Public SEMINAR

Committee on National Statistics

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

Friday, May 6, 2016, 2:00–4:30 pm
National Academy of Sciences Main Building
2101 Constitution Ave. NW, Washington, DC – Lecture Room

2:00 pm   ***Light refreshments for Seminar Guests (East Court)***

2:30   *Welcome*
            —**Lawrence Brown**, CNSTAT Chair and the University of Pennsylvania

2:35   *Developments at the OMB Statistical and Science Policy Office*
            —**Katherine Wallman**, Chief Statistician of the U.S.

2:45   *Featured Topic:*
            ***Combining Information from Survey and Non-Survey Data Sources for Policy
            Research: Challenges and Opportunities***
            —**Sharon Lohr**, Westat, and **Trivellore Raghunathan**, University of Michigan

Abstract: *Social media, electronic health records, credit card transactional and administrative data, web scraping, and numerous other ways of collecting information have changed the landscape for addressing policy-relevant research questions. At the same time, traditional sources of data for policy-relevant research, such as large-scale surveys, have suffered setbacks due to increasing nonresponse and data collection costs. The non-survey data usually contain detailed information on certain behaviors for a large number of individuals (such as all credit card transactions) but very little background information (such as important covariates to address the policy-relevant question). In contrast, the survey data may contain detailed information on covariates but not on the behaviors. Both data sources are likely imperfect for the target population. Combining information from multiple data sources may allow us to draw inferences that extend beyond the individual components. We outline some proposed statistical methods for combining information from multiple data sources, discuss their limitations and challenges, and identify areas for future research. We provide a case-study of analyzing health care costs by developing an explicit modeling framework to construct inferences for various "building block" parameters and using them to dissect costs as a function of covariates and health conditions.*

3:45   *Discussants*
            —**Thomas Louis,** outgoing Associate Director for Research and Methodology,
            U.S. Census Bureau, and Professor of Biostatistics, Johns Hopkins Bloomberg
            School of Public Health
            —**Linda Young**, Director of Research and Development, National Agricultural
            Statistics Service

4:15   *Floor discussion*

4:30   ***Reception – East Court***