What Trump's victory taught us about predictive analytics

By Jerry Berger | January 26, 2017

Few people last fall would have expected that, come this time, Donald Trump would be the president of the United States. The polls didn't predict it, by a long shot.

That spectacular fail was a black eye for predictive analytics, a relatively new science that is also at the heart of the rush to use "big data" in healthcare. Today, predictive analytics is used to study everything, from which patients will show up for appointments to how to diagnose more effectively in the emergency room. Data analysis is a foundational capability of value-based reimbursement; you can't manage what you can't measure.

And the lesson of the 2016 election for healthcare, experts say, is that your ability to measure is only as good as the quality of your data.

"In a fledgling science like this, there are always things you don't know you don't know," says Pradeep Mutalik, M.D., a medical research scientist at the Yale Center for Medical Informatics who wrote about the flaws in the 2016 election predictions in the New York Times. "The model works when you cover all the phenomena you expect to encounter," he says.

The pitfalls that resulted in flawed predictions in the U.S. — and in the United Kingdom's "Brexit" election — can be avoided, healthcare analysts say, as long as data scientists mind their models, don't give in to hype, and don't rely on data that's incomplete.

That wasn't possible, Mutalik says, for election prognosticators, whose data often only went back to 1972. Healthcare researchers won't face that sampling problem if they work with a robust dataset, he says.

But even then, Mutalik adds, "you have to understand there are always things that didn't happen" in the past, citing the rise in obesity, opioid addiction and new strains of infections.

"The key is not to think you're always going to be accurate," he says. "Always modify the model. It's a moving target and you have to keep an eye on the ball."

Leo Celi, M.D., a principal investigator at the MIT Laboratory for Computational Physiology and an assistant professor at Harvard Medical School, says he and his colleagues are doing just that as they build a database to provide intensive care unit clinicians with guidance on how to treat patients.

Using records from in the U.S., Brazil, the U.K., and France, Celi and his team will track blood pressure, blood sugar levels, and heart rate. But Celi — who still works ICU shifts — says providers also need to understand non-clinical behaviors, such as whether a patient is actually taking prescribed medications, to reach an accurate diagnosis.

"If you want to understand health and disease, you need to integrate non-clinical information," he says.

And you need a human touch, researchers say, to combat the tendency to rely too much on algorithms to model behavior. Steve Horng, M.D., an instructor in emergency medicine at Harvard Medical School, is working to add statistical probability formulas to his models to allow physicians to infer which disease goes with the symptoms, resulting in what he says is more sound diagnosis.

"Deep learning is a phenomenal technique, but it is dangerous for healthcare because you can't know what's going on behind the scenes," says Horng. "Data alone without interpretation can learn the wrong lesson. The goal is not to replace the human but to give humans better tools."

athenainsight

A daily news hub reporting from the heart of the health care internet, with access to a comprehensive data set of health care transactions from athenahealth's nationwide network. We equip leaders with actionable insight and inspiration for making health care work as it should.

Stay in the know

Sign up for weekly data and news:
insight.athenahealth.com/newsletter

Predictive analytics can also be a valuable tool in non-life and death situations that have huge financial ramifications. By one estimate, inefficiencies such as chronic patient no-shows, some of them due to flaws in the scheduling process, cost the U.S. healthcare system more than $150 billion per year.

That was the premise behind the Smart Scheduling, a company developed at an MIT "hackathon" a few years ago. Its system predicts which patients might be the greatest cancellation risks — and helps clinicians reach out to patients to stem the tide.

Now part of athenahealth, the model looks at patient no-shows by examining clinical and administrative records, including how often patients cancel, how frequently individual providers have cancellations, the reason for the visit, along with gender, age, and how easy the patient is to reach. The dataset currently holds about 20 million records from 800 providers.

The goal is to add all 80,000 athenahealth providers, says Chris Moses, the company's director of product innovation. A study done prior to acquisition found that, on average, providers added two patients per month per provider using the model.

Moses notes that not all wrong predictions carry the same cost. In the case of predicting no-shows, Moses notes, "the worst outcome could be double booking and a slightly longer wait."

The stakes are clearly higher when it comes to patient care. And when it comes to predicting diagnoses and interacting with patients, Yale's Mutalik cautions, it's important to avoid reducing everything to a number.

"It's a probability, not a score," he says. "There's a chance it could be spectacularly wrong."

That's why it's important to manage expectations upfront, he says — whether you're predicting a diagnosis, a no-show, or an election.

Jerry Berger is a writer based in Boston.

Artwork by Flashpop/Getty Images.