

The Role of Data Collection in Population Science: Contemporary Studies from ABCD to HBCD

Abstract

Recently nationwide consortiums of multiple research sites have conducted multi-modal, longitudinal cohort studies and provided unprecedented data sources for population science research. For example, the Adolescent Brain Cognitive Development (ABCD) Study has collected data from 11,880 children ages 9-10 across 21 U.S. research sites, as the largest long-term study of brain development and child health; and the Healthy Brain and Child Development (HBCD) Study will enroll 7,500 pregnant women across 25 research sites and follow them from pregnancy through early childhood, as the largest long-term study of early brain and child development in the U.S. Both studies aim to reflect the sociodemographic diversity of the target population to enable characterization of natural variability and trajectories. Without probability sampling as the touchstone for randomization-based inferences, the data quality and analysis validity require rigorous evaluations and potentially rely on untestable assumptions. The data collection process also presents various challenges during practical operation.

In this talk, I look into both inference and design schemes to study the impact of data collection on population science. First, using the ABCD study as an example of secondary data analysis, I discuss inference approaches focusing on multilevel regression and poststratification for population generalizability and latent subgroup detection for population heterogeneity in brain activity and association studies. Second, I introduce the HBCD study design. HBCD also aims to include individuals demographically and behaviorally similar to those in the substance exposure group, but without exposure, to enable valid causal inference in a non-experimental study design. I discuss our proposed weighting, matching, and modeling strategies to leverage analysis goals to inform the design and dashboard monitoring for adaptive sample enrollment.

Bio

Yajuan Si is a Research Associate Professor in the Institute for Social Research at the University of Michigan. Dr Si's research lies in cutting-edge methodology development in streams of Bayesian statistics, linking design- and model-based approaches for survey inference, missing data analysis, confidentiality protection involving the creation and analysis of synthetic datasets, and causal inference with observational data.