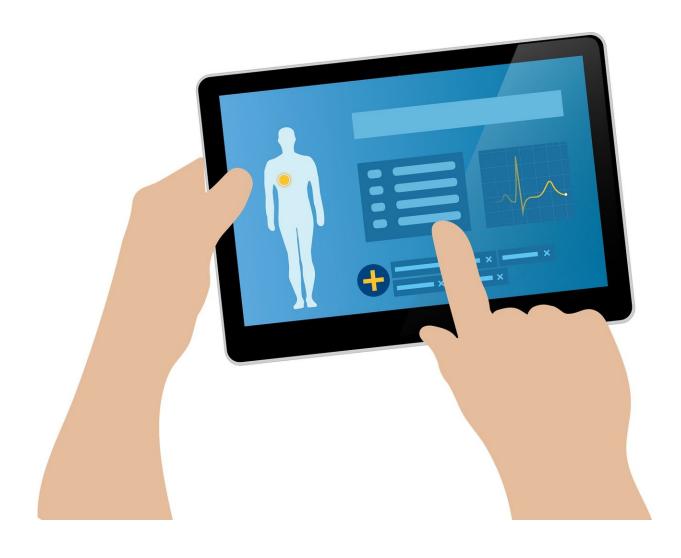# Solving rare disease mysteries using a new method for linking personal records while protecting privacy

July 10 2023

Credit: CC0 Public Domain

Macquarie University researchers have demonstrated a new way of linking personal records and protecting privacy. The first application is in identifying cases of rare genetic disorders. There are many other potential applications across society.

The research will be presented at the 18th ACM ASIA Conference on Computer and Communications Security in Melbourne on 12 July.

A five-year-old boy in the US has a mutation in a gene called GPX4, which he shares with just 10 other children in the world. The condition causes skeletal and central nervous system abnormalities. There are likely to be other children with the disorder recorded in hundreds of health and diagnostic databases worldwide, but we do not know of them, because their privacy is guarded for legal and commercial reasons.

But what if records linked to the condition could be found and counted while still preserving privacy? Researchers from the Macquarie University Cyber Security Hub have developed a technique to achieve exactly that. The team includes Dr. Dinusha Vatsalan and Professor Dali Kaafar of the University's School of Computing and the boy's father, software engineer Mr. Sanath Kumar Ramesh, who is CEO of the OpenTreatments Foundation in Seattle, Washington.

"I am very excited about this work," says Mr. Ramesh, whose foundation initiated and supported the project. "Knowing how many people have a condition underpins economic assumptions. If a condition was previously thought to have 15 patients and now we know, having pulled in data from diagnostic testing companies, that there are 100 patients, that increases market-size hugely."

"It would have a significant economic impact. The valuation of a company working on the condition would go up. Product costing would go down. How insurance companies account for medical costs would

change. Diagnostic companies would target [the condition] more. And you can start to do epidemiology more precisely."

Linking and counting data records might seem simple but, in reality, it involves many issues, says Professor Kaafar. First, because we are dealing with a rare disease, there is no centralized database, and the records are sprinkled across the world. "In this case in hundreds of databases," he says. "And from a business perspective, data is precious, and the companies holding it are not necessarily interested in sharing."

Then, there are technical issues of matching data that is recorded, encoded, and stored in different ways, and accounting for individuals who are double-counted in and between different databases. And, on top of all that, are the privacy considerations. "We are dealing with very, very sensitive health data," Professor Kaafar says.

This personal data isn't needed for a simple estimate of the number of patients and for epidemiological purposes. But, until now, it was needed to ensure that records are unique and can be linked.

Dr. Vatsalan and her colleagues used a technique known as Bloom filter encoding with differential privacy. They devised a suite of algorithms which deliberately introduces enough noise into the data to blur precise details to the point where they cannot be extracted from individual records, but it still allows the patterns of records of the same disease condition to be matched and clustered.

The accuracy of their technique was then evaluated using North Carolina voter registration data. And the results showed the method led to a negligible error rate with a guarantee of a very high level of privacy, even on highly corrupted datasets. The technique significantly outperforms existing methods.

In addition to detecting and counting [rare diseases](#), the research has many other applications; for determining awareness of a new product in marketing, for instance, or in cybersecurity for tracking the number of unique views of particular social media posts.

But it is the application to rare diseases about which the Macquarie University researchers are passionate. "There is no better feeling for a researcher than seeing the technology they've been developing having a real impact and making the world a better place," says Professor Kaafar. "In this case, it is so real and so important."

"The Foundation wanted to make this project completely [open source](#) from the very beginning," Dr. Vatsalan adds. "So the algorithm we implemented is being published openly."

More information: Nan Wu et al, Privacy-Preserving Record Linkage for Cardinality Counting, *Proceedings of the ACM Asia Conference on Computer and Communications Security* (2023). [DOI: 10.1145/3579856.3590338](#)

Provided by Macquarie University