Variable inclusion strategies through directed acyclic graphs to adjust health surveys subject to selection bias for producing national estimates

Along with the rapid emergence of web surveys to address time-sensitive priority topics, various propensity score (PS)-based adjustment methods have been developed to improve population representativeness for nonprobability- or probability-sampled web surveys subject to selection bias. Conventional PS-based methods construct pseudo-weights for web samples using a higher-quality reference probability sample. The bias reduction, however, depends on the outcome and variables collected in both web and reference samples. A central issue is identifying variables for inclusion in PS-adjustment. Directed acyclic graphs (DAG), a common graphical tool for causal studies but largely under-utilized in survey research, are used to examine and elucidate how different types of variables in the causal pathways impact the performance of PS-adjustment. While past literature generally recommends including all variables, our research demonstrates that only certain types of variables are needed in PS-adjustment. Our research is illustrated by NCHS Research and Development Survey, a probability-sampled web survey with potential selection bias, PS-adjusted to the National Health Interview Survey, to estimate U.S. asthma prevalence. Findings from this research can be used to design questionnaires with variables that improve web-samples population representativeness and to release more timely and accurate estimates for priority topics.