

Psychological Scientists Call for Paradigm Shift in Data Practices

July 10, 2013

Fabricating data to support an *a priori* hypothesis is the ultimate sin in scientific research. But what about throwing out an “outlier” or two? Or reporting some, but not all, of the measures you tested?

These questionable research practices tend to fly under the radar, but they present a real challenge to the rigor and replicability of science. Scientists discussed these practices and the steps that can be taken to combat them during the “Good Data Practices” symposium at the 25th APS Annual Convention. The symposium was part of the “Building a Better Psychological Science: Good Data Practices and Replicability” theme program.

“Questionable research practices — exploitation of the gray zone of acceptable practice — are sometimes justified, but often not,” said panelist Leslie John of Harvard Business School. “Because of this duality, these practices provide considerable latitude for rationalization.”

Even, it seems, for Nobel Prize-winning researchers.

“I always had a rule to replicate [my findings] before publishing [them], but I had two findings — two correlation matrices — that I couldn’t replicate,” said panelist Daniel Kahneman of Princeton University, who was awarded the Nobel Memorial Prize in Economic Sciences in 2002.

After years of trying to understand why he couldn’t replicate the findings, he figured it out:

“It turned out my samples were ridiculously small,” he said, “which was a bit shameful, actually, because I was teaching statistics at the time.”

This insensitivity to sample size, one of the first cognitive biases that Kahneman documented with his longtime collaborator Amos Tversky, can get researchers into serious trouble.

Underscoring this point, panelist Uri Simonsohn of the Wharton School at the University of Pennsylvania noted that the median study has about 20 people per condition. This would be enough to detect that men are, on average, taller than women and that people above the median age are closer to retirement. But it’s not sufficient to detect that people who like spicy food tend to like Indian food, for example. It’s not even sufficient to detect that men tend to weigh more than women, which requires about 46 people in each condition. This signals that researchers ought to increase their sample sizes. But doing so isn’t necessarily an easy feat.

“When you can run a whole classroom at once, then you don’t have a problem with sample size,” Kahneman observed. “But there are many who don’t have that luxury — it may take them an hour to collect one data point for one subject.”

And yet, the problem is too big to ignore. As Kahneman noted, researchers can't simply combine several underpowered studies and assume that they amount to a robust finding. So what *should* they do?

"My recommendation, for what it's worth, is that there should be one study — in which people present the hypothesis — that is the flagship study with sufficient power," said Kahneman. The flagship study can then be complemented with additional studies that have smaller samples.

"We must beef up the story of the data to make it comprehensible," he concluded.

But sample size isn't the only concern when it comes to questionable research practices — poor data documentation brings a whole other set of issues, as Jelte Wicherts of Tilburg University in The Netherlands noted.

Once a researcher has successfully recruited participants and collected data, how he or she chooses to label his or her variables in SPSS may seem relatively inconsequential. And yet, improper data documentation can significantly hinder science. In two years, the researcher may not remember what the data for "VAR001" represent, what measures were used to collect the data, or whether the data might have been transformed in some way. And someone else looking at the data certainly wouldn't be able to tell. Thus, a seemingly inconsequential practice can actually have a major impact on scientific accuracy and transparency.

"Poor data handling practices make the chance of sharing data later smaller and increase considerably the chance of making errors," Wicherts observed.

A major problem, as all of the panelists noted, is that the incentive structure of most scientific fields isn't necessarily designed to support good data practices, including data documentation, data sharing, and conducting direct replications.

"I think that the vast majority of researchers are really sincerely motivated to conduct sound research," said John. "But [the] inherent ambiguity [in research practices], plus the tremendous incentives to publish positive results, means we have a massive conflict of interest on our hands."

Just like organic farmers, researchers who subscribe to good data practices often end up with less perfect and lower-yield products; how then can they possibly compete? They need to label their product, Simonsohn argued.

"If you're already playing by the rules, you're shooting yourself in the foot if you don't label it," he said. "You're not letting people know."

Researchers may be hesitant to share data that doesn't show a clear and strong effect, but openness and transparency are key to furthering knowledge in the field, the panelists said.

"There's no such thing as a failed study," Wicherts concluded. "The only failed studies are when the functional MRI machine blows up or your participants run screaming out of the lab."