Statistical Methods to Preserve Patient Privacy When Sharing and Analyzing Data

Principal investigator

Darren Toh, ScD, BPHAR, MS

Organization

Harvard Pilgrim Health Care

What was the research about?

Sometimes a study can get better results using data from different sites. In these cases, researchers may want to share patient data, including personal and private information such as dates of birth and addresses. However, researchers may not want to share data across sites because of worries about patient privacy. Some statistical methods can change patients' sensitive individual data into summary data that hides individuals' personal information. These privacy-protecting methods, or PPMs, make it safe to share data across sites. But researchers don't know if PPMs produce accurate results.

In this study, the research team compared combinations of PPMs with methods that use patients' individual data.

What were the results?

PPMs provided similar results to methods that used patients' individual data. Some PPMs worked better than others.

What did the research team do?

The research team used a computer program to create data that mimicked real patient data. The team

used three ways to change the individual data into summary data. Then they analyzed the summary data using a combination of PPMs. The team compared the results they got using summary data with results from individual data. Finally, the team tested the PPMs using real-world data.

Patients, researchers, and hospital staff provided input during this study.

What were the limits of the study?

The research team studied only some PPMs and types of data. Results may differ with other PPMs or data.

Future research could test PPMs in other research studies with data from different sources.

How can people use the results?

Researchers can use the results when considering how to share data across research sites.

To learn more about this project, visit www.pcori.org/Toh210.