VIEWPOINT

Questions for Artificial Intelligence in Health Care

Thomas M. Maddox, MD, MSc
Healthcare Innovation Lab, BJC Healthcare/Washington University School of Medicine, St Louis, Missouri; and Division of Cardiology, Washington University School of Medicine, St Louis, Missouri.

John S. Rumsfeld, MD, PhD
American College of Cardiology, Washington, DC.

Philip R. O. Payne, PhD
Institute for Informatics, Washington University School of Medicine, St Louis, Missouri.

+ Viewpoint

Artificial intelligence (AI) is gaining high visibility in the realm of health care innovation. Broadly defined, AI is a field of computer science that aims to mimic human intelligence with computer systems.[1] This mimicry is accomplished through iterative, complex pattern matching, generally at a speed and scale that exceed human capability. Proponents suggest, often enthusiastically, that AI will revolutionize health care for patients and populations. However, key questions must be answered to translate its promise into action.

What Are the Right Tasks for AI in Health Care?

At its core, AI is a tool. Like all tools, it is better deployed for some tasks than for others. In particular, AI is best used when the primary task is identifying clinically useful patterns in large, high-dimensional data sets. Ideal data sets for AI also have accepted criterion standards that allow AI algorithms to "learn" within the data. For example, *BRCA1* is a known genetic sequence linked to breast cancer, and AI algorithms can use that as "the source for truth" criterion when specifying models to predict breast cancer. With appropriate data, AI algorithms can identify subtle and complex associations that are unavailable with traditional analytic approaches, such as multiple small changes on a chest computed tomographic image that collectively indicate pneumonia. Such algorithms can be reliably trained to analyze these complex objects and process the data, images, or both at a high speed and scale. Early AI successes have been concentrated in image-intensive specialties, such as radiology, pathology, ophthalmology, and cardiology.[2,3]

> These tools and related needs may simply replace current costs with different, and potentially higher, costs.

However, many core tasks in health care, such as clinical risk prediction, diagnostics, and therapeutics, are more challenging for AI applications. For many clinical syndromes, such as heart failure or delirium, there is a lack of consensus about criterion standards on which to train AI algorithms. In addition, many AI techniques center on data classification rather than a probabilistic analytic approach; this focus may make AI output less suited to clinical questions that require probabilities to support clinical decision making.[4] Moreover, AI-identified associations between patient characteristics and treatment outcomes are only correlations, not causative relationships. As such, results from these analyses are not appropriate for direct translation to clinical action, but rather serve as hypothesis generators for clinical trials and other techniques that directly assess cause-and-effect relationships.

What Are the Right Data for AI?

AI is most likely to succeed when used with high-quality data sources on which to "learn" and classify data in relation to outcomes. However, most clinical data, whether from electronic health records (EHRs) or medical billing claims, remain ill-defined and largely insufficient for effective exploitation by AI techniques. For example, EHR data on demographics, clinical conditions, and treatment plans are generally of low dimensionality and are recorded in limited, broad categorizations (eg, diabetes) that omit specificity (eg, duration, severity, and pathophysiologic mechanism). A potential approach to improving the dimensionality of clinical data sets could use natural language processing to analyze unstructured data, such as clinician notes. However, many natural language processing techniques are crude and the necessary amount of specificity is often absent from the clinical record.

Clinical data are also limited by potentially biased sampling. Because EHR data are collected during health care delivery (eg, clinic visits, hospitalizations), these data oversample sicker populations. Similarly, billing data overcapture conditions and treatments that are well-compensated under current payment mechanisms. A potential approach to overcome this issue may involve wearable sensors and other "quantified self" approaches to data collection outside of the health care system. However, many such efforts are also biased because they oversample the healthy, wealthy, and well. These biases can result in AI-generated analyses that produce flawed associations and insights that will likely fail to generalize beyond the population in which they are generated.[5]

What Is the Right Evidence Standard for AI?

Innovations in medications and medical devices are required to undergo extensive evaluation, often including randomized clinical trials and postmarketing surveillance, to validate clinical effectiveness and safety. If AI is to directly influence and improve clinical care delivery, then an analogous evidence standard is needed to demonstrate improved outcomes and a lack of unintended consequences. The evidence standard for AI tasks is currently ill-defined but likely should be proportionate to the task at hand. For example, validating the accuracy of AI-enabled imaging applications against current quality standards for traditional imaging is likely sufficient for clinical use. However, as AI applications move to prediction, diagnosis, and treatment, the standard for proof should be significantly higher.[1] To this end, the US Food and Drug Administration is actively considering how best to regulate AI-fueled innovations in care delivery,

Corresponding Author: Thomas M. Maddox, MD, MSc, Cardiovascular Division, Washington University School of Medicine/BJC Healthcare, Campus Box 8086, 660 S Euclid, St Louis, MO 63110 (tmaddox@wustl.edu).

attempting to strike a reasonable balance between innovation, safety, and efficacy.

Using AI in clinical care will need to meet particularly high standards to satisfy clinicians and patients. Even if the AI approach has demonstrated improvements over other approaches, it is not (and never will be) perfect, and mistakes, no matter how infrequent, will drive significant, negative perceptions. An instructive example can be seen with another AI-fueled innovation: driverless cars. Although these vehicles are, on average, safer than human drivers, a pedestrian death due to a driverless car error caused great alarm. A clinical mistake made by an AI-enabled process would have a significant chilling effect. Thus, ensuring the appropriate level of oversight and regulation is a critical step in introducing AI into the clinical arena.

In addition to demonstrating its clinical effectiveness, evaluation of the cost-effectiveness of AI is also important. Huge investments into AI are being made with promised efficiencies and assumed cost reductions in return, similar to robotic surgery. However, it is unclear that AI techniques, with their attendant needs for data storage, data curation, model maintenance and updating, and data visualization, will significantly reduce costs. These tools and related needs may simply replace current costs with different, and potentially higher, costs.

What Are the Right Approaches for Integrating AI Into Clinical Care?

Even after the correct tasks, data, and evidence for AI are addressed, realization of its potential will not occur without effective integration into clinical care. To do so requires that clinicians develop a facility with interpreting and integrating AI-supported insights in their clinical care. In many ways, this need is identical to the integration of more traditional clinical decision support that has been a part of medicine for the past several decades. However, use of deep learning and other analytic approaches in AI adds an additional challenge. Because these techniques, by definition, generate insights via unobservable methods, clinicians cannot apply the face validity available in more traditional clinical decision tools (eg, integer-based scores to calculate stroke risk among patients with atrial fibrillation). This "black box" nature of AI may thus impede the uptake of these tools into practice.

AI techniques also threaten to add to the amount of information that clinical teams must assimilate to deliver care. While AI can potentially introduce efficiencies to processes, including risk prediction and treatment selection, history suggests that most forms of clinical decision support add to, rather than replace, the information clinicians need to process. As a result, there is a risk that integrating AI into clinical workflow could significantly increase the cognitive load facing clinical teams and lead to higher stress, lower efficiency, and poorer clinical care.

Ideally, with appropriate integration of AI into clinical workflow, AI can define clinical patterns and insights beyond current human capabilities and free clinicians from some of the burden of integrating the vast and growing amounts of health data and knowledge into clinical workflow and practice. Clinicians can then focus on placing these insights into clinical context for their patients and return to their core (and fundamentally human) task of attending to patient needs and values in achieving their optimal health.[6] This combination of AI and human intelligence, or augmented intelligence, is likely the most powerful approach to achieving this fundamental mission of health care.

A Balanced View of AI

AI is a promising tool for health care, and efforts should continue to bring innovations such as AI to clinical care delivery. However, inconsistent data quality, limited evidence supporting the clinical efficacy of AI, and lack of clarity about the effective integration of AI into clinical workflow are significant issues that threaten its application. Whether AI will ultimately improve quality of care at reasonable cost remains an unanswered, but critical, question. Without the difficult work needed to address these issues, the medical community risks falling prey to the hype of AI and missing the realization of its potential.

ARTICLE INFORMATION

Published Online: December 10, 2018. doi:10.1001/jama.2018.18932

Conflict of Interest Disclosures: Dr Maddox reports employment at the Washington University School of Medicine as both a staff cardiologist and the director of the BJC HealthCare/Washington University School of Medicine Healthcare Innovation Lab; grant funding from the National Center for Advancing Translational Sciences that supports building a national data center for digital health informatics innovation; and consultation for Creative Educational Concepts. Dr Rumsfeld reports employment at the American College of Cardiology as the chief innovation officer. Dr Payne reports employment at the Washington University School of Medicine as the director of the Institute for Informatics; grant funding from the National Institutes of Health, National Center for Advancing Translational Sciences, National Cancer Institute, Agency for Healthcare Research and Quality, AcademyHealth, Pfizer, and the Hairy Cell Leukemia Foundation; academic consulting at Case Western Reserve University, Cleveland Clinic, Columbia University, Stonybrook University, University of Kentucky, West Virginia University, Indiana University, The Ohio State University, Geisinger Commonwealth School of Medicine; international partnerships at Soochow University (China), Fudan University (China), Clinica Alemana (Chile), Universidad de Chile (Chile); consulting for American Medical Informatics Association (AMIA), National Academy of Medicine, Geisinger Health System; editorial board membership for JAMIA, JAMIA Open, Joanna Briggs Institute, Generating Evidence & Methods to improve patient outcomes, BioMed Central Medical Informatics and Decision Making; and corporate relationships with Signet Accel Inc, Aver Inc, and Cultivation Capital.

REFERENCES:

1. Stead WW. Clinical implications and challenges of artificial intelligence and deep learning. *JAMA*. 2018;320(11):1107-1108. doi:10.1001/jama.2018.11029

2. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410. doi:10.1001/jama.2016.17216

3. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation*. 2018;138(16):1623-1635. doi:10.1161/CIRCULATIONAHA.118.034338

4. Harrell F. Is medicine mesmerized by machine learning? Statistical Thinking website. http://fharrell.com/post/medml/. Published February 1, 2018. Accessed October 26, 2018.

5. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med*. 2018;178(11):1544-1547. doi:10.1001/jamainternmed.2018.3763

6. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. *JAMA*. 2018;319(1):19-20. doi:10.1001/jama.2017.19198