

When “representative” surveys fail: Can a non-ignorable missingness mechanism explain bias in estimates of COVID-19 vaccine uptake?

Recently, attention was drawn to the failure of two very large internet-based probability surveys to correctly estimate COVID-19 vaccine uptake in the U.S. in early 2021. Both the Delphi-Facebook COVID-19 Trends and Impact Survey (CTIS) and Census Household Pulse Survey (HPS) overestimated vaccine uptake substantially (14 and 17 points in May 2021) compared to retroactively available CDC benchmark data. These surveys had large numbers of respondents but very low response rates (<10%), and thus non-ignorable nonresponse could have substantially impacted estimates. Specifically, it is plausible that “anti-vaccine” individuals were less likely to complete a survey about COVID-19; we might also hypothesize that “anti-vaccine” individuals could be suspicious of the government and thus less likely to respond to an official government-sponsored survey. In this talk we use *proxy pattern-mixture models* (PPMMs) to retrospectively estimate the proportion of adults (18+) who received at least one dose of a COVID-19 vaccine, using data from the CTIS and HPS, under a non-ignorable nonresponse assumption. We compare these estimates to the true benchmark uptake numbers and show that the PPMM could have detected the direction of the bias and have provided meaningful bias bounds. We also use the PPMM to estimate vaccine hesitancy, a measure without a benchmark truth, and compare to the direct survey estimates. We conclude with discussion of how the PPMM could be prospectively as part of an assessment of nonresponse and/or selection bias, factors that would facilitate such analyses in the future, and ongoing work to extend the PPMM to novel areas.

Rebecca Andridge is an Associate Professor of Biostatistics at The Ohio State University College of Public Health. She conducts methodologic work in imputation methods for missing data, primarily in large-scale probability samples, and measures of selection bias for nonprobability samples. In particular, she works on methods for imputing data when missingness is driven by the missing values themselves (missing not at random). She teaches introductory graduate and undergraduate biostatistics and won the College's Outstanding Teaching Award in 2011 and is a Fellow of the American Statistical Association.