DEEP MEDICINE

How Artificial Intelligence Can Make Healthcare Human Again

ERIC TOPOL

BASIC BOOKS
New York

CONTENTS

To my family—Susan, Sarah, Evan, Antonio, Julian, and Isabella—who provided unconditional support and the deep inspiration for me to pursue this work

BASIC BOOKS

FOREWORD

> Life can only be understood backwards; but it must be lived forwards.
>
> —Søren Kierkegaard

Among the many characteristics that make us human and that distinguish us from other animals must be our urge to look back. It is hard to imagine that other species brood late at night about the one that got away or a job they could have had. But we also do it as a form of scholarship, looking back at ourselves as a species, as if we were the Creator, poring through recorded history, charting the milestones of progress, from the harnessing of fire to the microchip. Then we try to make sense of it.

Kierkegaard's thesis that we live life forward but understand it backward might mean nothing more than we remember the past, and at best we have an (inaccurate) record of it. But with apologies to him and to George Santayana, understanding history does not provide immunity to repeating it. A cursory scan of the news shows this to be true. In short, even as a guide to what to avoid, the past is unreliable. Only the future is certain because it is still ours to make.

Which brings us to futurists, like the author of this wonderful book. Such individuals, on hearing that the Wright brothers became airborne, can foresee budget airlines, airline hubs, and humans walking on the moon. These historians of the now begin with the study of what *is* today, asking not how to avoid the perils of the past but how to maximize the advantages of the future. Pencil and paper, or tablet, in hand, they patrol the frontiers of science and tech and interview those at the cutting edge, including those who have tumbled over. They seek out innovators, scientists, mavericks, and dreamers. They listen, they monitor, they filter, and they synthesize knowledge across many disciplines to make sense of it all for the rest of us. As *Deep Medicine* will show you, theirs is a formidable intellectual task and an extraordinarily creative one. It involves as much right brain as left, and it invokes the muses, because what is in this book is as much inspiration as it is exposition.

Deep Medicine is Eric Topol's third exploration of what *will* be. The previous books, examined in the light of where we are now, reveal his prescient vision. In *Deep Medicine,* Eric tells us we are living in the Fourth Industrial Age, a revolution so profound that it may not be enough to compare it to the invention of steam power, the railroads, electricity, mass production, or even the computer age in the magnitude of change it will bring. This Fourth Industrial Age, revolving around artificial intelligence (AI), robotics, and Big Data, heralds a profound revolution that is already visible in the way we live and work, perhaps even in the way we think of ourselves as humans. It has great potential to help, but also to harm, to exaggerate the profound gap that already exists between those who have much and those who have less each passing year.

This revolution will overtake every human endeavor, medicine not least among them. Medicine itself is at a moment of crisis. As a profession, for all the extraordinary advances in the art and science of medicine in the last four decades, we have too often failed our patients. We fail to follow proven guidelines, and we fail in the art by not seeing the unique *person* in front of us. We know their genome, but by not *listening* to their story, we don't register their broken heart. We fail to see the neurofibroma that are raising lumps all over their skin, a finding that is relevant to their paroxysmal hypertension but that does need the gown to come off during the exam, does need our attention to be on the body and not on the screen; we miss the incarcerated hernia that explains an elderly patient's vomiting and have to wait for an expensive CAT scan and a radiologist to tell us what was before our eyes. Countries with the biggest expenditures on healthcare lag behind those that spend much less in basic rankings such as infant mortality. I think it is very telling that *Deep Medicine* opens with a profound, personal, revealing anecdote of the author's own painful and harrowing medical encounter that was a result of not being seen as an individual, someone with an uncommon disorder.

It should not surprise us that technology, despite the dramatic way it has altered our ability to image the body, to measure and monitor its molecular structure, can also fail just as badly as humans fail. The glaring example is in the electronic healthcare record systems (EHRs) currently in use in most hospitals. These EHRs were designed for billing, not for ease of use by physicians and nurses. They have affected physician well-being and are responsible for burnout and attrition; moreover, they have forced an inattentiveness to the patient by virtue of an intruder in the room: the screen that detracts from the person before us. In *Intoxicated by My Illness,* a poignant memoir about a man's ultimately fatal prostate cancer, Anatole Broyard articulates a wish that his urologist would "brood on my situation for perhaps five minutes, that he would give me his whole mind just once, be bonded with me for a brief space, survey my soul as well as my flesh, to get at my illness, *for each man is ill in his own way.*"[1] This poignant declaration, from the era just before electronic medical records, expresses the fundamental need of a sick human being; it is timeless, I believe, resistant to change, even as the world around us changes. It bears emphasizing: *each man and woman is ill in his or her own way.*

I am excited about the future, about the power to harness Big Data. By their sheer capacity to plow through huge datasets and to learn as they go along, artificial intelligence and deep learning will bring tremendous precision to diagnosis and prognostication. This isn't to say they will replace humans: what those technologies will provide is a recommendation, one that is perhaps more accurate than it has ever been, but it will take a savvy, caring, and attentive physician and healthcare team to tailor that recommendation to—and with—the individual seated before them. Over 2,000 years ago, Hippocrates said, "It is more important to know what sort of person has [a] disease than to know what sort of disease a person has." In a 1981 editorial on using a computer to interpret risk after exercise stress testing, Robert Califf and Robert Rosati wrote, "Proper interpretation and use of computerized data will depend as much on wise doctors as any other source of data in the past."[2] This is a timeless principle, so long as it is humans we are discussing and not brake parts on an assembly line.

We come back in the end to the glorious fact that we are human, that we are embodied beings, a mind with all its complexities in a body that is equally complex. The interplay between one and the other remains deeply mysterious. What is not mysterious is this: when we are ill, we have a fundamental need to be *cared* for; disease infantilizes us, particularly when it is severe, and though we want the most advanced technical skills, scientific precision, the best therapy, and though we would want our physicians to "know" us (and unlike the time of Hippocrates, such knowing includes the genome, proteome, metabolome, transcriptome, predictions driven by AI, and so on), we badly want it to be expressed in the form of a caring, conscientious physician and healthcare team. We want the physician—a caring individual and not a machine—to give us time, to perform an attentive exam if for no other reason than to acknowledge the locus of disease on our body and not on a biopsy or an image or a report, to validate our personhood and our complaint by touching where it hurts. As Peabody said years ago, the secret of caring for patients is in caring for the patient.

We want those who care for us to know our hearts, our deepest fears, what we live for and would die for.

That is, and it always will be, our deepest desire.

Abraham Verghese, MD
Department of Medicine
Stanford University

chapter one

INTRODUCTION TO DEEP MEDICINE

> By these means we may hope to achieve not indeed a brave new world, no sort of perfectionist Utopia, but the more modest and much more desirable objective—a genuinely human society.
>
> —ALDOUS HUXLEY, 1948

"YOU SHOULD HAVE YOUR INTERNIST PRESCRIBE ANTI-DEPRESSION medications," my orthopedist told me.

My wife and I looked at each other, bug-eyed, in total disbelief. After all, I hadn't gone to my one-month post-op clinic visit following a total knee replacement seeking psychiatric advice.

My knees went bad when I was a teenager because of a rare condition known as osteochondritis dissecans. The cause of this disease remains unknown, but its effects are clear. By the time I was twenty years old and heading to medical school, I had already had dead bone sawed off and extensive reparative surgery in both knees. Over the next forty years, I had to progressively curtail my physical activities, eliminating running, tennis, hiking, and elliptical exercise. Even walking became painful, despite injections of steroids and synovial fluid directly into the knee. And so at age sixty-two I had my left knee replaced, one of the more than 800,000 Americans who have this surgery, the most common orthopedic operation. My orthopedist had deemed me a perfect candidate: I was fairly young, thin, and fit. He said the only significant downside was a 1 to 2 percent risk of infection. I was about to discover another.

After surgery I underwent the standard—and, as far as I was told, only—physical therapy protocol, which began the second day after surgery. The protocol is intense, calling for aggressive bending and extension to avoid scar formation in the joint. Unable to get meaningful flexion, I put a stationary bicycle seat up high and had to scream in agony to get through the first few pedal revolutions. The pain was well beyond the reach of oxycodone. A month later, the knee was purple, very swollen, profoundly stiff, and unbending. It hurt so bad that I couldn't sleep more than an hour at a time, and I had frequent crying spells. Those were why my orthopedist recommended antidepressants. That seemed crazy enough. But the surgeon then recommended a more intensive protocol of physical therapy, despite the fact that each session was making me worse. I could barely walk out of the facility or get in my car to drive home. The horrible pain, swelling, and stiffness were unremitting. I became desperate for relief, trying everything from acupuncture, electro-acupuncture, cold laser, an electrical stimulation (TENS) device, topical ointments, and dietary supplements including curcumin, tart cherry, and many others—fully cognizant that none of these putative treatments have any published data to support their use.

Joining me in my search, at two months post-op, my wife discovered a book titled *Arthrofibrosis.* I had never heard the term, but it turned out to be what I was suffering from. Arthrofibrosis is a complication that occurs in 2 to 3 percent of patients after a knee replacement—that makes the condition uncommon, but still more common than the risk of infection that my orthopedist had warned me about. The first page of the book seemed to describe my situation perfectly: "Arthrofibrosis is a disaster," it said. More specifically, arthrofibrosis is a vicious inflammation response to knee replacement, like a rejection of the artificial joint, that results in profound scarring. At my two-month post-op visit, I asked my orthopedist whether I had arthrofibrosis. He said absolutely, but there was little he could do for the first year following surgery—it was necessary to allow the inflammation to “burn out” before he could go back in and remove the scar tissue. The thought of going a year as I was or having another operation was making me feel even sicker.

Following a recommendation from a friend, I went to see a different physical therapist. Over the course of forty years, she had seen many patients with osteochondritis dissecans, and she knew that, for patients such as me, the routine therapeutic protocol was the worst thing possible. Where the standard protocol called for extensive, forced manipulation to maximize the knee flexion and extension (which was paradoxically stimulating more scar formation), her approach was to go gently: she had me stop all the weights and exercises and use anti-inflammatory medications. She handwrote a page of instructions and texted me every other day to ask how “our knee” was doing. Rescued, I was quickly on the road to recovery. Now, years later, I still have to wrap my knee every day to deal with its poor healing. So much of this torment could have been prevented.

As we’ll see in this book, artificial intelligence (AI) could have predicted that my experience after the surgery would be complicated. A full literature review, provided that experienced physical therapists such as the woman I eventually found shared their data, might well have indicated that I needed a special, bespoke PT protocol. It wouldn’t only be physicians who would get a better awareness of the risks confronting their patients. A virtual medical assistant, residing in my smartphone or my bedroom, could warn me, the patient, directly of the high risk of arthrofibrosis that a standard course of physical therapy posed. And it could even tell me where I could go to get gentle rehab and avoid this dreadful problem. As it was, I was blindsided, and my orthopedist hadn’t even taken my history of osteochondritis dissecans into account when discussing the risk of surgery, even though he later acknowledged that it had, in fact, played a pivotal role in the serious problems that I encountered.

Much of what’s wrong with healthcare won’t be fixed by advanced technology, algorithms, or machines. The robotic response of my doctor to my distress exemplifies the deficient component of care. Sure, the operation was done expertly, but that’s only the technical component. The idea that I should take medication for depression exemplifies a profound lack of human connection and empathy in medicine today. Of course, I was emotionally depressed, but depression wasn't the problem at all: the problem was that I was in severe pain and had Tin Man immobility. The orthopedist's lack of compassion was palpable: in all the months after the surgery, he never contacted me once to see how I was getting along. The physical therapist not only had the medical knowledge and experience to match my condition, but she really cared about me. It's no wonder that we have an opioid epidemic when it's a lot quicker and easier for doctors to prescribe narcotics than to listen to and understand patients.

Almost anyone with chronic medical conditions has been "roughed up" like I was—it happens all too frequently. I'm fortunate to be inside the medical system, but, as you have seen, the problem is so pervasive that even insider knowledge isn't necessarily enough to guarantee good care. Artificial intelligence alone is not going to solve this problem on its own. We need humans to kick in. As machines get smarter and take on suitable tasks, humans might actually find it easier to be more humane.

AI in medicine isn't just a futuristic premise. The power of AI is already being harnessed to help save lives. My close friend, Dr. Stephen Kingsmore, is a medical geneticist who heads up a pioneering program at the Rady Children's Hospital in San Diego. Recently, he and his team were awarded a Guinness World Record for taking a sample of blood to a fully sequenced and interpreted genome in only 19.5 hours.[1]

A little while back, a healthy newborn boy, breastfeeding well, went home on his third day of life. But, on his eighth day, his mother brought him to Rady's emergency room. He was having constant seizures, known as status epilepticus. There was no sign of infection. A CT scan of his brain was normal; an electroencephalogram just showed the electrical signature of unending seizures. Numerous potent drugs failed to reduce the seizures; in fact, they were getting even more pronounced. The infant's prognosis, including both brain damage and death, was bleak.

A blood sample was sent to Rady's Genomic Institute for a rapid whole-genome sequencing. The sequence encompassed 125 gigabytes of data, including nearly 5 million locations where the child's genome differed from the most common one. It took twenty seconds for a form of AI called natural-language processing to ingest the boy's electronic medical record and determine eighty-eight phenotype features (almost twenty times more than the doctors had summarized in their problem list). Machine-learning algorithms quickly sifted the approximately 5 million genetic variants to find the roughly 700,000 rare ones. Of those, 962 are known to cause diseases. Combining that information with the boy's phenotypic data, the system identified one, in a gene called ALDH7A1, as the most likely culprit. The variant is very rare, occurring in less than 0.01 percent of the population, and causes a metabolic defect that leads to seizures. Fortunately, its effects can be overridden by dietary supplementation with vitamin B6 and arginine, an amino acid, along with restricting lysine, a second amino acid. With those changes to his diet made, the boy's seizures abruptly ended, and he was discharged home thirty-six hours later! In follow-up, he is perfectly healthy with no sign of brain damage or developmental delay.

The key to saving this boy's life was determining the root cause of his condition. Few hospitals in the world today are sequencing the genomes of sick newborns and employing artificial intelligence to make everything known about the patient and genomics work together. Although very experienced physicians might eventually have hit upon the right course of treatment, machines can do this kind of work far quicker and better than people.

So, even now, the combined efforts and talents of humans and AI, working synergistically, can yield a medical triumph. Before we get too sanguine about AI's potential, however, let's turn to a recent experience with one of my patients.

"I want to have the procedure," my patient told me on a call after a recent visit.

A white-haired, blue-eyed septuagenarian who had run multiple companies, he was suffering from a rare and severe lung condition known as idiopathic—a fancy medical word for "of unknown cause"—pulmonary fibrosis. It was bad enough that he and his pulmonologist had been considering a possible lung transplant if it got any worse. Against this backdrop he began to suffer a new symptom: early-onset fatigue that left him unable to walk more than a block or swim a lap. He had seen his lung doctor and had undergone pulmonary function tests, which were unchanged. That strongly suggested his lungs weren't the culprit.

He, along with his wife, then came to see me, very worried and depressed. He took labored, short steps into the exam room. I was struck by his paleness and look of hopelessness. His wife corroborated his description of his symptoms: there had been a marked diminution of his ability to get around, to even do his daily activities, let alone to exert himself.

After reviewing his history and exam, I raised the possibility that he might have heart disease. A few years previously, after he began to suffer calf pain while walking, he had stenting of a blockage in his iliac artery to the left leg. This earlier condition raised my concern about a cholesterol buildup in a coronary artery, even though he had no risk factors for heart disease besides his age and sex, so I ordered a CT scan with dye to map out his arteries. The right coronary artery showed an 80 percent narrowing, but the other two arteries were free of significant disease. It didn't fit together. The right coronary artery doesn't supply very much of the heart muscle, and, in my thirty years as a cardiologist (twenty of which involved opening coronary arteries), I couldn't think of any patients with such severe fatigue who had narrowing in only the right coronary artery.

I explained to him and to his wife that I really couldn't connect the dots, and that it might be the case of a "true-true, unrelated"—that the artery's condition might have nothing to do with the fatigue. His underlying serious lung condition, however, made it conceivable that the narrowing was playing a role. Unfortunately, his lung condition also increased the risk of treatment.

I left the decision to him. He thought about it for a few days and decided to go for stenting his right coronary artery. I was a bit surprised, since over the years he had been so averse to any procedures and even medications. Remarkably, he felt energized right after the procedure was done. Because the stent was put in via the artery of his wrist, he went home just a few hours later. By that evening, he had walked several blocks and before the week's end he was swimming multiple laps. He told me he felt stronger and better than he had for several years. And, months later, the striking improvement in exercise capacity endured.

What's remarkable about this story is that a computer algorithm would have missed it. For all the hype about the use of AI to improve healthcare, had it been applied to this patient's data and the complete corpus of medical literature, it would have concluded not to do the procedure because there's no evidence that indicates the opening of a right coronary artery will alleviate symptoms of fatigue—and AI is capable of learning what to do only by examining existing evidence. And insurance companies using algorithms certainly would have denied reimbursement for the procedure.

But the patient manifested dramatic, sustained benefit. Was this a placebo response? That seems quite unlikely—I've known this man for many years, and he tends to minimize any change, positive or negative, in his health status. He seems a bit like a Larry David personality with curbed enthusiasm, something of a curmudgeon. Ostensibly, he would be the last person to exhibit a highly exaggerated placebo benefit.

In retrospect, the explanation likely does have something to do with his severe lung disease. Pulmonary fibrosis results in high pressures in the pulmonary arteries, which feed blood to the lungs, where the blood becomes oxygenated. The right ventricle is responsible for pumping that blood to the heart; the high blood pressure in the arteries meant that it would have taken a lot of work to force more blood in. That would have stressed the right ventricle; the stent in the right coronary artery, which supplies the right ventricle, would have alleviated the stress on this heart chamber. Such a complex interaction of one person's heart blood supply with a rare lung disease had no precedent in the medical literature.

This case reminds us that we're each a one-of-a-kind intricacy that will never be fully deconvoluted by machines. The case also highlights the human side of medicine: We physicians have long known that patients know their body and that we need to listen to them. Algorithms are cold, inhumane predictive tools that will never know a human being. Ultimately, this gentleman had a sense that his artery narrowing was the culprit for his symptoms, and he was right. I was skeptical and would certainly not have envisioned the magnitude of impact, but I was thrilled he improved.

AI HAS BEEN sneaking into our lives. It is already pervasive in our daily experiences, ranging from autocomplete when we type, to unsolicited recommendations based on Google searches, to music suggestions based on our listening history, to Alexa answering questions or turning out the lights. Conceptually, its roots date back more than eighty years, and its name was coined in the 1950s, but only recently has its potential impact in healthcare garnered notice. The promise of artificial intelligence in medicine is to provide composite, panoramic views of individuals' medical data; to improve decision making; to avoid errors such as misdiagnosis and unnecessary procedures; to help in the ordering and interpretation of appropriate tests; and to recommend treatment. Underlying all of this is data. We're well into the era of Big Data now: the world produces zettabytes (sextillion bytes, or enough data to fill roughly a trillion smartphones) of data each year. For medicine, big datasets take the form of whole-genome sequences, high-resolution images, and continuous output from wearable sensors. While the data keeps pouring out, we've really processed only a tiny fraction of it. Most estimates are less than 5 percent, if that much. In a sense, it was all dressed up with nowhere to go—until now. Advances in artificial intelligence are taming the unbridled amalgamation of Big Data by putting it to work.

There are many subtypes of AI. Traditionally machine learning included logistic regression, Bayesian networks, Random Forests, support vector machines, expert systems, and many other tools for data analysis. For example, a Bayesian network is a model that provides probabilities. If I had a person's symptoms, for example, such a model could yield a list of possible diagnoses, with the probability of each one. Funny that in the 1990s, when we did classification and regression trees to let the data that we collected speak for itself, go into "auto-analyze" mode, without our bias of interpretation, we didn't use the term "machine learning." But now that form of statistics has undergone a major upgrade and achieved venerability. In recent years, AI tools have expanded to deep network models such as deep learning and reinforcement learning (we'll get into more depth in Chapter 4).

The AI subtype of deep learning has gained extraordinary momentum since 2012, when a now-classic paper was published on image recognition.[2]

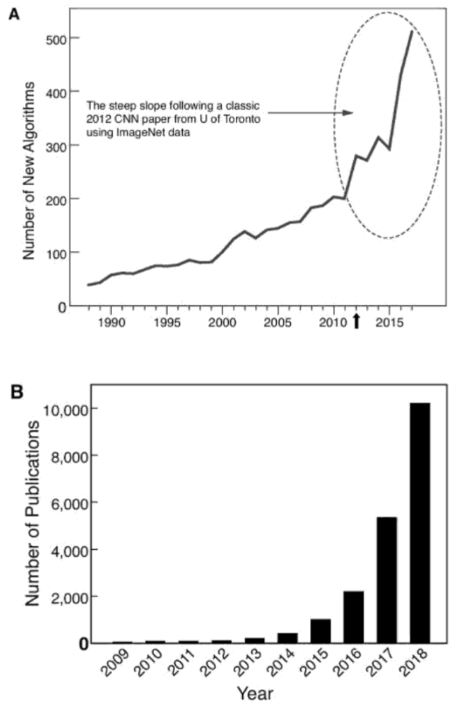


FIGURE 1.1: The increase in deep learning AI algorithms since the 2012 image recognition paper. Sources: Panel A adapted from A. Mislove, "To Understand Digital Advertising, Study Its Algorithms," *Economist* (2018): www.economist.com/science-and-technology/2018/03/22/to-understand-digital-advertising-study-its-algorithms. Panel B adapted from C. Mims, "Should Artificial Intelligence Copy the Human Brain?" *Wall Street Journal* (2018): www.wsj.com/articles/should-artificial-intelligence-copy-the-human-brain-1533355265?mod=searchresults&page=1&pos=1.

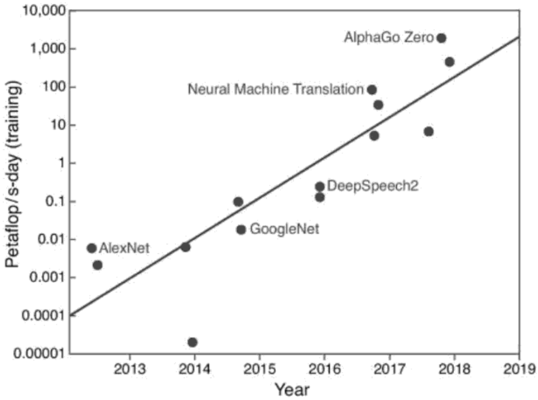


FIGURE 1.2: The exponential growth in computing—300,000-fold—in the largest AI training runs. Source: Adapted from D. Hernandez and D. Amodei, "AI and Compute," *OpenAI* (2018): https://blog.openai.com/ai-and-compute/.

The number of new deep learning AI algorithms and publications has exploded (Figure 1.1), with exponential growth of machine recognition of patterns from enormous datasets. The 300,000-fold increase in petaflops (computing speed equal to one thousand million million [10^{15}] floating-point operations per second) per day of computing used in AI training further reflects the change since 2012 (Figure 1.2).

In the past few years, several studies relying on deep learning have been published in leading peer-reviewed medical journals. Many in the medical community were frankly surprised by what deep learning could accomplish: studies that claim AI's ability to diagnose some types of skin cancer as well as or perhaps even better than board-certified dermatologists; to identify specific heart-rhythm abnormalities like cardiologists; to interpret medical scans or pathology slides as well as senior, highly qualified radiologists and pathologists, respectively; to diagnose various eye diseases as well as ophthalmologists; and to predict suicide better than mental health professionals. These skills predominantly involve pattern recognition, with machines learning those patterns after training on hundreds of thousands, and soon enough millions, of examples. Such systems have just gotten better and better, with the error rates for learning from text-, speech-, and image-based data dropping well below 5 percent, whizzing past the human threshold (Figure 1.3). Although there must be some limit at which the learning stops, we haven't reached it yet. And, unlike humans who get tired, have bad days, may get emotional, sleep deprived, or distracted, machines are steady, can work 24/7 without vacations, and don't complain (although both can get sick). Understandably, this has raised questions about the future role of doctors and what unforeseen impact AI will have on the practice of medicine.

Outperform doctors at all tasks
Diagnose the undiagnosable
Treat the untreatable
See the unseeable on scans, slides
Predict the unpredictable
Classify the unclassifiable
Eliminate workflow inefficiencies
Eliminate hospital admissions and readmissions
Eliminate the surfeit of unnecessary jobs
100% medication adherence
Zero patient harm
Cure cancer

TABLE 1.1: The outlandish expectations for AI in healthcare, a partial list.

I don't believe that deep learning AI is going to fix all the ailments of modern healthcare, but the list in Table 1.1 gives a sense of how widely the tool can be applied and has been hyped. Over time, AI will help propel us toward each of these objectives, but it's going to be a marathon without a finish line.

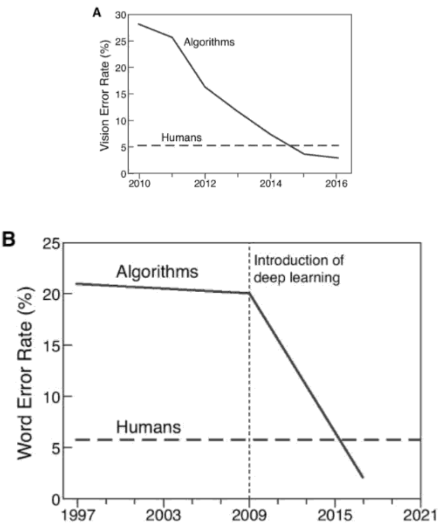


FIGURE 1.3: The increase in machine AI accuracy for image (A) and speech (B) interpretation, both now exceeding human performance for narrow tasks in labeled datasets. Sources: Panel A adapted from V. Sze et al., "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," *Proceedings of the IEEE* (2017): 105(12), 2295–2329. Panel B adapted from "Performance Trends in AI," *Word Press Blog* (2018): https://srconstantin.wordpress.com/2017/01/28/performance-trends-in-ai/.

The deep learning examples are narrow: the depression predictor can't do dermatology. These neural network algorithms depend on recognizing patterns, which is well-suited for certain types of doctors who heavily depend on images, like radiologists looking at scans or pathologists reviewing slides, which I call "doctors with patterns." To a lesser but still significant extent, all clinicians have some patterned tasks in their daily mix that will potentially be subject to AI algorithmic support.

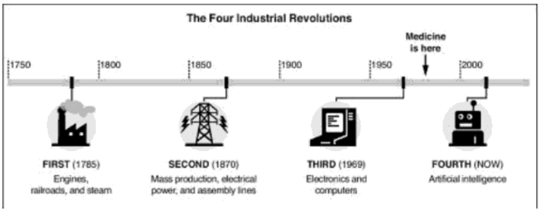


FIGURE 1.4: The four Industrial Revolutions. Source: Adapted from A. Murray, "CEOs: The Revolution Is Coming," *Fortune* (2016): http://fortune.com/2016/03/08/davos-new-industrial-revolution.

Most of the published deep learning examples represent only *in silico,* or computer-based, validation (as compared to *prospective* clinical trials in people). This is an important distinction because analyzing an existing dataset is quite different from collecting data in a real clinical environment. The in silico, retrospective results often represent the rosy best-case scenario, not fully replicated via a forward-looking assessment. The data from retrospective studies are well suited for generating a hypothesis, then the hypothesis can be tested prospectively and supported, especially when independently replicated.

We're early in the AI medicine era; it's not routine medical practice, and some call it "Silicon Valley–dation." Such dismissive attitudes are common in medicine, making change in the field glacial. The result here is that although most sectors of the world are well into the Fourth Industrial Revolution, which is centered on the use of AI, medicine is still stuck in the early phase of the third, which saw the first widespread use of computers and electronics (Figure 1.4). That MP3 files are compatible with every brand of music player, for example, while medicine has yet to see widely compatible and user-friendly electronic medical records exemplifies the field's struggle to change.

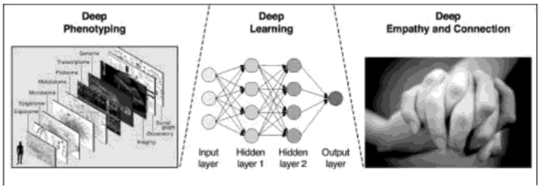


FIGURE 1.5: The three principal components of the deep medicine model. Source (left panel): Adapted from E. Topol, "Individualized Medicine from Prewomb to Tomb," *Cell* (2014): 157(1), 241–253.

This isn't the first time I've noted medicine's reluctance to adopt new technologies. This is the third book that I've written on the future of medicine. In the *Creative Destruction of Medicine,* I mapped out how sensors, sequencing, imaging, telemedicine, and many other technological opportunities enabled us to digitize human beings and achieve the digital transformation of medicine. In *The Patient Will See You Now,* I made the case for how medicine could be democratized—that medical paternalism would fade as consumers didn't simply generate their information but owned it, had far greater access to their medical data, and ultimately could (if they chose to) take considerably more charge of their care.

This book represents the next phase, the third *D* after digitizing and democratizing, and it's the most far-reaching one. Despite whatever impression you might get from my interest in new technology, it has always been a dream of mine to galvanize the essential human element of medical practice. With this third *D* of deep learning, we will have a framework to nurture medicine's roots: the human-human bond. Although we haven't even yet achieved digitization or democratization in medicine, they are slowly progressing, and I believe that we will not just complete them but bring AI into the heart of medicine as well. The culmination of this process is something I call "deep medicine."

Deep medicine requires three deep components (Figure 1.5).

First is the ability to deeply define each individual (digitizing the medical essence of a human being), using all relevant data. This might include all of one's medical, social, behavioral, and family histories, as well as one's biology: anatomy, physiology, and environment. Our biology has multiple layers—our DNA genome, our RNA, proteins, metabolites, immunome, microbiome, epigenome, and more. In the biomedical research community, the term that is frequently used is "deep phenotyping"; we saw an example of the approach in the case of the newborn boy with status epilepticus. Deep phenotyping is both thick, spanning as many types of data as you can imagine, and long, covering as much of our lives as we can, because many metrics of interest are dynamic, constantly changing over time. A few years ago, I wrote a review in which I said we needed medical data that spanned "from prewomb to tomb."[3] A former mentor told me that I should have called the span "from lust to dust." But you get the idea of deep and long data.

Second is deep learning, which will play a big part of medicine's future. It will not only involve pattern recognition and machine learning that doctors will use for diagnosis but a wide range of applications, such as virtual medical coaches to guide consumers to better manage their health or medical condition. It will also take on efficiency in the hospital setting, using machine vision to improve patient safety and quality, ultimately reducing the need for hospital rooms by facilitating remote, at-home monitoring. Although deep learning's outputs in medicine have considerable potential and have been accelerating in the past few years, we're still in the nascent phase. Nearly fifty years ago, William Schwartz published an article in the *New England Journal of Medicine* titled "Medicine and the Computer."[4] He speculated that, in the future, computers and physicians would engage "in frequent dialogue, the computer continuously taking note of history, physical findings, laboratory data, and the like, alerting the physician to the most probable diagnoses and suggesting the appropriate, safest course of action." What do we have to show for this projection from fifty years ago? Surprisingly, not too much. While there are certainly anecdotes about a Google search helping make difficult diagnoses, simple symptom lookup certainly has not been validated as an accurate means of diagnosis—instead, all too often, it serves as the groundwork for inducing anxiety and cyberchondria.

METRIC	1975	NOW
Number of healthcare jobs	4 million	> 16 million (#1 US economy)
Healthcare spending per person	$550/yr.	> $11,000/yr.
Time allotted for office visits	60 min. new, 30 min. return	12 min. new, 7 min. return
% of GDP healthcare	< 8	18
Hospital daily room charge (avg.)	~ $100	$4,600
Miscellaneous	None of these	Relative value units, EHRs, PBMs, "health systems"

TABLE 1.2: Selected metrics of healthcare in the United States that have changed in the past forty-plus years.

One can imagine that AI will rescue medicine from all that ails it, including diagnostic inaccuracy and workflow inefficiencies (such as mundane tasks like billing or coding charts), but none of these have been actualized yet. It's an extraordinary opportunity for entrepreneurs working with clinicians, computer scientists, and researchers in other disciplines (such as behavioral science and bioethics) to help fashion the right integration of AI and healthcare.

The third, and most important, component is deep empathy and connection between patients and clinicians. In the more than four decades since I started medical school, I've watched the steady degradation of the human side of medicine, outlined in Table 1.2. Over that span of time, healthcare became not just a big business but, by the end of 2017, the biggest. It is now the largest employer in the United States, towering over retail. By every metric, the amount of money spent on healthcare has exploded. Yet, even with all the employment in the sector and all the money expended per person, the time spent between doctors and patients has steadily dwindled, whether for office visits or in the hospital. Doctors are much too busy. The exorbitant charge of almost $5,000 for a day in the hospital might only include a few minutes of your doctor coming by to visit (for which there's another charge). Consumed by patient care, physicians were passive while major new changes took hold in the business of healthcare, including electronic health records, managed care, health maintenance organizations, and relative value units. Now, the highest-ever proportion of doctors and nurses are experiencing burnout and depression owing to their inability to provide real care to patients, which was their basis for pursuing a medical career.

What's wrong in healthcare today is that it's missing care. That is, we generally, as doctors, don't get to really care for patients enough. And patients don't feel they are cared for. As Francis Peabody wrote in 1927, "The secret of the care of the patient is caring for the patient."[5] The greatest opportunity offered by AI is not reducing errors or workloads, or even curing cancer: it is the opportunity to restore the precious and time-honored connection and trust—the human touch—between patients and doctors. Not only would we have more time to come together, enabling far deeper communication and compassion, but also we would be able to revamp how we select and train doctors. We have prized "brilliant" doctors for decades, but the rise of machines will heighten the diagnostic skills and the fund of medical knowledge available to all clinicians. Eventually, doctors will adopt AI and algorithms as their work partners. This leveling of the medical knowledge landscape will ultimately lead to a new premium: to find and train doctors who have the highest level of emotional intelligence. My friend and colleague, Abraham Verghese, whom I regard as one of the great humanists of medicine, has emphasized these critical points in the foreword of the book, which I hope you have taken the time to read carefully. This is what deep medicine offers.

To develop the conceptual framework of deep medicine, I'll start with how medicine is practiced now and why we desperately need new solutions to such problems as misdiagnosis, errors, poor outcomes, and runaway costs. That, in part, hinges on the basics of how a medical diagnosis is made today. To understand the reward and risk potential of AI, we will explore the AI precedents, the accomplishments ranging from games to self-driving cars. Of equal, and perhaps even greater, importance will be an exploration of AI's liabilities, such as human bias, the potential for worsening inequities, its black-box nature, and concerns for breaches of privacy and security. The transfer of tens of millions of people's personal data from Facebook to Cambridge Analytica, who then used AI to target individuals, illustrates one critical aspect of what could go wrong in the healthcare context.

Then we're ready to move on to the new medicine that will integrate the tools of AI. We'll assess how machine pattern recognition will affect the practice of radiologists, pathologists, and dermatologists—the doctors with patterns. But AI will cut across all disciplines of medicine, even "clinicians without patterns" and surgeons. One field that is especially in urgent need of new approaches is mental health, with a profound mismatch of the enormous burden of conditions like depression and the limited number of trained professionals to help manage or prevent it. AI will likely prove to have a critical role in mental health going forward.

But AI, and specifically deep learning, won't just affect the practice of medicine. In a complementary way, it will also transform biomedical science. For example, it will facilitate the discovery of new drugs. It will also extract insights from complex datasets, such as millions of whole genome sequences, the intricacies of the human brain, or the integrated streaming of real-time analytics from multiple biosensor outputs. These endeavors are upstream from the care of patients, but catalyzing advances in basic science and drug development will ultimately have a major effect in medicine.

AI can also revolutionize other aspects of our lives that are, in one sense or another, upstream from the clinic. A huge one is how we eat. One of the unexpected and practical accomplishments of machine learning to date has been to provide a potential scientific basis for individualized diets. That's conceivably an exciting advance—the idea of knowing what specific foods are best for any given person. We can now predict in healthy people, without diabetes, what particular foods will spike their blood sugar. Such advances far outstrip whatever benefits might accrue from following a diet for all people, such as the classic food pyramids, or fad diets like Atkins or South Beach, none of which ever had a solid evidence basis. We'll review that fascinating body of data and forecast where smart nutrition may go in the future. Many of these at-home advances will come together in the virtual medical coach. It most likely will be voice mediated, like Siri, Alexa, and Google Home, but unlikely to remain a cylinder or a squiggle on a screen. I suspect they're more apt to come in the form of a virtual human avatar or hologram (but simply text or e-mail if one prefers). The virtual medical coach is the deep learning of all of one's data, seamlessly collected, continuously updated, integrated with all biomedical knowledge, and providing feedback and coaching. Such systems will initially be condition specific, say for diabetes or high blood pressure, but eventually they'll offer a broad consumer health platform to help prevent or better manage diseases.

All this potential, however, could be spoiled by misuse of your data. This encompasses not just the crimes we've seen all too much of so far, such as cybertheft, extortion (hospitals having their data held for ransom), and hacking, but the nefarious and large-scale sale and use of your data. The new, worrisome, unacceptable wrinkle could be that insurance companies or employers get hold of all your data—and what has been deep learned about you—to make vital decisions regarding your health coverage, your premiums, or your job. Avoiding such dreadful scenarios will take deliberate and intense effort.

This book is all about finding the right balance of the patients, doctors, and machines. If we can do that—if we can exploit machines' unique strengths to foster an improved bond between humans—we'll have found a vital remedy for what profoundly ails our medicine of today.

I hope to convince you that deep medicine is both possible and highly desirable. Combining the power of humans and machines—intelligence both human and artificial—would take medicine to an unprecedented level. There are plenty of obstacles, as we'll see. The path won't be easy, and the end is a long way off. But with the right guard rails, medicine can get there. The increased efficiency and workflow could either be used to squeeze clinicians more, or the gift of time could be turned back to patients—to use the future to bring back the past. The latter objective will require human activism, especially among clinicians, to stand up for the best interest of patients. Like the teenage students of Parkland rallying against gun violence, medical professionals need to be prepared to fight against some powerful vested interests, to not blow this opportunity to stand up for the primacy of patient care, as has been the case all too often in the past. The rise of machines has to be accompanied by heightened humaneness—with more time together, compassion, and tenderness—to make the “care” in healthcare real. To restore and promote care. Period.

Let’s get started.

chapter two

SHALLOW MEDICINE

> Imagine if a doctor can get all the information she needs about a patient in 2 minutes and then spend the next 13 minutes of a 15-minute office visit talking with the patient, instead of spending 13 minutes looking for information and 2 minutes talking with the patient.
>
> —Lynda Chin

"He told me I need a procedure to plug the hole in my heart," my patient, whom I'll call Robert, said at the beginning of our first encounter. Robert is a fifty-six-year-old store manager who had been healthy until a few years ago, when he had a heart attack. Fortunately, he received timely treatment with a stent, and there was very little damage to his heart. Since that time, he had markedly improved his lifestyle, losing and keeping off more than twenty-five pounds while exercising regularly and rigorously.

So, it was devastating for him when, out of the blue one afternoon, he began to have trouble seeing and developed numbness in his face. During an evaluation in the emergency room of a nearby hospital, the symptoms continued while he had an urgent head CT scan, some blood tests, a chest X-ray, and an electrocardiogram. Over the course of the day, without any treatment, his vision gradually returned to normal, and the numbness went away. The doctors told him that he had suffered “just” a ministroke, or transient ischemic attack, and that he should continue taking an aspirin each day as he had been since the heart attack. The lack of any change in strategy or new medication left him feeling vulnerable to another event. He set up an appointment with a neurologist for a couple of weeks later. Robert thought maybe that way he would get to the bottom of the problem.

The neurologist did some additional tests, including an MRI of his brain and an ultrasound evaluation of the carotid arteries in his neck, but he did not find anything to explain the transient stroke. He referred Robert to a cardiologist. The heart doctor did an echocardiogram that showed a patent foramen ovale (PFO). That’s a tiny hole in the wall that separates the heart’s two atria, the collecting chambers. Present in all fetuses (because it keeps blood from flowing to the lungs before we need to breathe), it typically closes when we take our first breaths; nevertheless, it remains open in about 15 to 20 percent of adults. “A-ha!” the cardiologist exclaimed to Robert. “This echo cinched the diagnosis.” The cardiologist thought that a blood clot must have moved across the heart chambers and trekked up to his brain, where it caused the ministroke. To avoid any future strokes, he said, Robert needed to have a procedure to plug up that hole. It was scheduled for ten days later.

Well, Robert wasn’t so sure about this explanation or the need for the procedure. He spoke to a mutual friend and soon came to see me for a second opinion. I was alarmed. Robert’s PFO anatomy was far too common to be the definitive cause of the stroke based on such a minimal evaluation. Before invoking the hole as the cause of the stroke, a physician needs to exclude every other diagnosis. Plenty of people have such holes in their hearts and strokes, and they’re not at all connected. If they were, a lot more of the one in five of us with PFOs would be suffering strokes. Furthermore, multiple randomized trials have tested the effectiveness of the treatment for cryptogenic strokes so-called because they have no known cause. Although these trials showed a consistent reduction in the number of subsequent stroke events, the implant and procedure lead to enough complications that the net benefit was marginal. And for Robert that was even more questionable because he did not have a full stroke, and his evaluations were not extensive enough to force us to fall back on the cryptogenic, default diagnosis yet.

Together, he and I developed a plan to hunt for other possible causes of the ministroke. One very common cause is a heart-rhythm disorder known as atrial fibrillation. To investigate that possibility, I ordered an unobtrusive Band-Aid-like patch called a Zio (made by iRhythm) for Robert to wear on his chest for ten to fourteen days. A chip in the patch captures an electrocardiogram of every heartbeat during the period in which it is worn. Robert wore his for twelve days. A couple of weeks later I got the results. Sure enough, Robert had several, otherwise asymptomatic, bouts of atrial fibrillation during that time. He didn't have any other symptoms because the heart rate never got too fast, and a few of the episodes occurred while he was asleep. The atrial fibrillation was a much more likely cause of the ministroke than the hole in his heart. We could use a blood thinner to hopefully prevent future events, and there was no need to move ahead with plugging the hole. Yes, there was a small risk of bleeding complications from the new medicine, but the protection from a future stroke warranted that trade-off. Robert was relieved when we discussed the diagnosis, treatment, and prognosis.

I don't present Robert because we were able to nail down the likely diagnosis. Although his story has a happy ending, it also represents everything wrong with medicine today. His experience, from the emergency room through the first visit to a cardiologist, is what I call shallow medicine. Rather than an emotional connection between patients and doctors, we have an emotional breakdown, with disenchanted patients largely disconnected from burned-out, depressed doctors. At the same time, there is a systemic problem with mistaken and excessive diagnosis, both of which can result in significant economic waste and human harm. In fact, the deficiencies in the patient-doctor relationship and errors in medical practice are interdependent: the superficial contact with patients promotes incorrect diagnoses and the reflexive ordering of tests or treatments that are unnecessary or unsound.

Misdiagnosis in the United States is disconcertingly common. A review of three very large studies concluded that there are about 12 million significant misdiagnoses a year.[1] These mistakes result problem is treatment that flies in the face of evidence, but another part of the problem is the evidence used to make the decisions to treat. In medicine, we often rely on changes in the frequency of so-called surrogate endpoints instead of the frequency of endpoints that really matter. So with heart disease we might treat based on changes in blood pressure because we have no evidence about whether the treatment actually changes the frequency of heart attacks, strokes, or death. Or we might measure the treatment of diabetes by monitoring changes in glycosylated hemoglobin (A1c) instead of life expectancy and widely accepted quality-of-life measures. Although the surrogate symptoms might seem like reasonable stand-ins for the overarching goals, very few of them have held up to scrutiny. Nevertheless, the flimsy evidence that caused physicians to monitor the surrogates in the first place has bred overuse of tests, procedures, and medications.

Shallow evidence, either obtained from inadequate examination of an individual patient like Robert, or from the body of medical literature, leads to shallow medical practice, with plenty of misdiagnoses and unnecessary procedures. This is not a minor problem. In 2017 the American Heart Association and American College of Cardiology changed the definition of high blood pressure, for example, leading to the diagnosis of more than 30 million more Americans with hypertension despite the lack of any solid evidence to back up this guideline.[7] This was misdiagnosis at an epidemic scale.

Figure 2.1: Word cloud describing doctors.

Source: Adapted from B. Singletary et al., "Patient Perceptions About Their Physician in 2 Words: The Good, the Bad, and the Ugly," *JAMA Surg* (2017): 152(12), 1169–1170.

Yet even without such central dictates, the way medical practice works at a one-to-one level is a setup for misdiagnosis. The average length of a clinic visit in the United States for an established patient is seven minutes; for a new patient, twelve minutes. This preposterous lack of time is not confined to America. When I visited the Samsung Medical Center in South Korea a couple of years ago, my hosts told me that the doctor visits averaged only two minutes. Is it any wonder that there are so many mistaken diagnoses? Both patients and doctors believe that doctors are rushed. Recently, for example, the medical center at the University of Alabama at Birmingham asked patients what two words best describe its doctors.[8] The response, plotted as a word cloud in Figure 2.1, is telling.

It's not just the length of a visit. Because of electronic health records, eye contact between the patient and doctor is limited. Russell Phillips, a Harvard physician said, "The electronic medical record has turned physicians into data entry technicians."[9] Attending to the keyboard, instead of the patient, is ascribed as a principal reason for the medical profession's high rates of depression and burnout. Nearly half of doctors practicing in the United States today have symptoms of burnout, and there are hundreds of suicides per year.[10] In a recent analysis of forty-seven studies involving 42,000 physicians, burnout was associated with a doubling of risk of patient safety incidents, which sets up a vicious cycle of more burnout and depression.[11] Abraham Verghese nailed this in the book's foreword, the role of the "intruder" and its impact on doctors' mental health, which beyond clinicians has potential impact on patient care.

The use of electronic healthcare records leads to other problems. The information that they contain is often remarkably incomplete and inaccurate. Electronic records are very clunky to use, and most—an average of 80 percent—of each note is simply copied and pasted from a previous note.[12] Any mistakes made on one visit are very likely to be propagated to the next. And getting records from other doctors and health systems is exceptionally difficult, in part because of proprietary issues: software companies use file formats that do not work on competitors' software, and health systems take advantage of proprietary file formats to help lock patients in. As my radiologist friend, Saurabh Jha, aptly put it on Twitter: “Your ATM card works in Outer Mongolia, but your electronic health record can’t be used in a different hospital across the street.”[13]

The incompleteness of the records is accentuated by one-off medicine. By “one-off,” I’m not just referring to the brevity or rarity of the interaction. We haven’t had access to patients in their real world, on the go, at work, while asleep. The data doctors access are from the contrived setting of a medical office, constrained by the temporal limits of the visit itself. A patient wearing a patch like I had Robert do is exceedingly rare. For the most part, we have no idea of what any given individual’s real-life medical metrics—such as blood pressure, heart rate and rhythm, or level of anxiety and mood—actually are. In fact, even if we did know this for someone, we wouldn’t have a means to make useful comparisons, as we don’t even yet know what is normal for the population as a whole in a real-world context. This is worsened by the outmoded means of communication doctors do—or don’t—use to communicate with patients outside the clinic. Outside of medicine, people have learned how to maintain a close relationship with their family and friends through e-mail, texting, and video chats, even when they are in remote parts of the world. But more than two-thirds of doctors still do not leverage digital communication to augment their relationship with patients. The unwillingness to e-mail or text has been attributed to lack of time, medicolegal concerns, and lack of reimbursement, but I see them as another example of physicians’ thin connection with their patients.

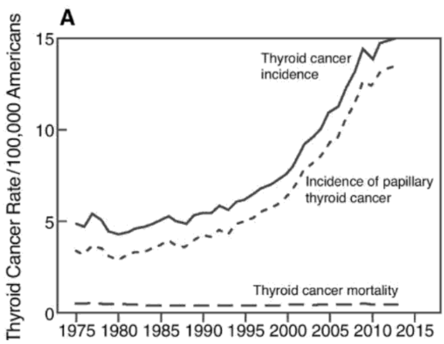
This is where we are today: patients exist in a world of insufficient data, insufficient time, insufficient context, and insufficient presence. Or, as I say, a world of shallow medicine.

THE OUTGROWTHS OF shallow medicine are waste and harm. Let’s take the example of medical screening today. In the United States, mammography is recommended annually for women in their fifties. The total cost of the screening alone is more than $10 billion per year. Worse, if we consider 10,000 women in their fifties who have mammography each year for ten years, only five (0.05 percent) avoid a breast cancer death, while more than 6,000 (60 percent) will have at least one false positive result.[14] The latter might result in the harm and expense of a number of unnecessary procedures, including biopsies, surgery, radiation, or chemotherapy; at the very least, it results in considerable fear and anxiety.

Remarkably parallel to mammography is the use of prostate-specific antigen (PSA) screening for prostate cancer in men. Despite the American Urological Association's 2013 recommendation against the routine use of PSA screening, it is still widely practiced. Each year about 30 million American men are screened, 6 million have an elevated PSA, and 1 million have prostate biopsies. Approximately 180,000, or 18 percent, do get a diagnosis of prostate cancer, but an equal number of men have prostate cancer that the biopsy missed.[15] Added to this issue is the well-established, but frequently ignored, fact that most prostate cancer is indolent and will never threaten a patient's life. Multiple studies have validated genomic markers of the tumor that indicate aggressiveness and increased propensity to spread, but this information is still not incorporated into clinical practice.[16] Overall, the result is that there will be one prostate cancer death averted per one thousand men screened.[17] If you're overly optimistic, you could conclude that this benefit is twice as much as mammography (0.5 per 1,000)! Another way to look at the data: a man is 120 to 240 times more likely to be misdiagnosed from an abnormal PSA and 40 to 80 times more likely to have unnecessary radiation therapy or surgery than to have his life saved.

Cancer screening exemplifies almost every problem with shallow medicine. Back in 1999, South Korea started a national screening program for many types of cancer. The program was free of charge or involved a nominal copay for people with above-average income, which meant that a great many people participated in it. One of the tests was an ultrasound of the thyroid. In just over a decade, the rate of thyroid cancer diagnosis increased fifteenfold, making it the most common form of cancer in South Korea, with more than 40,000 people carrying the diagnosis. This might sound like a victory, but it was a meaningless diagnosis; there was no change of outcomes, including no difference in mortality related to thyroid cancer in South Korea, despite the widespread detection.[18]

This thyroid cancer screening story was replicated in the United States. A decade ago there were ads to "check your neck" with text: "Thyroid cancer doesn't care how healthy you are. It can happen to anyone, including you. That's why it's the fastest growing cancer in the U.S."[19] That turned out to be a self-fulfilling prophecy, leading to a big spike in incidence, as seen in Figure 2.2. More than 80 percent of the people diagnosed underwent thyroid gland removal and had to take medication to replace the hormones that the thyroid normally produces; almost half had radiation therapy of their neck. As was seen in South Korea, there was no sign that this aggressive diagnosis and treatment had any impact on outcomes. That's besides the danger of unnecessary radiation therapy itself.



increased chance they will be diagnosed with kidney cancer and have surgery to remove the kidney. That may sound absurd, but 4 percent of those patients die within ninety days from the surgery itself. What's more, there is no improvement in overall cancer survival in those who do survive the surgery.[23]

No test should be done on a willy-nilly, promiscuous basis, but rather its appropriateness should be gauged by the individual having some risk and suitability to be tested.

In the United States, we are now spending more than $3.5 trillion per year for healthcare. As seen in Table 2.1, for 2015, the number one line item is hospitals, accounting for almost a third of the costs.[24] The proportion attributable to doctors has remained relatively constant over many decades, at approximately one-fifth the costs. Prescription drugs are on a runaway course, accounting for well over $320 billion in 2015 and projected to reach over $600 billion by 2021.[25] New specialty drugs for cancer and rare diseases are routinely launched with price tags starting at $100,000 per treatment or year and ranging up to nearly $1 million per year.

CATEGORY	DOLLARS SPENT
Hospital care	1.0 trillion
Physician and clinical services	635 billion
Prescription drugs	325 billion
Next cost health insurance	210 billion
Nursing home and continuing care	157 billion
Dental services	118 billion
Structures and equipment	108 billion
Home healthcare	89 billion
Other professional services	88 billion
Government and public health activities	81 billion
Other durable medical products	59 billion
Research	47 billion
Government administration	43 billion

Table 2.1: Healthcare spending in the United States in 2015.

Part of this growth is fueled by a shared belief among both patients and physicians that medications, and in particular very expensive ones, will have remarkable efficacy. When doctors prescribe any medication, they have a cognitive bias that it will work. Patients, too, believe the medicine will work. From an enormous body of randomized clinical trials, patients assigned to the placebo arm consistently have more treatment effect than expected, given that they are taking an inert substance.

A few years ago, Nicholas Schork, a former faculty member at Scripps Research with me, put together the responsiveness—the intended clinical response—of the top ten drugs by gross sales.[26] As seen in Figure 2.3, the proportion of people who don't respond to these drugs is well beyond the common perception. Taking Abilify as an example, only one in five patients is actually deriving clinical benefit from the drug. Overall, 75 percent of patients receiving these leading medications do not have the desired or expected benefit. With several of these drugs with sales of more than $10 billion per year (such as Humira, Enbrel, Remicade), you can quickly get a sense of the magnitude of waste incurred.

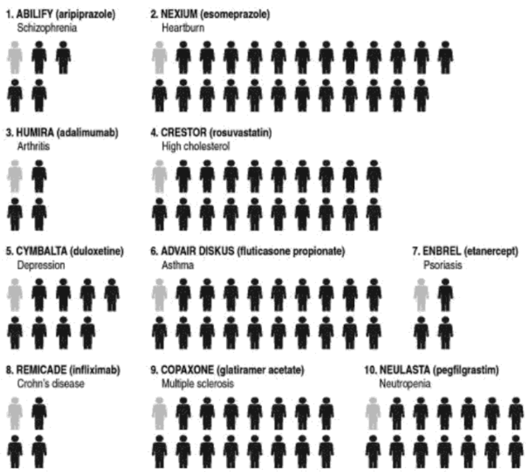


FIGURE 2.3: Schematic showing the number of people with clinical responsiveness to top ten drugs by gross sales in 2014. The gray schematic people represent clinical responders, the black nonresponders. Source: Adapted from N. Schork, "Personalized Medicine: Time for One-Person Trials," *Nature* (2015): 520(7549), 609–611.

These data do not simply illustrate that medicines don't work or are some kind of profiteering racket. Rather, in most cases these drugs don't work because physicians have not honed an ability to predict what sort of person will respond to a treatment or acquired adequate knowledge about an individual to know whether the patient is among those people who will respond positively to a treatment. It adds to the continuum, from unintelligent diagnosis to treatment, of pervasive medical miscues, unnecessary interventions, and overuse problems that plague clinical practice today.

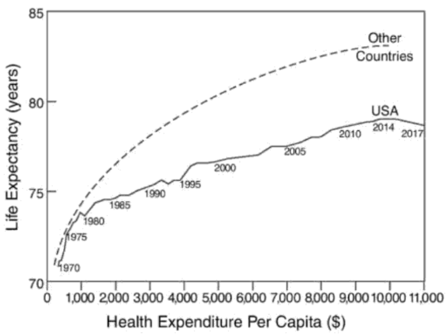
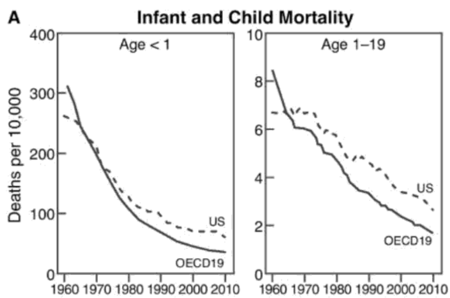


FIGURE 2.4: Life expectancy for twenty-four countries compared with the United States, plotted against health expenditures per person, from 1970 to 2017. Source: Adapted from M. Roser, "Link Between Health Spending and Life Expectancy: US Is an Outlier," *Our World in Data* (2017): https://ourworldindata.org/the-link-between-life-expectancy-and-health-spending-us-focus.

With all the unnecessary testing and treatment, misdiagnosis, and incidental findings that are chased down (and may do harm), we can look at perhaps the three most important measures of the efficacy of a healthcare system: longevity, infant/childhood mortality, and maternal mortality. They all look bad in the United States, and distinctly worse than the eighteen other member countries of the Organization for Economic Cooperation and Development (OECD) and beyond this group of countries (Figures 2.4 and 2.5). There are certainly other explanations that account for these outliers, such as the striking socioeconomic inequities in the United States, which continue to increase. For example, that appears to be a highly significant factor for the *alarming* and *disproportionate* maternal mortality rate among Black women.[27] I am not suggesting that the other countries are practicing deep medicine. In fact, my contention is that we, in the United States, are overindulged in shallow medicine. The evidence for overutilization, which is not the case for individuals of low socioeconomic status (who have problems with basic access), is quite compelling and contributory. That our life expectancy is the singular one declining, while at the same time our healthcare spending is increasing, is deeply concerning.



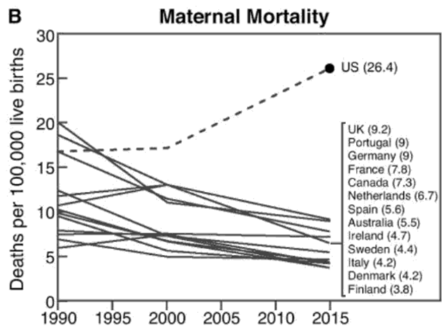


FIGURE 2.5: Reports showing the outlier status of the United States for key outcomes of (A) infant and child mortality and (B) maternal mortality. Sources: Panel A adapted from A. Thakrar et al., "Child Mortality in the US and 19 OECD Comparator Nations: A 50-Year Time-Trend Analysis," *Health Affairs* (2018): 37(1), 140–149. Panel B adapted from GBD Maternal Mortality Collaborators, "Global, Regional, and National Levels of Maternal Mortality, 1990–2015: A Systematic Analysis for the Global Burden of Disease Study 2015," *Lancet* (2016): 388(10053).

For many years, healthcare economists have been talking about "bending the curve," meaning reducing costs to achieve the same or better outcomes. But with longevity declining in past years in the United States, along with the continued steep increase in expenditures, we are indeed bending the curve, but in the wrong direction!

I hope that I've been able to convince you that the shallow medicine we practice today is resulting in extraordinary waste, suboptimal outcomes, and unnecessary harm. Shallow medicine is unintelligent medicine. This recognition is especially apropos in the information era, a time when we have the ability to generate and process seemingly unlimited data about and for any individual. To go deep. To go long and thick with our health data.

black wing-tip shoes. His goal for this morning was to teach us the basics of how to make a diagnosis.

He went over to the chalkboard (there were no whiteboards in 1977) and started writing some features about a patient.

First he wrote, “Sixty-six-year-old male presents to the emergency room.”

Then he asked, “What is in our differential diagnosis?”

Now this would seem peculiar since there was so little information to go on. But Dr. Moss’s point was that every time a physician assesses a case, we need to process every nugget of information—be it a symptom, a sign, or a lab test result—and quickly come up with the most common causes that fit the picture.

The responses he got from our group of uninitiated doctor-wannabees were heart attack, cancer, stroke, and accident.

Then he added another feature: chest pain.

The group concluded that he must be having a heart attack.

Dr. Moss, looking at us wryly, told us that we were all wrong. We need to be thinking of other reasons for chest pain in such a patient. Then we came up with other possibilities like aortic dissection, esophageal spasm, pleurisy, pericarditis, and a contusion of the heart.

He now wrote on the board that the chest pain was radiating to the neck and the back. We zeroed in on heart attack and aortic dissection. Then he added that the patient had briefly passed out, which led to our final diagnosis of aortic dissection. Moss smiled and said, “Correct.” He told us to never forget that possibility of aortic dissection when you see a patient with chest pain. That it was too often not considered and missing it could prove to be a fatal mistake.

Next it was time for a more difficult challenge. After erasing the board, he wrote, “Thirty-three-year-old woman is admitted to the hospital.”

We responded with breast cancer, complication of pregnancy, accident. Moss was disappointed we didn’t have more ideas to offer. The next feature he gave us was rash.

Our differential now extended to an infection, an adverse reaction to a drug, an insect or animal bite, a bad case of poison ivy. Our mentor, again looking a bit discouraged with us, had to add another feature to help us: facial rash. That didn’t seem to get us on the right track. We were stuck with the same differential list. So, he added one more descriptor about his made-up patient: she was African American.

One of our group whispered, “Lupus?”

That was the right answer. She nailed it, knowing that lupus is much more common in young women of African ancestry and one of its hallmarks is a butterfly facial rash.

This is how we learned to make a medical diagnosis. It was top-down, immediately reacting to a few general descriptors and quickly coming up with a short list of hypotheses, conjectures, tentative conclusions. We were instilled with the mantra that common things occur commonly, an outgrowth of the same logic that underlies Bayes's theorem. We were getting programmed to use our intuitive sense of recognition rather than our analytical skills. But Bayes's theorem relies on priors, and, because we, as inexperienced medical students, had visited so many books but so few patients, we didn't have much to go on. The method would leave aged physicians, who had seen thousands of patients, in far better stead.

The diagnostic approach we were being taught is an example of what Danny Kahneman would one day classify as System 1 thinking—thinking that is automatic, quick, intuitive, effortless.[1] This system of thinking uses heuristics, or rules of thumb: the reflexive, mental shortcuts that bypass any analytic process, promoting rapid solutions to a problem. System 2 thinking, in contrast, is a slow, reflective process involving analytic effort. It occurs in a different area of the brain and even has distinct metabolic requirements. One might think that master diagnosticians would rely on System 2 thinking. But no, multiple studies have shown that their talent is tied to heuristics admixed with intuition, experience, and knowledge. Indeed, more than forty years ago, System 1 thinking, represented by the rapid, reflexive hypothesis generation method every physician is taught, was shown to be the prototype for getting the right diagnosis. If a doctor thought of the correct diagnosis within five minutes of seeing a patient, the accuracy was a stunning 98 percent. Without having the diagnosis in mind by five minutes, the final accuracy was only 25 percent.[2]

One medical environment stands out for the challenge—the emergency room, where physicians must assess each patient quickly and either admit them to the hospital or send them home. A wrong diagnosis can result in a person's death soon after being discharged, and, with almost 20 percent of the population of the United States visiting an emergency room each year, the population at risk is huge. A large study of ER evaluations of Medicare patients showed that each year more than 10,000 people died within a week of being sent home, despite not having a previously diagnosed illness or being diagnosed with a life-threatening one.[3] This isn't simply a problem in the emergency room. There are more than 12 million serious diagnostic errors each year in the United States alone,[4] and, according to a landmark report published in 2015 by the National Academy of Sciences, most people will experience at least one diagnostic error in their lifetime.[5]

These data point to the serious problems with how physicians diagnose. System 1—what I call fast medicine—is malfunctioning, and so many other of our habitual ways of making an accurate diagnosis can be improved. We could promote System 2 diagnostic reasoning. Kahneman has argued that "the way to block errors that originate in System 1 is simple in principle: recognize the signs that you are in a cognitive minefield, slow down and ask for reinforcement from System 2."[6] But to date, albeit with limited study, the idea that we can supplement System 1 with System 2 hasn't held up: when doctors have gone into analytic mode and consciously slowed down, diagnostic accuracy has not demonstrably improved.[7] A major factor is that the use of System 1 or System 2 thinking is not the only relevant variable; other issues come into play as well. One is a lack of emphasis on diagnostic skills in medical education. Of the twenty-two milestones of the American Board of Internal Medicine Accreditation Council for Graduate Medical Education, only two are related to diagnostic skills.[8] Once trained, doctors are pretty much wedged into their level of diagnostic performance throughout their career. Surprisingly, there is no system in place for doctors to get feedback on their diagnostic skills during their careers, either. In *Superforecasting,* Philip Tetlock observes, "If you don't get feedback, your confidence grows much faster than your accuracy."[9] The lack of emphasis on diagnostic skills during and after medical school, however, seems to be overshadowed by the lack of appreciation of deep cognitive biases and distortions that can lead to diagnostic failure. They're not even part of teaching diagnosis today in medical school.

In *The Undoing Project: A Friendship That Changed Our Minds,* Michael Lewis wrote about Donald Redelmeier, a Canadian physician who as a teenager was inspired by Amos Tversky and Danny Kahneman.[10] At Sunnybrook Hospital's trauma center, he asked his fellow physicians to slow down, tame System 1 thinking, and try to avoid mental errors in judgment. "You need to be so careful when there is one simple diagnosis that instantly pops into your mind that beautifully explains everything all at once. That's when you need to stop and check your thinking."[11] When a patient was misdiagnosed to be hyperthyroid for her irregular heartbeat but instead was found to have fractured ribs and a collapsed lung, Redelmeier called this error an example of the representativeness heuristic, which is a shortcut in decision making based on past experiences (first described by Tversky and Kahneman). Patterns of thinking such as the representativeness heuristic are an example of the widespread problem of cognitive bias among physicians. Humans in general are beset by many biases—*Wikipedia* lists 185, for example—but I want to highlight only a few of those that impair diagnostic accuracy.[12] It's important to emphasize that these embedded cognitive biases in medicine are simply human nature, not at all specific to making a diagnosis or being sure about recommending a treatment. But what is different here is that medical decision making can have profound, even life-and-death, consequences.

Some of the cognitive biases that lead to errors in diagnosis are quite predictable. There are about 10,000 human diseases, and there's not a doctor who could recall any significant fraction of them. If doctors can't remember a possible diagnosis when making up a differential, then they will diagnose according to the possibilities that are mentally "available" to them, and an error can result. This is called the availability bias.

A second bias results from the fact that doctors deal with patients one at a time. In 1990, Redelmeier and Tversky published a study in the *New England Journal of Medicine* that showed how individual patients, especially patients that a doctor has recently seen, can shape medical judgment, simply because each doctor only ever sees a relatively small number of patients.[13] Their personal experience as doctors can override hard data derived from much larger samples of people, say, about the likelihood that a patient has some rare disease, simply because an earlier patient with similar symptoms had that rare disease. Like when I saw a patient with a stroke who had a very rare tumor on a heart valve (called papillary fibroelastoma) and thought of it as a potential culprit in many subsequent patients. Compounding this is the fact that, as Redelmeier has found, 80 percent of doctors don't think probabilities apply to their patients.

One example of this bias from my experience comes to mind. Inserting a coronary stent has a small chance of inducing a heart attack in the patient. These heart attacks are rarely accompanied by any symptoms but can be diagnosed with blood test enzymes that prove there has been some damage to heart muscle cells. When my colleagues and I published a series of papers in the 1990s about this issue, known as periprocedural myocardial infarction, the reaction of most cardiologists was that we were wrong, that the problem was