

# Can Big Data Tell Us What Clinical Trials Don't?

OCT. 3, 2014 By VERONIQUE GREENWOOD

[http://www.nytimes.com/2014/10/05/magazine/can-big-data-tell-us-what-clinical-trials-dont.html?ref=magazine&\\_r=1](http://www.nytimes.com/2014/10/05/magazine/can-big-data-tell-us-what-clinical-trials-dont.html?ref=magazine&_r=1)

When a helicopter rushed a 13-year-old girl showing symptoms suggestive of kidney failure to Stanford's Packard Children's Hospital, Jennifer Frankovich was the rheumatologist on call. She and a team of other doctors quickly diagnosed lupus, an autoimmune disease. But as they hurried to treat the girl, Frankovich thought that something about the patient's particular combination of lupus symptoms — kidney problems, inflamed pancreas and blood vessels — rang a bell. In the past, she'd seen lupus patients with these symptoms develop life-threatening blood clots. Her colleagues in other specialties didn't think there was cause to give the girl anti-clotting drugs, so Frankovich deferred to them. But she retained her suspicions. "I could not forget these cases," she says.

Back in her office, she found that the scientific literature had no studies on patients like this to guide her. So she did something unusual: She searched a database of all the lupus patients the hospital had seen over the previous five years, singling out those whose symptoms matched her patient's, and ran an analysis to see whether they had developed blood clots. "I did some very simple statistics and brought the data to everybody that I had met with that morning," she says. The change in attitude was striking. "It was very clear, based on the database, that she could be at an increased risk for a clot."

The girl was given the drug, and she did not develop a clot. "At the end of the day, we don't know whether it was the right decision," says Chris Longhurst, a pediatrician and the chief medical information officer at Stanford Children's Health, who is a colleague of Frankovich's. But they felt that it was the best they could do with the limited information they had.

A large, costly and time-consuming clinical trial with proper controls might someday prove Frankovich's hypothesis correct. But large, costly and time-consuming clinical trials are rarely carried out for uncommon complications of this sort. In the absence of such focused research, doctors and scientists are increasingly dipping into enormous troves of data that already exist — namely the aggregated medical records of thousands or even millions of patients to uncover patterns that might help steer care.

The Tatonetti Laboratory at Columbia University is a nexus in this search for signal in the noise. There, Nicholas Tatonetti, an assistant professor of biomedical informatics — an interdisciplinary field that combines computer science and medicine — develops algorithms to trawl medical databases and turn up correlations. For his doctoral thesis, he mined the F.D.A.'s records of adverse drug reactions to identify pairs of medications that seemed to cause problems when taken together. He found an interaction between two very commonly prescribed drugs: The antidepressant paroxetine (marketed as Paxil) and the cholesterol-lowering medication pravastatin were connected to higher blood-sugar levels. Taken individually, the drugs didn't affect glucose levels. But taken together, the side-effect was impossible to ignore. "Nobody had ever thought to look for it," Tatonetti says, "and so nobody had ever found it."

The potential for this practice extends far beyond drug interactions. In the past, researchers noticed that being born in certain months or seasons appears to be linked to a higher risk of some diseases. In the Northern Hemisphere, people with multiple sclerosis tend to be born in the spring, while in the Southern Hemisphere they tend to be born in November; people with schizophrenia tend to have been born during the winter. There are numerous correlations like this, and the reasons for them are still foggy — a problem Tatonetti and a graduate assistant, Mary Boland, hope to solve by parsing the data on a vast array of

outside factors. Tatonetti describes it as a quest to figure out “how these diseases could be dependent on birth month in a way that’s not just astrology.” Other researchers think data-mining might also be particularly beneficial for cancer patients, because so few types of cancer are represented in clinical trials.

As with so much network-enabled data-tinkering, this research is freighted with serious privacy concerns. If these analyses are considered part of treatment, hospitals may allow them on the grounds of doing what is best for a patient. But if they are considered medical research, then everyone whose records are being used must give permission. In practice, the distinction can be fuzzy and often depends on the culture of the institution. After Frankovich wrote about her experience in *The New England Journal of Medicine* in 2011, her hospital warned her not to conduct such analyses again until a proper framework for using patient information was in place.

In the lab, ensuring that the data-mining conclusions hold water can also be tricky. By definition, a medical-records database contains information only on sick people who sought help, so it is inherently incomplete. Also, they lack the controls of a clinical study and are full of other confounding factors that might trip up unwary researchers. Daniel Rubin, a professor of bioinformatics at Stanford, also warns that there have been no studies of data-driven medicine to determine whether it leads to positive outcomes more often than not. Because historical evidence is of “inferior quality,” he says, it has the potential to lead care astray.

Yet despite the pitfalls, developing a “learning health system” — one that can incorporate lessons from its own activities in real time — remains tantalizing to researchers. Stefan Thurner, a professor of complexity studies at the Medical University of Vienna, and his researcher, Peter Klimek, are working with a database of millions of people’s health-insurance claims, building networks of relationships among diseases. As they fill in the network with known connections and new ones mined from the data, Thurner and Klimek hope to be able to predict the health of individuals or of a population over time. On the clinical side, Longhurst has been advocating for a button in electronic medical-record software that would allow doctors to run automated searches for patients like theirs when no other sources of information are available.

With time, and with some crucial refinements, this kind of medicine may eventually become mainstream. Frankovich recalls a conversation with an older colleague. “She told me, ‘Research this decade benefits the next decade,’ ” Frankovich says. “That was how it was. But I feel like it doesn’t have to be that way anymore.”