Guest Blog: More on the Problems with Opt-in Internet Surveys

September 28, 2009
By GARY LANGER

In a Sept. 1 post I reported on a groundbreaking study by a team of researchers led by David Yeager and Jon A. Krosnick of Stanford University, finding significant data quality problems in surveys of people who sign up to click through online questionnaires – so-called “opt-in” panels. Their study, laudably, was accompanied by highly detailed methodological disclosure.

Postings challenging its conclusions followed – one here from Prof. Douglas Rivers, CEO of an opt-in online company; another here from Joel Roblson, chief research officer of the Advertising Research Foundation, many of whose members conduct or purchase such studies. At my invitation, Yeager, Krosnick and one of their co-authors, Harold Javitz of SRI International, a nonprofit research institute, have written a reply. It follows.

A bit of background: A professor of communication, political science and psychology, Krosnick is one of the nation’s leading academics in the field of survey research; among other honors, he recently was elected to the American Academy of Arts and Sciences. He’s authored or co-authored more than 100 published articles or book chapters and six books on issues in survey research. Yeager is completing a PhD at Stanford. Both report no financial interest in any company that sells survey data, nor in any particular method used to collect survey data.

More on the Problems with Opt-in Internet Surveys

By David Yeager and Jon A. Krosnick, Stanford University
and Harold A. Javitz, SRI International, Inc.

We are delighted that our new paper comparing the quality of data obtained from RDD telephone surveys, probability sample Internet surveys, and non-probability sample Internet surveys has been the focus of some discussion across the country and may help providers and purchasers of survey data to understand survey research methods better.

During the weeks since our paper was released, a number of reasonable questions have been asked about the paper’s methods and findings (see note 1), so we are pleased to have the opportunity to answer some of those questions here.

For those not familiar with our paper, here is a brief summary. We commissioned nine firms to administer the same survey questionnaire in 2004/2005 via (1) RDD telephone interviewing with a probability sample of American adults, (2) Internet data collection from a probability sample of American adults, and (3) Internet data collection from samples of American adults who volunteered to do surveys for money or prizes and were not randomly sampled from the American adult population (we refer to the latter as “opt-in” samples).

Our principal findings include:

(1) The probability sample surveys done by telephone or the Internet were consistently highly accurate.

(2) The opt-in sample surveys done via the Internet were always less accurate and were sometimes strikingly inaccurate.

(3) Best practices weighting of the opt-in samples sometimes improved their accuracy and sometimes reduced their accuracy but never made them as accurate as the RDD telephone and probability sample Internet surveys.

Some of the questions that have been raised about our study include: (1) aren’t the data too old to be informative about research being done today? (2) aren’t the differences between methods so small as to be inconsequential? (3) were the data collected and analyzed even-handedly using best practices? (4) isn’t this a re-release of an old paper, first distributed almost 5 years ago?

Some of these questions were addressed in our original paper (as we describe below), and
Even if your data tell us about the accuracy of surveys done in 2004/2005, don't recent improvements in opt-in survey methodology and declines in the performance of telephone surveys make your findings irrelevant to today?

No. The figure above compares three of the 2004/2005 surveys we evaluated in our paper with surveys done by the same firms in 2009. For each firm, we computed average accuracy using a set of measures administered identically in that firm's 2004/2005 and 2009 surveys.

In no instance is a firm's 2009 average error significantly different from its 2004/2005 average error (see note 3). Thus, we see no evidence that the accuracy of probability sample surveys by those firms declined or that the accuracy of this opt-in survey firm's data improved since 2004.

We hope to conduct more such studies to assess whether average errors of data from other firms have changed over time.

Didn't you find that opt-in surveys were only slightly less accurate than the probability sample surveys?

No. The average errors for the two probability samples were 3.5% and 3.3%, whereas the average errors for the opt-in surveys ranged from 4.9% to 10.0% and averaged 6.0%.

Thus, the average error of the opt-in surveys was almost twice as big as for the probability samples.

But that is just average error. The largest unweighted error for a single item for the probability samples was 12 percentage points, whereas the largest such error for the opt-ins was 35.5 points. The comparable weighted numbers are 9% and 18%, respectively. And the standard deviation of the errors was nine times larger for the opt-in sample surveys than for the RDD telephone survey.

This is the basis for our conclusion that probability sample surveys are very consistently accurate, while opt-in surveys occasionally produce data points that are close to accurate, but usually don't, and often produce measurements that are strikingly inaccurate.

The Advertising Research Foundation (ARF) has just completed a study like yours, but using data collected in 2008. Does their study show that opt-in survey accuracy has improved dramatically since 2004/2005?

No. Combining 100,000 opt-in Internet survey respondents in that study, average error computed using what ARF calls "best practice weighting" was 6.2 percentage points, almost exactly the average error we found in 2004/2005 (6.0%) (see note 4).

Of course, few if any customers buy data collected from 100,000 respondents by 17 different opt-in survey companies. So we look forward to seeing the ARF results separately for each company, to assess the accuracy of samples that real customers have been buying — in terms of average error, individual item error, and standard deviations alike, and with details of the ARF methodology.

Isn't the apparent accuracy of the probability sample surveys that you commissioned illusory, because they were done using unusually expensive and high-quality methods under the watchful eye of academic researchers?

No. We were concerned about this possibility, so our paper reports analyses of data from RDD and probability Internet surveys we did not commission in addition to the ones we did commission. From a public archive of surveys, we drew a random sample of 6 national RDD surveys done at about the same time as the one we commissioned. And we drew a random sample of 6 national surveys from all those done by the probability sample Internet survey firm at that time as well.

The RDD surveys' methods were all much less elaborate than those of the RDD survey we commissioned. For example, among the 6 additional RDD surveys we did not commission, data were collected in just two to seven days, in contrast to the months-long field period for the RDD survey we commissioned. Yet the 6 other RDD studies were, on average, more accurate than the RDD survey we commissioned. Likewise, the 6 other probability sample Internet surveys were more accurate than the one we commissioned.

All this suggests that, if anything, the probability sample surveys we commissioned for our paper were slightly less accurate than the populations of surveys being done with those methods at the same time.
Were the opt-in panel samples you examined unrepresentative because the firm failed to balance their samples on basic demographic variables, especially race and education?

No. All of the opt-in Internet panel firms conducted stratified random sampling of their panels using gender and age. Some of the opt-in firms also used race, education, region, and income. The opt-in firm that used race and education in addition to gender, age, and income had the largest average error of the opt-in surveys we examined. And weighting to correct the imbalances in terms of basic demographics did not eliminate the more substantial error typical of the opt-in surveys.

All nine firms that provided data for our study were given identical instructions: to provide “1,000 completed surveys with a census-representative sample of American adults 18 years and older residing in the 50 United States.” Each firm chose the methods to be implemented to achieve this objective.

Were the two probability samples balanced on race, education, and other demographics, thus advantaging them unfairly?

No. The RDD telephone survey was conducted with randomly generated telephone numbers and imposed no quotas or any other “balancing.” No other survey in our study was significantly more accurate than that one, even in terms of race and education.

Was the probability sample Internet survey’s superior accuracy in your study an illusion, because the firm cheated by weighting the probabilities of selection to match Census benchmarks?

No. Even when weighting was done to the opt-in survey data to correct for their demographic errors, the probability sample Internet survey’s average error on variables not used for weighting or selection (3.4%) was significantly better than the opt-ins’ error (which ranged from 4.5% to 6.6%).

Was the weighting method that you used suboptimal, because it capped weights at 5?

No. The weighting method used was developed by a committee of illustrious statistical experts, chaired by Professor Douglas Rivers. That committee recommends capping weights at 5. When we reran the analyses not imposing any caps on the weights, the opt-in surveys we evaluated became less accurate in terms of variables not used in the weighting, so weight capping is not the source of the inaccuracy in opt-in survey data that we described.

Is your paper old news dressed up to look new? More specifically, have the data appeared previously in a paper entitled “Web Survey Methodologies: A Comparison of Survey Accuracy,” authored by Kroscnick and Rivers? Are the only truly new elements of the “new” paper some standard error calculations, some late arriving data, and a new set of weights?

No. In fact, no paper by the above title or any other title has ever been written and released by the team of scholars who designed and implemented this project in 2004/2005.

A presentation at AAPOR in 2005 reported an initial partial set of analyses that were described at the conference as “extremely preliminary.” Kroscnick and Rivers decided that they were insufficient to merit releasing in a paper.

Subsequent, far more detailed analysis yielded the conclusions presented in the new paper, which is the first formal write-up of these data. None of the numbers in that paper have been released in any paper before now, and none of the analysis done in 2004 and 2005 was used to generate this new paper.

In sum, we agree with Humphrey Taylor, chairman of the Harris Poll, when he said that “the trust we have in opinion polls and the different methods they use (whether in person, telephone, or online) should be based on empirical evidence of their track record.” The empirical track record we see from our work indicates continued superior accuracy from probability sampling, and considerably less accuracy of opt-in surveys.

As we said in our paper, we see tremendous value in opt-in survey data. One hundred years worth of terrifically useful social science theory testing has been accomplished using nonrepresentative samples (e.g., laboratory experiments done by psychologists and college undergraduate participants).

But there is no theoretical basis for claiming that an opt-in sample is representative of the general population. And our studies suggest that in practice, probability samples are still more accurate than opt-in samples.
Many industries offer products at various levels of quality, and survey data collection firms do as well.

Our research is intended to illuminate the quality of probability and opt-in samples, so purchasers and users of data can make informed choices between these methods.

Notes


2. A detailed memo describing the methodology of the new analyses reported here will be posted on Professor Krosnick’s webpage shortly.

3. The three firms whose data appear in the figure are the only ones of the nine from which 2009 data are available to us. We commissioned the 2009 probability and non-probability sample Internet surveys for studies we are currently conducting for other purposes. We selected the 2009 RDD survey randomly from among a set of surveys that the RDD firm did for clients other than us during 2009. Average errors shown in the figure are based on all common variables available; some of the 2004/05 variables used in our paper were not available in the 2009 datasets, and some of the available variables in the 2009 RDD study are different from those in the 2009 probability and opt-in Internet surveys. Thus, the 2004/05 RDD average is directly comparable to the 2009 RDD average, and the probability and opt-in Internet survey averages are directly comparable both to each other and across time.


Study Group Issues a Warning on Opt-in Online Surveys

Dissecting the "Bradley Effect"
Melinin Trump Can Be Better First Lady Than Michelle Obama - Here's Why (scolopolitics.com)

Guest Blog: Survey Accuracy - Revisiting the Benchmarks

Recommended by 

How This App Teaches You A Language In 3 Weeks (Babble)

These Photos Of Trump And Ivanka Will Make You Cringe (Wizened)

Best Xbox 360 Games For Kids (Parenting.com)

Illegally Smuggled NKorea Photos They Don't Want You To See (True Activist)

JOIN THE DISCUSSION

107 Comments

ABC News

Login

Sort by Oldest

Recommend Share

Elina ponting • a year ago

Great blog. Keep sharing.

Reply • Share

Simply deplorable, clayusmcre • a year ago

Cover 80 years of established history on the matter of not confirming a nominee in a president's final year. How do the democrats undo this? Oh yeah, simply pretend it didn't happen.

Reply • Share

wuffwuffwuff • Simply deplorable, clayusmcre • a year ago

"not confirming a nominee in a president's final year"

Actually, that's false. I know you're just parroting when you've heard, but it's