

Cancer

An AI tool to make clinical trials more inclusive

Chunhua Weng & James R. Rogers

An artificial-intelligence tool called Trial Pathfinder can run clinical-trial emulations using health-care data from people with cancer, and can learn how to optimize trial-inclusion eligibility criteria, while maintaining patient safety. **See 629**

Clinical trials are the main way to determine whether new treatments are safe and effective. Trial success can depend on the timely enrolment of a representative sample of individuals who meet the eligibility criteria. However, enrolling enough people to draw a statistically significant conclusion about a trial result can be a problem. On page 629, Liu *et al.*¹ present a software tool that offers a data-driven way to optimize the inclusiveness and safety of eligibility criteria by learning from the real-world clinical data of people with cancer.

Most trials use eligibility criteria that restrict participants to those with low-risk profiles, such as healthy or young people, and exclude those who are pregnant, are elderly, or have other diseases (co-morbidities) besides the condition of interest. The exclusions are mainly to remove from the sample people who are physically vulnerable, or who might have weak immune systems or low tolerance

to drug toxicity. Such features in these individuals might compromise the uniformity of the study sample and provide confounding data. Yet this approach prevents the inclusion of some people who could potentially benefit from the trial treatment. Moreover, exclusions can contribute to a shortfall in participants that might delay a trial, compromise it because of its limited generalizability to the excluded subgroups, or cause it to be terminated because it failed to recruit enough participants.

Researchers are increasingly recognizing that eligibility criteria for clinical trials should be simplified, be made less restrictive and be better justified clinically than is currently the case². However, making eligibility criteria inclusive in a clinically meaningful way is a challenge because of a lack of evidence-based approaches that can be easily used when making these decisions. Conventional approaches

for setting eligibility criteria depend largely on the reuse of criteria from past trials or on arbitrary decisions by trial designers.

The widespread adoption of electronic health records (EHRs) has made people's clinical data available on a larger scale than was previously possible. A study published this year used EHR data to evaluate how modifications of eligibility criteria could enlarge the pool of people able to take part, and so improve the statistical power of clinical trials³. However, an accessible software tool to enable the systematic evaluation of eligibility criteria by emulating clinical trials using EHR data has been lacking.

Liu and colleagues address this lack by creating an open-source artificial-intelligence (AI) tool they call Trial Pathfinder. This tool can use EHR data to compare the survival outcomes of individuals who did or did not receive a particular approved drug treatment. Trial emulation such as this can be used to assess the effects of including or omitting eligibility criteria from the original clinical trial (Fig. 1). This offers a way to understand how eligibility criteria can be optimized by assessing the effectiveness of the treatment and the trade-offs between trial inclusiveness and participant safety.

The authors' study used the Flatiron Health EHR-derived database, which includes data from 61,094 individuals with advanced non-small-cell lung cancer at about 280 cancer clinics in the United States. Liu and colleagues focused on ten clinical trials for drugs approved for this type of cancer. Trial Pathfinder emulated these trials by identifying people who met the eligibility criteria used in the original trial in this real-world data set.

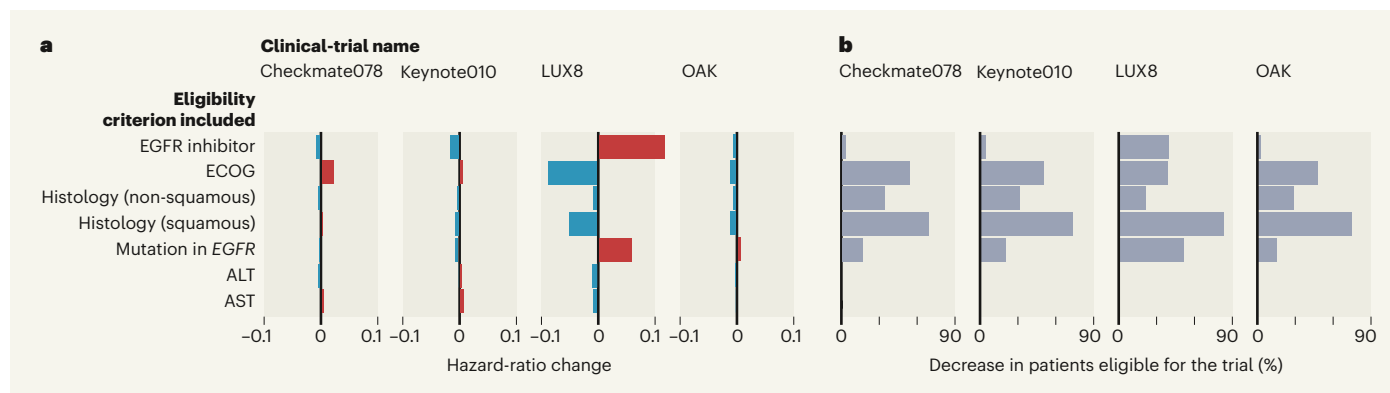


Figure 1 | Assessing the design of clinical trials. Eligibility criteria reduce the pool of potential trial participants and are not always chosen using a data-driven approach. Liu *et al.*¹ present an artificial-intelligence tool that evaluated the eligibility criteria for clinical trials of approved cancer treatments. Using a database of patient records to emulate clinical trials, the tool compared the outcomes of those who did and didn't receive a treatment associated with a particular clinical trial, and assessed the effect of each eligibility criterion used in the original trial. **a**, The tool calculated how each criterion affected the overall survival of the patients, represented by a metric called a hazard ratio (HR), which indicates a quantitative risk for an outcome. If criterion inclusion raised the HR, then removing this criterion

would be beneficial. **b**, The effect of each eligibility criterion on the number of potential participants was also assessed. These examples from the authors' study show the consequences of including certain eligibility criteria in the trial emulations. Some criteria (histology, non-squamous; and histology, squamous) decreased the number of potential participants and had minimal unwanted effects on HR. Other criteria (a scale indicating a person's ability to function normally, called ECOG; use of a therapy called an EGFR inhibitor; or a mutation in the *EGFR* gene) decreased potential participants and raised the HR of some trials. The level of the enzymes alanine aminotransferase (ALT) or aspartate aminotransferase (AST) had minimal effects on participants and HR.

On the basis of their treatment information, these eligible individuals were assigned either to the emulated treatment group (for example, those who received the immunotherapy tested in the clinical trial) or to a control group (for example, those who received a particular standard chemotherapy drug). For each trial, at least 250 individuals in the Flatiron database matched the eligibility criteria and drug treatments used in either the treatment or control groups of the original clinical trial.

Trial Pathfinder compared the treatment and control populations by calculating a value termed the overall-survival hazard ratio. This provides an assessment of whether the treatment of interest affected the probability of individuals in the treatment group surviving the time frame studied (27 months after therapy began, in this case). The lower the hazard ratio, the greater the treatment's benefit.

In real-world data, biases can arise because of physicians' or patients' judgement of disease severity, prognosis and expected treatment effect, resulting in differences in how patients are assigned treatments (for example, if those with more severe illness usually receive drug A rather than drug B). In clinical trials, randomization is a common approach to addressing treatment-selection biases. For these real-world data, the treatment is already assigned and thus randomization cannot be applied. To address this concern, Liu and colleagues used a technique called inverse probability of treatment weighting to generate less-biased estimates of the treatment effects.

The tool then ran variations of the trial emulation in which some of the original eligibility criteria were no longer included, and calculated the hazard ratio. An AI metric called the Shapley value measures the weighted average of the effect on the hazard ratio of including each criterion, and this value was used to determine the effect on inclusiveness and safety of using specific eligibility criteria.

Using this data-driven approach to select a smaller subset of the original eligibility criteria would increase the eligible population in this database from 1,553 to 3,209, on average, while achieving a lower overall-survival hazard ratio. For example, the results suggest that more women and more older adults could have been included in the trials. Comparing further trials along with the original ten, Liu *et al.* examined treatments in the same class of therapy. They found that if the eligibility criteria were standardized to align with those of the trials that had had successful recruitment and had used more-relaxed laboratory thresholds (for blood levels of molecules such as haemoglobin, for example), this would enhance trial diversity in general.

Liu and colleagues used several complementary analyses to evaluate the robustness of Trial Pathfinder. Their findings remained consistent if they used a different end point

of a clinical-trial emulation – progression-free survival (an individual's tumour does not grow). Liu and colleagues could also identify restrictive criteria that did not benefit a trial when they analysed trials for other types of cancer, such as colorectal cancer, advanced melanoma and metastatic breast cancer. The Trial Pathfinder tool provides an estimate that the population eligible for trial participation for those with other types of cancer could be increased by 53%, on average, while achieving a lower overall-survival hazard ratio, by having less-restrictive eligibility criteria.

The authors analysed toxicity follow-up and evaluation information from a further 22 trials of cancer treatments. Despite differences in the eligibility criteria used across these trials, the authors' work suggests that it is promising and feasible to consider changing some commonly used laboratory-test-based eligibility criteria and relaxing the eligibility thresholds without increasing the toxicity risk to participants. This was demonstrated by monitoring eligibility-criteria differences and finding that the omission of some criteria was associated with minimal to no changes in the number of

“The results suggest that more women and more older adults could have been included in the trials.”

treatment withdrawals from these trials owing to adverse events.

The Trial Pathfinder tool enables a scalable evaluation of the effects of relaxing specific eligibility criteria on treatment efficacy and on the size of the eligible population using retrospective data from a real-world setting. This provides actionable guidance that could be used to make improvements that have a clinical justification. Moreover, Liu and colleagues' work will encourage researchers to embrace the use of EHR data and data-driven algorithms when trying to enhance the diversity of trial populations and maintain safeguards for participants.

This study underscores the advances that are being made in evidence-based precision design of clinical-trial eligibility criteria. It might inspire AI-driven optimal participant selection for clinical trials for diseases other than cancer. However, for that to occur, major challenges would need to be overcome regarding the limitations in the quality of EHR data.

These include problems arising as a result of data complexity owing to variations in the methods used to assess and record outcomes (for example, the use of laboratory tests compared with questionnaires, or whether tests are available to quantitatively measure clinical improvement). Another issue is the lack of accessibility of the full clinical-trial protocols,

which are often treated as confidential business secrets. The Flatiron database is carefully curated and uniformly coded, whereas data from other EHR systems are often more variable and have differing levels of completeness and accuracy, and can be subject to idiosyncrasies in the data-coding practices used.

Trial Pathfinder might benefit from adopting the best practices for clinical data standardization recommended by the global open-science consortium OHDSI (Observational Health Data Sciences and Informatics). This could be achieved by Trial Pathfinder using the widely adopted OMOP (Observational Medical Outcomes Partnership) Common Data Model standardization approach, which would improve its interoperability with the vast number of different types of EHR data⁴. Health-care policymakers should consider the opportunities provided by AI tools such as Trial Pathfinder. Perhaps they could create policies that encourage clinical-trial sponsors to share their full trial protocols and to improve consistency between the full protocol and the condensed clinical-trial summaries available in public repositories such as clinicaltrials.gov.

Chunhua Weng and **James R. Rogers** are in the Department of Biomedical Informatics, Columbia University, New York, New York 10032, USA.
e-mail: chunhua@columbia.edu

1. Liu, R. *et al.* *Nature* **592**, 629–633 (2021).
2. National Academies of Sciences, Engineering, and Medicine. *Strategies for Ensuring Diversity, Inclusion, and Meaningful Participation in Clinical Trials: Proceedings of a Workshop* (National Academies, 2016).
3. Kim, J. H. *et al.* *J. Am. Med. Inform. Assoc.* **28**, 14–22 (2021).
4. Hripcsak, G. *et al.* *Stud. Health Technol. Inform.* **216**, 574–578 (2015).

This article was published online on 7 April 2021.