

The Intelligent Clinical Research Organization: A Conceptual Framework for Artificial Intelligence and Federated Learning in Trial Operations

Zomana MAJID^{1,2}

¹ ReliSource, Inc., 110 Great Road, Bedford, MA 01730, USA

² Boston University, 1010 Commonwealth Avenue, Boston, MA 02215, USA

E-mail: zomana@relisource.com

* Author to whom correspondence should be addressed;

Received: 3 February 2026 / Accepted: 23 September 2026 / Published online: 30 September 2026

Abstract

The use of AI (Artificial Intelligence) applications in clinical trials run by CROs (Clinical Research Organizations) has increased significantly, but studies have not yet provided consistent evidence of its effectiveness. This paper discusses AI applications in three key areas: patient selection and screening, genomics turnaround time, and biomarker-based patient stratification. The paper also examines industry analyses of AI implementation in biopharmaceutical development. Some of the findings from the research are: a 34% to 41% decrease in screening time with the help of Natural Language Processing -based eligibility criteria, turnaround times in genomics of several hours because of ultra-rapid sequencing, and AUC results of about 0.65 to 0.809 when predicting the treatment response with the help of AI as opposed to the single biomarker test. Based on these findings, the paper proposes the Intelligent CRO model, which includes NLP-based screening, automated genomic processing, and federated learning.

Keywords: Artificial Intelligence (AI) in Healthcare; Federated Learning; Natural Language Processing (NLP); Clinical Trial Operations; Precision Medicine.

Introduction

Artificial intelligence is widely promoted as a solution to the operational problems of running clinical trials, yet peer-reviewed evidence for specific, measured claims stays scattered across separate fields of study: NLP (Natural Language Processing) research, clinical genomics, and oncology biomarker studies rarely cite one another despite addressing the same underlying problem of trial velocity [1]. Industry reports describe a similar picture from the sponsor and CRO (Clinical Research Organization) side. Patient enrichment, recruitment, and site selection are among the leading use cases for AI adoption in biopharma research and development (R&D). Yet the most recent published industry analysis shows AI adoption remains uneven across the clinical development ecosystem, and end-to-end development timelines have not gotten shorter despite growing AI investment [2,3]. This gap between interest in adoption and proven operational benefit is exactly the split between fields this paper addresses.

The peer-reviewed literature consistently identifies two key limitations. One of these limitations is patient identification and screening, which remain time-consuming processes in which clinical coordinators must manually review patient information in EHRs, pathology, and genomic databases to identify eligible patients; this can result in eligible patients being missed when relevant information is documented only in clinical notes [4,5]. Second, genomic interpretation, particularly the classification of variants of uncertain significance, has for years required

turnaround times measured in weeks, a delay that matters clinically for patients with aggressive or rapidly progressing disease [6].

The paper bridges the gap through an integration of quantitative data collected from various peer-reviewed studies instead of conducting any experiment from scratch. The studies are divided into four sections, which are rarely reviewed together, including NLP patient screening, Automated genomic pipeline, Biomarker stratification through artificial intelligence, and Federated Learning. Industry reports on the adoption of AI technologies are also addressed independently. On the basis of all the information provided, an operational conceptual model known as the Intelligent CRO is proposed, which explains how the separate elements can be integrated into a single CRO workflow. The objective of this paper is to clarify the existing scientific evidence, what is unknown and what more evidence is required for evaluating the integrated system.

Evidence Synthesis Approach

This review is a narrative synthesis of the existing literature rather than a systematic review and does not provide original empirical findings; it is a narrative literature synthesis of articles found through targeted searching of PubMed, JMIR, and Nature Portfolio articles published from 2019 to 2026 reporting quantitative results regarding AI or NLP for patient screening, genomic variant calling and turnaround times, and prediction of immunotherapy responses compared to traditional single-biomarker analysis. Real-world clinical applications, such as prospective, pilot, and multi-site studies, have been prioritized over retrospective and simulation studies since these studies provide more insight into the potential advantages that could be realized in practice. Reports from industry, including Deloitte Insights and the IQVIA Institute, have been considered in a separate section to describe trends in AI technology adoption and have been kept separate from the peer-reviewed literature. All figures presented in the paper are cited, and no figures were calculated, extrapolated, or combined.

This method has one major flaw, and it is worth stating up front rather than treating it as a footnote. The peer-reviewed studies reviewed here were conducted independently, at different institutions, with different patient populations, disease areas, and system setups. Because of that, the gains reported by each study cannot be directly compared. No single dataset in the published literature evaluates NLP-based screening, automated genomic pipelines, and AI-assisted biomarker stratification together within one CRO's real operations. The proposed Intelligent CRO model is meant to make this gap clear rather than overlook it.

Research Evidence

Patient Identification and Screening Velocity

A prospective twelve-month evaluation of an NLP- and machine-learning-based eligibility screener integrated into the clinical research coordinator workflow at a pediatric emergency department found that automated screening reduced patient screening time by 34%, with a statistically significant result ($P < .001$) [4]. A separate deployment of a hybrid rules-based and NLP-based system across 29 oncology trials, involving more than 98,000 patients, reduced screening time per chart by 41% and reduced chart-review workload by roughly tenfold compared with manual review, while maintaining high agreement with gold-standard chart review [5].

Ultra-rare diseases represent another challenge, as patients suffering from ultra-rare diseases usually do not have a specific diagnostic code and can be described only using symptom descriptions. An AI-based phenotyping approach applied to more than 350,000 retrospective EHRs across five hospitals addressed this: within one month of screening, it flagged 104 patients for review, of whom 19 remained clinically plausible after specialist adjudication, a much faster way to find patients than conventional coded-record search [7]. Another phenotyping algorithm (PheIndex) designed to capture cases of rare genetic conditions during early childhood reached 90% sensitivity and 97% specificity, showing the potential of symptom pattern identification to attain comparable levels of accuracy with the diagnosis coding-based approaches and identify patients who may be missed by coding-based methods [8].

Genomic Turnaround Time

Rapid whole-genome sequencing (rWGS) pipelines have greatly reduced genomic turnaround time compared with the multi-week timelines previously associated with variant classification. An ultra-rapid nanopore sequencing workflow that combined optimized sample preparation, real-time base calling, and faster variant filtering identified a candidate disease-causing variant in under eight hours from sample preparation in two clinical cases, about twice as fast as previously reported rapid sequencing methods [6]. This result came from a small proof-of-concept study involving two cases. A larger program-level study provides a useful comparison: a citywide rWGS program serving critically ill pediatric patients across neonatal and pediatric intensive care units reported a median turnaround time of 3.4 days for 100 patients undergoing routine trio sequencing, with an overall diagnostic yield of 53%, rising to 80% among consanguineous families [13]. The difference between the sub-eight-hour proof-of-concept result and the multi-day median at program scale shows that early feasibility results do not necessarily apply to routine, full-service turnaround times.

Artificial Intelligence -Assisted Biomarker Stratification

A number of studies have compared AI- or machine-learning-based prediction of immunotherapy response against single-biomarker assessment using PD-L1 (programmed death-ligand 1) immunohistochemistry alone. Within the PD-L1-negative subgroup, where PD-L1 testing is by definition uninformative, a machine-learning model assessing tumor-infiltrating lymphocytes alone achieved an AUC of 0.77 for predicting response to immune checkpoint inhibition in non-small cell lung cancer, compared with an AUC of 0.65 for tumor mutation burden (TMB) alone, indicating that TIL assessment can outperform TMB as a single-marker approach when PD-L1 offers no signal [9]. A separate multimodal deep learning model integrating pathology, radiomic, and clinical features achieved an AUC of 0.809 for predicting immunotherapy response in esophageal cancer [10]. These studies as a whole provide evidence for the broad idea that AI- or machine-learning-based prediction, whether from a single marker assessed computationally or from multimodal integration, can outperform conventional single-biomarker assessment, with reported AUC values in the range of roughly 0.65-0.809, though the exact size of the improvement varies a good deal by cancer type, model architecture, and validation cohort.

Federated Learning as an Enabling Infrastructure

Federated learning is becoming a possible solution for training and deployment of the above-mentioned screening and stratification strategies without centralization of patient data, which is required because of the Health Insurance Portability and Accountability Act, the General Data Protection Regulation, and similar legislation [11]. Reviewing federated learning in health care has shown that there are both advantages in terms of privacy and regulatory compliance, as well as some unresolved challenges such as differences in data across organizations, overhead of communication, security concerns such as model inversion and poisoning, and lack of studies on clinical validation of federated learning [11]. This warning applies to any recommendation, such as the one recommended here, that treats federated learning as a developed system rather than an area for research.

Industry-Analyst Perspective on Adoption

While peer-reviewed medical journals offer insights into adoption trends, industry reports are derived from different sources and are discussed separately. An industry report that is often cited identifies patient enrichment, recruitment and enrollment, investigator/site selection, and operations- and data-based analytics as the most active domains for adoption of AI in biopharma research and development. The same report also finds that CROs with data science capabilities are becoming more strategically important to biopharma sponsors, which aligns with the intelligent CRO model put forth in this paper [2]. Another study warns about this trend in the adoption of new technology: so far, timelines within the ecosystem have not decreased. Early evidence does show higher success rates among individual AI-enabled development programs. This points to a gap between the component-level gains documented in the peer-reviewed studies above and the system-level outcomes tracked at the ecosystem level [3].

Table 1. Efficiency Benefits of Artificial Intelligence Reported in Published Literature.

Operational Area	Intervention	Reported Gain	Source
Patient Screening Time	NLP-based eligibility screener (pediatric ED)	34% reduction ($P < .001$)	Ni et al. [4]
Patient Screening Time	Hybrid rules/NLP prescreening (29 oncology trials)	41% reduction per chart	Gong et al. [5]
Genomic Turnaround Time	Ultra-rapid nanopore sequencing pipeline	Under 8 hours (2 cases)	Goenka et al. [6]
Genomic Turnaround Time	Citywide rWGS program (100 patients)	Median 3.4 days; 53% yield (80% consanguineous)	Rabea et al. [13]
Rare-Disease Case Identification	AI phenotyping of retrospective EHRs (Pompe disease)	104 flagged; 19 clinically plausible after review	Lin et al. [7]
Rare-Disease Case Identification	PheIndex phenotyping algorithm (ages 0-3)	90% sensitivity, 97% specificity	Webb et al. [8]
Immunotherapy Response Prediction	ML-based TIL assessment, PD-L1-negative subgroup (NSCLC)	AUC 0.77 vs. 0.65 (TMB alone)	Rakaee et al. [9]
Immunotherapy Response Prediction	Multimodal deep learning (esophageal cancer)	AUC 0.809	Liu et al. [10]

The figures in the table were obtained from separate studies and therefore cannot be directly compared. The data in the table does not come from a single database.

Intelligent Clinical Research Organization: An Operational Framework

Based on the evidence presented in the previous sections, the Intelligent CRO is proposed as a conceptual framework for CROs. It consists of three layers: an NLP-based filter layer that identifies candidates from structured and unstructured EHR data regardless of the level of coding; an automated genomic pipeline layer that speeds up variant calling and classification before review by a geneticist; and a federated learning layer that enables models to be trained at both participating sites while patient data remains at each site. The three aspects correspond to three separate stages of the pipeline where the peer-reviewed literature shows capacity-related bottlenecks: identifying eligible patients, confirming patient eligibility at the genomic level, and scaling either process without centralizing patient information. These aspects are paired based on the structure of the problem, not on the evidence, as each addresses a different stage of the same process. The above framework combines elements that have been individually validated through the literature reviewed above; however, this framework has not yet been tested or used as an integrated process. There is no assertion of the cumulative impact of the three elements working together.

The proposed model can be illustrated with a hypothetical example, presented as an illustration rather than a reported case. Consider a rare-disease trial in which manual chart review over several months identifies only a small number of eligible patients based on coded diagnoses. Applying an NLP screening layer similar to the one validated in the rare-disease phenotyping studies above [7,8] to the same patient population would likely identify more candidates based on the patterns reported in those studies. These candidates would have eligibility information documented only in symptom descriptions rather than coded diagnoses. Each candidate would still need the same confirmatory genetic testing used in the underlying studies. This example is based on the general findings reported across the cited studies; it does not describe an actual trial, patient count, or outcome and is not intended to be interpreted as one.

Discussion

Evidence-Supported Claims

These studies provide evidence for several well-supported claims: NLP screening has been shown to save time in practice by reducing screening time and workload [4,5]; genomics automation can significantly reduce turnaround time to days or even hours in well-equipped laboratories [6]; and AI-enabled prediction using single or multimodal biomarkers can be more effective than traditional single-biomarker testing when assessing treatment response, although the results vary significantly depending on the cancer type and model [9,10]. Each of these findings is based on a particular, citable, peer-reviewed article that presents original findings instead of being based on an aggregation of statistics; a general overview of AI technology in immunotherapy for lung cancer provides context for these studies [12].

Claims Not Yet Supported by the Evidence

There is no evidence in this literature review of any study that examines the combined use of NLP screening, an automated genomic pipeline, and federated learning in the processes of a CRO across various trials and therapeutic areas. Hence, the value of this combination cannot be determined. It is unknown whether these technologies interact or reinforce each other, or whether this interaction reduces the overall benefit. This paper does not address this question. The reader must not view the various percentages quoted above as components of an obtainable total.

Limitations

The present synthesis is a narrative rather than a systematic review, as it does not follow a PRISMA methodological framework or include a detailed description of the search methodology used in all relevant databases. The differences in patient samples, disease areas, settings, and levels of maturity of systems make it difficult to compare the figures obtained from the included studies. In the field of federated learning, there are only limited systems that have been clinically validated and deployed compared to the ones that are suggested or simulated [11]. Thus, the Intelligent CRO concept introduced here must be taken as a future validation direction rather than as a solution. Since this paper reviews literature and does not involve any analysis on a new set of data, no clinical trial registration number can be provided, as discussed in detail in the cover letter.

Generalizability

These screenings and biomarker studies cited above have been conducted in specific institutional and disease settings: pediatric emergency department, oncology trials network, populations of non-small cell lung cancer and esophageal cancer, and rare diseases population. The studies mentioned here have not shown how far these gains can be generalized to other trial stages, other diseases, and other health system settings, and generalization would require dedicated prospective evaluation.

Conclusion

Peer-reviewed evidence from independent studies, rather than a single study, shows that NLP for screening, AI for genomic pipeline processes, and AI for biomarker stratification can measurably improve clinical trials. This is largely in line with, although more reserved about, the industry's reporting on the adoption of AI in biopharmaceutical R&D. The Intelligent CRO Model, proposed here, is a theoretical model for the integration of all these elements based on the above-mentioned evidence; however, taking into consideration the fact that no such aggregate effect of all these elements has ever been proven in the scientific literature. An empirical study analyzing an integrated system through a prospective multicenter trial would be a reasonable next step.

List of Abbreviations: AI - Artificial Intelligence; AUC - Area Under the Curve; CRO - Clinical Research Organization; ED - Emergency Department; EHR - Electronic Health Record; GDPR - General Data Protection Regulation; HIPAA -

Health Insurance Portability and Accountability Act; NLP - Natural Language Processing; NSCLC - Non-Small Cell Lung Cancer; PD-L1 - Programmed Death-Ligand 1; R&D - Research and Development; rWGS - Rapid Whole-Genome Sequencing; TIL - Tumor-Infiltrating Lymphocyte; TMB: Tumor Mutation Burden.

Author Contributions: As the sole author, ZM conceived the conceptual framework, conducted the literature synthesis, and drafted the manuscript. ZM approved the final version and agrees to be accountable for the work's accuracy and integrity.

Funding: This research did not receive any external funding.

Data Availability Statement: This paper does not analyze any original or proprietary dataset. All figures cited are taken from publicly available peer-reviewed sources and industry reports listed in the References section, with each source identified by its full citation and, where available, its digital object identifier (DOI).

Conflict of Interest: The authors declare no conflict of interest.

References

1. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med.* 2019;25(1):30-36. <https://doi.org/10.1038/s41591-018-0307-0>
2. Properzi F, Cruz MJ. Intelligent clinical trials: transforming through AI-enabled engagement. Deloitte Insights. 2020 Feb 9. Available from: <https://www.deloitte.com/us/en/insights/industry/life-sciences/artificial-intelligence-in-clinical-trials.html>
3. IQVIA Institute for Human Data Science. Global R&D Trends 2026: Advancing Innovation in a Changing Landscape. IQVIA; 2026 Mar 25. Available from: <https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/global-r-and-d-trends-2026>
4. Ni Y, Bermudez M, Kennebeck S, Liddy-Hicks S, Dexheimer J. A real-time automated patient screening system for clinical trials eligibility in an emergency department: design and evaluation. *JMIR Med Inform.* 2019;7(3):e14185. <https://doi.org/10.2196/14185>
5. Gong G, Liu J, Pandya S, Taborda C, Wiesendanger N, Price N, et al. Clinical trial patient matching: a real-time, common data model and artificial intelligence-driven system for semiautomated patient prescreening in cancer clinical trials. *JCO Clin Cancer Inform.* 2026;10:e2500262. <https://doi.org/10.1200/CCI-25-00262>
6. Goenka SD, Gorzynski JE, Shafin K, Fisk DG, Pesout T, Jensen TD, et al. Accelerated identification of disease-causing variants with ultra-rapid nanopore genome sequencing. *Nat Biotechnol.* 2022;40:1035-1041. <https://doi.org/10.1038/s41587-022-01221-5>
7. Lin S, Nateqi J, Weingartner-Ortner R, Gruarin S, Marling H, Pilgram V, et al. An artificial intelligence-based approach for identifying rare disease patients using retrospective electronic health records applied for Pompe disease. *Front Neurol.* 2023;14:1108222. <https://doi.org/10.3389/fneur.2023.1108222>
8. Webb BD, Lau LY, Tsevdos D, Shewcraft RA, Corrigan D, Shi L, et al. An algorithm to identify patients aged 0-3 with rare genetic disorders. *Orphanet J Rare Dis.* 2024;19(1):183. <https://doi.org/10.1186/s13023-024-03188-9>
9. Rakaee M, Adib E, Ricciuti B, Sholl LM, Shi W, Alessi JV, et al. Association of machine learning-based assessment of tumor-infiltrating lymphocytes on standard histologic images with outcomes of immunotherapy in patients with NSCLC. *JAMA Oncol.* 2023;9(1):51-60. <https://doi.org/10.1001/jamaoncol.2022.4933>
10. Liu H, Bai Y, Wang Z, Yin S, Gong C, Wang B. Multimodal deep learning for predicting PD-L1 biomarker and clinical immunotherapy outcomes of esophageal cancer. *Front Immunol.* 2025;16:1540013. <https://doi.org/10.3389/fimmu.2025.1540013>
11. Abbas SR, Abbas Z, Zahir A, Lee SW. Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration. *Healthcare.* 2024;12(24):2587. <https://doi.org/10.3390/healthcare12242587>
12. Gao Q, Yang L, Lu M, Jin R, Ye H, Ma T. The artificial intelligence and machine learning in lung cancer immunotherapy. *J Hematol Oncol.* 2023;16(1):55. <https://doi.org/10.1186/s13045-023-01456-y>
13. Rabea F, Aljasmii I, Jain R, et al. Citywide implementation of a rapid whole-genome sequencing program for critically ill pediatric patients. *Nat Med.* 2026;32:3359–3371. <https://doi.org/10.1038/s41591-026-04598-x>