Trials and Challenges of Clinical Trial Emulation

David Ouyang, MD[1], Jeffrey M. Drazen, MD[2,3]

1. Division of Research, Kaiser Permanente Northern California
2. Division of Pulmonary and Critical Care, Department of Medicine, Brigham and Women's Hospital
3. Distinguished Parker B. Francis Professor of Medicine, Harvard Medical School

Clinical trials are costly, labor intensive, and have long lead times prior to results. However, clinical trials are still one of the strongest forms of medical evidence available and able to address questions of causality that other analyses cannot. Randomization balances out unknown confounders while blinding minimizes bias among physicians and patients[...]. In this setting, it has been a holy grail of evidence generation to use real world evidence to confirm the result of or even obviate the need to run clinical trials[1]. Electronic health records (EHRs) are abundant, rich in clinical context, and provide a promising testbed to conduct clinical trial emulation.

In this issue of *NEJM AI*, Gonzalez et al present a framework of using large language models (LLMs) to organize and extract information from clinical notes, perform causal inference on select cancer patients and their respective treatments, and assess treatment effect. The TRIALSCOPE program was able to closely emulate the results of clinical trials using data gleaned from electronic health records (EHRs) of patients who either received or did not receive a specific treatment for advanced non-small cell lung cancer or metastatic pancreatic cancer. In a wide range of completed historical clinical trials, TRIALSCOPE matched trial results after identifying patients that meet the inclusion and exclusion criteria of the studies.

Although the results are impressive, do these findings suggest that in the future that TRIALSCOPE, or similar applications, could generate sufficient high quality clinical evidence such that new treatments could be released for widespread use after a few small, controlled trials (think late Phase II trials) are favorably completed? We think the answer is both "yes" and "no". The optimistic answer derives from the disease conditions studied – specifically cancer patients receiving recently released treatments. Cancer differs from many other conditions in a few important ways. First, the diagnosis of cancer is most often based on laboratory evaluation of affected tissues, such that these diagnoses are often more certain and precise than other heterogenous syndromic diagnoses present in other specialties. Second, because cancer treatments are often quite expensive, insurers commonly limit the use of cancer treatments to patients whose clinical condition closely align with the inclusion and exclusion criteria of the clinical trials that led to the treatment's approval. Third, the outcomes used to declare efficacy are usually unequivocal, e.g. death or advancing metastasis and easily adjudicated from the EHR. As a result, the clinical characteristics of people who receive novel treatments outside of the clinical trial setting will likely mirror those of people who received the treatment in the clinical

trial setting. The uniformity of diagnosis, treatment and outcome makes cancer the perfect condition for clinical trial emulation.

The cautious answer reflects the fact that the “real world” outside of novel cancer agents, is not the clean “real world” of TRIALSCOPE. Consider conditions, e.g. asthma, headache, heart failure, inflammatory bowel disease, pain and many others, where the both the diagnostic criteria and outcome measures can be more subjective than objective. Oncology is in a unique position of having many therapeutics repurposed and re-evaluted for many other cancers other than the initial targets. Such a framework is ripe for clinical trial emulation as there are often patients taking medications off label in the real-world that can inform clinical evidence. Many areas of medicine already have a paucity of well-conducted clinical trials as well as few treatment options. For such cases, it would be hard to confirm the accuracy of TRIALSCOPE. For TRIALSCOPE to prove its real worth, it will need to show that “real world data” can be used to successfully emulate clinical trials for conditions outside of cancer.

1. Liu R, Rizzo S, Whipple S, et al. Evaluating eligibility criteria of oncology trials using real-world data and AI. Nature 2021;592(7855):629–33.