



# Evaluating the Utility of a Differentially Private Behavioral Science Dataset

**Raquel L. Hill**

**Indiana University**

**ralhill@indiana.edu**

Social and behavioral scientists often collect and maintain datasets that are high-dimensional, including some combination of demographic, medical, sexual, and other personal information), which presents opportunities to characterize participants in unique ways. The conventional wisdom for protecting the privacy of such participants is to either not ask certain questions or to remove or recode potentially identifiable information. The premise of the research discussed here is that neither approach may be sufficient for preventing the (re)identification of participants in large and/or multidimensional datasets. In this work, I present new results of a use-case analysis that evaluates *Differential Privacy* (hereafter referred to as DP) as a technique to protect behavioral science datasets while preserving their research utility. DP is a data perturbation technique that provides strong and formal privacy guarantees. The essential goal is to prevent a possible adversary from discovering whether or not some specific individual's data is present in a differentially private dataset, given some risk threshold.

## Biography

Raquel Hill is an Associate Professor of Computer Science in the School of Informatics and Computing at Indiana University. Her primary research interests are in the areas of trust and security of distributed computing environments and data privacy with a specific interest in privacy protection mechanisms for medical-related datasets. Dr. Hill's research is funded by various sources, including the National Science Foundation. She holds B.S. and M.S. degrees in Computer Science from the Georgia Institute of Technology and a Ph.D. in Computer Science from Harvard University.