

TOTTO RENNA

Kevin Hughes needed volunteers. It was 1994, and the breast-cancer surgeon was starting a randomized, controlled trial at Massachusetts General Hospital in Boston. He and his colleagues wanted to test the efficacy of a treatment regimen commonly followed by people with a certain type of early-stage breast cancer: surgery followed by the drug tamoxifen and radiation therapy. Despite being an established protocol, it wasn't clear whether the radiotherapy was beneficial for all women — and, in particular, those who were older.

The researchers sought volunteers over the age of 70 whose tumours were of a particular size and type. Of the roughly 40,000 women in the United States each year who could have qualified, they managed to enrol 636 people. That was enough for the study, but it took five years to find them.

Recruitment is just one of many bottlenecks in conducting clinical trials. “Medical research is remarkably inefficient in so many different ways,” says Eric Topol, director of the Scripps Research Translational Institute in La Jolla, California. An analysis of clinical-trial data from January 2000 up to April 2019 estimated that only around 12% of drug-development programmes ended in success¹ (see ‘The state of clinical trials’). Most clinical trials fail because they don't demonstrate the efficacy or safety of an intervention. Others flop because of a flawed study design, a shortage of money, participant drop-outs or a failure to recruit enough volunteers in the first place. Whether entering and transferring data or ensuring that participants take the correct dosage, delays, inaccuracies and inefficiencies abound.

To improve clinical trials, researchers in academia and the pharmaceutical industry are turning to artificial intelligence (AI). Fuelled by the rapidly increasing amounts of medical data that are available to researchers, including those provided by electronic health records (see page S114) and wearable devices, sophisticated machine-learning algorithms have the potential to save billions of dollars, to speed up medical advances and to expand access to experimental treatments. “Improving clinical trials would be a huge deal,” Hughes says.

UNDERSTANDING LANGUAGE

The trial led by Hughes was one of the successful ones². Although the extra step of radiotherapy reduced the rate of breast-cancer recurrence, it didn't affect the overall survival rate. For older women, at least, the added financial cost and risk of radiotherapy might outweigh the potential benefit. A follow-up study reached the same conclusion³. Had he and his colleagues found people faster, Hughes says, they might have arrived at their conclusions sooner — and then could have begun to better inform women earlier. It would have also enabled the researchers to move on to other burning questions.

The recruitment process is often the most time-consuming and expensive step of a trial.

CLINICAL STUDIES

Trial by artificial intelligence

A combination of big data and machine-learning algorithms could help to accelerate clinical testing.

BY MARCUS WOO

According to a 2016 study⁴, 18% of cancer trials that launched between 2000 and 2011 as part of the US National Cancer Institute's National Clinical Trials Network failed to find even half the number of patients they were seeking after three or more years of trying, or had closed entirely after signing up only a few volunteers. An estimated 20% of people with cancer are eligible to participate in such trials, but fewer than 5% do⁵. "Recruitment is the number one barrier to clinical research," says Chunhua Weng, a biomedical informaticist at Columbia University in New York City.

Many are hoping that AI can make a difference. One branch of AI, called natural language processing (NLP), enables computers to analyse the written and spoken word. When applied to medicine, such techniques could allow algorithms to search doctors' notes and pathology reports for people who would be eligible to participate in a given clinical trial.

The challenge is that the text in such documents is often free flowing and unstructured, and valuable information might only be implicit, requiring some background knowledge or context to understand. Doctors, for instance, have several ways of describing the same concept — a heart attack might be referred to as a myocardial infarction, a myocardial infarct or even just 'MI'. But an NLP algorithm can be trained to spot all such synonyms by exposure to sample medical records that have been annotated by researchers. The algorithm can then apply that knowledge to interpret unannotated records.

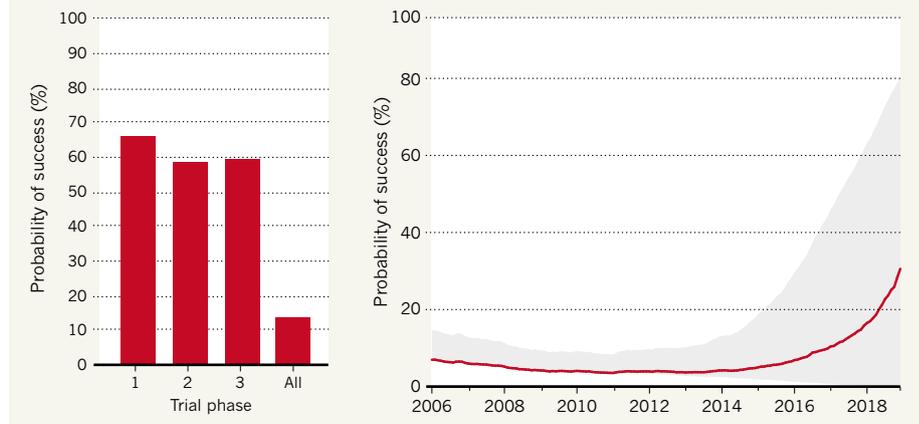
Efforts are being made to make it easier for computers to interpret the descriptions of clinical trials. The inclusion and exclusion criteria of trials are commonly written in plain text. So that hospitals can search patient databases for people who are eligible to take part, these criteria must first be translated into a standardized, coded query format that the database can understand. Weng and her colleagues built an open-source web tool called Criteria2Query that uses NLP to do just that — enabling researchers and administrators to search databases without needing to know a database query language⁶.

AI can also help patients to look for clinical trials by themselves. Typically, people rely on their doctors to inform them about suitable studies. Some patients search the website ClinicalTrials.gov, which lists more than 300,000 studies that are being conducted in the United States and 209 other countries. Daunting scale aside, the often highly technical eligibility criteria can be incomprehensible to the public. "It's pretty overwhelming," says Edward Shortliffe, a physician and biomedical informaticist at Columbia University.

To help patients to make sense of eligibility criteria, Weng and her colleagues developed another open-source web tool, called DQueST. The software reads trials on ClinicalTrials.gov and then generates plain-English questions such as "What is your BMI?" to assess users' eligibility. An initial evaluation⁷ showed that after 50 questions, the tool could filter out 60–80%

THE STATE OF CLINICAL TRIALS

The success rates of clinical trials are low. Between January 2000 and October 2015, just 13.8% of drug candidates successfully navigated all three stages of clinical testing (left). Success rates declined for much of this period, but since 2015 it seems the trend has begun to reverse (right). Shaded area represents the estimated probabilities of success if all clinical trials yet to be completed failed (lower bound) or if all succeeded (upper bound).



of trials that the user was not eligible for, with an accuracy of a little more than 60%.

COMMERCIAL INTEREST

Tools such as those developed by Weng have plenty of room for improvement. Machine-learning algorithms rely on being fed training data from which they can learn — and to reach their potential, they need plenty. But labelling important features in these data, as is required to train NLP algorithms, is time consuming. The problem in academia, Weng says, is that both data and people power are limited.

Industry might be better placed to overcome those obstacles, and the past few years have seen a burst of activity. For example, digital-health company Antidote in New York City has developed a tool that helps people to search for trials. Other companies are working with health-care providers to find participants for trials in patient data held by these providers. Software developed by Deep 6 AI, an AI-based trials recruitment company in Pasadena, California, was used by researchers at Cedars-Sinai Smidt Heart Institute in Los Angeles, California, to find 16 suitable participants for a trial in one hour. A conventional approach had turned up only two people in six months.

Similarly, in a pilot study⁸ conducted by Mayo Clinic in Rochester, Minnesota, IBM's Watson for Clinical Trial Matching system, which is powered by the company's Watson supercomputer, increased the average monthly enrolment for breast-cancer trials by 80%. And although many purported clinical applications of Watson have yet to come to fruition, matching participants to clinical trials is one that has. In March, IBM signed an agreement with Health Quest Systems (now part of Nuvance Health), a non-profit network of four hospitals in New York and Connecticut, that will enable the group to use the computing giant's trial-matching system.

Although many of these technologies

seem impressive, they still have limitations. "They're not as magical as they sound," says Noemie Elhadad, a biomedical informaticist at Columbia University.

For instance, there is no replacement yet for the manual annotation of data that is needed to train NLP algorithms. Such algorithms are also honed for use by specific health-care providers and particular diseases. "Right now, there is no such thing as an NLP engine that takes any clinical notes written from any physician and can understand what the notes say," Elhadad asserts — the variation between medical fields and institutions is just too great. "We're all working on this, but we have a long way to go for this kind of universal understanding of clinical text."

Not everyone is convinced that the amount of effort being invested in finding participants for trials is worthwhile. "Patient matching gets a lot of hype," says Craig Lipset, former chief of innovation at drug company Pfizer in New York City. "But truth be told, many clinical trials don't need the intelligence in order to drive the match." The eligibility criteria of most studies aren't that complex, he says. And even if an AI algorithm can identify suitable people faster than would conventional methods, or can find people that might otherwise have been missed, researchers who are using third-party tools will then have to navigate the challenge of contacting individuals without violating privacy policies.

But some researchers think that getting these systems right will provide a considerable payoff. In 2014, 86% of clinical-trial participants worldwide were white people⁸. And a 2019 study found that 79% of genomic data comes from people of European descent⁹, even though they only comprise 16% of the world's population. AI-powered patient-matching algorithms could lead to more-diverse trial cohorts, by giving anyone in need a chance to participate — not just those who know the right doctor or who live near large health-care institutions. "It's really going to democratize

access to care,” says Elhadad.

BETTER BY DESIGN

Another area in which AI is being applied is the design of clinical trials. Every clinical trial follows a protocol that describes exactly how the study will be run. Any problems that arise during the trial and that require amendments to the protocol can lead to months of delays and add hundreds of thousands of dollars to the cost. “When protocols are right, drug development is faster and cheaper,” Lipset says.

When designing a trial, researchers lean on information from numerous sources, including comparable studies, clinical data and regulatory information. AI-powered software can not only process all of that information faster, but also collate more data than a person could read. “It just screams as being an opportunity to use AI,” Lipset says.

Trials.ai, a start-up company in San Diego, California, describes its AI tool as a data-driven guide to designing better trials protocols. It uses NLP and other AI techniques to collect and analyse publicly available data such as journal papers and drug labels, as well as private data owned by the drug or medical-device companies with which Trials.ai works. From those data, the company’s software can help determine how aspects of the customer’s proposed trial, such as the strictness of its eligibility criteria, might affect outcomes such as cost, length or participant retention. “We want to see what’s associated with different measures of success,” says David Fogel, chief scientist at Trial.ai.

If a customer wants to test a drug for diabetes, for example, adjusting the minimum level of glycated haemoglobin (a blood protein used to diagnose diabetes) that is required for people to participate could lead to different trial outcomes. When the eligibility threshold is too low, improvements owing to the drug can be harder to detect. But when the threshold is too high, there might not be enough people who are qualified to participate. By searching the literature, Trials.ai’s algorithm can quickly find population-wide diabetes statistics to help the protocol writer to identify an appropriate level.

Eventually, AI software might provide more than just guidance. The ultimate goal, Lipset says, would be for the first draft of the trial protocol to be written by the machine.

Even trials with well-designed protocols must rely on participants to follow instructions. A simple mistake such as forgetting to take a pill at the correct time could threaten the accuracy of a study’s results. AiCure, a data-analysis company in New York City, is developing a potential solution. It offers a platform that enables people to use their smartphones to record videos of themselves taking medication. By analysing those images using computer-vision algorithms, AiCure’s software can identify the person and the pill, and confirm whether it was taken. A study in people with schizophrenia showed that around 90% of people who used the AiCure platform took their medication as prescribed,



Open-source web tool Criteria2Query helps researchers to search databases more easily.

compared with about 72% of those who were periodically monitored by a person when taking medication¹⁰. The company says that its software can even measure people’s facial expressions to track how they respond to treatment, which could guide the development of therapies.

SEEKING VALIDATION

Much of the promise of AI in clinical trials — and in health care, in general — is fuelled by hype. “A lot of these things are in the theoretical realm,” Topol says. This points to a main challenge in the field: how to show that AI technology does, in fact, improve trials.

“Validation is critical,” Lipset says. “We need to know it’s reproducible. We need to show the evidence back to regulators so they have confidence as well.”

“The real possibility of AI is to do away with clinical trials.”

But other than a few pilot studies and case studies, assessments of how AI can improve clinical trials are rare. Even for more-developed AI technologies, such as those used in medical-image analysis, rigorous, large-scale trials are lacking, Topol says. There’s still a big gap between promise and proof. “Hopefully, it won’t be long before we fill in that gap,” he says.

Companies are making moves to assess the performance of their AI tools. Trials.ai, for instance, is trying to quantify how its technology improves trial designs, says Kim Walpole, the company’s chief executive and co-founder. She hopes that the information will enable Trials.ai to calculate how much money and time the software could save for potential customers.

However, Weng says that the lack of a shared framework for evaluating AI tools is an issue. Although her patient-matching software is open source, most companies retain ownership of their tools, and it is difficult to compare and assess such technologies in a standardized way.

If these technologies live up to their potential, the impact could be enormous. Even randomized trials — the gold standard in

clinical trials — could become outdated, says Hughes. If data from hundreds of millions of people were available, and AI tools could accurately analyse them, studies such as his breast-cancer trial wouldn’t need to recruit anyone — the data would already exist.

Testing treatments might still require controlled trials. But even then, it is possible that AI systems with access to huge data sets such as electronic health records might be able to simulate how a cohort is likely to respond to a therapy. A virtual clinical trial of this nature could prevent a pharmaceutical company from embarking on a large real-world trial that’s doomed to failure, Topol says. Such simulations are largely theoretical, but the beginnings can be found in, for example, statistical models that are being used to simulate how virtual patients with irregular heartbeats might respond to a type of blood-thinning drug.

After a drug has been approved, Hughes explains, electronic health records would show how the wider population responds — superseding the initial trial, virtual or not. As AI systems and data availability continue to improve, more clinical research might happen outside the framework of randomized trials. “The real possibility of AI,” Hughes says, “is to do away with clinical trials.” ■

Marcus Woo is a science writer based in San Jose, California.

1. Wong, C. H., Siah, K. W. & Lo, A. W. *Biostatistics* **20**, 273–286 (2019).
2. Hughes, K. S. et al. *N. Engl. J. Med.* **351**, 971–977 (2004).
3. Hughes, K. S. et al. *J. Clin. Oncol.* **31**, 2382–2387 (2013).
4. Bennette, C. S. et al. *J. Natl Cancer Inst.* **108**, djv324 (2016).
5. Haddad, T. C. et al. *J. Clin. Oncol.* **36** (suppl.), abstr. 6550 (2018).
6. Yuan, C. et al. *J. Am. Med. Inform. Assoc.* **26**, 294–305 (2019).
7. Liu, C. et al. *J. Am. Med. Inform. Assoc.* <https://doi.org/10.1093/jamia/ocz121> (2019).
8. Knepper, T. C. & McLeod, H. L. *Nature* **557**, 157–159 (2018).
9. Martin, A. R. et al. *Nature Genet.* **51**, 584–591 (2019).
10. Bain, E. E. et al. *JMIR mHealth uHealth* **5**, e18 (2017).