Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

Preserving the Clinical Question When Artificial Intelligence Reshapes Trial Participation: A Methodological Framework Linking Eligibility Optimization, Treatment-Effect Heterogeneity, Estimand Integrity, and Transportability

Rajiv Sharma[1], Sophie Dubois[2]*, Kenji Nakamura[3], Priya Nair[4]

1. ***Department of Clinical Trials and AI-Based Patient Selection, Indian Institute of Management Calcutta, Kolkata, India.***
2. ***Department of Biostatistics and Causal Inference, HEC Paris, Paris, France.***
3. ***Department of Clinical Epidemiology and Trial Methodology, Hitotsubashi University, Tokyo, Japan.***
4. ***Department of Pharmaceutical Data Science and Clinical Trials, Indian Institute of Technology Bombay, Mumbai, India.***

ARTICLE INFO

Received:
03 June 2026
Received in revised form:
14 October 2026
Accepted:
18 October 2026
Available online:
28 October 2026

Keywords: Clinical trial eligibility, Artificial intelligence, Treatment-effect heterogeneity, Estimand, Transportability, External validity

ABSTRACT

Artificial intelligence is increasingly capable of extracting eligibility criteria, matching patients to clinical trials, ranking potential participants, and supporting protocol-level decisions about enrollment. These capabilities are usually evaluated as problems of screening accuracy, recruitment efficiency, or operational scalability. Yet optimizing who enters a trial can also alter the population in which treatment effects are observed and, consequently, the clinical question the trial can legitimately answer. This article develops a methodological framework for preserving that question when artificial intelligence influences trial participation. The analysis links four domains that are commonly treated separately: eligibility design, treatment-effect heterogeneity, estimand specification, and transportability. The central distinction is between operational optimization that leaves the intended inferential target substantively intact and selection that changes the distribution of treatment-effect modifiers, narrows empirical support for the target population, or creates discordance between the population named in the estimand and the population represented by enrolled participants. Eligibility optimization is therefore treated as a scientific intervention on the evidence-generating population rather than as a neutral administrative improvement. The framework distinguishes legitimate enrichment, transparent redefinition of the target question, and unacknowledged clinical-question drift. It further separates model-level validation of matching or eligibility decisions from inferential validation of the resulting trial population. The resulting approach places population–estimand alignment, heterogeneity assessment, and transportability considerations upstream in the design of AI-supported recruitment rather than leaving them to post hoc interpretation.

*This is an **open-access** article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, and build upon the work non commercially.*

To Cite This Article: Sharma R, Dubois S, Nakamura K, Nair P. Preserving the Clinical Question When Artificial Intelligence Reshapes Trial Participation: A Methodological Framework Linking Eligibility Optimization, Treatment-Effect Heterogeneity, Estimand Integrity, and Transportability. Pharmacophore. 2026;17(5):140-50. https://doi.org/10.51847/WjioHKFBSn

Introduction

Artificial intelligence is moving rapidly into the processes that determine which patients are considered for clinical trials. Contemporary systems can retrieve candidate studies, interpret complex eligibility rules, compare patient characteristics with trial requirements, and rank apparently suitable trial opportunities [1]. Large language models extend this capacity by decomposing eligibility into criterion-level judgments and integrating those judgments into patient–trial recommendations [2]. The resulting technological shift is often described as an answer to well-known operational problems: manual screening is labor intensive, eligibility language can be difficult to interpret consistently, and potentially eligible patients may never be identified. Those problems are important, but they do not exhaust the methodological consequences of AI-mediated participation.

Corresponding Author: Sophie Dubois, Department of Biostatistics and Causal Inference, HEC Paris, Paris, France. E-mail: sophie.dubois@hec.fr.

Eligibility rules themselves are increasingly computational objects. Large language models can extract structured trial criteria from protocol text [3], while smaller or distilled models can be adapted to patient–trial matching tasks [4]. This transformation matters because a computational representation is not merely a faster route to a fixed decision. Different representations can affect which characteristics are recognized, how ambiguous concepts are operationalized, and which candidates are placed near or beyond an eligibility boundary. Recent work has also moved upstream from matching existing protocols toward analysis of eligibility-criteria structure and clustering, creating a potential role for AI in the design or revision of the rules themselves [5]. At the implementation end of the same continuum, AI-enabled systems linked to standardized clinical data can perform near-real-time prescreening of patients for oncology trials [6]. Trial participation can therefore be influenced at multiple points: during protocol construction, criterion interpretation, case finding, ranking, prescreening, and referral.

The usual technical question is whether such systems identify eligible patients accurately. The methodological question developed here is different: what clinical question is being preserved when the system changes who enters the trial? An eligibility algorithm can have high predictive accuracy relative to the written protocol and still contribute to a population that differs systematically from the population for which the treatment question was intended. Conversely, an AI system may broaden practical access without materially altering the treatment-effect distribution or the inferential target. The scientific problem is therefore not whether algorithmic selection is intrinsically harmful or beneficial. It is how to distinguish efficiency gains that preserve the intended question from selection changes that redefine it.

This distinction is especially important because randomized treatment assignment protects internal comparison within the randomized trial population, but it does not make the composition of that population irrelevant. The population in which treatment is randomized determines which combinations of baseline risk, biological state, prior treatment, comorbidity, concomitant therapy, demographic characteristics, and care context contribute information to the estimated treatment effect. When these variables are only prognostic, changing their distribution may affect event rates or precision without necessarily changing the underlying treatment contrast. When they modify treatment effect, however, changing their distribution can change the average effect observed in the enrolled population. The same enrollment change can therefore be operationally beneficial, statistically neutral, or inferentially consequential depending on what characteristics are shifted and how they relate to treatment response.

The methodological problem also extends beyond heterogeneity. An estimand defines the treatment effect intended to answer a particular clinical question, including the population in which that effect is defined. Transportability asks under what assumptions evidence obtained in the trial can inform a separate target population. If AI-driven selection changes empirical representation while the trial continues to invoke the original estimand and external target, a mismatch can emerge between the population named by the research question and the population that generated the observed contrast. In other circumstances, intentional enrichment may appropriately narrow the population and yield a more precise answer to a deliberately narrower question. The distinction is not broad eligibility versus narrow eligibility; it is preserved alignment versus unacknowledged divergence.

This article develops that distinction as an original methodological framework. Its purpose is not to prescribe a universal set of eligibility criteria, require demographic representativeness as an end in itself, or assume that AI introduces selection problems absent from conventional recruitment. Instead, AI provides a particularly important case because algorithmic systems can apply rules at scale, combine numerous eligibility variables, prioritize candidates according to explicit or implicit optimization targets, and feed operational information back into trial-design decisions. The analysis therefore links AI-mediated participation to the scientific structure of the treatment question. The next section examines how eligibility optimization can alter the evidence-generating population. Section 3 then connects these changes to treatment-effect heterogeneity, estimand integrity, and transportability.

How AI-Driven Trial Participation Can Alter the Clinical Question

Eligibility criteria define more than the administrative border of a trial. They construct the population in which the randomized comparison will be observed. This is evident when criteria are systematically modified. In a large real-world analysis of oncology trials, computational relaxation of selected eligibility criteria substantially expanded the pool of potentially eligible patients while permitting investigation of how the estimated treatment effect behaved under alternative eligibility configurations [7]. The important methodological lesson is not that criteria should always be broadened. It is that the eligibility specification is capable of changing the population from which the trial's evidence is produced and must therefore be treated as part of the scientific design.

Empirical work on eligibility broadening reinforces this point from several directions. Broader eligibility criteria can change the demographic and clinical composition of the patients who would qualify, recommendations concerning previously excluded populations have changed trial inclusion practices over time, and restrictive eligibility remains a material issue across clinical research settings [8-10]. These observations establish a first distinction that is essential for AI-supported recruitment: an algorithm can be perfectly faithful to a restrictive protocol and still reproduce a narrow evidence-generating population, while an algorithm designed to identify more candidates can expand the eligible pool without necessarily preserving the same inferential structure. Matching accuracy and population appropriateness are related but different properties.

The scientific significance of a criterion also depends on what it represents. Performance-status requirements can exclude individuals on the basis of functional state even when the relevance of a particular threshold to safety or treatment response is uncertain [11]. Washout periods and restrictions on concomitant medications can encode assumptions about pharmacological

interference, toxicity, or interpretability that vary across therapies and settings [12]. Requirements concerning prior therapy likewise embody assumptions about treatment history, resistance, sequencing, and expected benefit–risk [13]. AI-based optimization cannot therefore treat every eligibility rule as an interchangeable obstacle to recruitment. Some rules protect patient safety, some reduce confounding or interpretive ambiguity, some identify the population for whom the intervention is intended, and some may persist despite weak contemporary justification.

This heterogeneity creates several mechanisms by which AI-mediated participation can alter the clinical question. First, an AI system may apply existing criteria differently, for example by resolving ambiguous clinical language, integrating scattered information, or assigning borderline cases more consistently. Second, it may change the practical reach of existing criteria by locating patients who would otherwise be missed. Third, it may prioritize among eligible patients, producing an enrolled population that differs from the nominally eligible population even when the formal criteria are unchanged. Fourth, AI-supported protocol analysis may influence which criteria are retained, relaxed, combined, or newly introduced. Only the first of these mechanisms can plausibly be described as purely administrative, and even then only when altered interpretation does not systematically change who crosses the enrollment boundary.

The key methodological distinction is between the intended trial population, the protocol-eligible population, and the actually enrolled population. These populations may overlap extensively, but they are not identical by definition. The intended population is specified by the clinical question. The protocol-eligible population is generated by applying inclusion and exclusion criteria. The enrolled population is further shaped by identification, referral, geography, site selection, patient preference, investigator judgment, competing trials, data availability, algorithmic ranking, and other participation mechanisms. AI can operate between each of these stages. Consequently, optimizing the transition from potential participant to enrolled participant can change the distribution of characteristics represented in the trial without formally changing a single line of the protocol.

This is where clinical-question drift can occur. In this article, clinical-question drift means a substantive change in the relationship between the intended treatment question and the population that empirically generates its answer, occurring without a corresponding explicit revision of the target question. The term does not apply whenever the enrolled sample differs from the source population; exact representativeness is neither necessary nor generally attainable. Drift becomes methodologically relevant when selection changes characteristics that matter to the treatment contrast, the estimand, or the population to which the result is intended to apply. A deliberate enrichment strategy that openly targets a biomarker-defined subgroup can therefore be methodologically coherent, whereas an opaque optimization procedure that produces the same narrowing while the trial continues to claim a broader target may not be.

This distinction also prevents a false opposition between inclusion and rigor. Broadening criteria can improve access and make the trial population more similar to clinical practice, yet broadening is not inherently equivalent to greater transportability. Conversely, restricting enrollment can increase homogeneity or protect safety without necessarily invalidating the external interpretation of the trial. What matters is the reason for the restriction, the characteristics affected, whether those characteristics modify treatment response or intervention feasibility, and whether the inferential target has been defined consistently with the resulting population.

The principal mechanisms are summarized in **Table 1**. The table treats the unit of analysis as a selection transformation rather than as a generic barrier or facilitator, because the same technological intervention can have different inferential consequences depending on where it acts and which characteristics it changes.

Table 1. Selection transformations through which AI-mediated participation can preserve, narrow, or alter the trial question

Selection transformation	What AI changes operationally	Population consequence to examine	Potential effect on treatment contrast	Potential effect on the clinical question	Preservation condition
Criterion interpretation	Converts narrative criteria and patient data into eligibility judgments	Which borderline or ambiguously documented patients cross the eligibility boundary	Usually indirect; material if affected features modify treatment response	The formal question may remain unchanged while its empirical population shifts	Interpretation rules remain clinically faithful and do not systematically exclude or favor relevant response strata
Candidate discovery	Identifies eligible patients who conventional workflows would miss	Representation within the protocol-eligible population	May improve support across clinically relevant strata	Can strengthen alignment between intended and enrolled populations	Discovery expands access without substituting a different target population
Eligibility-rule revision	Supports removal, relaxation, combination, or redesign of criteria	Changes the protocol-eligible population itself	May change the mixture of prognostic factors or treatment-effect modifiers	Can legitimately redefine the population attribute of the question	Any population change is prespecified and reconciled with the estimand
Candidate prioritization	Ranks or triages eligible patients for referral or screening	Alters which eligible candidates	Can change the observed average effect when	May create divergence between nominal	Ranking objective is evaluated against the

		are enrolled first or at all	prioritization tracks effect modifiers	eligibility and realized evidence population	intended inferential target
Site- or workflow-linked prescreening	Applies eligibility logic preferentially where structured data and implementation infrastructure are available	Creates differential participation by site, documentation, access, or care pathway	Treatment contrast may change if implementation-linked characteristics modify response	Can restrict the practical population beyond the written protocol	Coverage and data availability do not create unrecognized inferential exclusions
Enrichment optimization	Deliberately increases enrollment of patients predicted to be informative or responsive	Concentrates particular biological or clinical strata	Can increase or change the average treatment effect	The clinical question becomes narrower if the intended treatment population is narrowed	Enrichment is explicit, justified, and reflected in estimand and target-population language
Proposed preservation check	Compares intended population, eligibility logic, and resulting enrollment distribution	Detects whether selection has changed decision-relevant population support	Determines whether HTE could make the shift consequential	Distinguishes efficiency improvement from clinical-question redefinition	Population differences are interpreted through effect modification, estimand attributes, and transportability rather than through resemblance alone

The final row is an original synthesis proposed in this article. It anticipates a central implication of the framework: AI-mediated enrollment should not be judged only by whether it improves screening yield or criterion concordance. It should also be assessed for whether the population transformation it produces is scientifically compatible with the question the trial is intended to answer.

Eligibility, Heterogeneity, Estimands and Transportability

The inferential consequences of AI-mediated participation become clearer when eligibility is connected explicitly to estimands. The estimand framework requires investigators to define the treatment effect corresponding to the trial objective, including the population, treatment conditions, endpoint and handling of intercurrent events [14-16]. This matters because the "population" component is not background metadata. It is part of what makes one treatment effect different from another. A trial designed to estimate an effect in all patients with a disease is not asking exactly the same question as one designed for patients with a defined biomarker, treatment history, risk profile, or functional status.

Estimand integrity therefore depends on more than writing an estimand statement in the protocol. Principles for estimand construction emphasize alignment among trial objective, design, conduct, analysis and interpretation [17]. Formal approaches similarly show that an estimand represents a precisely defined target quantity rather than a retrospective label attached to whatever estimate a trial happens to produce [18]. Complex clinical processes, including competing events and alternative definitions of cumulative incidence, further demonstrate that apparently similar treatment questions can correspond to different estimands and different interpretations [19]. The implication for AI-supported eligibility is that a change in enrollment cannot be evaluated solely by asking whether the randomized comparison remains unbiased within the selected participants. One must also ask whether those participants still provide the empirical support required for the prespecified target quantity.

Treatment-effect heterogeneity provides the most direct mechanism through which population composition can change the observed average treatment effect. Contemporary reviews show that machine-learning approaches to heterogeneous treatment effects use substantially different modeling strategies and rely on different assumptions about effect modification [20]. Structured workflows for drug development likewise emphasize that heterogeneity analyses require planning, appropriate data, multiplicity awareness, and careful interpretation rather than uncritical discovery of responsive subgroups [21]. These cautions matter because AI-driven recruitment creates a separate selection problem: the enrollment model may change the distribution of the very covariates later used to investigate heterogeneity.

The principle is straightforward. If an eligibility change increases enrollment of patients with a higher baseline event risk but treatment effect is constant on the relevant effect scale, the overall event rate may change while the causal treatment contrast remains stable. If the same change increases enrollment of patients whose biological state modifies treatment response, the average treatment effect in the enrolled population may change. Neither consequence can be inferred from demographic or prognostic differences alone. What is required is knowledge—or at least a plausible, prespecified account—of which variables may modify treatment effect and whether the enrollment process changes their distribution.

This issue did not begin with large language models. Earlier deep-learning systems already demonstrated the feasibility of automated cohort selection from clinical records [22], while algorithmic approaches to recruitment have optimized site or investigator selection using historical enrollment information [23]. Such systems illustrate a broader methodological point: the optimization target used by a recruitment algorithm need not be the same as the inferential target of the trial. An algorithm

may be rewarded for finding eligible patients, maximizing predicted enrollment, minimizing screening burden, or increasing recruitment speed. None of these objectives necessarily penalizes the model for altering the distribution of a treatment-effect modifier unless that property is explicitly incorporated into design and evaluation.

The Predictive Approaches to Treatment effect Heterogeneity guidance reinforces why such distinctions must be handled cautiously [24, 25]. Heterogeneity can be assessed through baseline-risk modeling, individual-effect modeling, interactions, or other predictive approaches, but apparent subgroup differences may reflect modeling instability, limited sample size, multiplicity, or scale dependence rather than reproducible causal variation. For AI-supported enrollment, this means the preservation problem should not be operationalized as "make the recruited sample identical on every subgroup variable." That would replace one methodological error with another. The relevant task is to identify characteristics with credible clinical or statistical grounds for modifying the treatment contrast and then determine whether optimization materially changes empirical support across those characteristics.

Transportability adds a second layer. Methods for extending trial results to an external population are explicitly target-population dependent [26]. A target population should therefore be chosen to match the population about which a clinical or policy decision is actually being made, rather than treated as an abstract notion of the "real world" [27]. This distinction is critical when AI changes recruitment. A trial can be internally valid and highly efficient yet poorly informative for a particular decision population because important combinations of effect modifiers are absent or rare. Conversely, a trial need not mirror the target population proportionally if the required conditional treatment-effect relationships can be estimated and appropriately transported.

Effect scale further complicates this relationship. Transportability assumptions that may be plausible for a relative effect measure need not be plausible for an absolute effect measure, and effect modification can differ according to the scale on which it is defined [28]. Consequently, the same AI-induced population shift may have little consequence for one target quantity but substantial consequences for another. A population change cannot therefore be declared inferentially harmless simply because one summary treatment ratio appears stable. Preservation is necessarily estimand-specific.

These relationships support a distinction among four eligibility states. Preserved selection occurs when AI changes screening efficiency or access without materially altering the support needed for the specified estimand and target population. Transparent enrichment occurs when selection deliberately changes the population and the trial question is correspondingly narrowed or modified. Recoverable mismatch occurs when the enrolled and target populations differ but adequate measured effect modifiers, overlap, and target-population data support defensible standardization or transport. Unresolved drift occurs when enrollment changes the relevant population structure while the original clinical question is retained and the available data cannot establish that the treatment effect is transportable to that intended target.

Table 2 specifies the conditions separating those states. The analytical unit is the relationship between an eligibility-induced population change and the information required to interpret or transport the corresponding treatment effect.

Table 2. Eligibility-induced population changes and the conditions governing estimand integrity and transportability

Eligibility-induced change	Heterogeneity question	Estimand consequence	Transportability consequence	Evidence needed before claiming preservation	Interpretation when evidence is insufficient
Primarily prognostic shift	Do the shifted characteristics change baseline outcome risk without modifying treatment response on the relevant scale?	Population composition changes, but the target treatment contrast may remain substantively compatible	Absolute outcomes may shift even when a relative treatment effect is stable	Prespecified assessment of plausible effect modification and effect scale	Do not infer effect distortion solely from prognostic imbalance
Shift in plausible treatment-effect modifiers	Does optimization change the distribution or support of variables linked to differential treatment response?	The average treatment effect for the enrolled population may diverge from the prespecified population estimand	Direct application to the target becomes less defensible without adjustment	Adequate modifier measurement, overlap, and treatment-effect information	Treat preservation as unresolved rather than assuming invariance
Deliberate enrichment	Is the enriched characteristic part of the intended scientific target or only an operational optimization objective?	Legitimately changes the population attribute when the question is explicitly narrowed	Transport to the broader population requires additional assumptions or evidence	Explicit revised target population and estimand plus rationale for enrichment	Do not present the enriched estimate as answering the original broader question
Loss of target-population support	Are combinations of clinically relevant modifiers absent or	The nominal estimand may remain stated, but	Weighting or standardization may fail because	Target-population comparison and support	Redefine the supported target or

	extremely sparse after selection?	empirical identification becomes weak for parts of its population	positivity/support is inadequate	assessment before inference	obtain additional evidence
Change concentrated in variables unrelated to treatment effect	Is there credible evidence that shifted characteristics do not modify the relevant treatment contrast?	Usually limited effect on the causal contrast itself	Population resemblance changes without necessarily changing transportability of the effect	Scientifically defensible effect-modification assessment	Avoid equating demographic difference automatically with inferential failure
Difference addressable with measured target data	Are all major effect modifiers required for transport observed in trial and target datasets with adequate overlap?	The trial estimand and target estimand remain distinct but can be related analytically	Standardization, weighting, or other transport methods may support target inference	Explicit identification assumptions, common variables, and adequate positivity	Report transported inference as assumption-dependent
Difference involving unmeasured or unsupported modifiers	Could the enrollment process select on determinants of treatment response that are unavailable or unsupported in target data?	Alignment between empirical trial estimate and intended target estimand cannot be verified	Transport may be unidentified or highly assumption-sensitive	Additional measurement, redesign, sensitivity analysis, or new evidence	Preserve uncertainty; do not use technical adjustment as a substitute for missing support
Proposed preservation state	Do target population, AI selection logic, modifier support, estimand definition and target inference remain mutually compatible?	Estimand interpretation remains aligned with the evidence-generating population	External interpretation is supported only to the extent required assumptions remain credible	Joint population–selection–heterogeneity–estimand–transport audit	Classify the intervention as question-preserving, transparently question-changing, or unresolved

The final row represents the article's proposed integrative logic. It does not imply that transportability must be established perfectly before recruitment begins. Rather, it requires that the trial's target population and plausible sources of treatment-effect heterogeneity inform evaluation of the AI system before operational performance becomes the dominant criterion of success.

A further implication follows from combining the eligibility, estimand, heterogeneity, and transportability literatures. The relevant causal chain is not “AI changes enrollment, therefore AI changes the treatment effect.” That claim would be too strong. The more defensible structure is conditional: AI changes the selection mechanism; the selection mechanism may change the enrolled distribution of clinically relevant characteristics; some of those characteristics may modify the treatment effect or affect support for the target population; and those changes may alter the interpretation or transportability of the specified estimand. Each link can fail, weaken, or become irrelevant in a particular trial. Clinical-question preservation therefore requires locating the point at which the chain either remains intact or breaks.

This conditional structure also distinguishes formal estimand stability from empirical estimand support. A protocol may retain exactly the same estimand wording after an AI tool is introduced, yet the selected population may provide thinner support for parts of that population definition. Conversely, a transparent change to eligibility and estimand language may constitute a methodologically coherent redesign rather than a failure of preservation. The scientific concern is not terminological consistency alone. It is whether the evidence generated by the trial remains capable of supporting the treatment-effect claim represented by the stated question.

The relationship among these elements is summarized in **Figure 1**. The central analytical issue is the alignment between the intended population and the population produced by an AI-mediated selection mechanism; heterogeneity determines when compositional change can affect the observed treatment contrast, while estimand specification and transportability determine which downstream claims remain justified.

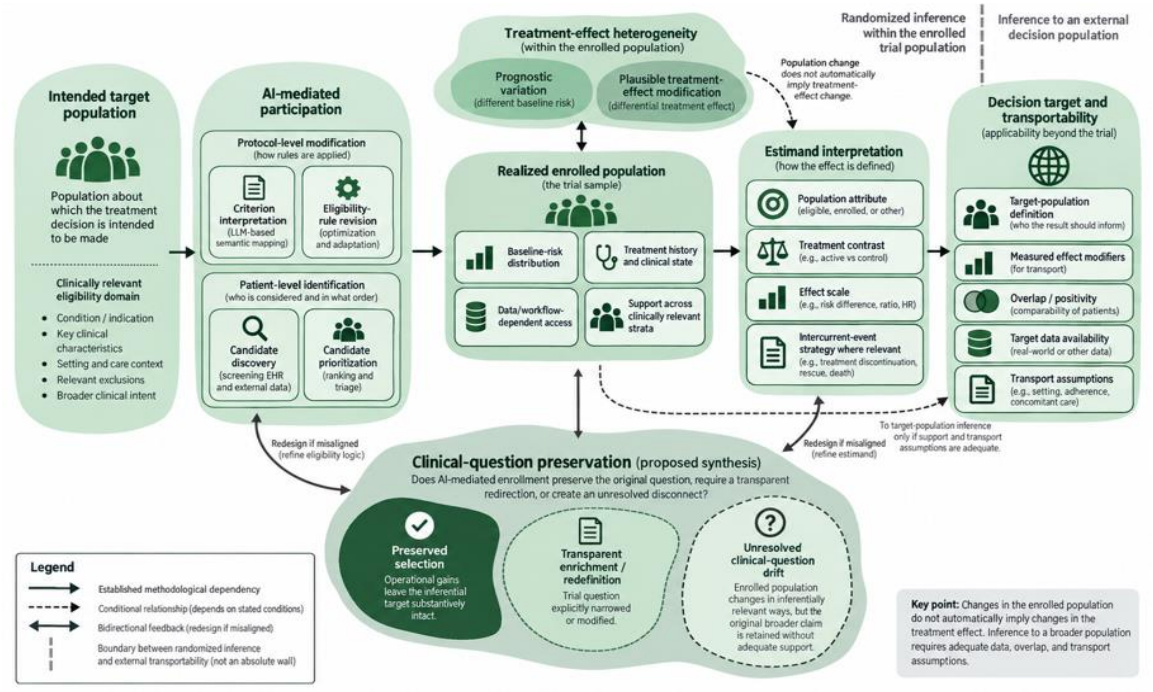


Figure 1. Where AI-mediated enrollment can preserve, redirect, or disconnect the clinical question.

Proposed Clinical-Question Preservation Framework

Clinical-question preservation requires more than checking whether algorithmically selected participants satisfy protocol criteria. Trial participation itself can introduce differences between participants and the population in which treatment decisions will ultimately be made, and those differences may matter even after conventional transportability variables are considered [29]. The appropriate object of evaluation is therefore the complete relationship among the intended population, the AI-supported selection mechanism, the population actually enrolled, the treatment effect estimated within that population, and the external population to which the result will be applied.

The first preservation check is selection–heterogeneity compatibility. Eligibility optimization should be examined for whether it redistributes characteristics plausibly associated with treatment-effect heterogeneity. Empirical work comparing treatment effects across trial-eligible and broader target populations demonstrates why this question cannot be reduced to overall representativeness [30]. Methods that explicitly extend subgroup inference toward defined target populations further show that subgroup interpretation depends on the relationship between the observed trial population and the target in which the subgroup claim is intended to hold [31]. An AI system can therefore improve recruitment while remaining inferentially neutral when it changes only characteristics unrelated to the treatment contrast, but the same operational gain becomes scientifically consequential when it reduces support across clinically important effect-modifying strata.

The second check is estimand–population compatibility. The population described by the estimand must remain connected to the population generated by recruitment. This does not require identical distributions. It requires sufficient empirical support for the treatment effect represented by the stated clinical question. If enrichment intentionally narrows that population, the appropriate response is not to conceal the narrowing through downstream adjustment but to redefine the population attribute and resulting interpretation transparently.

The third check is transport-support compatibility. Transport methods can estimate target-population quantities when appropriate trial and target data are available, including settings involving time-to-event outcomes [32]. Methods using aggregate target-population information can expand the range of feasible analyses, but they still require explicit assumptions and sufficient information about the characteristics governing trial-to-target differences [33]. Transportability should therefore be treated as a design consideration rather than a universal post hoc repair mechanism. Selection that removes empirical support for relevant population strata cannot necessarily be rescued after recruitment has ended.

These checks form a proposed clinical-question preservation framework. Its purpose is to classify AI-supported enrollment according to the relationship between operational improvement and inferential consequence. Question-preserving optimization improves identification, screening, or recruitment while maintaining adequate alignment among the intended population, relevant effect modifiers, estimand, and decision target. Transparent question modification occurs when deliberate enrichment or revised eligibility changes the population but the estimand and intended interpretation are explicitly updated. Unresolved question drift occurs when AI-mediated selection changes inferentially relevant population structure while the original broader question remains unchanged and available evidence cannot establish that the resulting treatment effect applies to that target.

The framework is operationalized through the checks in **Table 3**. Each check is deliberately framed as a comparison between two components of trial design rather than as a generic quality criterion.

Table 3. Clinical-question preservation checks across AI-supported trial design, recruitment, analysis, and interpretation

Preservation check	Design object compared	Question requiring resolution	Evidence that can support preservation	Signal requiring redesign, qualification, or explicit question change
Target–eligibility alignment	Intended treatment population vs protocol eligibility	Do the criteria operationalize the population named by the clinical objective?	Clinically justified restrictions and adequate coverage of intended strata	Criteria exclude decision-relevant strata without corresponding change in target question
Eligibility–AI alignment	Written eligibility logic vs computational implementation	Does the AI system apply the intended clinical meaning of the criteria?	Validated criterion interpretation and review of ambiguous/borderline cases	Systematic misclassification or operational reinterpretation of clinically meaningful rules
AI–enrollment alignment	Protocol-eligible population vs realized enrolled population	Does ranking, data availability, or workflow change who is actually recruited?	Comparable support across clinically relevant eligibility dimensions	Algorithmic prioritization creates additional unacknowledged selection
Enrollment–heterogeneity alignment	Enrolled population vs plausible treatment-effect modifiers	Has optimization altered support across variables that may modify treatment response?	Prespecified modifier assessment with adequate representation and overlap	Important response strata become sparse or absent
Population–estimand alignment	Realized population vs population attribute of estimand	Can the observed comparison answer the prespecified treatment-effect question?	Sufficient support for the stated population and treatment contrast	Formal estimand remains broad while empirical evidence supports only a narrower population
Trial–target alignment	Enrolled trial population vs clinical decision population	Are observed differences compatible with justified transport assumptions?	Measured effect modifiers, adequate overlap, relevant target data	Positivity failure, unmeasured modifiers, or major unsupported target strata
Performance–inference separation	Matching performance vs scientific validity	Does technical success preserve the inferential target rather than merely improve throughput?	High technical performance plus preserved population–estimand–target alignment	Recruitment gains coexist with untested or weakened external interpretation
Reassessment rule	Emerging enrollment pattern vs original design assumptions	Should eligibility, estimand, recruitment logic, or intended claim be revised?	Continuing concordance among design components	Persistent discordance requiring transparent redesign or narrower claims

Figure 2 represents the framework as a reassessment architecture rather than a one-way recruitment process. Recruitment optimization affects the population from which treatment effects are learned; observed heterogeneity and estimand interpretation then feed back into whether enrollment logic and external claims remain defensible.

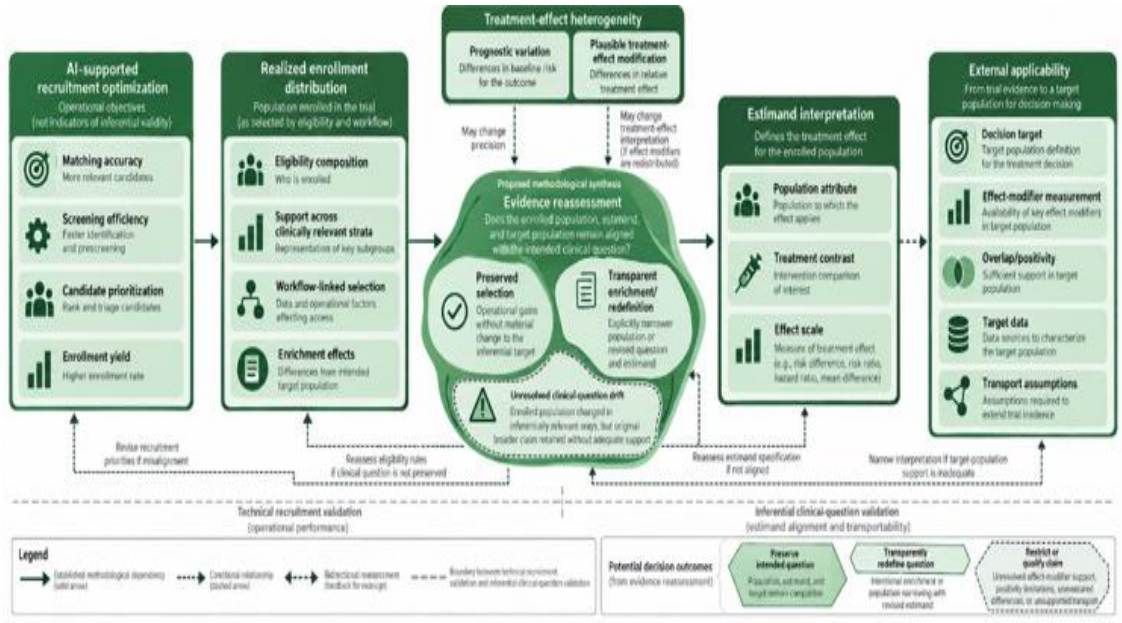


Figure 2. When recruitment optimization should trigger preservation, revision, or restriction of the trial claim

Implications for Trial Design and Evaluation

The practical implication is that AI-supported recruitment requires two distinct validation questions. The first concerns whether the system correctly identifies or prioritizes patients according to its operational specification. The second concerns whether the resulting enrollment process preserves the population required to answer the clinical question. These questions should not be collapsed into a single performance measure.

External applicability can be evaluated empirically by comparing trial participants with clinically relevant target populations and examining how adjustment changes target-population treatment-effect estimates. Trial-to-registry analyses illustrate that such comparisons can make transportability an observable methodological problem rather than an assumption attached to the discussion section [34]. The appropriate comparator, however, is not necessarily the largest available registry or a generic real-world population. It is the population for whom the trial result is intended to inform a decision.

Sampling-based generalization methods similarly demonstrate that the mechanism governing participation belongs within the inferential analysis [35]. Their usefulness depends on identification assumptions, measurement of relevant effect modifiers, and adequate overlap between trial and target populations. AI should therefore not be evaluated as though recruitment ended when a participant was successfully matched. For trials in which algorithmic selection materially shapes enrollment, evaluation should track whether the emerging population continues to support the intended estimand and whether external inference remains plausible.

A practical implementation would prospectively specify clinically credible treatment-effect modifiers, monitor their support as recruitment proceeds, compare the realized enrollment distribution with the intended decision population, and review any substantial eligibility or ranking changes against the estimand. No universal numerical threshold is warranted. A difference becomes methodologically important when it undermines the information needed for the intended treatment-effect interpretation.

This approach also generates a discriminating test of the framework. Two recruitment algorithms may achieve similar matching accuracy and enrollment yield yet differ substantially in inferential consequence. If one preserves adequate support across clinically relevant effect modifiers while the other concentrates enrollment within a narrower response profile, their operational performance may be equivalent while their capacity to answer the original clinical question is not. Conversely, conspicuous demographic differences need not establish failure when those differences do not alter the relevant treatment effect and the target inference remains supported.

The broader design principle is therefore to separate optimization performance from question preservation. AI can legitimately broaden participation, focus enrollment, or accelerate identification. What must remain explicit is whether those changes leave the original population–estimand relationship intact, deliberately replace it with a new one, or create uncertainty that requires a narrower claim.

Conclusion

Artificial intelligence can improve trial recruitment without changing the clinical question, but that outcome cannot be inferred from matching accuracy or enrollment efficiency alone. Eligibility rules determine who can contribute evidence, and AI can modify how those rules are interpreted, implemented, prioritized, or redesigned. The resulting population change becomes inferentially important when it alters treatment-effect-modifier support, weakens alignment with the population specified by the estimand, or reduces the credibility of inference to the intended decision population.

Clinical-question preservation therefore requires prospective alignment among the target population, eligibility logic, realized enrollment, plausible treatment-effect heterogeneity, estimand specification, and transportability assumptions. Deliberate enrichment is methodologically coherent when the corresponding question is explicitly narrowed. Operational optimization is question-preserving when the required inferential support remains intact. When AI-mediated selection changes relevant population structure without transparent redesign or adequate evidence for transport, the appropriate response is to restrict the resulting claim rather than assume that recruitment efficiency has preserved its meaning.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Abdallah M, Nakken S, Georges M, Bierkens M, Galvis J, Groppi A, et al. TrialMatchAI: an end-to- clinical trial recommendation system to streamline patient-to-trial matching. Nat Commun. 2026;17(1):4472. doi:10.1038/s41467-026-70509-w

2. Jin Q, Wang Z, Floudas CS, Chen F, Gong C, Bracken-Clarke D, et al. Matching patients to clinical trials with large language models. Nat Commun. 2024;15(1):9074. doi:10.1038/s41467-024-53081-z
3. Datta S, Lee K, Paek H, Manion FJ, Ofoegbu N, Du J, et al. AutoCriteria: a generalizable clinical trial eligibility criteria extraction system powered by large language models. J Am Med Inform Assoc. 2024;31(2):375-85. doi:10.1093/jamia/ocad218
4. Nievas M, Basu A, Wang Y, Singh H. Distilling large language models for matching patients to clinical trials. J Am Med Inform Assoc. 2024;31(9):1953-63. doi:10.1093/jamia/ocae073
5. Bornet A, Khlebnikov P, Meer F, Haas Q, Yazdani A, Zhang B, et al. Analysis of eligibility criteria clusters based on large language models for clinical trial design. J Am Med Inform Assoc. 2025;32(3):447-58. doi:10.1093/jamia/ocae311
6. 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(1). doi:10.1200/CCI-25-00262
7. Liu R, Rizzo S, Whipple S, Pal N, Lopez Pineda A, Lu M, et al. Evaluating eligibility criteria of oncology trials using real-world data and AI. Nature. 2021;592(7855):629-33. doi:10.1038/s41586-021-03430-5
8. Kaur M, Frahm F, Lu Y, Ascha MS, Guadamuz JS, Dotan E, et al. Broadening eligibility criteria and diversity among patients for cancer clinical trials. NEJM Evid. 2024;3(4). doi:10.1056/EVIDoa2300236
9. Xiao H, Vaidya R, Hershman DL, Unger JM. Impact of broadening trial eligibility criteria on the inclusion of patients with brain metastases in cancer clinical trials: Time series analyses for 2012-2022. J Clin Oncol. 2024;42(16):1953-60. doi:10.1200/JCO.23.01777
10. Heirali A, Heybati K, Sereeyotin J, Khan F, Yarnell C, Krewulak K, et al. Eligibility criteria of randomized clinical trials in critical care medicine. JAMA Netw Open. 2025;8(1). doi:10.1001/jamanetworkopen.2024.54944
11. Magnuson A, Bruinooge SS, Singh H, Wilner KD, Jalal S, Lichtman SM, et al. Modernizing clinical trial eligibility criteria: Recommendations of the ASCO-Friends of Cancer Research Performance Status Work Group. Clin Cancer Res. 2021;27(9):2424-9. doi:10.1158/1078-0432.CCR-20-3868
12. Harvey RD, Mileham KF, Bhatnagar V, Brewer JR, Rahman A, Moravek C, et al. Modernizing clinical trial eligibility criteria: Recommendations of the ASCO-Friends of Cancer Research Washout Period and Concomitant Medication Work Group. Clin Cancer Res. 2021;27(9):2400-7. doi:10.1158/1078-0432.CCR-20-3855
13. Osarogiagbon RU, Merino Vega D, Fashoyin-Aje L, Wedam S, Ison G, Atienza S, et al. Modernizing clinical trial eligibility criteria: Recommendations of the ASCO-Friends of Cancer Research Prior Therapies Work Group. Clin Cancer Res. 2021;27(9):2408-15. doi:10.1158/1078-0432.CCR-20-3854
14. Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: A primer on the ICH E9(R1) addendum. BMJ. 2024;384. doi:10.1136/bmj-2023-076316
15. Kahan BC, Devane D. Estimands: What they are and why we should use them. J Clin Epidemiol. 2026;189:112054. doi:10.1016/j.jclinepi.2025.112054
16. Kahan BC, Bretz F, Copas A, Harhay MO, Collins GS, Bahti M, et al. CRT-Estimands Framework: Consensus based extension of the ICH E9(R1) addendum for cluster randomised trials. BMJ. 2026;393. doi:10.1136/bmj-2025-089050
17. Mütze T, Bell J, Englert S, Hougaard P, Jackson D, Lanius V, et al. Principles for defining estimands in clinical trials—a proposal. Pharm Stat. 2025;24(1). doi:10.1002/pst.2432
18. Han S, Zhou XH. Defining estimands in clinical trials: A unified procedure. Stat Med. 2023;42(12):1869-87. doi:10.1002/sim.9702
19. Bühler A, Cook RJ, Lawless JF. Estimands and cumulative incidence function regression in clinical trials: Some new results on interpretability and robustness. Stat Med. 2024;43(29):5513-33. doi:10.1002/sim.10236
20. Inoue K, Adomi M, Efthimiou O, Komura T, Omae K, Onishi A, et al. Machine learning approaches to evaluate heterogeneous treatment effects in randomized controlled trials: A scoping review. J Clin Epidemiol. 2024;176:111538. doi:10.1016/j.jclinepi.2024.111538
21. Sechidis K, Sun S, Chen Y, Lu J, Zhang C, Baillie M, et al. WATCH: A workflow to assess treatment effect heterogeneity in drug development for clinical trial sponsors. Pharm Stat. 2025;24(2). doi:10.1002/pst.2463
22. Segura-Bedmar I, Raez P. Cohort selection for clinical trials using deep learning models. J Am Med Inform Assoc. 2019;26(11):1181-8. doi:10.1093/jamia/ocz139
23. Gligorijevic J, Gligorijevic D, Pavlovski M, Milkovits E, Glass L, Grier K, et al. Optimizing clinical trials recruitment via deep learning. J Am Med Inform Assoc. 2019;26(11):1195-202. doi:10.1093/jamia/ocz064
24. Kent DM, Paulus JK, van Klaveren D, D'Agostino R, Goodman S, Hayward R, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) statement. Ann Intern Med. 2020;172(1):35-45. doi:10.7326/M18-3667
25. Kent DM, van Klaveren D, Paulus JK, D'Agostino R, Goodman S, Hayward R, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) statement: Explanation and elaboration. Ann Intern Med. 2020;172(1). doi:10.7326/M18-3668
26. Inoue K, Hsu W. Transportability analysis—A tool for extending trial results to a representative target population. JAMA Netw Open. 2024;7(1). doi:10.1001/jamanetworkopen.2023.46302
27. Lund JL, Matthews AA. Identifying target populations to align with decision-makers' needs. Am J Epidemiol. 2024;193(11):1503-6. doi:10.1093/aje/kwae129

28. Dahabreh IJ, Robertson SE, Steingrimsson JA. Learning about treatment effects in a new target population under transportability assumptions for relative effect measures. Eur J Epidemiol. 2024;39(9):957-65. doi:10.1007/s10654-023-01067-4
29. Ung L, VanderWeele TJ, Dahabreh IJ. Generalizing and transporting causal inferences from randomized trials in the presence of trial engagement effects. Epidemiology. 2025;36(4):500-10. doi:10.1097/EDE.0000000000001863
30. Lugo-Palacios DG, Bidulka P, O'Neill S, Carroll O, Basu A, Adler A, et al. Going beyond randomised controlled trials to assess treatment effect heterogeneity across target populations. Health Econ. 2025;34(1):85-104. doi:10.1002/hec.4903
31. Webster-Clark M, Matthews AA, Ellis AR, Kinlaw AC, Platt RW. Using methods to extend inferences to specific target populations to improve the precision of subgroup analyses. J Clin Epidemiol. 2025;181:111716. doi:10.1016/j.jclinepi.2025.111716
32. Cao Z, Cho Y, Li F. Transporting randomized trial results to estimate counterfactual survival functions in target populations. Pharm Stat. 2024;23(4):442-65. doi:10.1002/pst.2354
33. Yan Y, Vuong Q, Metcalfe RK, Guan T, Shi H, Park JJH. Target Aggregate Data Adjustment Method for transportability analysis utilizing summary-level data from the target population. Pharm Stat. 2025;24(5). doi:10.1002/pst.70029
34. Wei L, Phillippo DM, Shah A, Cleland JGF, Lewsey J, McAllister DA. Transportability of two heart failure trials to a disease registry using individual patient data. J Clin Epidemiol. 2023;162:160-8. doi:10.1016/j.jclinepi.2023.08.019
35. Buchanan AL, Hudgens MG, Cole SR, Mollan KR, Sax PE, Daar ES, et al. Generalizing evidence from randomized trials using inverse probability of sampling weights. J R Stat Soc Ser A Stat Soc. 2018;181(4):1193-209. doi:10.1111/rssa.12357