

## Identity disclosure risk control in microdata release via post-randomization

Tapan K. Nayak

Department of Statistics, George Washington University, Washington, DC 20052, USA; e-mail:  
tapan@gwu.edu

Protecting the confidentiality of survey respondents' information in microdata release is a significant concern for most statistical agencies. To address the matter, agencies often publish data that has been perturbed using data swapping, noise addition, multiple imputation and other methods. However, agency reports and research papers in this area rarely state disclosure risk and disclosure protection goals clearly. In this paper, we propose a precise and strict identity disclosure protection goal and then present a post-randomization procedure for achieving that goal, at a minimal loss of data quality. The procedure can also allow agencies to retain selected marginal counts from original data and avoid undesirable changes to the original data.

**Keywords:** confidentiality protection; sample unique; key variable; correct match probability; data perturbation; data utility.