

# HOW TO TELL WHEN A SURVEY SAMPLE IS NOT VALID

RON SELLERS

GREY MATTER RESEARCH & CONSULTING

Originally published in *The NonProfit Times*, January 15, 2001

The headline in the local paper read *3Com Out to Poll World on Everything*. The story heralded plans for a huge “survey” of people worldwide, sponsored by 3Com. The project will allow participants to “see how their answers compare with those of people around the world,” according to the Associated Press.

The only problem with this groundbreaking study is that, as the article casually mentioned in the very last paragraph, “the poll will not be statistically valid.”

What this really means is that this study is actually representative of nothing and nobody. It’s a huge, glorified focus group, where we hear a lot of opinions with no way to quantify any of them in a way that represents the world’s population accurately. But when the “findings” from this “study” are released to the media, do you *really* think the emphasis is going to be on the fact that the data is not statistically valid?

Of course not. This little problem will be briefly explained in a footnote. But other than that, this study will most likely be treated as gospel truth. In short, this is a fabulous publicity idea. Unfortunately, for reasons too numerous to list here, it’s useless as serious research.

Lest you feel this is an attack solely on this upcoming “research,” stop and think for a moment how many similar “research” projects you see every day. Local newscasts ask people to call their 900-number and for only 75 cents they can vote in a telephone “poll.” A men’s magazine will survey its readers, and claim that the study is representative of all American men. Websites take “polls” all the time, and aside from an occasional footnote the resulting data is treated as if it’s a scientifically conducted research study. Companies and organizations use inadequate “research” methods to gather data all too frequently.

Unfortunately, many people treat surveys as if all you have to do is ask enough people a bunch of questions, and because you end up with a stack of data, somehow this means you have a study that reflects the public’s real opinions. Management leaders sometimes dismiss researchers’ concerns about “statistical significance” and “projectable data” as the cries of some anal-retentive academics who figure if they didn’t conduct the research themselves it must not be right.

Researchers face this problem frequently: “It doesn’t have to be scientific – it just has to be right!” (an actual quote from a research client). The problem is that the two are inseparable – either conduct it using statistically valid methods, or it won’t be right.

Clients want to cut corners, save time, and avoid all the technical jargon that researchers throw at them. Admittedly, researchers don’t always help themselves when they rely too much on that technical jargon, rather than trying to explain things so decision-makers realize what they’re giving up when they cut corners.

Above all, survey data must be *projectable* in order to be of any use at all. Simply put, this means that a survey is asking questions of a small portion that is supposed to represent a larger population. The portion who participate in the survey must accurately and fully represent the larger population in order for the research to be valid.

There are a number of different barriers to conducting a valid study:

***Ignoring response bias.*** Response bias is what happens when a certain type of person responds to a survey, while a different type does not. For instance, we know that women and older people are more likely to participate in a telephone survey than are men or younger people. Unless this factor is corrected, it will bias the survey results. Telephone surveys that attempt to get a random sample of the U.S. population but do not account for this factor often end up with 70% of the respondents being female. If women and men have different views on the survey topics, this means the data is being biased by this response rate difference.

Here’s a quick example. Let’s say 60% of men plan to vote for the Republican gubernatorial candidate in Delaware, while 60% of women plan to vote for the Democrat. If Delaware’s population is evenly split along gender lines, this means the actual vote will be split 50/50 between the two candidates. In a telephone survey that ends up 70% female and 30% male (as many surveys would without controlling this factor) the findings would show that 54% of all respondents support the Democrat, and only 46% support the Republican. This would provide inaccurate and misleading data.

This can be an even greater problem with e-mail and mail surveys. Response bias is more likely when the response rate (i.e. the proportion of people who complete the survey) is low. A mail survey that has a 5% response rate is probably useless, because the chances of response bias are very high when a response rate is that poor. The 5% who did respond are probably different from the 95% who did not respond.

***Using a non-representative sample.*** Generally, you don’t survey everyone in your target population, but you *sample* a small representation of that population. One of the keys to sampling is that the people you interview must be representative of the overall population in every way. If your donor base is 30% Spanish-speaking, you can’t say that a donor survey conducted only in English represents your donor base accurately.

People who attend your annual convention are probably not representative of all of your supporters. Donors who call in to complain about something are probably not representative of all donors. The readers of *Redbook* magazine are not representative of all American women. You cannot survey people in Portland and Pittsburgh and claim it's representative of all Americans.

A common mistake is trying to use a sample of convenience. You have website – why not just survey people who visit your website, and say they're representative of all of your donors? Because the truth is that they're probably *not* representative.

This continues to be one of the major obstacles to effective on-line research. People who use the internet still are not fully representative of the American population. Although this is gradually changing, today they still tend to be younger, better educated, higher income, and more likely to be male. If your donor base is heavily weighted towards older women, as many are, this will be a particular problem.

Viewing it internationally, the United States has a much higher proportion of the population on-line than do most other countries, which means a “worldwide” on-line survey will be dominated by responses from the U.S., unless the sample is very carefully controlled for this factor. Without taking these factors into account, an on-line survey will give you a sample of people who are not representative of your target.

***Using a self-selected sample.*** This is a major problem with call-in polls or surveys that are posted on a website – only those who really want to say something about the topic will participate. Do you really think someone who doesn't care about the issue of abortion will proactively take the time and spend the 75 cents to call in and vote pro-life or pro-choice in a 900-number telephone poll?

If a hotel gives you an 800-number to call to participate in a survey, are you more likely to do so if you had a decent experience staying at that hotel, or if you had a terrible experience and want to let someone know about it? Self-selected samples generally provide very polarized responses – people who love you or hate you will reply, but those who just don't care won't bother.

***Turning qualitative data into quantitative.*** This is a basic, unchangeable fact: you cannot pull quantitative, numeric data out of focus groups. Unfortunately, this is also one of the most common research mistakes. Focus groups are extremely useful for understanding how people perceive something. Focus groups are completely useless for finding out *how many* people feel a particular way. If nine out of ten people in a focus group really liked your new prospecting mailing, this does not mean that a majority of people you send it to will feel the same way. The sample size is too small and the sampling methods too non-representative for focus groups to provide projectable data, even if you're conducting a number of groups.

Questions are raised all the time about whether researchers can interview 400 people and say it represents a population of millions. The answer is *yes*, as long as it's done

correctly. It's relatively easy to ignore these bogus call-in polls or poorly conducted research studies when they're reported in the media. But if your own organization is making some of the same mistakes in its research efforts, that's not so easy to ignore – and far more dangerous.

*THE NONPROFIT TIMES*

“Facts do not cease to exist because they are ignored.”

ALDOUS HUXLEY



**GREY MATTER**  
RESEARCH & CONSULTING

3145 E. CHANDLER BLVD. SUITE 110 - 327 PHOENIX, AZ 85048

480.245.6483 • [WWW.GREYMATTERRESEARCH.COM](http://WWW.GREYMATTERRESEARCH.COM)