The Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard Bioethics Collaborative

Monday, November 16th, 2020 | 12:00PM-3:00PM ET
Virtual Meeting

Artificial Intelligence in Clinical Research
Meeting Summary

1. Introduction

Among many uses and applications, artificial intelligence (AI) technologies may be deployed as an intervention under investigation in a clinical trial or an observational study or used simply to facilitate certain aspects of clinical trials (e.g., eligibility screening). The potential benefits of using AI in clinical research vary with the specific AI application, but some general benefits are worth surveying here. AI technologies that facilitate certain aspects of clinical trials may increase the likelihood that trials are completed, accelerate the pace of clinical research, decrease the costs of research, and improve participant adherence, inter alia.[1] AI interventions may be able to improve patient outcomes, and clinical research that evaluates AI interventions helps to validate these technologies.

The potential ethical concerns raised by AI technologies also vary with the specific AI application, but some general ethical concerns can be noted here. AI algorithms are trained and then validated or confirmed on datasets, but the representativeness of those datasets is often obscure, raising concerns about the generalizability of the AI algorithm and potential biases embedded in the algorithm. Even if the data are "representative," it is often unclear how AI "learns" on a dataset, and the processes by which AI technologies reach conclusions are often unknown, raising concerns of transparency and trust in the use of AI technologies in clinical research. AI technologies may raise questions of data privacy and confidentiality, given their ability to process large amounts of information and aggregate distinct datasets, as well as questions of appropriate oversight for the use of AI technologies in clinical research.

The November 16th, 2020, meeting of the MRCT Center Bioethics Collaborative convened attendees to examine the topic of AI in clinical research.

2. Meeting Summary

Bioethics Collaborative attendees noted that the word 'bias' has a different meaning depending on the context in which it is used. Bias is generally defined as an unfair prejudice either in favor of or against something or someone. This definition generally carries negative connotations. In

[1] Stefan Harrer et al., "Artificial Intelligence for Clinical Trial Design," *Trends in Pharmacological Sciences*, Special Issue: Rise of Machines in Medicine, 40, no. 8 (August 1, 2019): 577–91, https://doi.org/10.1016/j.tips.2019.05.005.

statistics, bias has a more neutral and scientific frame, referring to an estimate that is not representative of the true value being estimated. The approach to managing scientific bias identified in a dataset used to train or validate an AI algorithm depends on the nature of the problem under consideration. Bias in datasets may help individuals better understand the algorithms being trained and/or validated, but only if known and considered. Data provenance, and methods to account for and expose bias, therefore, should be transparently communicated to users of an AI technology. Attendees also noted that transparency regarding data provenance may increase public trust in the use of AI algorithms in clinical research.

As a repeated theme during the Bioethics Collaborative, attendees debated whether stakeholders should be more concerned with validating the predictive capabilities of AI algorithms or designing explainable AI algorithms. Current AI technologies are prediction tools, and many of these technologies can be effectively and ethically used on the basis of their predictive capabilities alone. However, the processes by which these AI technologies make predictions may not be clear. AI developers attempt to overcome this lack of transparency by designing explainable AI (i.e., AI that illuminates the variables and processes an algorithm uses to make predictions). Understanding the "how" of an AI technology may bring users of the technology closer to trusting the technology itself and perhaps to revealing elements of its learning that may be related to causality. Some attendees argued that we should focus less on how an AI algorithm works and focus more on validating prospectively that a model performs accurately and reliably. Other attendees placed more value on explainability, noting that explainability is necessary to prevent a continuously learning algorithm (i.e., an algorithm that evolves in response to each piece of data it processes) from changing erroneously.

One attendee appreciated the value of explainability but cautioned that explainable AI may draw researchers and clinicians' attention away from treating the patient and towards treating the variables that an AI algorithm uses to make predictions, a concern that is particularly acute when causality has not been established. Another attendee cautioned that explainability may generate an unreasonable level of trust in an AI model. In its simplest formulation, an AI model may be explainable despite being inaccurate and unhelpful.

Throughout the discussion on predictability and explainability, attendees were sensitized to the distinction between correlation and causality. Predictive AI algorithms establish correlation between variables, but they do not establish causality. Explainable AI sheds light on how the algorithm correlates variables, potentially bringing individuals closer to establishing causal links between these variables.

Bioethics Collaborative attendees also addressed the methods used to validate AI technologies. While prospective randomized controlled trials are the gold-standard for validating AI technologies, whether and when an AI technology is considered a quality improvement initiative—and therefore not require evaluation through a clinical trial or approval by an IRB—are unclear. Further, prospective evaluation(s) may not be helpful in assessing AI algorithms that continuously learn since the AI algorithm will evolve both during and after the evaluation.

Different methods used to validate AI technologies were discussed. Some AI developers split their dataset, training the algorithm on one half of the dataset and validating the algorithm's performance on the other half. This method of validation, however, will not uncover unintended biases in the dataset itself and/or of issues related to dataset curation. Regardless of how an AI algorithm is validated, attendees agreed that AI algorithms should be annotated with guidelines to help the user understand the population(s) for which an AI algorithm has been validated. Given the limitations inherent in methods of validation, attendees also suggested that, generally, AI algorithms should have a 'shadow period,' in which the relevant authorities (e.g., hospital administrators, IRBs, national regulators, etc.) follow the algorithm's performance (i.e., accuracy, reliability, fairness) upon its deployment in real-world settings.

Bioethics Collaborative attendees concluded the meeting by emphasizing that there is a person behind every data point, and both individuals and their data need to be protected and respected. Data scientists, AI developers, researchers using AI technologies, and other stakeholders should receive ethics training to learn how to use and protect individuals' data appropriately while developing and deploying AI technologies and to understand the consequences of such use. Attendees suggested that the ethics training should emphasize values of justice and access; populations should not be unfairly excluded from receiving the benefits of AI algorithms in clinical research and healthcare, particularly if their data are used to develop these algorithms.

3. Potential Future Work

- Design a strategy for communicating the generalizability of AI technologies to the users of these technologies
    - Design a method for 'attaching' or 'annotating' this information to AI algorithms through metadata or other means
    - Design language and a template for communicating the generalizability of an AI algorithm
- Create a framework for the disclosure of biases in datasets and algorithms and whether these biases were mitigated and/or used to learn
- Design data ethics and AI ethics training for key stakeholders
- Draft a position paper that argues for AI applications that do not exclude certain communities (e.g., rural communities, communities without reliable electricity, etc.)
- Create guidelines to help researchers determine how to validate an AI algorithm
    - The guidelines may include:
        - Recommendations on whether an AI algorithm should be evaluated through a prospective clinical trial with IRB review or evaluated as a quality improvement initiative
        - Recommendations on whether an AI algorithm is ready to be deployed in human participant research