Family trees hidden in medical records could predict your disease risk

Jocelyn Kaiser

Who is your emergency contact? The answer to that question, standard in every doctor’s office, has now been used to predict the role of genes in hundreds of conditions, from diabetes to high cholesterol. A new study combined the emergency contact information of 2 million New Yorkers with their medical data to form family trees of heritability—all without ever looking at a patient’s DNA. The approach could be used by clinics widely to predict a person’s disease risks, if patients agree to let their data be used in this way.

“It’s an interesting idea that you can generate family structures across very large data sets” compiled by health care providers “and infer something about the shared basis of disease,” says cardiovascular disease genetics researcher Dan Roden of Vanderbilt University in Nashville, who was not involved with the study.

Hoping to explore the genetics of drug reactions, graduate student Fernanda Polubriaginof and others working in the lab of biomedical informatics researcher Nicholas Tatonetti at Columbia University wanted to determine whether patients at the school’s affiliated NewYork-Presbyterian Hospital were related. “It occurred to us there’s some data that every hospital routinely collects every time a person is admitted,” Tatonetti says: an emergency contact, who also often happens to be a blood relative.

His team pulled those contacts from electronic health records of patients who had given consent to use their information in research. To the scientists’ surprise, about one-third of the emergency contacts had also come to Columbia’s hospital for treatment. They then used the names, addresses, phone numbers, and relationships of these contacts to build 223,000 family trees connecting blood relatives. The biggest had more than 100 patients spread across four generations, Polubriaginof says.

The Columbia team and collaborators at the city’s Weill Cornell Medicine and Mount Sinai health systems also constructed family trees from thier records, bringing the total to 1.9 million patients with 7.4 million relationships. The patients were a diverse mix including Latinos, blacks, and whites.

The researchers then overlaid those trees with information on each individual’s health conditions, gleaned from billing codes and lab tests, and used the combined data to estimate the heritability of about 500 traits and diseases. For many conditions, such as glaucoma, the results matched previous estimates based on twin studies. For other conditions, the huge database may resolve conflicting results, the team reports today in Cell.

For example, two small-scale studies have drawn different conclusions about the inherited risk for high cholesterol. The new study found that high levels of high-density lipoprotein, commonly considered the good kind of cholesterol, are 50% inherited, whereas high levels of low-density lipoprotein, the more dangerous kind, are 25% inherited.

Another 400 or so conditions in the paper hadn’t been studied much by geneticists. Sinus infections, for example, appeared to be 85% inherited, which matches up with anecdotal evidence that these infections run in families, Tatonetti says. (Here are the data sets for all 500 conditions.)

The method wasn’t perfect for so-called Mendelian disorders, diseases known to always occur if a person inherits a single copy of a flawed gene—in other words, they’re thought to be 100% inheritable. Sickle cell disease was 97% heritable according to the electronic records analysis, very close to what is expected. But for cystic fibrosis, another such condition, heritability was only 1%. That discrepancy could be explained by many factors that complicate the family tree analysis, including that families with the condition tend to be small, Polubriaginof says.

Family health histories are harder to collect from interviewing patients than you might think, says Roden, who also works with electronic health records. “This could be another way of producing the same information,” he says.

At the same time, the study could make people leery of sharing their emergency contact information, Roden suggests. The researchers stripped names and other identifying information from the patients’ health data after linking it to relatives’ records. But, “The idea that researchers are mining information about your family without letting you know, it does run the risk of alienating people. We have to be pretty careful to make sure the public stays a partner in efforts to grow these large data sets,” he says.

Because of privacy rules, the Columbia team doesn’t plan to share the family trees or disease risks with individual patients. However, Polubriaginof is using the overall results to suggest that Columbia’s physicians could do a better job of screening patients at risk for diabetes and celiac disease, an autoimmune disorder. And the New York City researchers aren’t alone: Academic health centers in Boston and Chicago, Illinois, are already using the team’s formula to trace their own patients’ family trees for research.