Speed and Accuracy of an AI-Automated Medical Record Screening Tool for Therapeutic Clinical Trials

#46

Kaushal M. Kulkarni, M.D., Meghan Russell, M.D., Sean Crosby, Ph.D., Michael Frantz, Ph.D.

Predoc, 895 Broadway, New York, NY 10003

Abstract

Background

The recruitment of medically qualified patients is a major bottleneck in therapeutic clinical trials. Traditional manual screening methods are cost and labor-intensive, requiring up to 8.8 hours per enrolled patient per study.

Objective

To assess the speed and accuracy of an AI-automated medical record screening tool for therapeutic clinical trials.

Design

Retrospective chart review.

Results

We observed an overall **question accuracy of 98.1%**, a false negative rate of 0.46%, a false positive rate of 7.41%, and a **median processing time of 2.60 minutes** for PDF records. For Health Information Exchange records, we observed an overall **question accuracy of 99.7%**, a false negative rate of 0.24%, a false positive rate of 0.77%, and a **median processing time of 4.31 minutes**.

Conclusion

Our platform demonstrates the potential to drastically reduce the time and cost associated with manual medical record review for screening and medical qualification of patients into therapeutic clinical trials.

Introduction

The recruitment of medically qualified patients is a major bottleneck in therapeutic clinical trials. Traditional manual screening methods in therapeutic clinical trials are labor-intensive and contribute to up to 30% of clinical trial costs.[1] The manual retrieval and review of heterogeneous medical records, including unstructured PDFs from prior providers and semi-structured health information exchange (HIE) data, consumes a disproportionate amount of these screening costs, which have been estimated to require 3.4–8.8 hours per enrolled patient per study, and cost $129–$2,508 per enrolled patient per site.[2]

Here, we assess an AI-automated medical record screening platform tailored for therapeutic trials. Following AI-assisted retrieval and aggregation of patient records, we then harness natural language processing (NLP) and machine learning (ML) to rapidly answer inclusion and exclusion (I/E) questions from diverse record formats, while attempting to maintain clinical accuracy.

Methods

We retrospectively reviewed data on medical records of patients which were retrieved and analyzed between August 1, 2025 to September 30, 2025 as part of the recruitment efforts across 61 distinct therapeutic protocols.

Medical records on patients were obtained through an AI-assisted process which is outside the scope of this presentation.

I/E criteria for each trial were manually extracted from the clinical trial protocol provided by the sponsor or site. The I/E criteria for each protocol were then transformed into a set of machine learning prompts or questions designed to be answered from the medical records with a binary “yes” or “no”. If the answer to the question was not available or found within the available medical record set, the question was answered as a “no.” These yes/no questions were then asked over many medical records for each I/E protocol.

Retrieved medical records consisted of 2 different record types: (1) unstructured PDFs and (2) semi-structured data retrieved from Health Information Exchanges (HIEs).

Overall, we analyzed 1,569 PDF records against one of 54 unique sets of I/E criteria consisting of between 10 and 206 yes/no questions each. Overall we assessed 400 distinct yes/no questions, each of which had at least 50 observations for a total of 68,035 question-answer pairs.

Similarly, we analyzed 1,118 HIE records against 7 unique sets of I/E criteria, consisting of between 10 and 206 yes/no questions each. Overall we assessed 192 distinct yes/no questions, each of which had at least 50 observations for a total of 76,957 question-answer pairs.

Each question-answer pair included in this analysis included citations to page sources within the medical record, and was reviewed for accuracy by a trained human clinical reviewer.

Processing time was estimated on a sub-sample of 200 HIE records and 705 PDF records by comparing timestamps in processing records. We also evaluated the size (number of rows) for each of the 200 HIE records to determine the relationship between document size and processing time.

Results

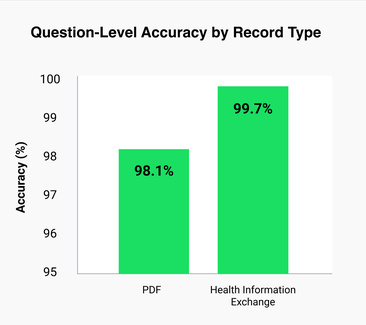

Figure 1. Question-level accuracy by record type

We observed an overall question accuracy of 98.1% for PDF records and 99.7% for HIE records (Figure 1).

Document Type	Answer	Accuracy (%)	N
Health Information Exchange	No	99.76	70081
Health Information Exchange	Yes	99.23	6500
PDF	No	99.54	55425
PDF	Yes	92.59	6737

Table 1: Accuracy by answer and record type

We also evaluated accuracy according to whether the answer to the clinical I/E criteria question was determined to be a “yes” or “no” by the machine learning tool, as measured against the final determination by a trained human clinical reviewer.

For HIE records, we observed a false negative rate of 0.24% and a false positive rate of 0.77%. For PDF records, we observed a false negative rate of 0.46%, and a false positive rate of 7.41% (Table 1).

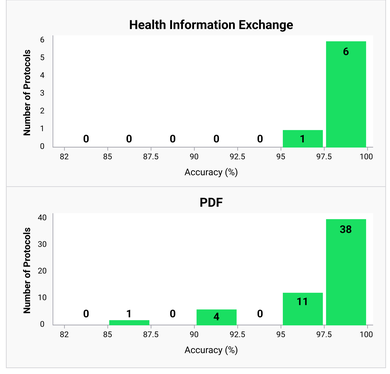

Figure 2: Histogram of accuracy by record type

We analyzed accuracy across protocols (sets of I/E criteria consisting of multiple yes/no questions). **For HIE records, 100% of protocols were answered with greater than 95% accuracy. For PDF records, 89% (45/54) were answered with greater than 95% accuracy (Figure 2).**

	PDF	Health Information Exchange
50th	2.6	4.31
95th	10.87	19.92
99th	17.6	25.24

Table 2: Distribution of processing time (in minutes) by record type

We also measured the distribution of processing times for different document types. **For PDF records, we saw a median processing time of 2.60 minutes, with a 99th percentile of 17.6 minutes**. **For HIE records, we saw a median processing time of 4.31 minutes with a 99th percentile of 25.24 minutes (Table 2)**.

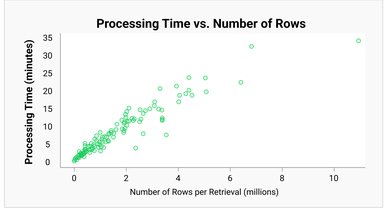

Figure 3: Scatterplot of processing time as a function of number of rows

By analyzing HIE records, we found that the main contributor to processing time was the size of the records (Figure 3).

Discussion

Our platform achieved 98.1% accuracy on unstructured PDF records and 99.7% on semi-structured HIE records, with median processing times of 2.6 and 4.3 minutes respectively—vastly outpacing the multi-hour[2] manual benchmarks that currently strain sites and coordinators, while preserving and likely exceeding the clinical accuracy of human clinicians.

The exceptionally low false-negative rates (0.24–0.46%) observed here are reassuring in that they minimize enrollment oversights.

We did observe an unusually high false positive rate (7.41%) across PDF records. These are instances where the machine learning tool answered “yes” to a clinical trial eligibility screening question, but that a trained human clinical later determined, upon manual review of the medical record, should have been answered as a “no”.

This high false positive rate is likely a result of the unstructured and varied nature of PDF based medical records. Given that this tool provides citations to page sources within the PDF medical record, and is used by coordinators as a tool to amplify and accelerate, but not replace the final clinical judgement of the coordinator or investigator, we do not expect this false positive rate on PDF records to result in incorrect randomization into a clinical trial.

We observed > 95% accuracy of the tool across 89–100% of questions and protocols, which demonstrates dependable performance across varied clinical protocols.

Limitations include medical records and data which were limited to that which was obtainable from previous providers and the HIEs, a limited and varied number of therapeutic areas and sites, lack of visibility into the final determination on enrollment for each patient, and a limited sample size and time window. Future efforts will require evaluation across an increased number of records, sites, clinical trials, and disease states.

These results demonstrate the potential of artificial intelligence to efficiently perform clinical trial screening in minutes instead of hours, and to greatly reduce the cost and increase the velocity of clinical trial enrollment.

References:

1. Unlu O, et al. Manual vs AI-Assisted Prescreening for Trial Eligibility Using Large Language Models—A Randomized Clinical Trial. JAMA 2025;333:1084-1087.
2. Penberthy LT, et al. Effort Required in Eligibility Screening for Clinical Trials. J Onc Pract 2012;8:365–370.

Disclosures:
The authors are compensated employees of Predoc.