DOI: 10.1017/cjn.2024.352
This is a manuscript accepted for publication in *Canadian Journal of Neurological Sciences*.
This version may be subject to change during the production process.

Automatic Matching Algorithms to Identify Eligible Participants for Stroke Trials: A Proof-of-Concept Study

Pattarawut Charatpangoon,[1] Nishita Singh,[2] Brian H Buck,[3] Federico Carpani,[4] Luciana Catanese,[5] Shelagh B. Coutts,[6] Thalia S. Field,[7] Gary Hunter,[8] Houman Khosravani,[9] Kanjana Perera,[10] Tolulope T. Sajobi,[11] Michel Shamy,[12] Jai Jai Shiva Shankar,[13] Aleksander Tkach,[14] Richard H. Swartz,[15] Mohammed A. Almekhlafi,[16] Bijoy K. Menon,[17] M. Ethan MacDonald*,[18] Aravind Ganesh*[19]

*Co-senior authors

Affiliations:

1. Departments of Biomedical Engineering, the Hotchkiss Brain Institute, University of Calgary, Calgary, Canada
2. Department of Internal Medicine, Neurology Division, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada.
3. Division of Neurology, Department of Medicine, University of Alberta, Edmonton, Canada
4. University Health Network (UHN) Stroke Program. Toronto Western Hospital. University of Toronto. Toronto, Canada.
5. Department of Medicine, Neurology Division, McMaster University, Hamilton, Ontario, Canada. Population Health Research Institute, Hamilton, Ontario.
6. Departments of Clinical Neurosciences, Radiology and Community Health Sciences. University of Calgary Cumming School of Medicine, Calgary, Canada
7. Vancouver Stroke Program, Division of Neurology, University of British Columbia, Vancouver, Canada.
8. University of Saskatchewan, Saskatoon, Canada
9. Division of Neurology, Department of Medicine, Hurvitz Brain Sciences Program, Sunnybrook Health Sciences Centre, University of Toronto, ON, Canada

10. Department of Medicine, Division of Neurology, McMaster University, Hamilton, Canada
11. Departments of Community Health Sciences & Clinical Neurosciences, the Hotchkiss Brain Institute, University of Calgary Cumming School of Medicine, Calgary, Canada
12. Department of Medicine, Ottawa Heart Research Institute, University of Ottawa, ON, Canada
13. Department of Radiology, University of Manitoba, Winnipeg, Canada.
14. Interior Health Stroke Network, Division of Neurology, Kelowna, British Columbia, Canada
15. Hurvitz Brain Sciences Program, Sunnybrook Health Sciences Centre, Department of Medicine (Division of Neurology), University of Toronto, Toronto, Canada
16. Departments of Community Health Sciences & Clinical Neurosciences, the Hotchkiss Brain Institute, University of Calgary Cumming School of Medicine, Calgary, Canada
17. Calgary Stroke Program, Departments of Clinical Neurosciences, Radiology and Community Health Sciences, the Hotchkiss Brain Institute, University of Calgary Cumming School of Medicine, Calgary, Canada
18. Departments of Biomedical Engineering, Electrical and Software Engineering, and Radiology, the Hotchkiss Brain Institute, Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Canada
19. Calgary Stroke Program, Departments of Clinical Neurosciences and Community Health Sciences, the Hotchkiss Brain Institute and the O'Brien Institute for Public Health, University of Calgary Cumming School of Medicine, Calgary, Canada

KEYWORDS: CLINICAL TRIALS, TRIAL ENROLLMENT, STROKE, ISCHEMIC STROKE, MATCHING ALGORITHM

Corresponding Author: Aravind Ganesh, MD DPhil(Oxon) FRCPC, Calgary Stroke Program, University of Calgary Cumming School of Medicine, HMRB Room 103, 3280 Hospital Drive NW, Calgary, AB T2N 4Z6, aganesh@ucalgary.ca Phone: (403) 220-3747 Fax: (403) 210-8113; Twitter: @draravindganesh

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

ABSTRACT

Background: Clinical trials often struggle to recruit enough participants, with only 10% of eligible patients enrolling. This is concerning for conditions like stroke, where timely decision-making is crucial. Frontline clinicians typically screen patients manually, but this approach can be overwhelming and lead to many eligible patients being overlooked.

Methods: To address the problem of efficient and inclusive screening for trials, we developed a matching algorithm using imaging and clinical variables gathered as part of the AcT trial (NCT03889249) to automatically screen patients by matching these variables with the trials' inclusion and exclusion criteria using rule-based logic. We then used the algorithm to identify patients who could have been enrolled in six trials: EASI-TOC (NCT04261478), CATIS-ICAD (NCT04142125), CONVINCE (NCT02898610), TEMPO-2 (NCT02398656), ESCAPE-MEVO (NCT05151172), and ENDOLOW (NCT04167527). To evaluate our algorithm, we compared our findings to the number of enrollments achieved without using a matching algorithm. The algorithm's performance was validated by comparing results with ground truth from a manual review of two clinicians. The algorithm's ability to reduce screening time was assessed by comparing it with the average time used by study clinicians.

Results: The algorithm identified more potentially eligible study candidates than the number of participants enrolled. It also showed over 90% sensitivity and specificity for all trials, and reducing screening time by over 100-fold.

Conclusions: Automated matching algorithms can help clinicians quickly identify eligible patients and reduce resources needed for enrolment. Additionally, the algorithm can be modified for use in other trials and diseases.

Highlights

- Clinical trials currently require manually intensive screening to find potentially eligible patients
- An automatic matching algorithm using imaging and clinical variables could quickly and accurately list eligible trials for 1,577 individual acute stroke patients
- This algorithm can be adapted to other diseases and integrated with imaging and health records data extraction modules for full automation.

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

Introduction

In recent years, significant advancements in healthcare research have been driven by emerging technologies, innovative methodologies, and the results of randomized clinical trials[1–4]. These developments have the potential to improve healthcare practices and patient outcomes. Patients who are admitted to hospitals that participate in clinical trials receive better care and have a lower mortality rate[5].

Clinical trials are generally designed to find an alternative treatment that will be superior to standard care. A higher rate of participant enrollment in clinical trials could result in faster medical advancement, which in the long term leads to better care and outcomes for the general population[6]. However, many clinical trials struggle to meet their enrollment goals[7–10]. A hospital may participate in many trials simultaneously, and it is often impractical for physicians to be aware of the inclusion and exclusion criteria for every trial enrolling patients at their hospital[11].

Stroke is an acute disease and a time-sensitive emergency. It is one of the leading causes of mortality, and 30-40% of survivors are disabled[6]. Rapid screening and identification of eligible patients is the key to efficient trial recruitment for acute stroke. Currently, acute stroke clinical trial recruitment is managed by physicians and research personnel who screen patients on a per-trial basis, most often using a manual approach that is time-consuming and complex. Physicians are appropriately focused on delivering patient care and may overlook eligibility for ongoing trials. Hiring research personnel to manually screen patients is expensive, and they may not have direct access to patients in clinics or the emergency room. In addition, some jurisdictions have limited specialists and knowledge about ongoing trials, and most of those who know about the trials are in larger urban medical research centers[12]. This is also a common issue among clinicians who do not engage in research studies and clinician-scientists.

From an equity lens, the cognitive biases of physicians may prevent many eligible patients from being enrolled in acute trials, with the consequence that women, older, Indigenous persons, and other ethnic minorities are underrepresented[13,14]. Such inequity also contributes to slower medical advancement through missed enrolment opportunities and enrolment of a study population that may not represent those affected by the disease in the general population[14,15].

This proof-of-concept study aimed to develop a matching algorithm using imaging and clinical variables to automatically screen patients by matching these variables with the inclusion

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

and exclusion criteria of the trials. The algorithm has been designed to incorporate advanced AI capabilities like image auto-interpretation and smart notifications. These tools will work together seamlessly to create an efficient and streamlined automatic recruiting process. We hypothesized that the number of potentially eligible patients identified using a matching algorithm would be higher than the number of patients who were enrolled by conventional recruitment methods and that the algorithm would achieve high accuracy in identifying eligible patients compared to expert clinical researchers.

Methods

Patient Data

We used imaging and clinical variables gathered as part of the AcT trial (NCT03889249, Alteplase Compared to Tenecteplase in Patients with Acute Ischemic Stroke). The ACT trial was an investigator-initiated, phase 3, pragmatic, multicenter, open-label, registry-linked, randomized, controlled, non-inferiority trial, with blinded end-point assessment (PROBE), comparing tenecteplase to alteplase in patients presenting with acute ischemic stroke[16]. Inclusion and exclusion criteria were informed by the Canadian Stroke Best Practice Recommendations (CSBPR 2018)[17] and are published elsewhere[18]. The trial used deferred consent procedures, details of which have already been published. Reuse of data for design and development of the algorithms was approved by the Conjoint Health Research Ethics Board of the University of Calgary (REB22-0592). Data have been disclosed to only researchers and clinicians involved in this study. The sample size was one of convenience, making use of all available data from the AcT dataset.

The data were collected from December 2019 to January 2022 from 1,577 patients. Available features included demographic, medical history, clinical and imaging data (with baseline imaging consisting of computed tomography (CT) and computed tomography angiography (CTA)). This dataset was selected to test our algorithms for two key reasons: 1) AcT had comprehensive characterization of patients with key clinical, imaging, and demographic variables, and 2) AcT was a pragmatic trial and therefore reflected patients with acute ischemic stroke seen in routine practice (with the notable exception that all the AcT patients had to be eligible for thrombolysis).

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

The key baseline and imaging characteristics of patients in the AcT dataset are given in Table 1.

Clinical Trials

We developed our matching algorithm to identify patients in the AcT dataset who would potentially be eligible for six exemplar stroke trials, including a variety of ischemic stroke mechanisms and intervention strategies: EASI-TOC (NCT04261478, Endovascular Acute Stroke Intervention - Tandem Occlusion Trial), CATIS-ICAD (NCT04142125, Combination Antithrombotic Treatment for Prevention of Recurrent Ischemic Stroke in Intracranial Atherosclerotic Disease), CONVINCE (NCT02898610, Colchicine for Prevention of Vascular Inflammation in Non-cardioembolic Stroke), TEMPO-2[19] (NCT02398656, A Randomized Controlled Trial of TNK-tPA Versus Standard of Care for Minor Ischemic Stroke With Proven Occlusion), ESCAPE-MEVO (NCT05151172, EndovaSCular TreAtment to imProve outcomEs for Medium Vessel Occlusions), and ENDOLOW (NCT04167527, Endovascular Therapy for Low NIHSS Ischemic Strokes). The first three – EASI-TOC, CATIS-ICAD, and CONVINCE – were used as proof-of-concept as these three trials were ongoing at the time of the AcT trial and permitted patients in the AcT trial to be co-enrolled, as was the case for EASI-TOC, or to be enrolled after the 90-day follow-up for AcT was completed, as with CATIS-ICAD and CONVINCE (Table 2). The last three – TEMPO-2, ESCAPE-MEVO, and ENDOLOW – were used to evaluate the capability of expanding the algorithm to trials that were currently enrolling patients but for which the AcT population could not, in fact, have been co-enrolled.

Matching Algorithm

The study clinicians started the pipeline by simplifying and adapting the original clinical trial inclusion and exclusion criteria to align with the available features in the dataset. Then, the algorithms were developed based on a rule-based method, which manually added all criteria using a cascade of if-else statements. The code was developed on Python, a widely used high-level programming language known for its simplicity and power in the data science field. The patient's clinical features, collected by the AcT research team when they were presented at the hospital, were used as input variables. The complete criteria can be reviewed from the

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

registrations published on clinicaltrials.gov. The validation was conducted by comparing with a manual screening on a subsample, as explained in the following section.

Evaluation

We used the matching algorithms to identify potentially eligible patients who could have been enrolled in these six trials: EASI-TOC (NCT04261478), CATIS-ICAD (NCT04142125), CONVINCE (NCT02898610), TEMPO-2 (NCT02398656), ESCAPE-MEVO (NCT05151172), ENDOLOW (NCT04167527). We then compared eligible patients identified by the matching algorithms to the number of enrollments from the AcT population that had been achieved in the three trials (EASI-TOC, CATIS-ICAD, CONVINCE), that allowed co-enrolment with AcT. The algorithm's performance was also validated by having study clinicians manually screen a 10% validation set, which rounded up to 200 patients from AcT, for eligibility into each of the six trials while blinded to the algorithm's results. The validation set was weighted more toward the patient group that was evaluated by the algorithm as not being eligible for any trial, as we wanted to specifically evaluate the risk of false negative classification by the algorithm, which is crucial to mitigate when deploying such an algorithm for screening patients for ongoing trials. The validation set therefore included 100 patients who were screened by the algorithm as not eligible for any trial. Another half were the patients eligible for 1 to 5 trials. Specifically, there were 50, 35, 10, 3, and 2 individuals for those deemed eligible by the algorithm to be eligible in 1, 2, 3, 4, and 5 trials, respectively. These numbers approximately represent 51%, 36%, 9%, 3%, and 1% of all eligible patients in each group, and we ensured that the characteristics of the validation group were representative of the entire dataset. The first study clinician (AG) reviewed the neuroimaging scans and available clinical data for every patient on the list, indicating which of the six trials (if any) each patient was potentially eligible for enrolment. This clinician was blinded to the algorithm's results, and the matching algorithm, of course, did not have access to the clinician's impression. In the spirit of efficiency, discrepancies between the first physician and the algorithm were adjudicated based on screening by a second clinician (NS) who was also blinded to the algorithm's output. Then, the combination of screening results from the first clinician and adjudicated results on the discrepancy list from the second clinician were used as ground truth to determine the performance metrics of the algorithm. Lastly, we calculated sensitivity, specificity, PPV, NPV, and accuracy.

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

Results

Figure 1 shows the number of potentially eligible patients identified by the algorithm after filtering by each criterion for all six trials. The range of potentially eligible patients varied from 51 in the ENDOLOW trial to 1,090 in the CONVINCE trial, as shown in Table 3 under the trial's name. For patients with missing data in critical features, any trial that required those features was excluded from the eligible list. However, the name of the trial and the missing features were shown in the algorithm's remarks to let the clinician know that it was possible to enroll if the criteria were met. The proportion of patients with missing data in key criteria in each trial ranged from 0.4% to 6.2%. For missing data in optional criteria, trials with missing data were still eligible to appear in the list with remarks indicating the missing features. Imputation methods were not applied to the algorithm to avoid altering the screening results.

The distribution of potentially eligible patients in each trial, compared to the entire patient population in the original dataset, is shown in Table 3. The median age range of participants in each trial was between 69.5 and 79, compared to 74 in the entire population. The distribution of sex was mostly balanced, with roughly equal numbers of male and female patients, except for the EASI-TOC trial, which had a higher proportion of female potentially eligible participants at 69.3%.

The algorithm results were compared with the actual number of enrollments achieved without utilizing the algorithm in the three proof-of-concept trials that allowed enrolment during the AcT trial study period: EASI-TOC, CATIS-ICAD, and CONVINCE. A summary of the comparison of enrollment rates is presented in Figure 2. The number of patients actually recruited was observed to be considerably lower when compared to the total number of patients who were identified by the algorithm, showing a more than 25 to 90-fold difference in all trials. In particular, the CONVINCE trial had only 12 patients who were actually recruited from the AcT sample, but 1,090 were identified as potentially eligible by the algorithm.

Comparing the time used between the algorithms and manual screening by the clinician, the algorithms could complete the screening process for all six trials in 2.14 seconds per patient, and it took 2.83 seconds for 200 patients in the validation set. In contrast, the study clinician spent more than 140 times longer to evaluate. The screening required at least 5 minutes per patient, and it took about 17 hours to complete the validation set of 200 patients.

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

The percent agreement between results from the algorithm and study clinicians (ground truth) are shown in Table 4. The results showed that the algorithm was highly accurate, achieving over 90% for all performance metrics in all trials except for some metrics in CATIS-ICAD and ESCAPE-MEVO. This implies that the algorithms generated only a few false positives and false negatives in most trials. CATIS-ICAD and ESCAPE-MEVO had a slightly higher number of false positives because of important limitations in real-world clinical aspects of patient selection for those studies, resulting in a lower positive predictive value (PPV) rate at 68% and 75%, respectively.

Discussion

In this proof-of-concept study, we developed algorithms to automatically match patients in an acute ischemic stroke dataset to six different clinical trials based on clinical and imaging features. The study solely compared the results of the algorithm with manual screening of a subset and with the actual number of enrollments because our team did not have any other available automated tools available to us in our routine practice. We opted not to use other non-rule-based algorithmic techniques for developing our automated screening technique because we wanted to ensure that the rules used by the algorithm were easily explainable and not subject to unanticipated distortions through ‘black-box’ AI methods. The trials had varying inclusion and exclusion criteria, resulting in different numbers of eligible patients. The CONVINCE trial had the most eligible patients due to its broad criteria, while EASI-TOC had stricter criteria, resulting in fewer eligible patients. CATIS-ICAD's requirement for specific ICAD locations further reduced eligible numbers. Although TEMPO-2, ESCAPE-MEVO, and ENDOLOW had similar criteria, ESCAPE-MEVO had more eligible patients due to focusing on those with NIHSS scores of 3 or higher. According to the performance metrics, the algorithm performed well in all aspects except the PPV in the ESCAPE-MEVO and CATIS-ICAD trials. We designed the algorithms by weighing more on the impact of false negatives, which resulted in a high NPV that was higher than 95% in all trials. For PPV, it was low in the ESCAPE-MEVO because the human readers also considered the technical feasibility of thrombectomy for the given patient’s neurovascular anatomy and specific clot location, which the algorithm could not evaluate. For CATIS-ICAD, when the readers reviewed the data alongside the imaging, they might have overlooked certain

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

vessels that had a less clinically significant burden of ICAD and also appeared to be more selective when considering the affected area of ICAD.

The algorithm could significantly reduce the time required for screening patients. Most of the algorithm's time was spent on initializing the software package and importing data, which is illustrated by a small difference in time used between 1 and 200 patients. Therefore, increasing the number of patients or clinical trials did not substantially influence the algorithm's run time. However, the impact of time-effectiveness depends on where the algorithms are implemented; in acute trials, screening case by case with a limited number of trials could significantly differ from screening in a large database in prevention trials. Additionally, these results relied on the assumption that all necessary data is accessible to the clinician, and the algorithms used the processed structural data. In real situations, several factors affecting screening time need to be considered, such as the time required to obtain information from the patient and the waiting time for imaging acquisition and interpretation. Addressing these aspects will be vital for future evaluations.

Another important consideration with such algorithms is their potential cost-effectiveness. Figure 3 compares the estimated cost of hiring researchers and clinicians with the cost of running the algorithm. This estimation was based on the time that our study clinicians used when screening the validation set, which was approximately 50 seconds per trial per patient. Hiring a research associate and a clinician to screen patients can cost around CAD$30/hour and CAD$200/hour. This can be contrasted with the cost of using automated algorithms like ours, which are expected to cost less than CAD$1.5/hour (based on the virtual machine price from the Google Cloud Compute Engine), with a running time less than 5 seconds for the entire dataset. Therefore, the cost of human raters quickly rises as the number of trials and patients increases while that of the automated algorithm remains the same. That being said, this comparison does not account for the fact that clinical staff would still need to prepare data for the algorithm, confirm eligibility to enrol, and approach the patient for enrollment; as such, prospective evaluation of the algorithm is needed to more formally evaluate its cost-effectiveness.

However, the algorithm was fast and accurate, comparable to experienced human screeners. In addition, the algorithm itself would not introduce biases because the screening relied only on each trial’s criteria. As shown in Table 3, there was no significant selection bias

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

regarding patient characteristics such as age, sex, weight, and time from onset to randomization. Moreover, since the screening algorithm does not require clinicians to actively consider each trial for a given patient, it can potentially mitigate cognitive biases from clinicians that arise in manual screening processes.

Previous studies have developed algorithms or software to match patients with clinical trials automatically[20–31]. Many studies used rule-based logic with inclusion and exclusion criteria, but some recent studies have tried to incorporate machine learning in the matching process. Penberthy and Kamal conducted studies aiming to use healthcare institute data and systems to design adaptable rule-based software for various diseases[20,21]. Their research focused on improving the screening time and increasing the enrollment of potentially eligible patients, but it did not mention the accuracy of their method. The studies conducted by Lucila and Musan were focused on AIDS and cancer[22,23]. Both researchers used logical rules and Bayesian networks to match patients and suggest additional data for informed decisions. Recent studies aimed to develop matching algorithms focused on extracting clinical variables from patient records. Hassanzadeh and Chen used natural language processing (NLP) and Medical Knowledge, respectively, to extract clinical variables from the records and then trained a deep learning model to match patients with trials[24,25]. Their study was based on the National NLP Clinical Challenges (N2C2) data and attempted to match the extracted variables with preset eligibility criteria, which was not a real-world trial. There are existing methods to extract clinical variables of patients based on oncology and use rule-based logic to match them with clinical trials[26,27]. Yuan and Ni proposed to matching both clinical variables and trial criteria from raw data[28,29]. Yuan also focused on stroke clinical trials, and their study yielded a sensitivity range of 0.41-0.98 for six trials. Kaskovich and colleagues used NLP to automatically extract inclusion and exclusion criteria from raw data of 216 leukemia-associated trials. The approach was to input patients' data to match with those trials[30]. However, during the N2C2 shared task, the rule-based method had the highest performance, and four of the top ten systems were rule-based[24,31]. Our research focused on matching structured data to specific criteria and applying this to clinical trial recruitment. Rather than the accuracy of the matching algorithm, the rule-based method was chosen for the reasons of its simplicity for maintenance and expansion. Unlike “black box” machine learning methods, the logic behind a matching algorithm is interpretable and easily understood, and it does not require retraining. Additional criteria could be added or removed

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

from the cascade statement for each trial. Adding other trials is as simple as adding another cascade of statements to the list. Moreover, implementing it as a Python software package makes it easy to add any future modules. This approach is valuable for increasing enrollment in stroke trials and simplifying the enrollment process in smaller healthcare settings. Other areas of acute care, like cardiac failure, could also benefit from this approach.

Automatic matching algorithms could mitigate critical limitations in current recruitment methods by quickly identifying eligible patients, allowing clinicians to focus on quality care. This approach reduces screening costs for hospitals and research centers, and benefits patients by considering them for appropriate treatment trials. The algorithms do not require a high-performance computing system due to the simplicity of a rule-based algorithm, even when used on a larger scope. Therefore, it is suitable to be implemented in remote areas. Combining the algorithm with advanced notification systems can help mitigate the shortage of specialized clinicians in rural areas by sending screening results to nearby specialists for timely care[4]. Commercial applications have used similar notification systems for stroke cases to speed up enrollment. These algorithms could be applied to other stroke trials or diseases and potentially improve the representation of underrepresented populations, but this remains to be demonstrated. However, when considering applying these algorithms to other trials and diseases, there are some challenges in adapting the original trial criteria to the nature of the available structural data, which requires collaboration between the technical and clinical teams in a given healthcare system.

Importantly, there are some limitations of the proposed method. First, some of the exclusion criteria for the trials, such as baseline pre-morbid function (e.g., pre-stroke modified Rankin Scale) and alternative stroke etiologies (e.g., atrial fibrillation for CATIS-ICAD), were not gathered in the AcT dataset, meaning that an unknown proportion of the patients flagged as eligible for the trials by our algorithm would likely be ultimately excluded from participation. This was especially the case for CONVINCE, which had several specific comorbidity- and medication tolerance-related exclusionary criteria that were simply unavailable in the routinely gathered clinical and imaging data in AcT. The study was conducted using only one dataset, which might not be reflective of the general stroke population. The high level of data completeness in the AcT randomized-controlled trial dataset does not reflect the missingness that is inevitable in routine clinical data. Therefore, our future plan involves utilizing datasets from

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

multiple sources to validate the generalizability and effectiveness of the algorithms. Missing data and features could hinder the real-world performance of the algorithm by reducing the number of potentially eligible patients. Even in the manual screening process, clinicians cannot decide whether to enroll patients if relevant data are missing. The list of missing data variables for specific trials shown in the algorithm's outputted remarks will nevertheless help alert clinicians to fill in remaining criteria to complete screening for otherwise potentially eligible patients.

Second, in some clinical trials, more nuanced clinical interpretation is required to determine whether a patient is eligible to participate. For instance, in the CATIS-ICAD trial, the treating physician would need to establish whether they consider the patient's ICAD (flagged by the algorithm) to be symptomatic or not. The absence of this information can lead to a lower algorithm performance. The data from the AcT trial was extracted from data available in an electronic data capture system. Real-world data is often a combination of free text, notes, and a wide variety of other data formats. In addition, imaging variables from CTA that were crucial selection criteria for these trials need to be gathered by specialized physicians; the AcT trial dataset benefited from a detailed review of key imaging features by study readers. In practice, this could lead to a delay in the availability of key information for the algorithm. By integrating with EMRs at the point-of-care, we can greatly enhance the utility of this approach. It will allow us to take advantage of real-time data entry, resulting in more efficient data collection. However, human interpretation of medical imaging could potentially confound the algorithm's performance due to reader biases. The same image can be interpreted differently by different readers, which might lead to misleading results. Another confounding factor could be the variations in data quality in different sites where the algorithm is deployed. Some variations directly impact the quality and homogeneity of the data, such as the protocol and image processing method, which can cause variations in assessment of certain stroke characteristics such as infarct core volume estimation.

Third, obtaining ethical permission to run a screening algorithm through patients' electronic medical records (EMRs) and imaging can be a potential challenge. For enrolling patients in clinical trials, both consent to use their data and participation in the trial are crucial. While having a higher number of eligible candidates may seem like it would lead to more people consenting to participate, this may not always be the case. In reality, only a proportion of eligible candidates will actually agree to participate[32]. This can be due to various reasons, such as a lack

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

of interest, concerns about side effects, or a desire for certainty in receiving a particular intervention. In real-life scenarios, clinicians are authorized to access the EMRs of patients they care for, evaluate which clinical trial would be appropriate for the patient, and then seek the patient's consent to participate. Rather than involving clinicians in the initial screening process, the algorithm directly reviews the EMR and generates a list of eligible trials. Then, clinicians are responsible for selecting a trial and obtaining consent from the patient before enrollment. Since the algorithm is technically not directly involved in the patient's care, privacy concerns may therefore be potentially raised about its use of patient data. Therefore, it is important to consider potential regulatory or data access barriers which may vary from one healthcare system to another, and develop strategies to overcome them. For example, the algorithm may require pre-approval of patients to access their EMRs to screen for trial participation. This could be facilitated by implementing patient-directed communication strategies and offering patients the option to provide advance consent for their data to be used for such screening purposes in medical emergencies when interacting with their family doctors or otherwise sharing information with health systems. These steps can increase the number of trial enrollments while still respecting patient privacy. However, given that it is a matching algorithm, the patient data can remain local to the site and does not need to be stored or transmitted, easing some of these concerns.

In the future, we envision this solution ideally being paired with other modules to achieve complete automation and mitigate human error and biases. Automated imaging analysis and data extraction algorithms for relevant clinical variables from electronic health records are important upstream modules that are increasingly being adopted by hospital systems worldwide to identify patients eligible for therapy, automatically gathering variables such as age, medications, occlusion presence/location and extent of ischemic changes as examples. Once the matching algorithm generates screening results, a smart notification system can be integrated into the smartphone system. This could notify either the attending physicians or research staff in the coverage area of any positive trial eligibility, ideally without interfering with patient care processes. This system can prompt them to take appropriate actions in terms of further evaluating and consenting the patient or a proxy decision-maker for the trial, leading to a timely and accurate screening process. Deferral of consent and advance consent processes could help facilitate the automatic flow of the entire process since patients are often incapacitated and may

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

not be accompanied by a proxy decision-maker[33–36]. Future studies should aim to evaluate trial enrolment rates achieved with such screening algorithms in the real world, before and after implementation, and across multiple sites. Our future work will include implementing and validating the algorithm at different stroke centers for point-of-care use. In particular, we plan to adapt the algorithm to guide patient selection for different domains of an upcoming platform trial for acute ischemic stroke, using trial-related checklists to capture relevant characteristics. However, this initial offline research was crucial to justify this novel enrolment method for future ethics applications.

Conclusion

We found that automated trial matching algorithms achieved fast and accurate performance in identifying patients eligible for six different stroke trials. Overall, this research has the potential to significantly improve clinical trial recruitment and thereby help accelerate the development of new treatments for time-sensitive diseases like stroke. Mitigating cognitive biases and ensuring equitable access to clinical trials are important benefits of these innovative strategies.

Sources of Funding

The project was funded by the Government of Canada's New Frontiers in Research Fund.

Competing Interests Disclosure

TF received personal fees from the Canadian Medical Protective Agency and plaintiff. TF participated as an advisory board and received personal fees from Bayer Canada, Novartis, HLS Therapeutics, and AstraZeneca. TF also had a role in the VGH/UBC Hospital Foundation Board and DESTINE Health. BM holds a stock in Circle CVI. MS received grants from the Canadian Institutes of Health Research (CIHR) for the iCATCHER trial and ACTION feasibility study, and from the New Frontiers in Research Fund (NFRF) for the ACTION feasibility study. Other authors have no conflict of interest to disclose.

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

Authors contributions

PC: Designing the algorithm, performing data analysis and interpretation, drafting and revising the manuscript. AG, MEM: Study conception, supervision, interpretation, revising the manuscript. NS: Performing data analysis and interpretation, revising the manuscript. Remaining authors: Obtaining data and revising the manuscript.

References

1. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2:230-243.

2. Junaid SB, Imam AA, Balogun AO, et al. Recent Advancements in Emerging Technologies for Healthcare Management Systems: A Survey. *Healthcare*. 2022;10:1940.

3. Shaheen MY. Applications of Artificial Intelligence (AI) in healthcare: A review. Published online September 25, 2021.

4. Broderick JP, Silva GS, Selim M, et al. Enhancing Enrollment in Acute Stroke Trials: Current State and Consensus Recommendations. *Stroke*. 2023;54:2698-2707.

5. Majumdar SR, Roe MT, Peterson ED, Chen AY, Gibler WB, Armstrong PW. Better Outcomes for Patients Treated at Hospitals That Participate in Clinical Trials. *Arch Intern Med*. 2008;168:657-662.

6. Virani SS, Alonso A, Benjamin EJ, et al. Heart Disease and Stroke Statistics—2020 Update: A Report From the American Heart Association. *Circulation*. 2020;141:e139-e596.

7. Fogel D. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp Clin Trials Commun*. 2018;11.

8. Bower P, Wallace P, Ward E, et al. Improving recruitment to health research in primary care. *Fam Pract*. 2009;26:391-397.

9. Simes RJ. Clinical trials and “real-world” medicine. *Med J Aust*. 2002;177:407-411.

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

10. Avis NE, Smith KW, Link CL, Hortobagyi GN, Rivera E. Factors Associated With Participation in Breast Cancer Treatment Clinical Trials. *J Clin Oncol*. 2006;24:1860-1867.

11. Dugas M, Lange M, Berdel WE, Müller-Tidow C. Workflow to improve patient recruitment for clinical trials within hospital information systems – a case-study. *Trials*. 2008;9:2.

12. Harrington RA, Califf RM, Balamurugan A, et al. Call to Action: Rural Health: A Presidential Advisory From the American Heart Association and American Stroke Association. *Circulation*. 2020;141:e615-e644.

13. Murthy VH, Krumholz HM, Gross CP. Participation in Cancer Clinical TrialsRace-, Sex-, and Age-Based Disparities. *JAMA*. 2004;291:2720-2726.

14. Carcel C, Harris K, Peters SAE, et al. Representation of Women in Stroke Clinical Trials: A Review of 281 Trials Involving More Than 500,000 Participants. *Neurology*. 2021;97.

15. Zhu JW, Le N, Wei S, et al. Global representation of heart failure clinical trial leaders, collaborators, and enrolled participants: a bibliometric review 2000–20. *Eur Heart J - Qual Care Clin Outcomes*. 2022;8:659-669.

16. Menon BK, Buck BH, Singh N, et al. Intravenous tenecteplase compared with alteplase for acute ischaemic stroke in Canada (AcT): a pragmatic, multicentre, open-label, registry-linked, randomised, controlled, non-inferiority trial. *The Lancet*. 2022;400:161-169.

17. Boulanger J, Lindsay M, Gubitz G, et al. Canadian Stroke Best Practice Recommendations for Acute Stroke Management: *Prehospital, Emergency Department, and Acute Inpatient Stroke Care, 6th Edition, Update 2018*. *Int J Stroke*. 2018;13:949-984.

18. Sajobi T, Singh N, Almekhlafi MA, et al. AcT Trial: Protocol for a Pragmatic Registry-Linked Randomized Clinical Trial. *Stroke Vasc Interv Neurol*. 2022;2:e000447.

19. Coutts SB, Ankolekar S, Appireddy R, et al. Tenecteplase versus standard of care for minor ischaemic stroke with proven occlusion (TEMPO-2): a randomised, open label, phase 3 superiority trial. *The Lancet*. doi:10.1016/S0140-6736(24)00921-8

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

20. Penberthy L, Brown R, Puma F, Dahman B. Automated matching software for clinical trials eligibility: Measuring efficiency and flexibility. *Contemp Clin Trials*. 2010;31:207-217.

21. J. Kamal, K. P, et el. USING AN INFORMATION WAREHOUSE TO SCREEN PATIENTS FOR CLINICAL TRIALS: A PROTOTYPE. In: *AMIA Annu Symp Proc 2005*. Vol 2005. ; 2005:1004.

22. Ohno-Machado L, Parra E, Henry SB, Tu SW, Musen MA. AIDS2: a decision-support tool for decreasing physicians' uncertainty regarding patient eligibility for HIV treatment protocols. *Proc Symp Comput Appl Med Care*. Published online 1993:429-433.

23. Musen MA, Tu SW, Das AK, Shahar Y. EON: A Component-Based Approach to Automation of Protocol-Directed Therapy. *J Am Med Inform Assoc*. 1996;3:367-388.

24. Hassanzadeh H, Karimi S, Nguyen A. Matching patients to clinical trials using semantically enriched document representation. *J Biomed Inform*. 2020;105:103406.

25. Chen CJ, Warikoo N, Chang YC, Chen JH, Hsu WL. Medical knowledge infused convolutional neural networks for cohort selection in clinical trials. *J Am Med Inform Assoc*. 2019;26:1227-1236.

26. Patel C, Cimino JJ, Dolby JT, et al. Matching Patient Records to Clinical Trials Using Ontologies. In: *ISWC/ASWC*. ; 2007.

27. Johnson T, Liebner D, Chen JL. Opportunities for Patient Matching Algorithms to Improve Patient Care in Oncology. *JCO Clin Cancer Inform*. 2017:1-8.

28. 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*. 2015;15:28.

29. Yuan J, Tang R, Jiang X, Hu X. Large Language Models for Healthcare Data Augmentation: An Example on Patient-Trial Matching. Published online August 4, 2023. Accessed March 25, 2024. http://arxiv.org/abs/2303.16756

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

30. Kaskovich S, Wyatt KD, Oliwa T, et al. Automated Matching of Patients to Clinical Trials: A Patient-Centric Natural Language Processing Approach for Pediatric Leukemia. *JCO Clin Cancer Inform*. 2023:e2300009.

31. Stubbs A, Filannino M, Soysal E, Henry S, Uzuner Ö. Cohort selection for clinical trials: n2c2 2018 shared task track 1. *J Am Med Inform Assoc*. 2019;26:1163-1171.

32. O'Neill ZR, Deptuck HM, Quong L, et al. Who says "no" to participating in stroke clinical trials and why: an observational study from the Vancouver Stroke Program. *Trials*. 2019;20:313.

33. Niznick N, Lun R, Dewar B, Perry J, Dowlatshahi D, Shamy M. Advance consent for participation in randomised controlled trials for emergency conditions: a scoping review. *BMJ Open*. 2023;13:e066742.

34. Shamy M, Dewar B, Niznick N, Nicholls S, Dowlatshahi D. Advanced consent for acute stroke trials. *Lancet Neurol*. 2021;20:170.

35. Udoh U, Dewar B, Nicholls S, et al. Advance Consent in Acute Stroke Trials: Survey of Canadian Stroke Physicians. *Can J Neurol Sci J Can Sci Neurol*. 2024;51:122-125.

36. Goyal M, Ospel JM, Ganesh A, Marko M, Fisher M. Rethinking Consent for Stroke Trials in Time-Sensitive Situations: Insights From the COVID-19 Pandemic. *Stroke*. 2021;52:1527-1531.

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

537 **Table 1.** Key characteristics of the 1,577 patients in the AcT dataset.

Features	**Median(IQR), n(%)**
Age [years]	74(63-83)
Sex: female	822(52.12%)
Weight [kilograms.]	75(65-89)
Baseline NIHSS	9(6-16)
Time from onset to randomization [minutes]	122(87-179)
ASPECTS	9(8-10)
Occlusion site on baseline CT angiography (n=1,558*)	
Intracranial internal carotid artery	135(8.56%)
M1 segment MCA	237(15.21%)
M2 segment MCA	315(20.22%)
Other distal occlusions	268(17.20%)
Vertebrobasilar arterial system	64(4.11%)
Cervical internal carotid artery	26(1.67%)
No visible occlusions	513(32.93%)
ICAD exists (n=1,558*)	373(23.94%)
Follow-up affected territory (n=1,577)	
MCA	1091(69.18%)
ACA	84(5.33%)
PCA	130(8.24%)
Vertebral	99(6.28%)
Anterior choroidal	3(0.19%)
No evident affected territory	170(10.78%)

*19 patients did not have a baseline CT angiography

ACA = Anterior Cerebral Artery; ASPECT = Alberta Stroke Program Early CT Score; CT = Computed Tomography; ICAD = Intracranial Atherosclerotic Disease; IQR = Interquartile Range; MCA = Middle Cerebral Artery; NIHSS = National Institutes of Health Stroke Scale; PCA = Posterior Cerebral Artery.

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

538 **Table 2.** Summary of key details of the selected clinical trials.

	EASI-TOC	CATIS-ICAD	CONVINCE	TEMPO-2	ESCAPE-MEVO	ENDOLOW
Target patient population	485	115	3,154	1,274	530	175
Design	A multi-center, prospective, randomized, open-label, blinded endpoint (PROBE) controlled trial (1:1 allocation).	Randomized, open-label, blinded endpoint, pilot trial	A multi-center international, Prospective, Randomized Open-label, Blinded-Endpoint assessment (PROBE) controlled Phase 3 clinical trial	A Phase 3, prospective, randomized controlled, open-label with blinded outcome assessment (PROBE) controlled trial.	Multicenter, prospective, randomized, open- label study with blinded endpoint evaluation (PROBE design)	Phase 2/3, prospective, randomized, open-label, blinded-endpoint (PROBE) adaptive two-stage design trial
Intervention	Acute ICA stenting during the thrombectomy procedure versus intracranial thrombectomy alone without ICA stenting	Rivaroxaban 2.5mg bid and ASA 81mg od versus ASA 81mg od	Colchicine 0.5 mg/day and usual care versus usual care alone (antiplatelet, lipid-lowering, antihypertensive treatment, lifestyle advice)	TNK-tPA versus Standard of Care for minor ischemic stroke with proven occlusion	EVT with Solitaire group of intracranial stent-retriever devices as the first line approach and standard medical care, versus standard medical care	iMT using EmboTrap Revascularization Device versus iMM
Key inclusion criteria*	- Occlusion in the intracranial carotid, M1, or M2	- Age >= 40 years - ICAD 30-99% - Affected	- Age >= 40 years - Atherosclerosis of the carotid or	- Within 12 hours of last seen normal	- Within 12 hours of last seen normal	- NIHSS (baseline) <= 5 - CT/CTA

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

	- Within 24 hours of last seen normal - Ipsilateral (same side) >= 70% cervical ICA stenosis or occlusion	territory correspond to the ICAD	vertebral or MCA or ACA or basilar artery	- NIHSS (baseline) <=5 - Any acute intracranial (not cervical ICA) occlusion, or CTP focal perfusion abnormality	- NIHSS (baseline) >= 3 - Occlusion in M2, M3, A2, A3, P2, or P3	occlusion of ICA, M1, or M2 - ASPECTS >= 6 or core volume < 70cc
Key exclusion criteria*	- Ipsilateral ICA stenosis or occlusion attributable to clinically or radiologically confirmed arterial dissection - Isolated cervical carotid occlusion without intracranial occlusion * Not applicable to AcT: pregnancy * Not available in AcT: pre-stroke mRS≥3, contraindication to angioplasty/stenting	- Intracranial arterial occlusion (e.g. 100% stenosis) responsible for the acute brain ischemia - Intracranial arterial stenosis secondary to causes other than atherosclerosis. - Intraluminal thrombus * Not available in AcT: Not available in AcT: Other indication for longterm dual	* Not applicable to AcT: pregnancy * Not available in AcT: cardioembolic etiology, drug use, venous thrombosis, hyper-coagulability states, migraine, myopathy, blood dyscrasia, impaired hepatic function, use of CYP3A4 inhibitors, symptomatic peripheral neuropathy or	- ASPECTS < 7 - Core volume > 10cc - Intracranial hemorrhage - Chronic intracranial occlusion * Not applicable to AcT: severe fatal or disabling illness preventing 90-day follow-up, Pregnancy, Exclusions for thrombolysis * Not available	- ASPECTS <= 5 - Lack of core:penumbra mismatch (based on available imaging combination)- Intracranial hemorrhage * Not applicable to AcT: Pregnancy * Not available in AcT: nursing care needs, major comorbidity like dementia or cancer	- Time from last seen normal >=8 hours - NIHSS ≥6 - Intracranial hemorrhage - Multifocal ICAD * Not applicable to AcT: Pregnancy * Not available in AcT: investigator judgement of futile recanalization, pre-stroke mRS ≥3, seizures at onset, baseline glucose <50

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

	or antiplatelet therapy	antiplatelet or anticoagulation, plans for carotid revascularization, atrial fibrillation, subdural hematoma in prior 12 months, prior hemorrhagic stroke, pre-stroke mRS ≥4	neuromuscular disease, inflammatory bowel disease, dementia, active malignancy, hepatitis B/C, HIV, dysphagia, poor medication compliance, colchicine allergy or sensitivity	in AcT: In-hospital stroke		mg/dL or >400 mg/dL, platelets < 100,000/uL, creatinine > 3.0 mg/dL, suspected bacterial endocarditis, intubation, drug or alcohol dependence, incarceration, acute COVID-19

*The inclusion and exclusion criteria have been adapted to align with the clinical features in the AcT dataset.

ACA = Anterior Cerebral Artery; ASA = Acetylsalicylic Acid; ASPECT = Alberta Stroke Program Early CT Score; cc = cubic centimeter; CT = Computed Tomography; CTA = Computed Tomography Angiography; CTP = Computed Tomography Perfusion; EVT = Endovascular Thrombectomy; HIV = Human Immunodeficiency Virus; ICA = Internal Carotid Artery; ICAD = Intracranial Atherosclerotic Disease; iMM = Initial Medical Management; iMT = Immediate mechanical thrombectomy; IQR = Interquartile Range; MCA = Middle Cerebral Artery; mg = milligrams; mg/dL = milligrams per decilitre; NIHSS = National Institutes of Health Stroke Scale; TNK = Tenecteplase; tPA = Tissue Plasminogen Activator; uL = microliters.

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

546 **Table 3.** The summary statistics of potentially eligible patients identified by the algorithm for each trial compared with the entire AcT
547 population.

Features	Median (IQR), n (%)						
	AcT Dataset (n=1,577)	**EASI-TOC** (n=140*)	**CATIS-ICAD** (n=105*)	**CONVINCE** (n=1,090*)	**TEMPO-2** (n=191*)	**ESCAPE-MEVO** (n=544*)	**ENDOLOW** (n=51*)
Age [years]	74 (63-83)	69.5 (60-79)	79 (70-87)	76 (67-84)	72 (60-83)	76 (66-84)	75 (65.5-84)
Sex: female	822 (52.12%)	97 (69.3%)	53 (50.5%)	586 (53.8%)	114 (59.7%)	279 (51.3%)	29 (56.9%)
Weight [kilograms]	75 (65-89)	80 (66.8-89)	75 (67-90)	75 (65-88)	77 (66-90)	75 (64-86)	81 (68-90.5)
Baseline NIHSS	9 (6-16)	16 (10-20)	8 (5-14)	10 (6-17)	4 (3-5)	9 (6-16)	4 (3-5)
Time from onset to randomization [minutes]	122 (87-179)	117 (85-168)	145 (97-208)	124 (88-181)	140 (101-201)	117.5 (86-173)	140 (98.5-203)
ASPECTS	9 (8-10)	8 (7-10)	10 (9-10)	9 (8-10)	10 (9-10)	9 (8-10)	10 (8.5-10)

548 *Number of potentially eligible patients identified by the algorithm

549 ASPECT = Alberta Stroke Program Early CT score; IQR = Interquartile Range; NIHSS = National Institutes of Health Stroke Scale.

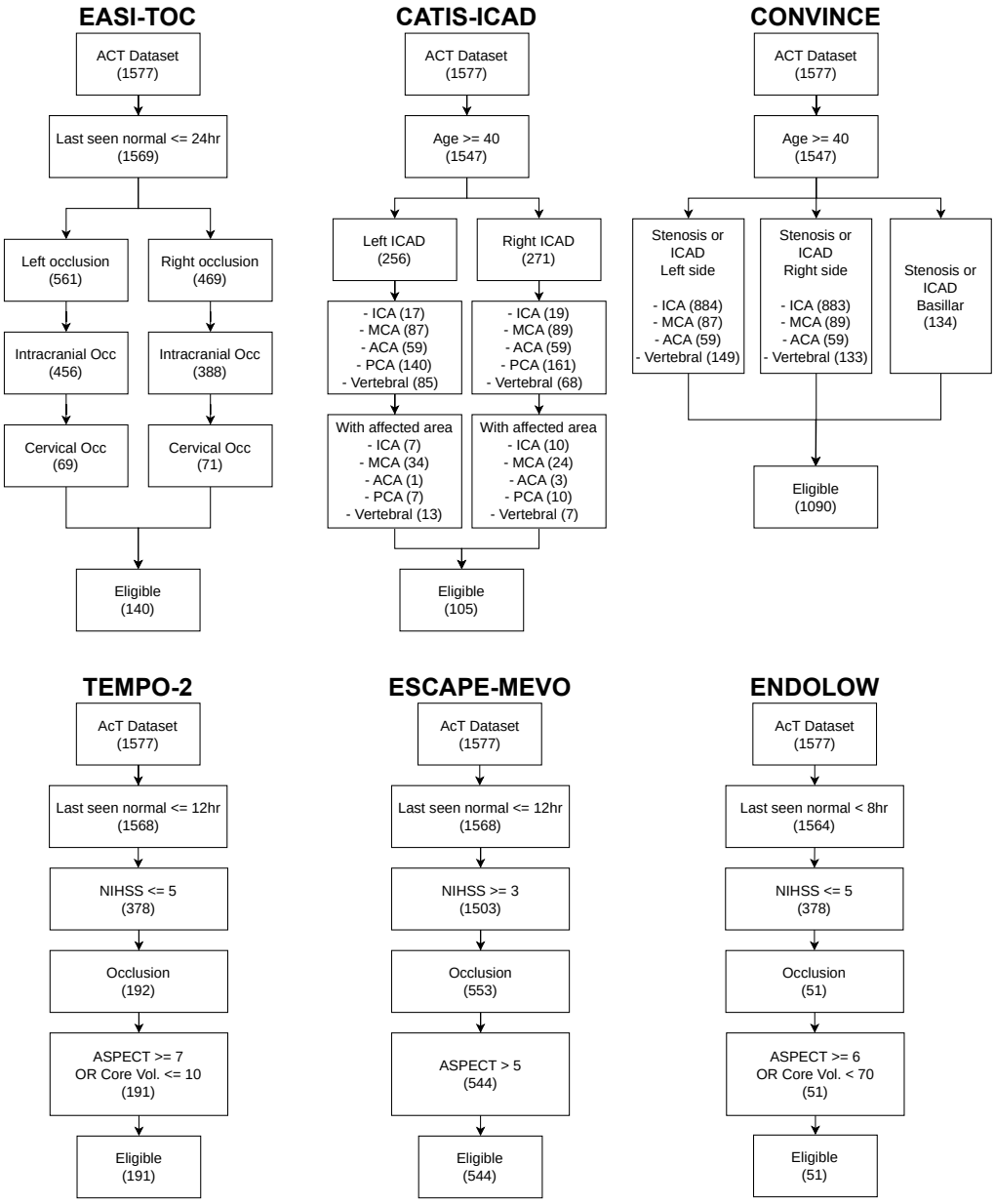
https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

Table 4. Classification performance metrics between the algorithm and ground truth.

Patient groups	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	Accuracy (95%CI)
EASI-TOC	0.92 (0.78-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	0.99 (0.98-1.00)	0.99 (0.99-1.00)
CATIS-ICAD	0.94 (0.82-1.00)	0.96 (0.93-0.99)	0.68 (0.49-0.88)	0.99 (0.98-1.00)	0.96 (0.93-0.99)
CONVINCE	0.93 (0.88-0.98)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	0.95 (0.91-0.99)	0.97 (0.95-0.99)
TEMPO-2	0.96 (0.89-1.00)	0.98 (0.96-1.00)	0.89 (0.78-1.00)	0.99 (0.983-1.00)	0.98 (0.96-1.00)
ESCAPE-MEVO	1.00 (1.00-1.00)	0.93 (0.89-0.97)	0.76 (0.64-0.88)	1.00 (1.00-1.00)	0.94 (0.91-0.97)
ENDOLOW	0.90 (0.71-1.00)	0.99 (0.98-1.00)	0.90 (0.71-1.00)	0.99 (0.98-1.00)	0.99 (0.98-1.00)

CI = Confidence Interval; NPV = Negative Prediction Value; PPV = Positive Predictive Value

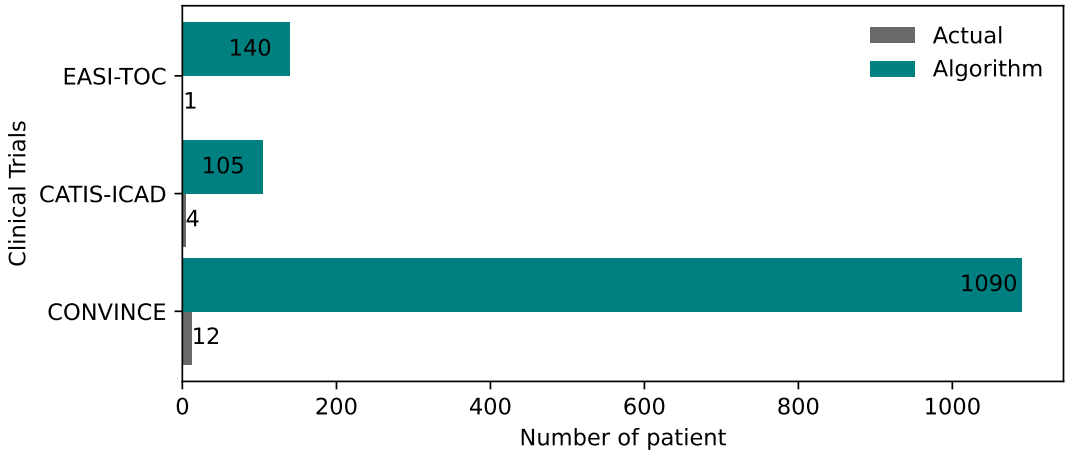
https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press



552

553 **Figure 1**. Potentially eligible patients identified for each trial according to key criteria used for
554 automatic matching. Each box displays the key criteria for inclusion and exclusion, along with
555 the number of potentially eligible patients up to that criterion in the blanket. ACA = Anterior
556 Cerebral Artery; ASPECT = Alberta Stroke Program Early CT Score; hr = hours; ICA = Internal
557 Carotid Artery; ICAD = Intracranial Atherosclerotic Disease; MCA = Middle Cerebral Artery;
558 NIHSS = National Institutes of Health Stroke Scale; PCA = Posterior Cerebral Artery; Vol. =
559 volume.

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press



560

561 **Figure 2**. Comparison of the total number of enrolled patients for each trial versus the number of

562 potential candidates identified by the algorithm.

563

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press

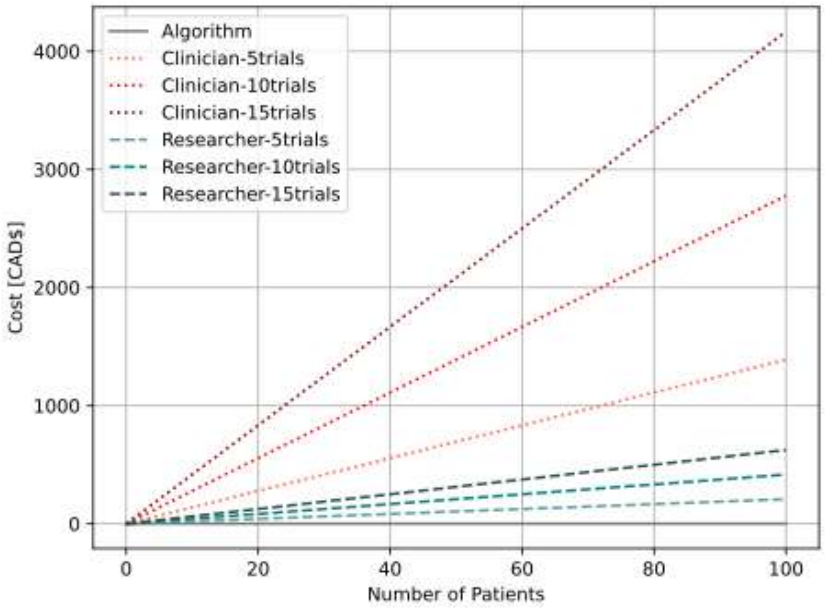


Figure 3. Cost comparison estimate between using a clinician or a research assistant versus our automatic algorithm for the trials screening process, using standard hourly rates and extrapolating from the comparative time data from our test sample.

https://doi.org/10.1017/cjn.2024.352 Published online by Cambridge University Press