

The role of big data in medicine

November 2015

Technology is revolutionizing our understanding and treatment of disease, says the founding director of the Icahn Institute for Genomics and Multiscale Biology at New York's Mount Sinai Health System.

Most companies make a conscious and deliberate decision to embrace digitization and the information revolution. Yet the role of big data in medicine seems almost to compel organizations to become involved. In this interview, Dr. Eric Schadt, the founding director of the Icahn Institute for Genomics and Multiscale Biology at New York's Mount Sinai Health System, tells McKinsey's Sastry Chilukuri how data-driven approaches to research can help patients, in what ways technology has the potential to transform medicine and the healthcare system, and how the Icahn Institute is building its talent base. An edited transcript of Schadt's remarks follows.

Evolution or revolution?

The role of big data in medicine is one where we can build better health profiles and better predictive models around individual patients so that we can better diagnose and treat disease.

One of the main limitations with medicine today and in the pharmaceutical industry is our understanding of the biology of disease. Big data comes into play around aggregating more and more information around multiple scales for what constitutes a disease—from the DNA, proteins, and metabolites to cells, tissues, organs, organisms, and ecosystems. Those are the scales of the biology that we need to be modeling by integrating big data. If we do that, the models will evolve, the models will build, and they will be more predictive for given individuals.

It's not going to be a discrete event—that all of a sudden we go from not using big data in medicine to using big data in medicine. I view it as more of a continuum, more of an evolution. As we begin building these models, aggregating big data, we're going to be testing and applying the models on individuals, assessing the outcomes, refining the models, and so on. Questions will become easier to answer. The modeling becomes more informed as we start pulling in all of this information. We are at the very beginning stages of this revolution, but I think it's going to go very fast, because there's great maturity in the information sciences beyond medicine.

The life sciences are not the first to encounter big data. We have information-power companies like Google and Amazon and Facebook, and a lot of the algorithms that are applied there—to

predict what kind of movie you like to watch or what kind of foods you like to buy—use the same machine-learning techniques. Those same types of methods, the infrastructure for managing the data, can all be applied in medicine.

How wearables are poised to transform medicine

Wearable devices and engagement through mobile health apps represent the future—not just of the research of diseases, but of medicine. I can be confident in saying that, because today in medicine, a normal individual who is generally healthy spends maybe ten minutes in front of a physician every year. What that physician can possibly score you on to assess the state of your health is very minimal. Unless something catastrophic is going on within you—lipid levels that are way off the charts or glucose levels or something extreme—they're not doing much to assess what your state of well-being is, and the information stored in medical records is not extensive enough.

What the wearable-device revolution provides is a way to longitudinally monitor your state—with respect to many different dimensions of your health—to provide a much better, much more accurate profile of who you are, what your baseline is, and how deviations from that baseline may predict a disease state or sliding into a disease state. That means we'll be able to intervene sooner to prevent you from that kind of slide. That sort of modeling would be impossible unless you could phenotype individuals on a longitudinal and long-term basis.

And while the wearable devices today are in this more recreational-grade state, they're changing incredibly rapidly into research grade and ultimately clinical grade. There are already glucose monitors that are FDA¹ approved that individuals can wear and that interface with digital apps, which then connect directly with healthcare providers based on what they're seeing with your glucose profiles. You'll see that kind of sensing get better and better, providing higher and higher grades and better and better profiles on individuals over time. I estimate that in five to ten years, accurate information about your health will exist more outside the health system than inside the health system. And that will force the engagement of that information by the medical community.

What big data means for patients, payers, and pharma

What I see for the future for patients is engaging them as a partner in this new mode of understanding their health and wellness better and understanding how to make better decisions around those elements.

Most of their data collection will be passive, so individuals won't have to be active every day—logging things, for example—but they'll stay engaged because they'll get a benefit from it. They'll agree to have their data used in this way because they get some perceived benefit. Ultimately, that'll be the number of doctor visits you require, the number of times you were sick, the number of times you progressed into a given disease state. All should diminish. And there's a benefit from being presented with the information, so they're looking at dashboards about themselves—they're not blind to the information or dependent on a physician to interpret it for them, they're able to see it every day and understand what it means.

¹US Food and Drug Administration.

A better understanding of Alzheimer's disease

One of the most fascinating experiences I've had creating this ecosystem—with lots of different area experts all coming together to solve a common problem and actually having a real impact on disease—came about through our Alzheimer's work. There's a lot of motivation to better understand that disease. What we wanted to do was to try to take a more objective look at what was going on in the brain of individuals with Alzheimer's disease.

For a long time, the plaque and tangles were the driving force for how people were seeking to understand Alzheimer's and to come up with preventative or more effective treatments. What we were able to do was engage modern technology—the genomics technologies—and go to some of the established brain banks and carry out a much deeper profiling in a completely data-driven way.

We didn't have to constrain ourselves by the plaques-and-tangles hypothesis. We could say, "We're going to sequence all the DNA in different brain regions. We're going to sequence the RNA," which is a more active sort of sensor of what's going on at the deep molecular level in different parts of the brain. And then, "We're going to try to reconstruct predictive or network models to understand how the millions of variables we're measuring are connected to one another in a cause-effect sort of way," and, "We're going to see how those models change between the disease state and the normal, nondemented state."

Think of these networks as a graphical model where the nodes in the network are different genes and clinical features and DNA variance, and the edges indicate relationships between those variables that we observe over the population of brains we profiled. What we were very surprised to find is that the most important network for Alzheimer's had nothing directly to do with tangles or plaques, but the immune system. We directly implicated microglial cells—which are sort of the macrophage-type cells of the brain that keep the brain healthy—as a key driver of Alzheimer's disease.

What enabled us to make that kind of connection was basically ignoring what the field thought it knew about Alzheimer's disease, taking a very data-driven, objective approach to construct models that could help us get our heads around the millions of variables that we were scoring, and then letting the data speak to us in terms of what the likely drivers of the disease are and the ways we can best prevent it. That work alone has led to a revolution—around novel therapeutics to target Alzheimer's—that is less about the tangles and plaques and more about how to modulate the immune system in the brain to have a benefit as opposed to damaging the brain.

I believe payers are perhaps among the top of the chain as far as who can benefit from this. Because, ultimately, payers want to constrain the cost of each patient. They care about the health of the patient, but they want to do whatever they can to motivate both the patients and the medical systems that treat them to minimize the cost through better preventative measures, better targeted therapies, and increased compliance for medication usage. So now, payers are getting a better benefit from drugs being taken, because they're able to see that the drug is being taken as prescribed or that it's not having the effect on the patient so the patient can be switched earlier to a more effective treatment. If you're able to intervene sooner in the course of a patient's health, before they slide into a disease state, then you're going to save money on those unexpected hospitalizations or emergency-room visits or even physician visits.

Then there's just the general risk profiling of patients. Of course, payers care a lot about understanding the overall risk of a patient and what they're likely to cost year over year. For example, say we're able to generate genomic information that tells us what the heritable cancer risk of every patient is; you don't need to wait until a lump is felt or the person's at a later stage of cancer, when it's much more expensive. Those better risk profiles will be an incentive for payers to pay attention and to actually be involved in that development.

For device makers, I just see this as a revolution that's theirs to lose if they don't embrace the development of consumer wearable devices or sensors, more generally, in environments where every person in the US or on the planet is buying a device versus one of a handful of medical systems. That's a better business model that's going to generate lots of revenue. And so it's up to the device maker to embrace that revolution and even start transforming some of the devices they're already making into consumer-grade devices that can be not just recreation grade but higher grade, on toward the clinical grade.

Finally, from the pharmaceutical standpoint, I think it's major. I mean, just look at Regeneron Pharmaceuticals and Geisinger engaging the Geisinger Health System and sequencing everybody in that population to create a better understanding of disease and protections against disease to do therapeutics. What you're seeing, at some level, is some embracing of this sort of information revolution by the pharmaceutical companies. It's doing it mainly from the genomics arena, but it's also approaching it from the standpoint of better understanding disease, having a better understanding of the causal players of disease, and using that or the causal protectants against disease to directly develop therapeutics.

Bringing together the right talent

One of the most fun aspects of creating the Icahn Institute—and growing it into the state it's in today and where it's heading—is creating the right kind of ecosystem that can be comprised of highly diverse individuals from the standpoint of different areas of expertise.

In the past three or four years, we've hired more than 300 people, spanning from the hardware side and big data computing to the sequence informatics and bioinformatics to the CLIA-certified² genomics core—to generate the information—to the machine-learning and predictive-modeling guys and the quantitative guys, to build the models. And then

²Clinical Laboratory Improvement Amendments.

we've linked that up to all the different disease-oriented institutes at Mount Sinai, and to some of the clinics directly, to start pushing this information-driven decision making into the clinical arena.

Not all the physicians were on board and, of course, there are lots of people who will try to cause all sorts of fear about what kind of world we're going to transform into if we are basing medical decisions on sophisticated models where nobody really understands what's happening. So it was all about partnering with individuals such as key physicians who were viewed as thought leaders—leading their area within the system—and carrying out the right kinds of studies with those individuals.

In all of these different areas, we're recruiting experts, and we view what we build as sort of a hub node that we want linked to all the different disease-oriented institutes to enable them to take advantage of this great engine. But you need people to help translate it, and that's what these key hires have done. They have a strong foot within the Icahn Institute, but they also care about disease. And so they form their whole lab around the idea of how to more efficiently translate the information from the big information hub out to the different disease areas. That's still done mainly by training individuals within those labs to be able to operate at a lower level. I think what needs to happen beyond that is better engagement through software engineering: user-interface designers, user-experience designers who can develop the right kinds of interfaces to engage the human mind in that information.

One of the biggest problems around big data, and the predictive models that could build on that data, really centers on how you engage others to benefit from that information. Beyond the tools that we need to engage noncomputational individuals in this type of information and decision making, training is another element. They've grown up in a system that is very counter to this information revolution. So we've started placing much more emphasis on the generation of coming physicians and on how we can transform the curriculum of the medical schools. I think it's a fundamental transformation of the medical-school curriculum, and even the basic life sciences, where it becomes more quantitative, more computational, and where everybody's taking statistics and combinatorics and machine learning and computing.

Those are just the tools you need to survive. And it has to start at that earlier stage, because it's very, very difficult to take somebody already trained in biology or a physician and teach them the mathematics and computer science that you need to play that game.

Dr. Eric Schadt is the founding director of the Icahn Institute for Genomics and Multiscale Biology at New York's Mount Sinai Health System. Sastry Chilukuri is a principal in McKinsey's New Jersey office.