

# Study applies game theory to genomic privacy

January 17 2017, by Paul Govern

---

It comes down to privacy—biomedical research can't proceed without human genomic data sharing, and genomic data sharing can't proceed without some reasonable level of assurance that de-identified data from patients and other research participants will stay de-identified after they're released for research.

Data use agreements that carry penalties for attempted re-identification of participants may be a deterrent, but they're hardly a guarantee of privacy. Genomic data can be partially suppressed as they're released, addressing vulnerabilities and rendering individual records unrecognizable, but suppression quickly spoils a data set's scientific usefulness.

A new study from Vanderbilt University presents an unorthodox approach to re-identification risk, showing how optimal trade-offs between risk and scientific utility can be struck as genomic data are released for research.

The study appears in the *American Journal of Human Genetics*.

Doctoral candidate Zhiyu Wan, Bradley Malin, Ph.D., and colleagues draw on game theory to simulate the behavior of would-be data privacy adversaries, and show how marrying data use agreements with a more sensitive, scalpel-like data suppression policy can provide greater discretion and control as data are released. Their framework can be used to suppress just enough genomic data to persuade would-be snoops that

their best privacy attacks will be unprofitable.

"Experts in the privacy field are prone to assume the worse case scenario, an attacker with unlimited capability and no aversion to financial losses. But that may not happen in the real world, so you would tend to overestimate the risk and not share anything," Wan said. "We developed an approach that gives a better estimate of the risk."

Malin agrees that failure to come to grips with real-world risk scenarios could stifle genomic data sharing.

"Historically, people have argued that it's too difficult to represent privacy adversaries. But the game theoretic perspective says you really just have to represent all the ways people can interact with each other around the release of data, and if you can do that, then you're going to see the solution. You're doing a simulation of what happens in the real world, and the question just becomes whether you've represented the rules of the game correctly," said Malin, associate professor of Biomedical Informatics, Biostatistics and Computer Science.

To date, no one has faced prosecution for attacking the privacy of de-identified genomic data. Privacy experts nevertheless assume a contest of computerized algorithms as de-identified data are released, with privacy algorithms patrolling the ramparts while nefarious re-identification algorithms try to scale them.

Re-identification attacks have occurred, but according to earlier research by Malin and colleagues, the perpetrators appear to be motivated by curiosity and academic advancement rather than by criminal self-interest. They're sitting at computers just down the hall, so to speak, overpowering your data set's de-identification measures, then publishing an academic paper saying just how they did it. It's all very bloodless and polite.

The new study is something different, more tough-minded, situating data sharing and privacy algorithms in the [real world](#), where people go to jail or are fined for violations. Here the envisaged privacy adversary doesn't wear elbow patches, lacks government backing and is simply out to make a buck through the illicit sale of private information.

De-identified genotype records are linked to de-identified medical, biometric and demographic information. In what the study refers to as "the game," the attacker is assumed already to have some named genotype data in hand, and will attempt to match this identified data to de-identified genotype records as study data are released.

To bring these prospective attackers out of the shadows, the authors present a detailed case study involving release of genotype data from some 8,000 patients. They painstakingly assign illicit economic rewards for the criminal re-identification of research data. Based on costs for generating data, they also assign economic value to the scientific utility of study data.

On the way to estimating risk and the attacker's costs, the authors estimate the likelihood that any named individual genotype record already held by the attacker is included in the de-identified data set slated for release; according to the authors, this key estimate is often neglected in re-identification risk assessments.

The authors measure the utility of a study's genomic data in terms of the frequencies of genetic variants: for a given variant, the greater the difference between its frequency in the study group and its frequency in the general population (based on available reference data), the greater its scientific utility. This approach to utility triumphed recently when Wan and Malin won the 2016 iDASH Healthcare Privacy Protection Challenge. Their winning algorithm proved best at preserving the scientific utility of a genomic data set while thwarting a [privacy](#) attack.

For any [genomic data](#) set, before any data are released in a game's opening move, the sharer can use the game to compare various data sharing policies in terms of risk and utility. In the case study, the game theoretic policy provides the best payoff to the sharer, vastly outperforming a conventional data suppression policy and edging out a data use agreement policy.

No matter where parameters are set regarding illicit financial rewards or information that's likely to be wielded by an attacker, the authors show that the game theoretic approach generally provides the best payoff to the sharer. They sketch how their approach could serve the release of data from other sources, including the federal government's upcoming Precision Medicine Initiative.

Provided by Vanderbilt University Medical Center

Citation: Study applies game theory to genomic privacy (2017, January 17) retrieved 10 April 2024 from <https://medicalxpress.com/news/2017-01-game-theory-genomic-privacy.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
---