Original Paper

TrialTriage, a Semiautonomous Prescreening Workflow for Resolving Ambiguity in Phase I Oncology Trial Eligibility: Development and Proof-of-Concept Study Using Synthetic Cases

Kenneth A Kern[1,2], MS, MPH, MD

[1]Affiliate Faculty, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, United States
[2]Oncology Breakthrough Advisory Group, LLC, San Diego, CA, United States

Corresponding Author:

Kenneth A Kern, MS, MPH, MD
Oncology Breakthrough Advisory Group, LLC
1804 Garnet Avenue, Suite 701
San Diego, CA 92109
United States
Phone: 1 (619) 354-1026
Email: info@oncologybreakthrough.com

Abstract

Background: Enrollment in phase I oncology trials remains low largely because potentially eligible patients are not identified and evaluated quickly enough. Current clinical trial matching systems can identify candidate patients from the electronic health record, but cases with missing or uncertain eligibility data are often routed for offline manual review. This delay impedes clarification and prolongs the final eligibility determination.

Objective: This study evaluated TrialTriage, a semiautonomous system built on the n8n platform and designed to resolve eligibility ambiguity during prescreening for phase I oncology trials. When eligibility information is missing or uncertain, TrialTriage emails the investigator, captures the reply, and reruns classification within the same workflow.

Methods: TrialTriage combined large language model–based variable extraction from free-text clinical narratives and investigator email replies with a deterministic rule engine applying a prespecified 7-criterion protocol. Each case was classified as eligible, not eligible, or ambiguous. Ambiguous cases triggered a structured email query to the investigator, followed by reclassification after a reply. Two requests were sent at 24-hour intervals; after 48 hours without a reply, the case was referred for manual review. The system was tested on 90 synthetic patient cases generated independently by Claude Sonnet 4.6, Gemini 3.1, and Grok 4, with 30 cases per model and balanced distributions of eligible, not eligible, and ambiguous cases. Answer keys were reviewed for accuracy before system execution. Five independent reviewers classified the Claude dataset using a uniform survey form.

Results: TrialTriage's classifications were 100% concordant with the author-confirmed ground truth in all 90 synthetic cases (95% CI 96.0%-100.0%). All ambiguous cases were correctly escalated to investigator query. The mean processing time was 2.3 (SD 0.5) minutes per 30-case dataset (range 1.8-2.8 min, approximately 3.5-5.5 s per case). The 5 reviewers achieved a mean accuracy of 96.7% (SD 3.3%), with a Fleiss κ of 0.910, and required a mean of 9.8 (SD 4.8) minutes to review 30 cases. In a subset test of 6 first-pass ambiguous cases, 4 of 6 were reclassified definitively after investigator response, while 2 remained ambiguous because the replies lacked actionable information.

Conclusions: TrialTriage demonstrates the feasibility of a semiautonomous prescreening workflow in which ambiguous cases trigger an immediate investigator email query and are reclassified after reply capture with new information within the same system. The main contribution is the integration of email ambiguity resolution into the workflow rather than immediate deferral to offline manual review. Because the evaluation used synthetic cases and label definitions aligned with the same protocol rules used to design the rule engine, these findings should be interpreted as proof of concept and implementation fidelity rather than evidence of real-world clinical performance. Prospective validation using data from real-world electronic health records would be a plausible next step.

JMIR Form Res 2026;10:e100779; doi: 10.2196/100779

Keywords: clinical trials; prescreening; phase I; oncology; patient eligibility; clinical trial matching; workflow automation; large language model; LLM; natural language processing; semiautonomous AI; feasibility study; patient accrual

Introduction

Phase I oncology trials are the initial clinical testing platform for new anticancer drugs, either alone or in combination [1-3]. These trials establish safety, tolerability, and preliminary dosing. Their successful completion is a prerequisite for progression to later-phase trials that support regulatory approval [4]. Robust interpretation of phase I data requires the enrollment of a predefined sample size [4,5]. Reaching that sample size depends on identifying and enrolling patients who meet trial entry criteria through a 2-step process known as eligibility screening.

Eligibility screening for phase I oncology trials is conducted in 2 sequential steps: prescreening followed by formal screening [5,6]. Prescreening is an initial review of the electronic health record (EHR) or oncologist's patient listing against a limited set of key trial entry criteria, conducted without patient contact. Patients identified as potential candidates are then contacted, the trial is explained, and those who express interest are asked to sign an informed consent document to allow formal screening. Formal screening establishes final eligibility through detailed disease assessment, including physical examination, laboratory testing of major organ function, and radiographic evaluation of disease extent. Patients who satisfy all criteria after formal screening are enrolled [7,8]. Both steps are challenging in phase I oncology because of strict entry criteria, the unpredictable safety profile of novel agents, and the narrow patient populations these trials require.

Fewer than 8% of eligible patients with cancer enroll in clinical trials [9-11]. This low accrual rate is not primarily driven by patient refusal: between 55% and 80% of patients identified as eligible and invited agreed to enroll [12-14]. The dominant barrier is the inefficient identification of the smaller subset of patients who meet entry criteria from among many patients who do not [6,15]. Among patients who are prescreened and reach formal screening, approximately 25% fail to meet eligibility criteria and are termed "screen failures" [16,17]. Because formal screening is expensive and resource-intensive, this rate raises the question of whether better prescreening could improve the screening success rate and increase trial enrollment.

Improving low enrollment rates would have significant consequences for early-phase trial completion. Many phase I studies terminate early due to low enrollment rates and failure to meet statistically defined sample sizes. In one analysis of terminated trials, 57% closed specifically because of insufficient accrual of patients rather than drug toxicity or lack of antitumor activity [10].

Phase I oncology trials typically last 2 to 3 years, with low enrollment rates as a primary contributor to delays [5,18]. For patients with malignant disease, these delays extend the time required for potentially effective therapies to reach them [19].

Prescreening is typically performed manually by clinical research staff, who review records, query databases for missing information, and escalate uncertain cases to clinicians. These steps are labor-intensive, and a substantial fraction of patients cannot be definitively classified at this stage. Prior studies of clinical trial matching (CTM) systems report that 20% to 25% of prescreened patients cannot be definitively classified because of incomplete or ambiguous data and require manual adjudication [20,21]. High model performance does not resolve this problem on its own. A prospective randomized trial of 20,707 patients showed no enrollment impact despite high model performance (area under the receiver operating characteristic curve=0.91), largely because 23.2% of notifications were rejected due to a lack of clinical factors key to protocol entry [22]. Delay in resolving ambiguity has direct clinical consequences: processing delays of 48 hours after imaging detects the progression of disease have resulted in 20% of patients starting alternative treatments before trial notification occurs [23]. Although multiple factors contribute to accrual delays, unresolved eligibility ambiguity at the prescreening stage is a key driver of enrollment inefficiency and a documented contributor to the screen-failure rate described above [16,17,24].

AI approaches have been developed to support patient identification through the analysis of structured and unstructured clinical data [25,26]. These systems, often described as CTM systems, analyze eligibility criteria and classify patients as eligible, not eligible, or ambiguous. In current CTM systems, cases classified as ambiguous are typically removed from the automated workflow and sent for manual review; information obtained during that review is generally not returned to the original classification process [27]. As a result, ambiguity is not resolved within the workflow itself but is shifted into a slower manual process, increasing delays and creating additional opportunities for inconsistency or error.

To address this limitation, we designed TrialTriage, a semiautonomous workflow for prescreening in phase I oncology trials. TrialTriage classifies patients as eligible, not eligible, or ambiguous and automatically sends an email to the investigator when required information is missing or unclear. The system's large language model (LLM) evaluates the investigator's reply, extracts any substantive clarifying information, and passes that information to the rule engine for reclassification within the same workflow. This design builds on the broader principle that corrective human input can improve classification accuracy, as reported in a study in which manual feedback increased performance from 88.0% to 92.7% [20]. TrialTriage extends that principle by automatically requesting feedback and clarification of ambiguous

cases by email within the workflow, rather than leaving those cases for manual review outside of it.

The aim of this study was to develop and evaluate TrialTriage, a semiautonomous prescreening workflow for phase I oncology trials, whose novel component is an immediate, iterative email exchange with the principal investigator (PI). There were 2 key objectives for this study. The first was to determine whether TrialTriage could perform the following functions: correctly classify synthetic phase I oncology cases as eligible, not eligible, or ambiguous; route a listing of ambiguous cases and the data needed to classify them to an investigator through immediate email query; apply an iterative process that uses PI responses to reclassify ambiguous cases to the correct eligibility category; and retain unresolved cases as ambiguous for manual classification. The second was to compare TrialTriage's classification accuracy and processing time against those of human reviewers evaluating the same cases.

We hypothesized that TrialTriage would classify cases in complete agreement with the predefined ground truth and do so in less time than human reviewers, even when those reviewers were experienced in phase I trials, and that TrialTriage would request additional data from the PI and reclassify ambiguous cases accurately. Because the evaluation used synthetic cases built from the same criteria used to design the rule engine, we framed these as tests of internal validity and implementation fidelity, not of real-world clinical performance. This approach aligns with calls to reduce screen failure through automation [28] and with current regulatory guidance emphasizing transparent, auditable outputs and structured human oversight for AI used in clinical trials [29, 30].

Methods

Study Design

This proof-of-concept study evaluated a semiautonomous workflow for prescreening synthetic patient cases against the eligibility criteria of a hypothetical phase I oncology trial. Synthetic cases were generated by LLMs to simulate patients with advanced pancreatic cancer, as described below. Pancreatic cancer was used as a representative disease to test eligibility in a hypothetical phase I oncology trial. The architecture of the workflow is not specific to any one tumor type and could be generalized to other phase I oncology trials.

Ethical Considerations

This study did not involve human subjects and did not use actual patient data. All clinical cases were synthetic narratives generated by LLMs. Because this study involved only synthetic patient data and no living individuals underwent data collection, it did not constitute human subjects research and did not require institutional review board review or approval under the US Common Rule (45 CFR 46.102). The 5 clinicians who classified cases for the manual comparison did so as expert colleagues evaluating synthetic cases as part of a methods comparison. Since they were not research subjects, no personal data about these reviewers were recorded beyond their professional role—physician or nonphysician clinician—and the time taken to complete the classification task.

System Architecture

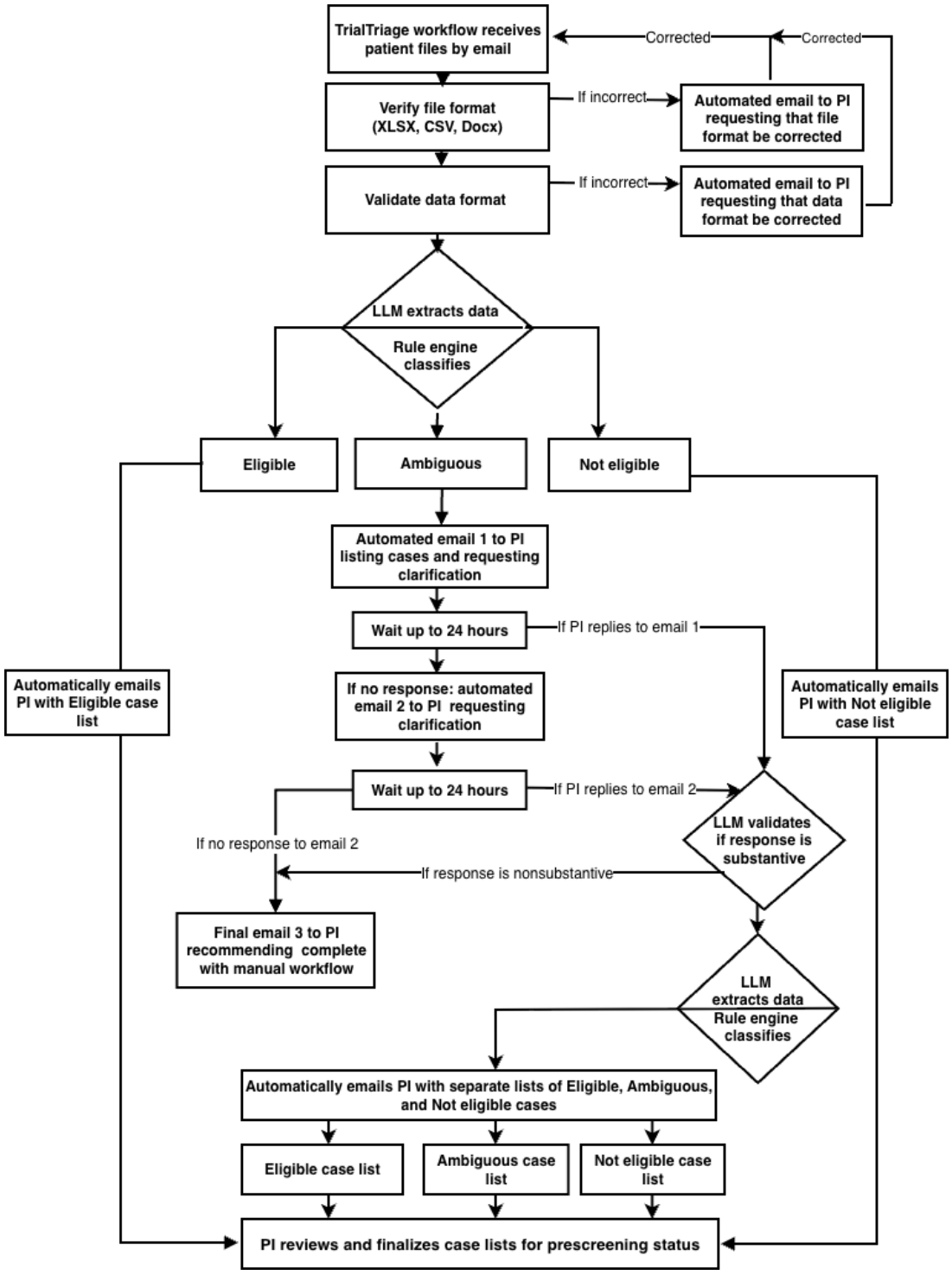
TrialTriage is a hybrid architecture orchestrated on the n8n workflow automation platform (n8n GmbH). In n8n, individual workflow steps, called "nodes," are connected through a visual interface to define a sequence of operations. The TrialTriage workflow comprises 3 functional components: intake and preprocessing, classification and iterative reclassification, and communication and output. These components are organized into 5 sequential stages: input processing, initial classification, first investigator email, ingestion of additional data for ambiguous cases, and final case reclassification and output. A simplified architecture for the TrialTriage workflow is shown in Figure 1.

The 5 stages are implemented using 45 deployed node instances drawn from 20 unique node types. Three of these node types incorporate LLMs: an Extract from File node, which extracts the 7 eligibility variables from free-text input; a Text Classifier node, which categorizes investigator replies as substantive or nonsubstantive; and an AI Agent node, which converts substantive reply text into structured data for reingestion by the rule engine. All 3 node types use Anthropic's Claude Sonnet 4.6 as the underlying language model. References to "Agent" elsewhere in this *Methods* section refer specifically to the n8n AI Agent node type.

Claude Sonnet 4.6 was selected as the model embedded within the 3 LLM-based nodes for 2 reasons. First, Claude was already integrated as a selectable model within the n8n platform, which simplified deployment and avoided the need for custom API handling. Second, the rule engine had been developed using GPT-5 (OpenAI) during Version 1 of TrialTriage, and choosing a different LLM for Version 2 provided some independence from that earlier development.

Screenshots of the workflow architecture are provided in Multimedia Appendix 1. A complete listing of the n8n node types used in the workflow, along with their deployed instance counts, is provided in Multimedia Appendix 2. The version of n8n used in this study retained timestamped records of each classification, email query, response ingestion, and reclassification event for up to 7 days, creating an audit trail of workflow activity. An enterprise version of n8n, not used in this report, can maintain longer-term storage.

Figure 1. Simplified TrialTriage architecture. Patient files are received by email and validated. The large language model (LLM) extracts the 7 eligibility variables; a deterministic rule engine classifies each case as eligible, ambiguous, or not eligible. Ambiguous cases trigger an automated email to the principal investigator (PI), with one escalation if no reply is received within 24 hours. Substantive replies are reextracted by the LLM and reclassified by the rule engine. Unresolved cases exit to a recommended manual workflow. Diamond shapes indicate decision steps.



Data Ingestion, Validation, and Initial Classification

When a patient file is received by email, an input validation node checks both the file format and data format. Accepted file formats are XLSX, CSV, and DOCX. If either check fails, the system sends an automated email to the PI identifying the format problem and requesting resubmission. Only files and data that pass validation proceed to downstream processing. After format validation, the system's LLM, Claude Sonnet 4.6 (Anthropic), extracts the patient ID and eligibility variables from the free-text clinical narrative for each patient. A deterministic rule engine then assigns each case to one of 3 categories: eligible, not eligible, or ambiguous.

Eligibility Criteria

Seven eligibility criteria, representative of those commonly used in prescreening patients with pancreatic cancer for a

phase I oncology trial, were defined for this study (Table 1). Thresholds were approximate and were not intended to reproduce any specific existing trial. The system was designed to accept "pancreatic cancer" as synonymous with "pancreatic adenocarcinoma" to avoid misclassification based on terminology alone. Cases were classified as eligible only if all 7 criteria were documented in the case narrative as meeting entry requirements.

Table 1. The 7 eligibility criteria used by the TrialTriage rule engine to classify synthetic phase I oncology trial cases.

Criterion	Threshold
Histology	Pancreatic adenocarcinoma (pancreatic cancer) confirmed by pathology within the previous 6 months
Age	18-80 years, inclusive
ECOG[a] performance status	0 or 1: ECOG functional scale, 0-4, where 0 indicates fully active and 4 indicates completely disabled and confined to bed or chair
AST[b] (SGOT[c])	≤40 U/L
ALT[d] (SGPT[e])	≤40 U/L
Total bilirubin	≤1.2 mg/dL
Creatinine	≤1.5 mg/dL

[a]ECOG: Eastern Cooperative Oncology Group.
[b]AST: aspartate aminotransferase.
[c]SGOT: serum glutamic-oxaloacetic transaminase.
[d]ALT: alanine aminotransferase.
[e]SGPT: serum glutamic-pyruvic transaminase.

Classification Categories

After the system extracts the eligibility variables and the rule engine classifies each case, the system groups the results by category and sends 3 separate emails to the PI: one listing all eligible cases, one listing all not eligible cases, and one listing all ambiguous cases. When a case is classified as ambiguous, the system sends the investigator a structured email requesting the specific missing or unclear information needed for reclassification. When a substantive response is received, the workflow ingests the investigator's reply and reruns extraction and classification to generate an updated classification. After reclassification, the system again sends 3 category emails to the PI containing the updated eligible, not eligible, and remaining ambiguous cases. A final email asks the PI to review and approve all classifications.

Closed-Loop Workflow Execution and Final Eligibility Decisions

The workflow executes a closed-loop sequence of data ingestion, classification, action, and reingestion without human intervention between steps. Specifically, it extracts narrative information, classifies cases through the rule engine, sends emails and schedules reminders, receives investigator responses, and reclassifies cases when new information is provided. The architecture is read-only with respect to any upstream clinical data source: the workflow does not write back to the EHR or modify original input data [30]. As shown in Figure 1, the workflow outputs case lists for eligible, not eligible, and ambiguous categories after both the initial classification and any iterative reclassification. Final eligibility decisions for all cases remain the responsibility of the PI.

Synthetic Dataset Generation

TrialTriage was built in 2 versions: Version 1, a limited prototype; and Version 2, the system presented in this report. Both versions used the same 7 eligibility criteria. In both versions, synthetic cases were generated using 4 LLMs, each prompted independently: ChatGPT-5.0 (OpenAI), Claude Sonnet 4.6 (Anthropic), Gemini-3.1 (Google), and Grok 4 (xAI). The same prompt structure was used across models (Multimedia Appendix 3). Each model was instructed to generate brief but realistic synthetic case narratives of patients with pancreatic cancer, 2 to 3 sentences in length, using the 7 protocol-specific eligibility criteria described above.

Version 1, built and tested using 50 ChatGPT-generated cases, performed initial classification only. It produced eligible, not eligible, or ambiguous category emails and did not include an iterative investigator email loop or reclassification. Version 1 is not reported here, and the Version 1 synthetic cases are not included in the multimedia appendix. In Version 2 of TrialTriage, evaluated in this report, additional functions included an iterative investigator email loop, reclassification of ambiguous cases, and a substantiveness check on investigator replies.

Version 2 of TrialTriage was evaluated on cases generated independently by Claude Sonnet 4.6, Gemini-3.1, and Grok 4. Each of the 3 models generated 30 clinical scenarios in randomized order, comprising 10 eligible cases, 10 not eligible cases, and 10 ambiguous cases. Each model also produced an answer key specifying the correct classification for each case. The 90 cases were all different clinical scenarios, produced using different permutations of the 7 eligibility criteria, with no case appearing in more than 1 model's dataset. The synthetic case datasets were generated in February 2026. All answer keys were checked by the author to confirm the ground truth. The full case listings, with ground-truth labels and classification outputs, are provided in Multimedia Appendix 4.

Evaluation Procedure

The workflow included a built-in evaluation branch that permitted either a regular workflow mode, triggered by a Gmail message with a specified subject line, or a separate

evaluation mode run from spreadsheet tabs stored in Google Sheets. Initial testing used the 30 Claude-generated cases. The Gemini and Grok datasets were then submitted to the system to assess concordance with independently generated inputs. TrialTriage classification of all 90 cases was performed in March 2026. TrialTriage classifications for each case are included alongside the case listings in Multimedia Appendix 4.

Ground Truth

Each model generated an answer key specifying the correct classification for each case. During the Grok evaluation, the author identified disagreements with the model-generated answer keys in more than 50% of cases. Investigation determined that the rule engine recognized only "pancreatic adenocarcinoma," whereas some Grok-generated cases used the medically synonymous term "pancreatic cancer." Both terms refer identically to cancer of the pancreas. The rule engine, which is deterministic and not a trained model, was modified to accept "pancreatic cancer" as equivalent. The final evaluation of all 90 cases was conducted on the modified rule engine. Classifications agreed with the author-confirmed ground truth in all 90 cases.

Because Claude served both as one of the 3 case generators and as the embedded LLM in the current TrialTriage, we checked whether TrialTriage favored cases produced by Claude by comparing its performance on cases it generated versus those generated by Gemini and Grok. Following the correction of the pancreatic cancer definition described above, classification was 100% concordant across the Claude, Gemini, and Grok cases. This single-pass evaluation was not designed to measure run-to-run variability, which remains untested and is noted as a limitation.

Investigator Clarification for Ambiguous Cases

Cases classified as ambiguous triggered immediate email communication, with the PI requesting clarification of the specific missing or uncertain data elements. The first email listed each ambiguous case by ID along with the relevant clinical narrative and contained a "Respond" button. When clicked, the button opened a reply email prompting the PI to provide the specific missing or unclear data for each identified case. If no reply was received within 24 hours, a second email was sent requesting a response. If no reply was received within an additional 24 hours, a third and final email informed the PI (or, in future designs, other designated research staff) that the case was being removed from the automated classification system and needed to be reviewed through a manual process. The workflow then took no further action on classifying the case and relied on the PI to have the case evaluated by the site's standard manual prescreening process.

Routing the case directly from the workflow to a named secondary reviewer or coordinator was not implemented in this study and is identified in the *Future Directions* section as a planned enhancement. Investigator replies were evaluated by the LLM and classified as substantive or nonsubstantive. Substantive replies contained actionable clinical information relevant to one or more ambiguous eligibility criteria. Nonsubstantive replies lacked actionable clinical information, such as "I do not have that information," "I will investigate it," or out-of-office replies.

Comparison of Manual Prescreening Performance to TrialTriage

Five reviewers manually classified the randomized 30-case Claude dataset using the same 7 eligibility criteria applied by TrialTriage. Reviewers were blinded to the answer key. Three reviewers were physicians with extensive drug development experience, while the remaining 2 were nonphysician clinicians with similarly extensive drug development experience. None had seen the cases prior to evaluation. The 5 reviewers completed the manual classification of the Claude dataset between March 4 and March 30, 2026. Each reviewer independently recorded classifications for all 30 cases on a separate survey sheet (Multimedia Appendix 5), assigning each case as eligible, not eligible, or ambiguous. At the top of each survey sheet, before the case listings, instructions defined the 3 classifications and stated that any case with missing, pending, vague, or nonnumeric data should be classified as ambiguous. Average classification time per case was calculated by dividing each reviewer's total classification time by 30. Manual review was limited to the Claude dataset; performance against the Gemini and Grok datasets was not assessed.

Statistical Analysis

Classification accuracy was reported with exact binomial 95% CIs. Agreement between each reviewer and the ground truth was quantified using Cohen κ with 95% CI. Interrater agreement across all 5 reviewers was quantified using Fleiss κ. The 95% CI was estimated by nonparametric bootstrap resampling over 5000 iterations, drawing from the table of how each reviewer classified each case. This was a post hoc analysis performed in Python 3.12.3 (Python Software Foundation) with NumPy 2.4.4 (NumFOCUS, Inc), outside the n8n workflow.

Total time required to classify the 30-case dataset was summarized descriptively. Reviewer times were reported as the mean (SD) across the 5 reviewers, while TrialTriage times were reported as the observed range across the 30 cases. TrialTriage and reviewer classification times were compared descriptively rather than inferentially because both sample sizes were small. This paper follows the TRIPOD-LLM reporting guideline, which specifies what studies that develop or evaluate LLMs in health care should report [31]. A completed checklist is provided in Checklist 1.

Results

Datasets and Evaluation

TrialTriage was evaluated on the 90 synthetic cases described in *Methods*, comprising three 30-case datasets generated by

Claude Sonnet 4.6, Gemini-3.1, and Grok 4. The Claude dataset served as the primary evaluation set. TrialTriage was applied to all 3 datasets, but the iterative investigator email clarification loop was applied only to the Claude dataset. The 5 reviewers also classified only the Claude dataset. The Gemini and Grok datasets were used to test the consistency of TrialTriage's performance across cases generated by different LLMs. Complete case narratives, ground-truth labels, TrialTriage classifications, and reviewer classifications are provided in Multimedia Appendix 4.

TrialTriage Classification Performance

TrialTriage's classifications were concordant with the author-confirmed ground truth in all 90 cases across the 3 evaluation datasets (90/90; exact binomial 95% CI 96.0%-100.0%), including all eligible, not eligible, and ambiguous cases. Concordance was 30/30 on the Claude dataset, 30/30 on the Gemini dataset, and 30/30 on the Grok dataset. Classification remained stable across datasets generated by different LLMs, despite differences in wording and case presentation. Because the synthetic cases were generated by LLMs prompted with the same 7 eligibility criteria from which the rule engine was constructed, this finding represents internal consistency of the workflow on prompt-faithful narratives rather than performance on the real clinical text.

System Processing Time

TrialTriage processed each 30-case dataset in 1.8 to 2.8 minutes, corresponding to 3.5 to 5.5 seconds per case. Total processing time across all 90 cases was 5.4 to 8.4 minutes. This timing included data ingestion, LLM extraction of structured variables, and rule-engine classification. Per-dataset processing times for the Claude, Gemini, and Grok datasets all fell within this range. All 90 cases were processed without the interruption of the n8n workflow. These figures reflect automated processing time only and do not include the 24-hour escalation intervals or investigator response time required for ambiguous cases entering the email clarification loop.

Iterative Reclassification of Ambiguous Cases

The iterative investigator feedback loop was evaluated on 6 of the 10 ambiguous cases in the Claude primary dataset, selected at random. The author served as a simulated PI and replied to the workflow's automated email queries; this single-responder design is acknowledged in the *Limitations* section. For 4 cases, the author provided additional clarifying data involving ECOG (Eastern Cooperative Oncology Group) status, confirmation of diagnosis, or laboratory values within the range. The rule engine reclassified all 4 as eligible after the LLM re-extracted variables from the reply, consistent with the data the author provided. For the remaining 2 cases, the author responded with nonsubstantive statements such as "data not available" or "I don't have that information." Both cases were retained as ambiguous cases and triggered the workflow's recommendation for manual classification, as specified in the *Methods* section. All replies in this test were submitted within 24 hours, so the second-email escalation path was not triggered or evaluated.

Manual Reviewer Performance

Reviewer accuracy on the 30-case Claude dataset ranged from 93.3% to 100.0%, with a mean of 96.7% (SD 3.3%); per-reviewer accuracy, classification time, and Cohen κ are shown in Table 2. Interrater agreement across all 5 reviewers was almost perfect according to the Landis and Koch benchmarks [32] (Fleiss κ=0.910; bootstrap 95% CI 0.813-0.980). The mean reviewer classification time was 9.8 (SD 4.8) minutes per 30 cases, about 20 seconds per case. TrialTriage classified the same 30 cases in 1.8 to 2.8 minutes, about 4 to 6 seconds per case, with no errors. TrialTriage was therefore about 4 times faster.

Table 2. Reviewer and TrialTriage performance on the 30-case Claude dataset.

Reviewer	Role	Time (min)	Time (s)	Correct	Accuracy (%)	Cohen κ[a] (95% CI)
Reviewer A	MD[b]	6.0	360	30/30	100.0	1.000[c]
Reviewer B	MD	18.0	1080	30/30	100.0	1.000[c]
Reviewer C	MD	6.9	415	29/30	96.7	0.950 (0.854-1.000)
Reviewer D	Non-MD	10.0	600	28/30	93.3	0.900 (0.766-1.000)
Reviewer E	Non-MD	8.0	480	28/30	93.3	0.900 (0.766-1.000)
Mean (SD)	—[d]	9.8 (4.8)	587 (290)	—	96.7 (3.3)	—
TrialTriage	—	1.8-2.8	105-165	30/30	100.0	1.000[c]

[a]Cohen κ is reported for each reviewer relative to the ground-truth answer key, with asymptotic 95% CI. Interrater agreement across all 5 reviewers: Fleiss κ=0.910 (bootstrap 95% CI 0.813-0.980, 5000 resamples).
[b]MD: physician.
[c]The CI is not calculable when κ=1.000, with no observed disagreement.
[d]Not applicable.

Manual Reviewer Errors

Five misclassifications occurred across the 150 reviewer classifications; each is listed by case, reviewer, and failure mode in Table 3. Four of the 5 errors reflected overinclusion of patients who did not meet the criteria or the premature resolution of ambiguity, and no reviewer classified an eligible case as not eligible or ambiguous. TrialTriage made no classification errors on the same dataset.

Table 3. Manual reviewer classification errors in the 30-case Claude dataset[a].

Case ID	Reviewer	Correct class	Reviewer response	Error	Failure mode
5	C	Not eligible	Eligible	Not eligible→Eligible	Age of 83 years exceeds the upper limit of 80 years; should be classified as not eligible
5	D	Not eligible	Eligible	Not eligible→Eligible	Age of 83 years exceeds the upper limit of 80 years; should be classified as not eligible
6	E	Ambiguous	Not eligible	Ambiguous→Not eligible	Bilirubin absent from the lab panel; missing data should be classified as ambiguous
16	D	Ambiguous	Eligible	Ambiguous→Eligible	No ECOG[b] numeric status, narrative only; should be classified as ambiguous
18	E	Not eligible	Eligible	Not eligible→Eligible	SGPT[c] 57 U/L exceeds ULN[d] of 40 U/L; should be classified as not eligible

[a]Error column shows the misclassification as correct class→reviewer response.
[b]ECOG: Eastern Cooperative Oncology Group.
[c]SGPT: serum glutamic-pyruvic transaminase.
[d]ULN: upper limit of normal.

Discussion

Principal Findings

TrialTriage met both objectives of this proof-of-concept evaluation. The first objective was to determine whether TrialTriage could classify synthetic phase I oncology cases, route ambiguous cases to the investigator, reclassify them based on the PI's response, and retain unresolved cases for manual classification. TrialTriage successfully classified every synthetic case in agreement with the predefined ground truth across datasets generated by 3 different LLMs, categorizing each as eligible, not eligible, or ambiguous. It routed every ambiguous case through an immediate, structured email query, reclassified the cases that received substantive replies, and retained the remaining cases as Ambiguous for manual classification. The second objective was to compare TrialTriage's classification accuracy and processing time against those of human reviewers evaluating the same cases. TrialTriage classified cases faster than the reviewers and without classification errors, whereas the reviewers' errors involved either overinclusion or premature resolution of ambiguous cases.

Comparison With Prior Work

Existing CTM systems and TrialTriage can be compared based on 2 points: what they do with the cases they cannot classify and how accurately they classify. Current systems focus their automated workflow primarily on identifying eligible patients, and they divert ambiguous cases for offline, manual resolution, which is often delayed [33-35]. When clinical observations are the only source for an eligibility variable, or when variables related to the past medical history are absent or unclear, automated CTMs cannot resolve the case definitively, and the case is routed for manual review. Manual review of such cases is time-consuming and resource-intensive [6]. TrialTriage operates downstream of these systems, prescreening identified candidates against a specific trial's criteria and resolving ambiguity within the same workflow.

We surveyed published descriptions of CTM and prescreening systems, including Tempus Tapp [36], TrialMatchAI [37], OncoLLM [38], Watson CTM [21], TrialGPT [39], and a natural language processing–based screen failure prediction model [19]. None of these systems describe an automated mechanism for resolving ambiguous cases within the workflow. Several do not address ambiguous cases at all, leaving it unclear whether such cases are excluded from analysis, defaulted to one classification, or routed for offline manual review.

In a cohort of 102 patients with non–small cell lung cancer, IBM Watson CTM had a median processing time of 15.5 seconds per case (range 7.2-37.8 s) and achieved 97% agreement with human review [40]. Nevertheless, the workflow required 8088 manual actions to enter data or obtain clinician interpretation for eligibility, and Watson CTM did not query for missing information during the automated workflow. Reported accuracies for CTMs reflect performance largely on cases with sufficient information already available because the data-incomplete cases are excluded from the accuracy denominator [38].

Implications

This proof-of-concept study holds several implications for the use of semiautonomous systems in real-world eligibility determinations. This type of workflow may address enrollment inefficiencies because prescreening and formal enrollment screening are interdependent. Inaccurate prescreening advances potential study participants who subsequently fail to enroll due to the more detailed formal screening process [27]. McKane et al [16] reported a 24.6% screen-failure rate in phase I trials, which rose 3-fold to 78.2% at a center that included inaccurate prescreening classification as part of the overall screen-failure rate. These failed screening efforts are also costly ($25,000 per patient) and resource-intensive (up to 8.5 h of staff time per patient) [41,42]. Overall, phase I trials commonly require accrual times 5-fold longer than originally planned because of the logistical obstacles created by complex eligibility criteria, such as biomarker and biopsy data [5,43,44]. Automated

workflows, such as TrialTriage, may help to reduce costs and resource time by classifying cases accurately and quickly and by resolving ambiguous cases through an immediate email query to the investigator, although these benefits remain to be confirmed with real clinical data.

Bias can enter TrialTriage at the points where LLMs are used: extracting variables from the case, judging whether an investigator's reply is substantive, and parsing substantive reply text into the data fields read by the rule engine. Several features of the architecture guard against errors at these points. Eligibility decisions are made by a deterministic rule engine applying threshold values, so the model does not rely on its own judgment. Investigator replies pass through a substantiveness check, and noninformative replies trigger an email recommending manual prescreening. The eligibility output is sent to the PI or research staff for final sign-off, and a timestamped audit trail records all classifications, queries, replies, and reclassifications. To prevent contamination of the EHR, TrialTriage reads only the source record and never writes back to it.

Some subjectivity and bias are likely to remain because some eligibility parameters, such as ECOG performance status (a graded measure of a patient's ability to carry out daily activities), rely strongly on clinical judgment and cannot be classified by objective measures alone.

Limitations

This study has several limitations, including the use of synthetic rather than real clinical cases, the limited independence of the test cases from the rules used to build the rule engine, small sample sizes, the limited set of eligibility criteria tested, and the absence of real-world data governance.

Synthetic Cases

The evaluation used synthetic case narratives rather than EHR-derived records. Synthetic cases are useful for testing workflow logic and rule implementation; however, because of their inherent simplicity, they cannot exactly reproduce the semantic and syntactic complexity of unstructured, real clinical documentation [38,45]. Actual medical records often contain inconsistencies, redundancies, conflicting documentation, and highly variable narrative quality, all of which complicate the extraction of clinically meaningful information [46-49]. Whether TrialTriage's performance on synthetic cases would be maintained on real-world EHR narratives remains untested. Because the synthetic cases were not designed to reflect any patient demographic mix, the evaluation also cannot demonstrate how TrialTriage performs across different patient groups.

Test-Case Independence

The Version 2 test cases generated by Claude, Gemini, and Grok were built from variations on the same 7 eligibility criteria used to construct the rule engine during Version 1 development. The use of 3 different LLMs broadened the diversity of how cases were described but did not create a fully independent evaluation set. Performance against an external standard independent of the rule-based development process remains to be established.

Sample Size

The ambiguity-resolution analysis was limited to 6 randomly selected ambiguous cases from the Claude dataset, with the author responding to the system's email queries as a simulated PI. The workflow handled those cases correctly; however, the sample is too small to support strong conclusions about how the system would perform with real investigator replies that vary in clarity, completeness, and timing. Whether real PIs or qualified study personnel would respond quickly enough to preserve the workflow's time advantage remains untested. The reviewer comparison involved 5 clinical reviewers evaluating one 30-case dataset. That sample was sufficient to reveal general error patterns, particularly overinclusion and premature classification of ambiguous cases as eligible or not eligible, but it was not large enough to establish stable benchmarks for manual prescreening performance or to identify which case features most reliably produce reviewer error. The reviewer comparison should therefore be understood as hypothesis-generating rather than definitive.

Data Governance

Because the study used only synthetic cases and deliberately did not include protected health information, it cannot show how actual patient data would be governed within the workflow. Before using a semiautonomous system like TrialTriage in a phase I clinical trial, procedures for handling uploaded patient data must be established to ensure confidentiality. These procedures would specify where identifiable patient data enter the workflow, which steps process or store data, how long data are retained, how data are encrypted in storage and in transit, and how the investigator email channel is protected.

Criteria Scope

The workflow was tested on 7 representative criteria rather than the full complexity of a phase I protocol. Formal trial screening commonly involves more criteria, including biomarker requirements, disease extent, prior treatment history, organ function, and many other features derived from physical examination, laboratory data, and imaging. Typical phase I oncology protocols have upward of 60 eligibility criteria that must be met during formal screening [40]. Whether the same approach remains effective as criterion count and complexity expand will require further study.

Future Directions

This proof-of-concept evaluation leaves 3 questions about the TrialTriage architecture unresolved, which could form the basis for future development.

Extraction Layer

The next stage of development would test TrialTriage on EHR-derived cases, with the goals of characterizing extraction performance on clinician-generated text and

determining whether immediate, iterative email communication can shorten the time to final prescreening classification under real-world conditions. Operational end points for such testing include time to final prescreening decision, frequency of ambiguity resolution within a defined time window after the first email request, percentage of workload shifted away from manual review, prescreening failure rate (the number of patients referred for formal screening who were subsequently rejected), and screen-failure rate (the percentage of consented patients who were never dosed).

Human Factors

The performance of the human in the loop should be measured by the time taken to respond to the iterative emails and by the quality of the data supplied in reply. Implementations could direct email not only to the PI but also to designated research staff or coordinators, allowing a comparison of which responders provide the greatest gain in efficiency. TrialTriage must ultimately be tested in its integrated form, combining the semiautonomous classification workflow and the human-in-the-loop email response, to determine the overall accuracy and timing of the complete system.

Scalability

How the rule engine behaves as eligibility criteria increase, perhaps by a factor of 10 or more, is unknown. Testing performance with substantially more eligibility criteria must also include the human responders, whose replies are essential to the iterative process.

Conclusions

TrialTriage demonstrated the feasibility of resolving prescreening ambiguity through immediate investigator email communication and iterative reclassification within the same automated workflow. If validated on EHR-derived cases and complex phase I trial criteria, the TrialTriage architecture could reduce manual prescreening workload and improve prescreening accuracy in early-phase oncology trials.

Acknowledgments

The author thanks Ryan Nolan, BSEE, for his implementation of the n8n workflow according to the author's specifications, along with the physicians and clinicians who contributed their clinical judgment to the manual classification survey. Draw.io was used to create Figure 1. Generative AI tools were used in the preparation of this manuscript. The literature search was carried out using SciSpace and Consensus. Claude Sonnet 4.6 assisted with text editing and generated Tables 1–3. The author reviewed and verified all AI-generated content, including literature references, and takes full responsibility for the manuscript. No AI tool is listed as an author.

Funding

The author declared that no financial support was received for this work. The platform vendor, n8n, did not sponsor this work, provide financial support, or compensate the consultant. The author used a standard self-paid subscription and personally funded all costs.

Data Availability

All data generated and analyzed during this study are included in the manuscript and its multimedia appendices.

Authors' Contributions

Conceptualization: KAK
Data curation: KAK
Formal analysis: KAK
Investigation: KAK
Methodology: KAK
Validation: KAK
Visualization: KAK
Writing – original draft: KAK
Writing – review and editing: KAK

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of the deployed TrialTriage n8n workflow, showing the complete workflow canvas and detailed views of each functional section.
[DOCX File (Microsoft Word File), 1522 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Table of n8n node types used in the TrialTriage workflow, with the function and number of deployed instances of each.
[DOCX File (Microsoft Word File), 17 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Prompt template used for generation of the synthetic pancreatic cancer case set across the 3 large language models reported in this study (Claude Sonnet 4.6, Gemini 3.1, and Grok 4).
[DOCX File (Microsoft Word File), 19 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Datasets generated and analyzed during the study, organized by source (large language model or human reviewer). Each tab contains the synthetic case classifications, the ground truth, and the corresponding TrialTriage workflow outputs.
[XLSX File (Microsoft Excel File), 40 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Survey instrument used to collect case classifications and completion times from human reviewers for comparison against the TrialTriage workflow.
[DOCX File (Microsoft Word File), 25 KB-Multimedia Appendix 5]

Checklist 1

TRIPOD-LLM checklist.
[DOCX File (Microsoft Word File), 35 KB-Checklist 1]

References

1. Araujo D, Greystoke A, Bates S, et al. Oncology phase I trial design and conduct: time for a change-MDICT Guidelines 2022. Ann Oncol. Jan 2023;34(1):48-60. [doi: 10.1016/j.annonc.2022.09.158] [Medline: 36182023]
2. Alotaibi H, Anis AM, Alloghbi A, Alshammari K. Oncology early-phase clinical trials in the Middle East and North Africa: a review of the current status, challenges, opportunities, and future directions. J Immunother Precis Oncol. Aug 2024;7(3):178-189. [doi: 10.36401/JIPO-23-25] [Medline: 39219998]
3. Weber JS, Levit LA, Adamson PC, et al. American Society of Clinical Oncology policy statement update: the critical role of phase I trials in cancer research and treatment. J Clin Oncol. Jan 20, 2015;33(3):278-284. [doi: 10.1200/JCO.2014.58.2635]
4. Rubin EH, Gilliland DG. Drug development and clinical trials—the path to an approved cancer drug. Nat Rev Clin Oncol. Feb 28, 2012;9(4):215-222. [doi: 10.1038/nrclinonc.2012.22] [Medline: 22371130]
5. Massett HA, Mishkin G, Rubinstein L, et al. Challenges facing early phase trials sponsored by the National Cancer Institute: an analysis of corrective action plans to improve accrual. Clin Cancer Res. Nov 15, 2016;22(22):5408-5416. [doi: 10.1158/1078-0432.CCR-16-0338] [Medline: 27401246]
6. Denicoff AM, Ivy SP, Tamashiro TT, et al. Implementing modernized eligibility criteria in US National Cancer Institute clinical trials. J Natl Cancer Inst. Nov 14, 2022;114(11):1437-1440. [doi: 10.1093/jnci/djac152] [Medline: 36047830]
7. 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. Jul 24, 2019;7(3):e14185. [doi: 10.2196/14185] [Medline: 31342909]
8. Xiang JJ, Roy A, Summers C, et al. Brief report: implementation of a universal prescreening protocol to increase recruitment to lung cancer studies at a Veterans Affairs Cancer Center. JTO Clin Res Rep. Jul 2022;3(7):100357. [doi: 10.1016/j.jtocrr.2022.100357] [Medline: 35815320]
9. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. Clin Trials. Feb 2015;12(1):77-83. [doi: 10.1177/1740774514558307] [Medline: 25475878]
10. Williams RJ, Tse T, DiPiazza K, Zarin DA. Terminated trials in the ClinicalTrials.gov results database: evaluation of availability of primary outcome data and reasons for termination. PLoS ONE. 2015;10(5):e0127242. [doi: 10.1371/journal.pone.0127242] [Medline: 26011295]
11. Wornow M, Lozano A, Dash D, Jindal J, Mahaffey KW, Shah NH. Zero-shot clinical trial patient matching with LLMs. NEJM AI. Jan 2025;2(1):AIcs2400360. [doi: 10.1056/AIcs2400360]
12. Unger JM, Xiao H, Vaidya R, et al. The cost of doing business: drug costs in federally sponsored cancer clinical trials. JCO Oncol Pract. Oct 2024;20:10. [doi: 10.1200/OP.2024.20.10_suppl.10]
13. Unger JM, Hershman DL, Till C, et al. "When offered to participate": a systematic review and meta-analysis of patient agreement to participate in cancer clinical trials. J Natl Cancer Inst. Mar 1, 2021;113(3):244-257. [doi: 10.1093/jnci/djaa155] [Medline: 33022716]

14. Canouï-Poitrine F, Lièvre A, Dayde F, et al. Inclusion of older patients with cancer in clinical trials: the SAGE prospective multicenter cohort survey. Oncologist. Dec 2019;24(12):e1351-e1359. [doi: 10.1634/theoncologist.2019-0166] [Medline: 31324663]
15. Calaprice-Whitty D, Galil K, Salloum W, Zariv A, Jimenez B. Improving clinical trial participant prescreening with artificial intelligence (AI): a comparison of the results of AI-assisted vs standard methods in 3 oncology trials. Ther Innov Regul Sci. Jan 2020;54(1):69-74. [doi: 10.1007/s43441-019-00030-4] [Medline: 32008227]
16. Mckane A, Sima C, Ramanathan RK, et al. Determinants of patient screen failures in phase 1 clinical trials. Invest New Drugs. Jun 2013;31(3):774-779. [doi: 10.1007/s10637-012-9894-7] [Medline: 23135779]
17. Campillo-Gimenez B, Buscail C, Zekri O, et al. Improving the pre-screening of eligible patients in order to increase enrollment in cancer clinical trials. Trials. Jan 16, 2015;16(1):15. [doi: 10.1186/s13063-014-0535-7] [Medline: 25592642]
18. Frankel PH, Chung V, Tuscano J, et al. Model of a queuing approach for patient accrual in phase 1 oncology studies. JAMA Netw Open. May 1, 2020;3(5):e204787. [doi: 10.1001/jamanetworkopen.2020.4787] [Medline: 32401317]
19. Delorme J, Charvet V, Wartelle M, et al. Natural language processing for patient selection in phase I or II oncology clinical trials. JCO Clin Cancer Inform. Jun 2021;5(5):709-718. [doi: 10.1200/CCI.21.00003] [Medline: 34197179]
20. Ferber D, Hilgers L, Wiest IC, et al. End-to-end clinical trial matching with large language models. arXiv. Preprint posted online on Jul 18, 2024. [doi: 10.48550/arXiv.2407.13463]
21. Haddad T, Helgeson JM, Pomerleau KE, et al. Accuracy of an artificial intelligence system for cancer clinical trial eligibility screening: retrospective pilot study. JMIR Med Inform. Mar 26, 2021;9(3):e27767. [doi: 10.2196/27767] [Medline: 33769304]
22. Mazor T, Farhat KS, Trukhanov P, et al. Clinical trial notifications triggered by artificial intelligence–detected cancer progression: a randomized trial. JAMA Netw Open. Apr 1, 2025;8(4):e252013. [doi: 10.1001/jamanetworkopen.2025.2013] [Medline: 40257799]
23. Kehl KL, Mazor T, Trukhanov P, et al. Identifying oncology clinical trial candidates using artificial intelligence predictions of treatment change: a pilot implementation study. JCO Precis Oncol. Mar 2024;8(8):e2300507. [doi: 10.1200/PO.23.00507] [Medline: 38513166]
24. Ozaki H, Miyawaki E, Miyazaki N, et al. Patient characteristics related to screening failure in phase I trials. BMC Cancer. Dec 4, 2025;26(1):69. [doi: 10.1186/s12885-025-15311-5] [Medline: 41339809]
25. Nievas M, Basu A, Wang Y, Singh H. Distilling large language models for matching patients to clinical trials. J Am Med Inform Assoc. Sep 1, 2024;31(9):1953-1963. [doi: 10.1093/jamia/ocae073] [Medline: 38641416]
26. Meystre SM, Heider PM, Cates A, et al. Piloting an automated clinical trial eligibility surveillance and provider alert system based on artificial intelligence and standard data models. BMC Med Res Methodol. Apr 11, 2023;23(1):88. [doi: 10.1186/s12874-023-01916-6] [Medline: 37041475]
27. Wu J, Yakubov A, Abdul-Hay M, et al. Prescreening to increase therapeutic oncology trial enrollment at the largest public hospital in the United States. JCO Oncol Pract. Apr 2022;18(4):e620-e625. [doi: 10.1200/OP.21.00629] [Medline: 34748371]
28. La Rosa A, Vaterkowski M, Cuggia M, et al. “The truth is, we must miss some”: a qualitative study of the patient eligibility screening process, and automation perspectives, for cancer clinical trials. Cancer Med. Dec 2024;13(23):e70466. [doi: 10.1002/cam4.70466] [Medline: 39624972]
29. Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products: guidance for industry and other interested parties. U.S. Food and Drug Administration; 2025. URL: https://www.fda.gov/media/184830/download [Accessed 2026-07-10]
30. Guiding principles of good AI practice in drug development. European Medicines Agency; 2026. URL: https://www.ema.europa.eu/en/documents/other/guiding-principles-good-ai-practice-drug-development_en.pdf [Accessed 2026-07-10]
31. Gallifant J, Afshar M, Ameen S, et al. The TRIPOD-LLM reporting guideline for studies using large language models. Nat Med. Jan 2025;31(1):60-69. [doi: 10.1038/s41591-024-03425-5] [Medline: 39779929]
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. Mar 1977;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
33. Embi PJ, Jain A, Clark J, Bizjack S, Hornung R, Harris CM. Effect of a clinical trial alert system on physician participation in trial recruitment. Arch Intern Med. Oct 24, 2005;165(19):2272-2277. [doi: 10.1001/archinte.165.19.2272] [Medline: 16246994]
34. Köpcke F, Prokosch HU. Employing computers for the recruitment into clinical trials: a comprehensive systematic review. J Med Internet Res. Jul 1, 2014;16(7):e161. [doi: 10.2196/jmir.3446] [Medline: 24985568]

35. Ni Y, Wright J, Perentesis J, et al. Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients. BMC Med Inform Decis Mak. Apr 14, 2015;15(1):28. [doi: 10.1186/s12911-015-0149-3] [Medline: 25881112]
36. Mallahan S, Gajra A, Blau S, et al. Optimizing clinical trial subject selection: insights from Exigent Research Network and the Tempus AI TIME Program collaboration. AI Precis Oncol. Dec 1, 2024;1(6):306-314. [doi: 10.1089/aipo.2024.0030]
37. Abdallah M, Nakken S, Georges M, et al. TrialMatchAI: an end-to-end AI-powered clinical trial recommendation system to streamline patient-to-trial matching. Nat Commun. Mar 25, 2026;17(1):4472. [doi: 10.1038/s41467-026-70509-w] [Medline: 41876500]
38. Gupta S, Basu A, Nievas M, et al. PRISM: patient records interpretation for semantic clinical trial matching system using large language models. NPJ Digit Med. Oct 28, 2024;7(1):305. [doi: 10.1038/s41746-024-01274-7] [Medline: 39468259]
39. Jin Q, Wang Z, Floudas CS, et al. Matching patients to clinical trials with large language models. Nat Commun. Nov 18, 2024;15(1):9074. [doi: 10.1038/s41467-024-53081-z] [Medline: 39562565]
40. Alexander M, Solomon B, Ball DL, et al. Evaluation of an artificial intelligence clinical trial matching system in Australian lung cancer patients. JAMIA Open. Jul 2020;3(2):209-215. [doi: 10.1093/jamiaopen/ooaa002] [Medline: 32734161]
41. Hernando-Calvo A, Nguyen P, Bedard PL, et al. Impact on costs and outcomes of multi-gene panel testing for advanced solid malignancies: a cost-consequence analysis using linked administrative data. EClinicalMedicine. Mar 2024;69:102443. [doi: 10.1016/j.eclinm.2024.102443] [Medline: 38380071]
42. Penberthy LT, Dahman BA, Petkov VI, DeShazo JP. Effort required in eligibility screening for clinical trials. J Oncol Pract. Nov 2012;8(6):365-370. [doi: 10.1200/JOP.2012.000646] [Medline: 23598846]
43. Dienstmann R, Garralda E, Aguilar S, et al. Evolving landscape of molecular prescreening strategies for oncology early clinical trials. JCO Precis Oncol. 2020;4:PO.19.00398. [doi: 10.1200/PO.19.00398] [Medline: 32923891]
44. Middleton G, Fletcher P, Popat S, et al. The National Lung Matrix Trial of personalized therapy in lung cancer. Nature. Jul 2020;583(7818):807-812. [doi: 10.1038/s41586-020-2481-8] [Medline: 32669708]
45. Beigi M, Shafquat A, Mezey J, Aptekar J. Simulants: synthetic clinical trial data via subject-level privacy-preserving synthesis. AMIA Annu Symp Proc. 2023;2022:231-240. [Medline: 37128411]
46. Hripcsak G, Albers DJ. Correlating electronic health record concepts with healthcare process events. J Am Med Inform Assoc. Dec 2013;20(e2):e311-8. [doi: 10.1136/amiajnl-2013-001922] [Medline: 23975625]
47. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. Jan 1, 2013;20(1):144-151. [doi: 10.1136/amiajnl-2011-000681] [Medline: 22733976]
48. Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. EGEMS (Wash DC). 2016;4(1):1244. [doi: 10.13063/2327-9214.1244] [Medline: 27713905]
49. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. Summit Transl Bioinform. Mar 1, 2010;2010:1-5. [Medline: 21347133]

Abbreviations

CTM: clinical trial matching
ECOG: Eastern Cooperative Oncology Group
EHR: electronic health record
LLM: large language model
PI: principal investigator

Edited by Luke MacNeill; peer-reviewed by Heber Anandan, Kinjal Satasiya; submitted 08.May.2026; final revised version received 30.Jun.2026; accepted 06.Jul.2026; published 05.Aug.2026

Please cite as:
Kern KA
TrialTriage, a Semiautonomous Prescreening Workflow for Resolving Ambiguity in Phase I Oncology Trial Eligibility: Development and Proof-of-Concept Study Using Synthetic Cases
JMIR Form Res 2026;10:e100779
URL: https://formative.jmir.org/2026/1/e100779
doi: 10.2196/100779