Protecting confidentiality of public use genetic data: What can we learn from the experiences of statistical agencies?

Dr. Jerome P. Reiter

Department of Statistical Science,

Duke University

Abstract:

When data are shared with the public, i.e. beyond the original investigators, the data disseminator is ethically and often legally obligated to protect the confidentiality of subjects' identities and attributes. At the same time, the data disseminator seeks to share high quality data that afford valid inferences for a wide range of analyses.

In this talk, I provide an overview of some of the methods used by statistical agencies to protect confidential data, with an emphasis on the use of multiple imputation. The discussion is in the context of protecting demographic or standard medical data, as there has been little research on methods for protecting genetic data shared with the public.

I propose some potential avenues of research for sharing genetic data.