Predictive model identifies patients for genetic testing

June 3 2021

(L-R) Douglas Ruderfer, Ph.D., and Theodore (Ted) Morley pose for a photo at Vanderbilt University Medical Center. Credit: Donn Jones/Vanderbilt University Medical Center.

Patients who, perhaps unbeknownst to their health care providers, are in need of genetic testing for rare undiagnosed diseases can be identified en

masse based on routine information in electronic health records (EHRs), a research team reported today in the journal *Nature Medicine*.

Findings from the Vanderbilt University Medical Center study suggest that, among the patients of any sizeable health care system, there are hundreds or thousands with undiagnosed rare diseases of the sort where a genetic [test](#) could lead to a diagnosis.

"Patients with rare genetic diseases often face years of diagnostic odyssey before getting a genetic test, if they get one at all. Our work could contribute to a more systematic and timely approach, alerting providers of patients that might benefit from a genetic test," said the leader of the study, geneticist Douglas Ruderfer, Ph.D., associate professor of Medicine at VUMC.

It's estimated that more than 70% of rare diseases are genetic in origin. According to authors of the study, rare genetic diseases may look quite different from one patient to the next and may go undiagnosed even when well characterized in the medical literature.

Using routine EHR data to directly identify patients with specific genetic diseases is, for now, quite challenging: according to the study authors, the current state of knowledge about disease-causing [genetic variation](#) is too wanting, and the genetic resolution of current clinical tests is too low. Given this state of affairs, the team developed their predictive model to instead detect patients who fit historical criteria for testing in the eyes of clinicians.

To begin, the team developed a range of competing predictive algorithms. For [training data](#), they used EHRs of patients for whom clinicians had ordered a type of genetic test called chromosomal microarray (1,818 cases), and similar patients whose records showed no history of genetic testing (7,326 matched controls). Across the training

and test sets, the average age of the patients represented was 8.

"We were really aiming to build a model that captured and automated clinical suspicion of a genetic disease," said Theodore Morley, a staff data scientist who worked closely with Ruderfer on the study.

With all traces of genetic testing removed from a test set of 2,286 records, a machine learning algorithm emerged as the best performer, correctly classifying 87% of cases and 96% of controls.

Also showing high accuracy were independent validation efforts at Boston's Massachusetts General Hospital and again at VUMC, with far greater numbers of longstanding patients, cases now defined by evidence in the EHR of interaction with a genetics provider.

Importantly, the algorithm also performed well in identifying patients who received genetic tests other than chromosomal microarray.

The study supports the hypothesis that, within any EHR population, patients who are to be suspected of having rare genetic diseases are distinguishable via computation thanks to the presence of multiple rare signs and symptoms—phenotypes, in the language of the study—that are subject to documentation in the EHR. The team's predictive algorithms used EHR diagnosis codes exclusively (the same codes that drive health care billing), amalgamated for prediction purposes into so-called phenotype codes.

"After extensive validation demonstrated high predictive performance, we were really interested in assessing how an implementation of our model might compare to the current status quo for who receives a test, and what the results of those tests are," Ruderfer said.

From a set of 6,445 deidentified EHRs corresponding to genotyped

patient specimens from BioVU, Vanderbilt's DNA biobank, the algorithm proved accurate in picking out individuals with pathogenic copy number variations, a type of genetic abnormality.

In collating this and other findings from the study, the authors estimate there are more than 2,000 patients at VUMC who have an unrecognized, potentially diagnostic copy number variation that could be identified with a genetic test.

Among close observers of the study is Josh Peterson, MD, MPH, director of VUMC's Center for Precision Medicine (where Ruderfer is an affiliated faculty member).

"The team's predictive model was shown to handily outperform current clinical performance, identifying patients who need testing more quickly and in greater numbers. Crucially, the model also proved portable to another health system. This paper's findings align very well with the push at VUMC to use data science to improve diagnosis of genomic syndromes," Peterson said.

Provided by Vanderbilt University Medical Center