nih.gov/articles/PMC6092479/

[2] Matthew Hutson, "How AI is being used to accelerate clinical trials," Nature, 2024. [Online]. Available: https://www.nature.com/articles/d41586-024-00753-x

[3] Morgan M Millar et al., "Clinical trial recruitment in primary care: exploratory factor analysis of a questionnaire to measure barriers and facilitators to primary care providers' involvement," BMC Primary Care, 2022. [Online]. Available:https://pmc.ncbi.nlm.nih.gov/articles/PMC9719201/

[4] PubMed, "Transforming Clinical Research in the United States: Challenges and Opportunities: Workshop Summary," Institute of Medicine (US) Forum on Drug Discovery, Development, and Translation, 2010. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/21210556/

[5] U.S Food & Drug Administration, "The Role of Artificial Intelligence in Clinical Trial Design and Research with Dr. ElZarrad," FDA Office of Clinical Pharmacology. [Online]. Available: https://www.fda.gov/drugs/news-events-human-drugs/role-artificial-intelligence-clinical-trial-design-and-research-dr-elzarrad

[6] Yamilé Molina et al., "Leveraging system sciences methods in clinical trial evaluation: An example concerning African American women diagnosed with breast cancer via the Patient Navigation in Medically Underserved Areas study," Contemporary Clinical Trials Communications, Volume 15, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2451865419300250

[7] Stefan Harrer et al., "Artificial Intelligence for Clinical Trial Design," Trends in Pharmacological Sciences, Vol. 40, No. 8, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165614719301300

[8] Mia S Tackney et al., "Digital endpoints in clinical trials: emerging themes from a multi-stakeholder Knowledge Exchange event," Trials, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11297702/

[9] Huanyu Long et al., "Digital health in chronic obstructive pulmonary disease," Chronic Diseases and Translational Medicine, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10249197/

[10] Thai Hau Koo et al., "Systematic Review of the Application of Artificial Intelligence in Healthcare and Nursing Care," Malaysian Journal of Medical Sciences, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11477473/