Title

Old-fashioned Intelligence Will Always Be Needed in Medicine

Permalink

https://escholarship.org/uc/item/6106f5vc

Journal

European Urology Focus, 7(4)

ISSN

2405-4569

Authors

Powell, Kerrington
Prasad, Vinay

Publication Date

2021-07-01

DOI

10.1016/j.euf.2021.03.022

Peer reviewed

Point of Focus Debate: Con

Old-fashioned Intelligence Will Always Be Needed in Medicine

Kerrington Powell [a], *Vinay Prasad* [b],*

[a] *College of Medicine, Texas A&M Health Science Center, College Station, TX 77843, USA;* [b] *Department of Epidemiology and Biostatistics, School of Medicine, University of California-San Francisco, San Francisco, CA, USA*

1. Introduction

A popular slogan in 2021 is "follow the science", which implies that the solution to complex challenges, such as the COVID-19 pandemic, is clear if we adhere to scientific principles. Of course, science is integral to tackling a pandemic, but science alone faces limits. Science cannot tell you how to prioritize your values. Science cannot tell you how to choose between trade-offs.

This, of course, represents the strongest case arguing that human beings—and our ability to weigh competing values and desires—will always be necessary in medicine. Artificial intelligence (AI) may someday be able to quantify precisely the harms and benefits of a surgery, medication, or device, but it cannot guide a patient towards the right choice for them. It cannot help a patient decide when the goal of treatment should shift from life extension to maximizing comfort. Artificial intelligence will never hold your hand when you are sick, and cannot console your loved one when you are gone.

2. Limits with the cutting edge of science

Beyond the hyperbolic view that physicians will be replaced by AI, there are many challenges for the implementation of AI into clinical care before there is robust evidence to do so. It is possible that AI may augment the work of physicians in the future, mitigating errors, improving physician time allocation, and improving diagnoses and detection [1]. However, before these technological innovations are adopted prematurely, they must prove their effectiveness for reliable clinical endpoints from well-constructed randomized controlled trials (RCTs). Failure to do so may lead to costly interventions and could put patients at risk with ineffective, or even harmful, therapies and screenings, which may someday be contraindicated [2].

Consider two case studies. A prospective RCT that tested polyp and adenoma detection rates showed that physicians who worked in conjunction with AI systems significantly increased the detection rate compared to unassisted endoscopists [3]. Another study found that an AI system yielded an absolute reduction in both false positives and false negatives in breast cancer identification [4]. An external research clinic demonstrated that this AI system outperformed radiologists, with a significant increase in the area under the receiver operating characteristic curve of 0.115 (95% confidence interval 0.055–0.175; $p = 0.0002$) [4]. From these studies, it is evident AI and machine learning may be able to play a role in the detection and diagnosis of cancer, but the question remains as to whether AI is being implemented correctly.

Cancer screening is complicated. It is simple to conclude that the goal of screening is to identify as many early cancerous lesions as possible. However, the true goal of screening is to identify lesions that (1) have not already spread and (2) are destined to spread if they are not identified and removed and (3) once they spread, are destined to shorten quality or quantity of life. The histopathologic characteristics we rely on—invasion of the basement membrane, appearance of cells on hematoxylin and eosin and other staining—provide a rough guide as to which lesions to target, but the truth is that this science remains imperfect. We routinely remove lesions that are destined not to spread or not to cause harm, so-called "overdiagnosed" cancers. We also routinely remove cancers that have already spread and are destined to result in metastatic disease in the future; in

* Corresponding author at: Department of Epidemiology and Biostatistics, School of Medicine, University of California-San Francisco, San Francisco, CA, USA. Fax: +1 415 5148150.
E-mail address: vinayak.prasad@ucsf.edu (V. Prasad).

other words, we subject patients to local therapy that does not increase their survival.

When we train AI algorithms to optimize the diagnosis of cancer on screening tests, we train the system to optimize the wrong endpoint. The system becomes highly efficient in finding more cancer or precancerous lesions on the basis of how a pathologist would score them, instead of becoming more efficient at finding the lesions we want to find, the cancers that are going to do bad things, but are neutralized by surgical removal.

This difference may sound trivial, but the result might even be counterproductive. If trained to solve the wrong problem, AI may even worsen the problem of overdiagnosis and impair (rather than enhance) the value of screening tests. How can we know for sure that AI is improving the lives of patients? The answer is that large, prospective randomized trials are needed to test the key question: does routine application of AI improve outcomes?

3. Conclusions

The idea that AI will make physicians expendable is hyperbole. Medicine is a moral profession, one that balances values and desires, and humans are irreplaceable in this context. Beyond this, AI is a tool, a hammer, and it is up to us to decide where and how to swing the hammer. We can pick a process and use AI to maximize its efficiency, such as the diagnosis of cancer, but this requires human beings to debate and decide if that is the problem we wish to solve.

If we want to apply AI to a field such as oncology, we must acknowledge the fundamental biological challenges and limitations in data [2]. Even in other fields such as radiology and pathology, it remains unclear who should be diagnosed or treated. The way to ensure that AI is doing what we hope it does is to conduct RCTs that measure meaningful endpoints such as overall mortality.

AI is new and enticing, and it is easy to think that improving the rate of diagnosis is the bar that AI must meet before adoption. Human intelligence is needed to remember the true goal of what we do, and hold AI accountable.

References

[1] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44–56.

[2] Prasad VK, Cifu AS. Ending medical reversal: improving outcomes, saving lives. Baltimore, MD: Johns Hopkins University Press; 2015.

[3] Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813–9.

[4] McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. Nature 2020;577:89–94.