

Putting decisions under the microscope

Ziad Obermeyer is an acting associate professor at the University of California, Berkeley, and an emergency physician. His teaching and research focus on how algorithms can aid in human decision-making in health care. Previously, he taught at Harvard Medical School, where he received the Early Independence Award, the most prestigious award for early-career scientists given by the US National Institutes of Health.

Ziad Obermeyer

From the outside, modern medicine looks like a high-performing science.

Entering medical school did nothing to dispel this illusion for me, starting with the white coat I was given in my first week. But as I memorized the facts underlying this science—the Krebs cycle, the five dangerous causes of chest pain—I remember feeling incredibly bored. Medicine, it seemed, was basically complete; further research seemed hopelessly incremental.

In fact, the most exciting problems started precisely where medical science left off: the inefficiency of healthcare. As a future physician, I was particularly interested in how physician behavior responded to incentives, just as health economics would predict, thus leading to overuse of care.

These issues were at the top of my mind as I entered clinical residency. I rolled my eyes when attending physicians—allegedly responsible for supervising my education—insisted on high-cost tests or interventions: the last thing the patient, or our health care system, needed was another incidental finding, another ineffective prescription.

I shudder internally when I think back to that time. Of the many mistakes I made, some were caught by others. Some were not, causing harm I think about to this day. At some point, the combined weight of these mistakes finally made me understand two things: The first was that I needed to get much better at being a doctor. The second was that, even as I improved, it was not enough.

A patient would present with a new disturbing symptom; after several hours and tests, neither of us knew why. My explanations were shallow (“Your back pain is musculoskeletal”); often, they were non-explanations (“Your chest pain is not a heart attack”). At home, I would lie awake agonizing: I should have ordered another test; I should not have sent her home.

This happens often enough to shake one’s faith in medical science. Doctors, it turns out, do not have it all figured out. The facts I learned in medical school and residency were helpful, pointing me in the right direction or preventing catastrophic misjudgments.



Credit: Ziad Obermeyer

But I was still making mistakes. Just like the attending physicians I had rolled my eyes at, I was ordering costly tests that, all too often, came out negative. Worse, I was also failing to order tests and treatments that would have helped—as I would sometimes learn later, when a patient returned with a diagnosis I had missed, or needing a treatment I didn’t think to deliver. Health economics has little to say about why doctors make these mistakes.

My first research project after residency studied patients who died unexpectedly after reassuring medical evaluations. On one level, I still wonder whether my funders at the Office of the Director of the National Institutes of Health wish they had spent their monies elsewhere: I published some papers in high-profile journals, but many of my hypotheses proved wrong, and many of the things I hoped to do proved impossible. On another level, though, those years of support were transformative. By giving me space to work on problems that I found interesting, I learned an enormous amount,

built a network of collaborators in medicine, economics and computer science, and launched several of the projects I am most excited about today.

One of those projects uses machine learning to study who is tested for heart attack—and who should be. My preliminary work on unexpected deaths suggested that many resulted from missed diagnoses. I realized that this was an ideal-use case for machine-learning tools: by forming highly accurate risk predictions and comparing them to testing decisions, we could put physician judgment under the proverbial microscope.

A striking finding is that physicians simultaneously over-test predictably low-risk patients (who go on to have negative tests) and under-test predictably high-risk patients (who go on to experience catastrophic outcomes). This finding echoes my own experience as a clinician and suggests that decision-making can be improved, perhaps, and we are now hoping to partner with a healthcare system to perform a randomized trial and test this hypothesis directly.

A natural next question is: why do physicians make these mistakes? Interestingly, physicians seem to use a model of risk that is far too simple: they make effective use of a handful of variables but neglect thousands of others. These variables, which better represent the full richness of patients’ medical histories, are the key to the algorithm’s advantage. In other words, these mistakes are driven not by bad incentives but by physicians’ inability to process the vast data that they need to make good decisions.

This, I believe, is a clue to the lasting contribution of machine learning to medicine: opening our eyes to the deep complexity of medical decision-making, and the science in which it is grounded. □

Ziad Obermeyer

School of Public Health, University of California at Berkeley, Berkeley, CA, USA.

e-mail: zobermeyer@berkeley.edu

Published online: 07 November 2019

<https://doi.org/10.1038/s41591-019-0591-3>