

# New statistical method exponentially increases ability to discover genetic insights

January 8 2021

---



Credit: CC0 Public Domain

Pleiotropy analysis, which provides insight on how individual genes result in multiple characteristics, has become increasingly valuable as medicine continues to lean into mining genetics to inform disease treatments. Privacy stipulations, though, make it difficult to perform comprehensive pleiotropy analysis because individual patient data often can't be easily and regularly shared between sites. However, a statistical

method called Sum-Share, developed at Penn Medicine, can pull summary information from many different sites to generate significant insights. In a test of the method, published in *Nature Communications*, Sum-Share's developers were able to detect more than 1,700 DNA-level variations that could be associated with five different cardiovascular conditions. If patient-specific information from just one site had been used, as is the norm now, only one variation would have been determined.

"Full research of pleiotropy has been difficult to accomplish because of restrictions on merging [patient data](#) from [electronic health records](#) at different sites, but we were able to figure out a method that turns summary-level data into results that are exponentially greater than what we could accomplish with individual-level data currently available," said the one of the study's senior authors, Jason Moore, Ph.D., director of the Institute for Biomedical Informatics and a professor of Biostatistics, Epidemiology and Informatics. "With Sum-Share, we greatly increase our abilities to unveil the [genetic factors](#) behind [health conditions](#) that range from those dealing with heart health, as was the case in this study, to mental health, with many different applications in between."

Sum-Share is powered by bio-banks that pool de-identified patient data, including genetic information, from electronic health records (EHRs) for research purposes. For their study, Moore, co-senior author Yong Chen, Ph.D., an associate professor of Biostatistics, lead author Ruowang Li, Ph.D., a post-doc fellow at Penn, and their colleagues used eMERGE to pull seven different sets of EHRs to run through Sum-Share in an attempt to detect the genetic effects between five cardiovascular-related conditions: obesity, hypothyroidism, type 2 diabetes, hypercholesterolemia, and hyperlipidemia.

With Sum-Share, the researchers found 1,734 different single-nucleotide polymorphisms (SNPs, which are differences in the building blocks of

DNA) that could be tied to the five conditions. Then, using results from just one site's EHR, only one SNP was identified that could be tied to the conditions.

Additionally, they determined that their findings were identical whether they used summary-level data or individual-level data in Sum-Share, making it a "lossless" system.

To determine the effectiveness of Sum-Share, the team then compared their method's results with the previous leading method, PheWAS. This method operates best when it pulls what individual-level data has been made available from different EHRs. But when putting the two on a level playing field, allowing both to use individual-level data, Sum-Share was statistically determined to be more powerful in its findings than PheWAS. So, since Sum-Share's summary-level data findings have been determined to be as insightful as when it uses individual-level data, it appears to be the best method for determining genetic characteristics.

"This was notable because Sum-Share enables loss-less data integration, while PheWAS loses some information when integrating information from multiple sites," Li explained. "Sum-Share can also reduce the multiple hypothesis testing penalties by jointly modeling different characteristics at once."

Currently, Sum-Share is mainly designed to be used as a research tool, but there are possibilities for using its insights to improve clinical operations. And, moving forward, there is a chance to use it for some of the most pressing needs facing health care today.

"Sum-Share could be used for COVID-19 with research consortia, such as the Consortium for Clinical Characterization of COVID-19 by EHR (4CE)," Yong said. "These efforts use a federated approach where the data stay local to preserve privacy."

Provided by Perelman School of Medicine at the University of Pennsylvania

Citation: New statistical method exponentially increases ability to discover genetic insights (2021, January 8) retrieved 6 May 2024 from <https://medicalxpress.com/news/2021-01-statistical-method-exponentially-ability-genetic.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.