

Professor Xiao-Li Meng
Harvard University, USA

TITLE:

Statistics Can Lie But Can Also Correct for Lies:

Reducing Response Bias in NLAAS via Bayesian Imputation

ABSTRACT:

This talk is based on the joint work with Liu, Chen and Alegria with the above title and the following abstract: "National Latino and Asian American Study (NLAAS) is a multi-million-dollar survey of psychiatric epidemiology, the most comprehensive survey of this kind and its data were made public in July 2007. A unique feature of NLAAS is its embedded experiments for estimating effect of alternative orderings of interview questions. The findings from the experiment are not completely unexpected, but nevertheless astonishing. Compared to the survey results from the widely used traditional ordering, the self-reported psychiatric service-use rates are often doubled or even tripled under a more sensible ordering introduced by NLAAS. These findings partially answer some perplexing questions in the literature, e.g., why the self-reported rates of using religion services were typically much lower than results from other empirical evidences? But at the same time, they also impose some grand challenges, for example, how can one assess racial disparities when different races were surveyed with different survey instruments (e.g., the exiting data on the White populations were collected using the traditional questionnaire ordering) that are now known to induce substantial differences? The project documented in this paper is part of the effort in addressing these questions, by creating models for imputing the correct responses had the respondents under the traditional survey had not taken advantage of the skip patterns to reduce interview time, which resulted in increased rates of untruthful negative responses over the course of the interview. The imputation modeling task is particularly challenging because of the complexity of the questionnaire, the small sample sizes for subgroups of interests, the existence of high-order interactions among variables, and above all, the need of providing sensible imputation for whatever subpopulation a future user might be interested in studying. This paper is intended to serve three purposes: (1) to provide a published record of the key steps and strategies adopted in creating the released multiple imputation for NLAAS, (2) to alert the potential users of the limitations of the imputed data, and (3) to provide a vivid demonstration of the type of challenges and opportunities typically encountered in modern applied statistics.