The Atlantic

Clinical Genetics Has a Big Problem That's Affecting People's Lives

Unreliable research can lead families to make health decisions they might regret.

Cheryl Ravelo / Reuters

ED YONG | DEC 16, 2015 | SCIENCE

For Heidi Rehm, it looked like a straightforward case. Her lab at Partners Healthcare offers tests for genetic diseases. They had received a blood sample from a fetus after a doctor conducting an ultrasound spotted signs of Noonan syndrome—an inherited disorder involving heart problems and stunted growth. The fetus turned out to have a mutation in PTPN11, a gene that affects the risk of Noonan syndrome.

Rehm found that another team of scientists had published on that very same mutation before. (Not every mutation of PTPN11 increases the rick of Noonan syndrome.) They found that it was more common among Noonan patients than in healthy people, and had billed it as "pathogenic"—that is, likely to cause disease. Rehm reported it as such to the doctor who sent her the sample.

Sometime later, she was listening to a talk by a colleague who had found the same mutation in a patient with Noonan syndrome and, based on the same published study, had also classified it as pathogenic. But this time, the patient—an adult—had contacted the researchers behind the paper. And they had admitted that their conclusions were wrong. In later work, they had found that the mutation is so common in certain ethnic groups that it couldn't possibly be responsible for a rare disease like Noonan syndrome. It wasn't pathogenic after all.

"I immediately contacted the physician to find out the story with that baby," Rehm says. "And that's when I found out that the parents had terminated it."

This story is unusual only in that Rehm is uncommonly open about it. Many geneticists have similar tales where mistakes in the scientific literature have led to wrong—and sometimes harmful—diagnoses.

In one study, Stephen Kingsmore at the National Center for Genome Resources

in Santa Fe found that a quarter of mutations that have been linked to childhood genetic diseases are debatable. In some cases, the claims were based on papers that contained extremely weak evidence. In other cases, the claims were plain wrong: The mutations turned out to be common, like the one in Rehm's anecdote, and couldn't possibly cause rare diseases.

Daniel MacArthur at Massachusetts General Hospital found a similar trend in a study of over 60,000 people, the results of which have been uploaded to a pre-print server. On average, each of these volunteers is walking around with 53 gene variants that are classified as "pathogenic" in two widely-used databases. When the team took a closer look at 200 of these variants, they found enough evidence to classify just *nine* of them as pathogenic.

This is an absurd situation, especially given the stakes. Over the last decade, there's been a lot of talk about reproducibility problems in science—about published results that turn out to be false alarms. In fields like psychology, neuroscience, and cell biology, these errors can send scientists down unproductive paths, waste time and money, and pollute headlines with misleading claims. "But I get much more exercised about reproducibility problems in clinical genetics, because those have massive and real-time consequences for thousands of families," says MacArthur.

People get abortions on the basis of mutations that are linked to severe congenital diseases. They get mastectomies on the basis of mutations in breast-cancer genes. They get monitoring devices surgically implanted in their chests on the basis of mutations in heart-disease genes. "This is absolutely an issue, and it's led to all sorts of problematic decision-making," says Rehm.

How did things get so bad? Everyone I spoke to said that studies used to hew to lower standards. Even just a decade ago, scientists would classify a variant as

pathogenic if they found it in a handful of patients with a disease but not in, say, 100 healthy peers. "That's sooooo not sufficient evidence," says Rehm. A study that small just won't tell you how common the variant in question really is in the general population.

"I think none of us really appreciated just how many rare, nasty-looking genetic variants exist in everyone's genome," admits MacArthur. That only became clear once geneticists acquired enough money, technological power, and collaborative will to do really big sequencing projects, like the 1,000 Genomes Project. Then, "it became abundantly clear that every single one of us is walking around with hundreds of genetic changes that look like they should cause disease, but actually don't. This means that every genome has 'narrative potential'—material that you could use to tell a story about diseases."

It didn't help that many older studies focused on people of European ancestry. A particular variant might be rare in those populations, but very common in other ethnic groups. It couldn't be responsible for rare diseases, but you'd never know if you only sequenced white people.

"Reproducibility problems in clinical genetics ... have massive and real-time consequences for thousands of families."

These problems are understandable in a historical light, but they are still around

today. Just last year, one paper linked to a severe recessive genetic disease to a variant that's carried by 70 percent of the people in large genetics databases. "That's just egregiously wrong," says MacArthur. "There's absolutely no excuse."

It's now very easy to cross-check any patient's variants to see how common or rare they are. Just use the ExAc Consortium, a collection of comprehensive genetic data from over 90,000 people, from a wide range of ethnicities. Geneticists certainly *are* making use of it: It had 2.5 million page views in the last year alone. "It's arguably one of the most useful resources for variant assessment that has ever existed," says Rehm.

It's also getting easier to do laboratory experiments that would actually confirm if a variant causes disease. As I reported last month, some scientists are developing ways of introducing every possible variant into a given gene, and working out exactly what they do. With precise gene-editing tools like CRISPR, these kinds of studies will become even easier.

Raising awareness of the problem is crucial, and not just among geneticists but among peer reviewers and journal editors. "Journals that focus on specific areas of medicine need to be particularly cautious," says Sharon Plon from Baylor College of Medicine. "I am very likely to reject a genetics article if it's in a journal that doesn't publish a lot of genetics."

That would prevent future errors, but it won't clean up the years of accumulated garbage in existing journals. Nothing much will—it's hard enough to publish contradictory information, let alone correct or retract misleading information that's already been published. "Even after you remove all the obvious errors from the databases, there's still a long tail of wrongness," says MacArthur.

Rehm says that the solution is for clinical geneticists to move away from published papers as their main source of information. "The primary literature is just what someone said at one portion in time," she says. "I'd rather rely on well-curated databases."

She is leading the development of ClinVar, an open database of genetic variants that *Nature* has billed as a "one-stop shop for disease genes." It includes information from existing papers, but also from other databases and genetic-testing labs. And it provides a place for researchers to enter information about the variants they find, under strictest of standards. Increasingly, when geneticists see a new variant and want to learn more about it, they search ClinVar first.

Meanwhile, the team behind a related initiative called ClinGen is slowly going through important genes, such as those involved in cystic fibrosis and breast cancer, and reassessing each variant based on the latest databases. They have also implemented a rating system to show how strongly any particular variant has been linked to a particular disease. They've created a measure of trust for genetic results.

"Here's what things will look like in ten years," says MacArthur. "We find a variant in a baby born with a severe disease, and within seconds we can show that this variant has never been seen in 10 million healthy controls, but *has* been observed in 12 cases of the same disease. And just in case that's not enough, we can look in another database for which a researcher has generated every possible variant in that gene and tested its effect, and shown that this specific variant has a catastrophic one. Boom: There's no ambiguity here."

"All we need to do to get there is convince researchers around the world to share their data, build the world's largest repository of genetic and clinical

information, and develop functional tests for every gene in the human genome," he adds. "Easy."

LATEST VIDEO

On Touch: A Visualization of Walt Whitman's Musings

An excerpt from "Song of Myself" is put in motion.

ABOUT THE AUTHOR

ED YONG is a staff writer at *The Atlantic*, where he covers science.