

washingtonpost.com

Yes, you can trust international surveys. Mostly.

By Michael Robbins



(Kacper Pempel/Reuters)

How do you know when research is based on falsified data? That's one of the challenges faced throughout statistical research. Noble Kuriakose and I authored a paper (forthcoming in the [Statistical Journal of the IAOS](#)) that outlines a technique to detect potential fraud in public opinion surveys. So far its merits have been debated [on this blog](#), on [Andrew Gelman's blog](#), and in the pages of [Science](#). The Pew Research Center has also published [several](#)

[rejoinders](#) to our paper.

We wanted to develop a way to detect a type of potential survey fraud that we suspected was widely unidentified — what we call “near matches.” Near matches occur when the answers of many survey respondents match on nearly every question. This, too, is suspicious, for reasons I will explain below.

Duplicates are suspicious. So are near-duplicates.

Here’s the basic intuition that inspired our method. Organizations that do international survey research perform checks on data quality, including removing duplicate responses, which are cases where the answers of one respondent match those of another respondent for every question. The reason is simple: Most major international survey projects have lengthy questionnaires, making it extremely unlikely that any two individuals will provide the exact same responses to all questions. Instead, a more likely cause of a pair of exact duplicates is that a firm accidentally entered responses from the same form twice, or, more perniciously, someone at the local firm hired to do the survey intentionally copied a valid interview to save time or money.

Yet, the likelihood of two individuals answering 99 percent of questions in the same way is, at least in statistical terms, virtually no different than having an exact match. Why then do researchers throw out exact matches but retain those that are extremely near matches? It seemed to us that there were two basic reasons.

First, there was no readily available program to test for near

matches, meaning they went undetected. To address this issue, we developed a new Stata program called [percentmatch](#), which is publicly available and free to download. Percentmatch identifies how similar each observation is to its most similar neighbor. Second, there was no proposed standard for how near a match would have to be to indicate possible fraud. We wrote our paper to address this second issue.

To develop a standard threshold for how near a match is too near, we ran hundreds of computer simulations. These simulations consistently yielded two expectations. First, it is relatively rare for two respondents to match on more than 85 percent of questions, at least under conditions found in most social science surveys. Second, regardless of the precise threshold, the distribution of percentmatch should converge on a [Gumbel distribution](#), a statistical distribution designed to model extreme values.

Now a word about the expectation that has garnered the most attention so far: our proposed 85 percent cutoff. While it is appealing to have one figure to rely on, as with most statistical thresholds, the 85 percent cutoff is not a magic number. While we think it is an appropriate threshold for the types of surveys we analyzed, we note in our article we do not believe our proposed test for identifying possible fraud should be applied to all surveys in all contexts.

Our tool flagged many suspect surveys. But “flagged” doesn’t mean they were fraudulent.

Since it is virtually impossible to fully simulate real-world survey

conditions, our next step was to test our expectations using more than 1,000 publicly available, cross-national data sets from a wide variety of survey projects. The good news is that the vast majority of surveys passed these basic tests. The bad news is that a substantial minority — nearly 1 in 5 by our measures — did not. Although our editors advised us not to publish the names of the surveys we examined due to potential legal implications, we found that most of the projects we analyzed had one or more surveys that were flagged by our program.

The key word here is flagged. An important caveat is a point we highlight throughout our paper: Our test is designed to flag signs of potential fabrication, rather than be a definitive test of falsified data. In some instances, further investigation may reveal that no fabrication has occurred. For example, the Pew Research Center has identified [one such example in a survey on religion in the United States](#).

Additionally, it is important to highlight that we believe fabrication is mostly happening at the local level. International survey organizations contract local firms or research units to conduct public opinion surveys. These local partners then hire interviewers to do the research. Our suspicion is that the interviewers or the local partners are the most likely source of the probable fabrication. In other words, the research organizations coordinating these efforts are unlikely to be the source of the fabrication. Instead, they too are victims.

Many international survey projects are now using the tool

Fortunately, many international survey projects have already adopted our tool, and there is evidence it is already yielding important gains in data quality. For example, thanks to our program, one major international survey project realized that an administrative error had resulted in the publication of an incorrect file. This mistake was easily corrected and the valid data set is now on its website.

In our paper, we provide another example of the benefits of using this tool. We describe how the project I direct, the [Arab Barometer](#), used the tool to identify a small subset of fabricated interviews in a country survey. Meanwhile, other researchers have reported that when the tool has been used to test previous surveys, it has been effective at flagging surveys that they already identified as having potential data quality problems through other, more costly methods. In short, growing evidence makes it clear that the program has become an efficient and effective tool for researchers seeking to prevent data fraud in survey research.

A useful byproduct of this process has been the open discussion of data quality issues among experienced survey researchers. Within the public sphere, however, interest in our paper has primarily centered on a single basic question: Can you trust international public opinion surveys? In general, at least by this new method, our response would be yes, since most surveys pass our test.

But a more nuanced answer is that it's complicated. Our findings highlight the need for ongoing vigilance and for research organizations to take additional, and often costly, steps to help prevent fabrication in the future. Most of these steps are rather

technical but generally well known to survey research practitioners, so they are not addressed here.

Monkey Cage newsletter

Commentary on political science and political issues.

Please provide a valid email address.

How can international surveys get even more trustworthy?

What merits additional discussion is how the field can improve the trustworthiness and quality of international survey research. An essential step toward better surveys is to increase transparency regarding survey methodology and results. The need for transparency is the reason we made our program publicly available and free to download. Such transparency requires making data publicly available as well.

Fortunately, many projects already make their data available for public scrutiny, including the [Global Barometer](#) project and its associated regional barometers, the [World Values Survey](#), the [Pew Research Center](#), the [International Social Survey Programme](#), the [Latin American Public Opinion Project](#), and the [European Social Survey](#), among others. These projects should be lauded for taking this step, which allows independent researchers to conduct a more complete assessment of their data quality.

Unfortunately, it is difficult to assess the quality and trustworthiness of the international survey research of other projects that publicize

their results but do not routinely make their survey data files available to researchers, or do so only for a prohibitive fee. Many such surveys are commonly cited in media and policy reports that inform the public debate — despite the fact that it is virtually impossible for researchers to perform independent assessments of data quality. Until such projects receive the same scrutiny as those that are publicly available, it is far more difficult to say whether we can trust their results.

Michael Robbins ([@mdhrobbins](#)) is a research fellow at the University of Michigan, a senior research specialist at Princeton University, and director of the [Arab Barometer](#) ([@arabbarometer](#)).