Submission for the 2022 IAOS Prize for Young Statisticians

Statistical data integration models to reconcile health official statistics

Dr. Andreea L. Erciulescu

Senior Statistician Westat

westat.com Rockville, MD, USA

Abstract

There are at least two sources of health insurance coverage estimates in the United States: the Behavioral Risk Factor Surveillance System (BRFSS) and the Small Area Health Insurance Estimates (SAHIE) program. This paper addresses the integration of BRFSS and SAHIE data using multilevel models that account for the different levels of aggregation at which data are available and for the different errors to which data are subject to. The precision of the initial state-level estimates available from BRFSS and SAHIE is improved by borrowing information from both surveys and across geographies. County-level model estimates are produced on both BRFSS and SAHIE scales, improving the usability of the BRFSS public-use data and inspiring possible extensions to estimation of BRFSS quantities other than health insurance coverage. The application uses 2018 public-use data. Parallels to small area estimation models and measurement error models are briefly discussed.

Keywords: different levels of aggregation, health insurance coverage, hierarchical Bayes, small area estimation, multilevel models