To find disease risk, genetics provides mother of all shortcuts

January 23 2017

Jonathan Mosley, Ph.D., M.D. Credit: John Russell/Vanderbilt University

In a new study in the journal *Circulation: Cardiovascular Genetics*, Vanderbilt University's Jonathan Mosley, M.D., Ph.D., and colleagues use genetic correlation to link two unrelated biomedical data sets, one

from a longstanding prospective epidemiological cohort and the other from electronic health records.

For answering questions about disease risks, there are major limitations inherent in each of these types of data, and this study sets out a strategy to overcome them. The method itself has interest, and here the results concern validation of risk factors for heart disease and diabetes.

One limitation of prospective epidemiology is that "Having enrolled your subjects and carefully gathered baseline data, you then have to wait 20 or 30 years to see who thrives and who gets sick," said Mosley, instructor in Medicine.

Meanwhile, the relative importance of disease risk factors may shift as habits change, therapies improve and longevity increases.

Conversely, the trouble with today's electronic health records is that, "While the records may contain large numbers of clinically significant outcomes, such as heart attacks, they are frequently spotty in terms of baseline data essential for epidemiology, and they don't contain data for novel or unproven risk factors and biomarkers," Mosley said.

Some academic medical centers and some epidemiological studies these days have accompanying biorepositories and genotype data. Mosley's study demonstrates how the increasing availability of these data presents opportunities to answer epidemiological questions in weeks rather than decades.

This mother of all shortcuts is provided by genetic correlation.

Because ischemic heart disease (IHD) and total-to-HDL cholesterol ratios correlate to a known degree, checking one helps predict the other. Correlated traits may also happen to overlap genetically, that is, they

may share genetic influences to some measurable degree, IHD and total-to-HDL cholesterol being only one example.

Genetically, think of humans as collectively defining a norm, an evolving statistical figment, call it the human genotype. At conception, we each land at some measurable distance from this norm, and there we live out our lives. If researchers genotype a large sample of the population, they can estimate the distance from the genetic norm for each individual in the sample, characterizing the sample's genetic variability from one pair of individuals to the next. Meanwhile, of course, they can also note the variability within the sample of any trait they care to measure.

"These two types of data let you estimate the effect of genes on the variability of traits. They won't tell you which genes affect the traits you've measured, but you'll be able to estimate the magnitude of genetic effects," Mosley said.

These are all the data Mosley needs for his next step, calculating the genetic correlation between traits. In this study, for example, he finds a 44 percent genetic correlation between IHD and total-to-HDL cholesterol. That means there's a 44 percent overlap in his sample between two unknown sets of causative genetic variants, one set partly explaining who gets high total-to-HDL cholesterol and the other set partly explaining who gets IHD.

If this seems mysterious, Mosley points out that researchers have long worked out genetic correlations without genotype data, measuring traits in sets of siblings, in sets of parents and children, learning the sheer size of genetic correlations if not getting their addresses and phone numbers.

"Genetic epidemiologists have been measuring genetic correlations between pairs of traits for decades, but this required using related individuals and measuring each trait in each individual. With newer

genetic technologies, we can use unrelated individuals and we only need to measure one of the traits in each individual," Mosley said.

Mosley's way is superior for finding traits genetically correlated with diagnoses.

In two traits with a known genetic correlation, where one is captured in a simple clinical observation or test result and the other is captured in a diagnosis for a chronic disease with a long preclinical onset, "There's no need to wait around for decades to learn what association may or may not obtain between the two. With genetic correlation you can have an immediate view into a chronic disease risk," Mosley said.

To work out probabilities in dice, you don't need any physics and you don't need to throw any dice.

Crucially, so long as there's a common well of genotype data from which to calculate the correlations, there's no reason the traits can't be measured in completely unrelated groups.

According to Mosley, the study at hand is the largest, most systematic demonstration of its type so far. He calculates genetic correlations between, on one hand, 37 baseline measurements taken decades ago in some 13,000 subjects enrolled in a heart disease epidemiological cohort, and, on the other hand, IHD and type 2 diabetes (T2D) as documented more recently within electronic health records of some 25,000 patients seen at various academic medical centers.

The 37 baseline measurements were gathered from 1987 to 1989 in four U.S. communities for the Atherosclerosis Risk in Communities cohort study (ARIC). The de-identified electronic health record data come from the eMERGE Network (Electronic Medical Records and Genomics) and the Vanderbilt Electronic Systems for Pharmacogenomic

Assessment cohort (VESPA).

After excluding related pairs of individuals and non-Europeans, Mosley's analysis includes data from around 27,000 subjects. They had been genotyped using different equipment, and around half a million common genetic variants from across the genome are captured in the merged intersection of the genotyping results.

In all, 14 of the baseline measures from the ARIC cohort bear genetic correlation with T2D in the electronic health record (EHR) group. For each of these, associations with T2D in ARIC were also measured, yielding 14 hazard ratios. Mosley finds that these paired hazard ratios and genetic correlations climb together, the greater the one, the greater the other. This confirms that, next time, Mosley can calculate disease risks based on genetic correlation alone, saving 25 years of longitudinal research.

Mosley's caveat: his method is on firmest ground when genetic influences on traits are greatest. In conventional prospective studies aimed at measuring general disease risk, a local environmental exposure may covertly exert significant influence on a trait in a localized population sample. Mosley's method isn't shielded from this uncertainty. If a local exposure were the predominant modulator of a trait's variability, it could compromise a risk calculation's applicability to the wider population.

Ischemic heart disease is a less well defined disease than T2D is, and the genetic correlation of IHD in the ARIC cohort with IHD in the EHR group is accordingly weaker than the T2D-T2D genetic correlation. In all, eight of the baseline measures from the ARIC cohort bear genetic correlation with IHD in the EHR group. Here again, paired hazard ratios and genetic correlations climb together, but the correlation of the two is weaker than with T2D.

Mosley widens the net, looking for genetic correlations between the ARIC baseline measures and other traits in the EHR group. He finds links to metabolic disorders that can influence IHD. The final picture is of shifting IHD risk between the ARIC cohort and the EHR group.

Compared to the ARIC cohort, IHD in the EHR group is more highly associated with the variability of triglycerides, blood pressure and HDL cholesterol, "Which really suggests to me a phenotype of metabolic syndrome, which we understand to be a toxic syndrome. This syndrome appears to be a stronger driver of IHD in the EHR population than was seen in ARIC," he said.

Many of the EHR group's IHD cases and controls were seen at Vanderbilt. "This really allows us to say, 'What if a study like ARIC had been done on our population? What unmet needs or what unmet risk are we not treating?' This allows us to create a risk profile localized to our institution," Mosley said.

Mosley looks forward to collaborating with other epidemiological studies, including upcoming research involving the well-known Framingham Heart Study.

"There are lots of groups doing large proteomic studies, large metabolomic studies, using brand new methods of measuring new proteins, new biomarkers, new everything ... and now they have to wait 25 years to see if those biomarkers predict anything.

"We might be able to give you your outcome right now."

Provided by Vanderbilt University Medical Center

Citation: To find disease risk, genetics provides mother of all shortcuts (2017, January 23)