ASA News
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®
732 North Washington Street, Alexandria, VA 22314 • (703) 684-1221 • Toll Free: (888) 231-3473 • www.amstat.org • www.twitter.com/AmstatNews

# Alarming Error Common in Survey Analyses

It is difficult to overstate the importance of survey data: They tell us who we are and—in the hands of policymakers—what to do.

It had long been apparent to Brady West, an expert on survey methodology at the University of Michigan, Ann Arbor, that the benefits of survey data coexisted with a lack of training in how to interpret them correctly, especially when it came to secondary analyses—researchers reanalyzing survey data that had been collected by a previous study.

"In my consulting work for organizations and businesses, people would come in and say, 'Well, here's my estimate of how often something occurs in a population,' such as the rate of a disease or the preferences for a political party. And they'd want to know how to interpret that. I would respond, 'Have you accounted for weighting in the survey data you're using—or, did you account for the sample design?' And I would say, probably 90 percent of the time, they'd look at me and have no idea what I was talking about. They had never learned about the fundamental principles of working with survey data in their standard Intro to Stats classes."

As a survey methodologist, West wondered whether his experience was indicative of a systemic problem. There wasn't much in the academic literature to answer the question, so he and his colleagues, Joseph Sakshaug and Guy Aurelien, sampled 250 papers, reports and presentations—all available online, all conducting secondary analyses of survey data—to see if these analytic errors were, indeed, common.

"It was quite shocking," says West. "Only about half of these analyses claimed to account for weighting, the impact of sample designs on variance estimates was widely misunderstood and there was no sign of improvement in these problems over time." But possibly worst of all, these problems were just as prevalent in the peer-reviewed literature in their sample as they were in technical reports and conference presentations. "That's what was really most shocking to me," says West. "The peer-review process was not catching these errors."

An alarming example of what can happen when you compute an estimate but ignore the survey weighting can be found in the 2010 National Survey of College Graduates (NSCG). "This is a large national survey of college graduates, and they literally say in their documentation that they're oversampling individuals with science and engineering degrees," says West. "If you take

account of the weighting, which corrects for this oversampling, about 30 percent of people are getting science and engineering degrees; if you forget about the weighting, you extrapolate the oversample to the entire population, and suddenly 55 percent of people have science and engineering degrees."

Ironically, better sampling of under-studied populations may be exacerbating the problem. "There's a lot of interest in under-represented populations, such as Hispanics," says West. "So, a lot of national surveys oversample these groups and others to create a big enough sample for researchers to adequately study. But when Average Joe Researcher grabs all the data—not just the data from the subpopulation they're interested in, but everybody, whites, African Americans, and Hispanics—and then they try to analyze all that data collectively, that's when oversampling can have a horrible effect on the overall picture if that feature of the sample design is not accounted for correctly in estimation."

There are many easy-to-use software tools that can easily account for the sampling and weighting complexities associated with survey data, but the fact they are not being used speaks to the underlying problem.

"This problem originates in the fact that the people publishing these articles just aren't told about any of this in their training," says West. "We've known about the importance of survey weighting for nearly a century—but somehow how to deal with weighted survey data hasn't penetrated the statistics classes that researchers take at the undergraduate or graduate level. We spend a fortune on doing national surveys—and who knows how much misinterpreting that data is costing us."

To solve that problem, West is helping design a MOOC (massive open online course) at the University of Michigan introducing statistics with the software Python. Weighting and correct survey analyses will be taught in the very first course of that specialization. "We're really focusing on making sure that before you jump into any analyses of survey data, you have a really firm understanding of how the data were collected and where they came from."

**JSM talk:**
http://ww2.amstat.org/meetings/jsm/2018/onlineprogram/AbstractDetails.cfm?abstractid=326973

**Study link:** http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0158120

**For further details, contact:** Brady West

**Email:** bwest@umich.edu

**Tel:** (734) 223-9793

**Webpage:** www.umich.edu/~bwest

**About JSM 2018**

JSM 2018 is the largest gathering of statisticians and data scientists in the world, taking place July 28–August 2, 2018, in Vancouver. Occurring annually since 1974, JSM is a joint effort of the American Statistical Association, International Biometric Society (ENAR and WNAR), Institute of Mathematical Statistics, Statistical Society of Canada, International Chinese Statistical Association, International Indian Statistical Association, Korean International Statistical Society, International Society for Bayesian Analysis, Royal Statistical Society and International Statistical Institute. JSM activities include oral presentations, panel sessions, poster presentations, professional development courses, an exhibit hall, a career service, society and section business meetings, committee meetings, social activities and networking opportunities. http://ww2.amstat.org/meetings/jsm/2018/index.cfm

**About the American Statistical Association**

The ASA is the world's largest community of statisticians and the oldest continuously operating professional science society in the United States. Its members serve in industry, government and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare. For additional information, please visit the ASA website at www.amstat.org.