Bedside to Bench — AI and the New Science of Medicine

Ziad Obermeyer, M.D.[1]

For centuries, doctors have tried to make sense of the body as a complex system. Today, medical research builds "up" an understanding of the whole from an increasingly large set of molecular parts — bench to bedside. This endeavor, as Francis Collins told the *Wall Street Journal* in 2020, is proving "more complicated than many people guessed it would be."

Artificial intelligence (AI) will reinvigorate an older approach that builds "down" from patient-level data to theory and understanding — bedside to bench. This type of clinical research was a major source of historic breakthroughs but has lain dormant for many years.

Complex systems such as the body operate on many levels, each with its own logic. Computers, another complex system, exist as transistors (zeros and ones), low-level instructions (machine code), high-level programming languages (Python or C), and user interfaces (programs). All these levels of abstraction are "accurate," but not all of them are useful for debugging a given issue.

Problems that appear hopelessly complex on one level may be easily solved on another. The Achilles' heel of many diseases lies at the molecular level.[1] Cancers driven by specific mutations respond to targeted therapies. Autoimmune diseases can sometimes be turned off with an antibody. Sickle cell disease can be cured by gene editing. Messenger RNA vaccines train the immune system to recognize individual viral proteins.

Now consider pain. Of course pain exists on the molecular level: genes predispose, nerves transmit, receptors modulate. But attacking the problem at this level has produced few breakthroughs — indeed, one could argue that it led directly to the opioid epidemic. This failure highlights the importance of psychological or social factors. But treating pain as a social science problem ignores its obvious physical basis. If the molecular level is too low and the psychosocial level too high, what is "just right"?

When Jonas Kellgren and John Lawrence studied pain in the 1950s, they relied on a recent innovation: the x-ray. Comparing miners with office workers in the same colliery, they linked pain to x-ray findings and developed the definition of osteoarthritis we still use today.

X-rays, like other clinical measurement tools, view a patient through the lens of a machine. They bridge levels of abstraction, projecting a knee onto a plane of bone, tendon, and tissue. Many breakthroughs arise in this "middle ground." Electrocardiography (ECG) revealed myocardial infarction and arrhythmic syndromes, ultrasonography documented fetal development, electroencephalography (EEG) enabled characterization of seizures and sleep — paving the way for elucidating mechanisms of disease. Whatever textbooks say about the scientific method, science often proceeds in this way: first fact, then theory.

This once-fertile research tradition has withered away — not because we lack mysterious phenomena to explain, but because the human mind struggles to find reliable signals in high-dimensional machine-derived data. Radiographic signs have a tenuous relationship to pain: even Kellgren and Lawrence noted that many patients with erosions on x-rays had no pain, and many with pain lacked erosions. The latter mismatch was a particular problem for Black patients, who often had far more pain than their radiographs suggested. Traditional statistics are of little help here: How do you fit a chest x-ray into a regression model? Without rigorous tools to advance it, the research foundered.

Four years ago, my colleagues and I trained an AI system to reexamine the relationship between x-ray findings and pain.[2] Our question was the same as Kellgren and Lawrence's, but we had two advantages: a far larger and more diverse cohort and machine learning tools for attacking images directly. The algorithm we trained found new sources of pain present in x-rays — signals that Kellgren, Lawrence, and generations of doctors after them had missed, and ones that proved particularly useful for explaining the "extra" pain reported by Black patients.

By allowing direct engagement with machine-derived clinical data, AI is finding fascinating new patterns. Algorithms can predict breast cancer from normal mammograms, atrial fibrillation from sinus-rhythm ECGs, brain activation from EEGs of unre-

sponsive patients, depression from functional magnetic resonance imaging (MRI). Massive datasets of laboratory results and vital signs are pointing to mysterious "set points" in individual patients' physiology, with large implications for mortality.

we cannot learn about mechanisms or potential new treatments.

Instead, AI will help humans develop understanding by generating hypotheses and guiding new data collection and experimentation. For example, in ongoing work, my colleagues and I apply AI to predict sudden cardiac deaths, many of which could be prevented by defibrillators — if they could be predicted in time. Linking ECG waveforms to death certificates from Sweden, we have trained a deep-learning model that outperforms ejection fraction in predicting defibrillator benefits and is generalizable internationally.

An AI-driven medical science, grounded in clinical data, will help us rediscover clinical ways of thinking about bodies at their own level of abstraction, as neither aggregates of molecules nor social constructs.

Surprising new facts can produce seismic shifts. Eighteenth-century physicists tried to build up a science of chemistry from Newtonian principles but were stymied by oddities: fixed ratios of elemental combinations, and later, specific angles of bonds, chirality. By setting aside contemporary physics and focusing on facts, John Dalton's work — initially dismissed as mere heuristics — built down to a new science of chemistry. This advance set the stage for Linus Pauling's quantum-theoretical account of chemical bonds and helped reinvent physics from above.[3]

Similarly, new facts from AI — statistical regularities in clinical data — will demand new theories to account for them. One might ask why we need theories. The fact that AI can reliably diagnose disease is useful enough on its own. But facts are not generative: they describe the world as it is but say nothing about underlying causes.[4] And without understanding causes

So far, this is a fairly "vanilla" prediction exercise: it might be clinically useful but says nothing about causes or solutions. So to build new understanding, we have trained a generative model that can produce "what-if" waveforms: starting with a low-risk ECG, it tweaks the waveform in a way that causes the predictive model to flag it as high-risk. This step has revealed a new biomarker for sudden cardiac death: "scattering" of the QRS complex over time — distinct from known phenomena such as prolongation or fragmentation — that correlates with subtle, diffuse left-ventricular fibrosis on cardiac MRI. Rather than simply predicting an outcome, AI is pointing us to potentially new sets of genes, proteins, and cellular architectures to probe and target at the bench, then translate back to the bedside — where we started. Thus, AI generates hypotheses, but those hypotheses are tested the old-fashioned way, avoiding the myriad statistical problems of relying on AI alone for discovery.

This approach will not only feed new hypotheses into molecular biology. It will bridge the gulf between medical research and clinical reasoning, thereby creating a new science. Today, medicine is not an independent scientific discipline: one cannot get a Ph.D. in it, to use a trivial definition. Medical researchers largely work below the clinical level of abstraction (genetics, molecular and cellular biology) or above it (epidemiology, biostatistics, social science); clinical thinking is left to clinicians. Concepts such as homeostasis, central in the clinic, are largely absent from research. Physiology departments focus on molecules. An AI-driven medical science, grounded in clinical data, will help us rediscover clinical ways of thinking about bodies at their own level of abstraction, as neither aggregates of molecules nor social constructs.

A preview of this new science comes from an elegant study of knee pain by Sharma and colleagues,[5] who begin with a simple observation: natural variation in the angle at which the femur exits the hip can create varus alignment, which shifts the leg's load-bearing axis medially; valgus alignment shifts it laterally. Empirically, even a few degrees of alignment shift have huge effects on the eventual development of osteoarthritis, exactly where the

physical model predicts higher loads. Rather than focusing on the component that failed, the study asks why it failed. By blending human intuition and simple physics, the investigators reduce a complex problem to three vectors connecting hip, knee, and ankle, which predict when and where osteoarthritis will develop — without knowing genetic codes, ZIP codes, or TNF-α levels.

AI will supercharge this approach. Every aspect of a knee can be quantified: alignment, tendon and ligament attachments, muscle volume. Patient-specific values can be fed into a model that calculates load and forms testable hypotheses about how to fix problems. Feeding massive datasets into cutting-edge tools, we can develop patient-tailored treatments, whose effects may be detectable in small (even N-of-1) studies. Sharma and colleagues found odds ratios for osteoarthritis of 4 to 5 by studying just 230 patients over a period of 18 months.

The science of medicine is ready for reinvention, for blending new tools with old ways of thinking about the bedside: not as an end point for research, but as a source.

Disclosure forms provided by the author are available at NEJM.org.

[1] University of California, Berkeley, Berkeley.

This article was published on December 6, 2025, at NEJM.org.

1. Rees J. Complex disease and the new clinical sciences. Science 2002;296:698-700.
2. Pierson E, Cutler DM, Leskovec J, Mullainathan S, Obermeyer Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. Nat Med 2021;27:136-40.
3. Chomsky N. The mysteries of nature: how deeply hidden? J Philos 2009;106:167-200.
4. Vafa K, Chang PG, Rambachan A, Mullainathan S. What has a foundation model found? Using inductive bias to probe for world models. August 14, 2025 (http://arxiv.org/abs/2507.06952). preprint.
5. Sharma L, Song J, Felson DT, Cahue S, Shamiyeh E, Dunlop DD. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. JAMA 2001;286:188-95.

DOI: 10.1056/NEJMp2510203

Raising a RUCkus

In Season 3, episode 7, of the Not Otherwise Specified podcast, host Lisa Rosenbaum and her guests dissect the common notion that the key to improving the primary care system is simply increasing reimbursement.