How are polls conducted?

by Frank Newport, Lydia Saad, David Moore
from Where America Stands, 1997 John Wiley & Sons, Inc.

Public opinion polls would have less value in a democracy if the public -- the very people whose views are represented by the polls -- didn't have confidence in the results. This confidence does not come easily. The process of polling is often mysterious, particularly to those who don't see how the views of 1,000 people can represent those of hundreds of millions. Many Americans contact The Gallup Organization each year

1. to ask how our results can differ so much from their own personal impressions of what people think,
2. to learn how we go about selecting people for inclusion in our polls, and
3. to find out why they have never been interviewed.

The public's questions indicate a healthy dose of skepticism about polling. Their questions, however, are usually accompanied by a strong and sincere desire to find out what's going on under Gallup's hood.

It turns out that the callers who reach Gallup's switchboard may be just the tip of the iceberg. Survey researchers have actually conducted public opinion polls to find out how much confidence Americans have in polls -- and have discovered an interesting problem. People generally believe the results of polls, but they do not believe in the scientific principles on which polls are based. In a recent Gallup "poll on polls," respondents said that polls generally do a good job of forecasting elections and are accurate when measuring public opinion on other issues. Yet when asked about the scientific sampling foundation on which all polls are based, Americans were skeptical. Most said that a survey of 1,500-2,000 respondents -- a larger than average sample size for national polls -- cannot represent the views of all Americans.

In addition to these questions about sampling validity, the public often asks questions about the questions themselves -- that is, who decides what questions to ask the public, and how those looking at poll results can be sure that the answers reflect the public's true opinion about the issues at hand.

The Sampling Issue

Probability sampling is the fundamental basis for all survey research. The basic principle: a randomly selected, small percent of a population of people can represent the attitudes, opinions, or projected behavior of all of the people, if the sample is selected correctly.

The fundamental goal of a survey is to come up with the same results that would have been obtained had every single member of a population been interviewed. For national Gallup polls, in other words, the objective is to present the opinions of a sample of people that are exactly the same opinions that would have been obtained had it been possible to interview all adult Americans in the country.

The key to reaching this goal is a fundamental principle called equal probability of selection, which states that if every member of a population has an equal probability of being selected in a sample, then that sample will be representative of the population. It's that straightforward.

Thus, it is Gallup's goal in selecting samples to allow every adult American an equal chance of falling into the sample. How that is done, of course, is the key to the success or failure of the process.

Selecting a Random Sample

The first one thousand people streaming out of a Yankees game in the Bronx clearly aren't representative of all Americans. Now consider a group compiled by selecting 1,000 people coming out of a Major League Baseball game in every state in the continental United States -- 48,000 people! We now have a much larger group -- but we are still no closer to representing the views of all Americans than we were in the Bronx. We have a lot of baseball fans, but, depending on the circumstances, these 48,000 people may not even be a good representative sample of all baseball fans in the country -- much less all Americans, baseball fans or not.

When setting out to conduct a national opinion poll, the first thing Gallup does is select a place where all or most Americans are equally likely to be found. That wouldn't be a shopping mall, or a grocery store, an office building, a hotel, or a baseball game. The place nearly all adult Americans are most likely to be found is in their home. So, reaching people at home is the starting place for almost all national surveys.

By necessity, the earliest polls were conducted in-person, with Gallup interviewers fanning out across the country, knocking on Americans' doors. This was the standard method of interviewing for nearly fifty years, from about 1935 to the mid-1980s, and it was a demonstrably reliable method. Gallup polls across the twelve presidential elections held between 1936 and 1984 were highly accurate, with the average error in Gallup's final estimate of the election being less than 3 percentage points.

By 1986, a sufficient proportion of American households had at least one telephone to make telephone interviewing a viable and substantially less expensive alternative to the in-person method. And by the end of the 1980s, the vast majority of Gallup's national surveys were being conducted by telephone. Today, approximately 95% of all households have a telephone and every survey reported in this book is based on interviews conducted by telephone.

Gallup proceeds with several steps in putting together its poll with the objective of letting every American household, and every American adult have an equal chance of falling into the sample.

* First, we clearly identify and describe the population that a given poll is attempting to represent
. If we were doing a poll about baseball fans on behalf of the sports page of a major newspaper, the target population might simply be all Americans aged 18 and older who say they are fans of the sport of baseball. If the poll were being conducted on behalf of Major League Baseball, however, the target audience required by the client might more specific, such as people aged twelve and older who watch at least five hours worth of Major League Baseball games on television, or in-person, each week.

In the case of Gallup polls that track the election and the major political, social and economic questions of the day, the target audience is generally referred to as "national adults." Strictly speaking the target audience is all adults, aged 18 and over, living in telephone households within the continental United States. In effect, it is the civilian, non-institutionalized population. College students living on campus, armed forces personnel living on military bases, prisoners, hospital patients and others living in group institutions are not represented in Gallup's "sampling frame." Clearly these exclusions represent some diminishment in the coverage of the population, but because of the practical difficulties involved in attempting to reach the institutionalized population, it is a compromise Gallup usually needs to make.

* Next, we choose or design a method that will enable us to sample our target population randomly
. In the case of The Gallup Poll, we start with a list of all household telephone numbers in the continental United States. This complicated process really starts with a computerized list of all telephone exchanges in America, along with estimates of the number of residential households those exchanges have attached to them. The computer, using a procedure called random digit dialing (RDD), actually creates phone numbers from those exchanges, then generates telephone samples from those. In essence, this procedure creates a list of all possible household phone numbers in America and then selects a subset of numbers from that list for Gallup to call.

It's important to go through this complicated procedure because estimates are that about 30% of American residential phones are unlisted. Although it would be a lot simpler if we used phone books to obtain all listed phone numbers in America and sampled from them (much as you would if you simply took every 38th number from your local phone book), we would miss out on unlisted phone numbers, and introduce a possible bias into the sample.

The Number of Interviews or Sample Size Required

One key question faced by Gallup statisticians: how many interviews does it take to provide an adequate cross-section of Americans? The answer is, not many -- that is, if the respondents to be interviewed are selected entirely at random, giving every adult American an equal probability of falling into the sample. The current U.S. adult population in the continental United States is 187 million. The typical sample size for a Gallup poll, which is designed to represent this general population, is 1,000 national adults.

The actual number of people that need to be interviewed for a given sample is to some degree less important than the soundness of the fundamental equal probability of selection principle. In other words -- although this is something many people find hard to believe -- if respondents are not selected randomly, we could have a poll with a million people and still be significantly less likely to represent the views of all Americans than a much smaller sample of just 1,000 people -- if that sample is selected randomly.

To be sure, there is some gain in sampling accuracy that comes from increasing sample sizes. Common sense -- and sampling theory -- tell us that a sample of 1,000 people probably is going to be more accurate than a sample of 20. Surprisingly, however, once the survey sample gets to a size of 500, 600, 700 or more, there are fewer and fewer accuracy gains that come from increasing the sample size. Gallup and other major organizations use sample sizes of between 1,000 and 1,500 because they provide a solid balance of accuracy against the increased economic cost of larger and larger samples. If Gallup were to -- quite expensively -- use a sample of 4,000 randomly selected adults each time it did its poll, the increase in accuracy over and beyond a well-done sample of 1,000 would be minimal, and generally speaking, would not justify the increase in cost.

Statisticians over the years have developed quite specific ways of measuring the accuracy of samples -- so long as the fundamental principle of equal probability of selection is adhered to when the sample is drawn.

For example, with a sample size of 1,000 national adults, (derived using careful random selection procedures), the results are highly likely to be accurate within a margin of error of plus or minus three percentage points. Thus, if we find in a given poll that President Clinton's approval rating is 50%, the margin of error indicates that the true rating is very likely to be between 53% and 47%. It is very unlikely to be higher or lower than that.

To be more specific, the laws of probability say that if we were to conduct the same survey 100 times, asking people in each survey to rate the job Bill Clinton is doing as president, in 95 out of those 100 polls, we would find his rating to be between 47% and 53%. In only five of those surveys would we expect his rating to be higher or lower than that due to chance error.

As discussed above, if we increase the sample size to 2,000 rather than 1,000 for a Gallup poll, we would find that the results would be accurate within plus or minus 2% of the underlying population value, a gain of 1% in terms of accuracy, but with a 100% increase in the cost of conducting the survey. These are the cost value decisions that Gallup and other survey organizations make when they decide on sample sizes for their surveys.

The Interview Itself

Once the computer has selected a phone number for inclusion into a sample, Gallup goes to extensive lengths to try to make contact with an adult American living in that household. In many instances, there is no answer or the number is busy on the first call. Instead of forgetting that number and going on to the next, Gallup typically stores the number in the computer where it comes back up to be recalled a few hours later, and then to be recalled again on subsequent nights of the survey period. This procedure corrects for a possible bias that could occur in if we included interviews only with people who answered the phone the first time we called their number. For example, people who are less likely to be at home, such as young single adults, or people who spend a lot of time on the phone, would have a lower probability of falling into the sample than an adult American who was always at home and rarely talked on his or her phone. The call-back procedure corrects for this possible bias.

Once the household has been reached, Gallup attempts to assure that an individual within that household is selected randomly -- for those households that include more than one adult. There are several different procedures that Gallup has used through the years for this within household selection process. Gallup sometimes uses a shorthand method of asking for the adult with the latest birthday. In other surveys, Gallup asks the individual who answers the phone to list all adults in the home based on their age and gender, and Gallup selects randomly one of those adults to be interviewed. If the randomly selected adult is not home, Gallup would tell the person on the phone that they would need to call back and try to reach that individual at a later point in time.

These procedures, while expensive and while not always possible in polls that are conducted in very short time periods, help to ensure that every adult American has an equal probability of falling into the sample. With those assurances in mind, the outside observer or researcher should dive into poll data with a critical mind. Interpretation of survey research results is most importantly dependent on context. What the American public may say about an issue is most valuable when it can be compared to other current questions or to questions asked across time. Where trend data exist, one should also look at changes over time and determine whether these changes are significant and important. Let's say, for example, that Bill Clinton has a job approval rating of 48%. Is this a good rating or a poor rating? The best way to tell is to look at history for context: compare it to Clinton's ratings throughout the rest of his presidency, then compare it to approval ratings for previous presidents. Did previous presidents with this rating at the equivalent point in time tend to get re-elected or not? Then it can be compared to approval ratings of Congress, of the Republican and Democratic congressional leaders.