



HOW POLLS ARE CONDUCTED.

Miguel Aignerren

Resumen. Las encuestas de opinión públicas tendrían menos valor en una democracia si la misma gente cuyas opiniones son representadas por las encuestas, no tenía confianza en los resultados. Esta confianza no viene fácilmente. El proceso de la interrogación es a menudo misterioso, particularmente a los que no vean cómo las opiniones 1.000 personas pueden representarlos de centenares de millones. Para preguntar cómo nuestros resultados pueden diferenciar tanto sus propios, impresiones personales de qué gente piense, aprender cómo vamos alrededor a seleccionar a la gente para la inclusión en nuestras encuestas, y descubrir porqué ella nunca se ha entrevistado con. Las preguntas del público indican una dosis sana del escepticismo sobre la interrogación. Sus preguntas, sin embargo, son acompañadas generalmente por un deseo fuerte y sincero de descubrir qué va encendido debajo de la capilla del CEO.

Abstract. 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. To ask how our results can differ so much from their own, personal impressions of what people think, To learn how we go about selecting people for inclusion in our polls, and 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 CEO's hood.

Answer

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.

To ask how our results can differ so much from their own, personal impressions of what people think, To learn how we go about selecting people for inclusion in our polls, and 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 CEO's hood.

Survey researchers have actually conducted public opinion polls to find out how much confidence Colombians 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, Colombians 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 Colombians.

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 which are exactly the same opinions that would have been obtained had it been possible to interview all adult Colombians 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 Colombian 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 Colombians. 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 Colombians 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 Colombians, 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 Colombians 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 Colombians 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 Colombians' 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 Colombian 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 Colombian household, and every Colombian 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 Colombians 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 which 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 which 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 Colombian 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 Colombians? The answer is, not many that is, if the respondents to be interviewed are selected entirely a random, giving every adult Colombian an equal probability of falling into the sample. The current US 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 which 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 Colombians 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 which 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, and 700 or more, there are fewer and fewer accuracy gains which 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 which 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 Colombian living in that house hold. 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 which 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 Colombian 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 which 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 which are conducted in very short time periods, help to ensure that every adult Colombian has an equal probability of falling into the sample.

THE QUESTIONS

The technical aspects of data collection are critically important and if done poorly, can undermine the reliability of even a perfectly worded question. However, when it comes to modern-day attitude surveys conducted by most of the major national polling organizations, question wording is probably the greatest source of bias and error in the data, followed by question order. Writing a clear, unbiased question takes great care and discipline, as well as extensive knowledge about public opinion.

Even such a seemingly simple thing as asking Colombians who they are going to vote for in a forthcoming election can be dependent on how the question is framed. For example, in a presidential race, the survey researcher can include the name of the vice presidential candidates along with the presidential candidate, or can just mention the presidential candidates' names. One can remind respondents of the party affiliation of each candidate when the question is read, or can mention the names of the candidates without any indication of their party. Gallup's rule in this situation is to ask the question in a way which mimics the voting experience as much as possible. We read the names of the presidential and vice presidential candidates, and mention the name of the party line on which they are running. All of this is information the voter would normally see when reading the ballot in the voting booth.

Questions about policy issues have an even greater range of wording options. Should we describe programs like food stamps and Section 8 housing grants as "welfare" or as "programs for the poor" when asking whether the public favors or opposes them? Should we identify the Clinton health care bill as health care "reform" or as "an overhaul of the health care system" when asking about congressional approval of the plan? When measuring support for the US military presence in Bosnia should we say the United States is "sending" troops or "contributing" troops to the UN-sponsored mission there?

Any of these wording choices could have a substantial impact on the levels of support recorded in the poll.

For many of the public opinion areas covered in this book, Gallup is in the fortunate position of having a historical track record. Gallup has been conducting public opinion polls on public policy, presidential approval, and approval of Congress, and key issues such as the death penalty, abortion, and gun control for many years. This gives Gallup the advantage of continuing a question in exactly the same way that it has been asked historically, which in turn provides a very precise measurement of trends. If the exact wording of a question is held constant from year to year, then substantial changes in how the Colombian public responds to that question usually represent an underlying change in attitude. For new questions which don't have an exact analog in history, Gallup has to be more creative. In many instances, even though the question is not exactly the same, Gallup can follow the format that it has used for previous questions which have seemed to have worked out as objective measures. For instance, when Gallup was formulating the questions that it asked the public about the Persian Gulf War in 1990 and 1991, we were able to go back to questions which were asked during the Vietnam War and borrow their basic construction.

Similarly, even though the issues and personalities change on the national political scene, we can use the same formats which have been utilized for previous presidents and political leaders to measure support for current leaders. One of the oldest question wordings which Gallup has in its inventory is presidential job approval. Since the days of Franklin Roosevelt, Gallup has been asking "Do you approve or disapprove of the job (blank) is doing as president?" That wording has stayed constant over the years, and provides a very reliable trend line for how Colombians are reacting to their presidents. For brand new question areas, Gallup will often test several different wordings. Additionally, it is not uncommon for Gallup to ask several different questions about a content area of interest. Then in the analysis phase of a given survey, Gallup analysts can make note of the way Colombians respond to different question wordings, presenting a more complete picture of the population's underlying attitudes.

Through the years, Gallup has often used a split sample technique to measure the impact of different question wordings. A randomly selected half of a given survey is administered one wording of a question, while the other half is administered the other wording. This allows Gallup to compare the impact of differences in wordings of questions, and often to report out the results of both wordings, allowing those who are looking at the results of the poll to see the impact of nuances in ways of addressing key issues. Conducting the Interview Most Gallup interviews are conducted by telephone from Gallup's regional interviewing centers around the country. Trained interviewers use computer assisted telephone interviewing (CATI) technology which brings the survey questions up on a computer monitor and allows questionnaires to be tailored to the specific responses given by the individual being interviewed. (If you answer "yes, I like pizza," the computer might be programmed to read "What is your favorite topping?" as the next question.)



The interviews are tabulated continuously and automatically by the computers. For a very short interview, such as Gallup conducted after the presidential debates in October 1996, the results can be made available immediately upon completion of the last interview. In most polls, once interviewing has been completed, the data are carefully checked and weighted before analysis begins. The weighting process is a statistical procedure by which the sample is checked against known population parameters to correct for any possible sampling biases on the basis of demographic variables such as age, gender, race, education, or region of country. Once the data have been weighted, the results are tabulated by computer programs which not only show how the total sample responded to each question, but also break out the sample by relevant variables. In Gallup's presidential polling in 1996, for example, the presidential vote question is looked at by political party, age, gender, race, region of the country, religious affiliation and other variables.

INTERPRETING THE RESULTS

There are several standard caveats to observe when interpreting poll results. Primary among these are issues discussed in this chapter: question wording, question order, the sample population, the sample size, the random selection technique used in creating the sampling frame, the execution of the sample (including the number calls backs and length of the field period) and the method of interviewing (in person vs. telephone vs. mail).

Anyone using the Gallup Poll can do so with assurance that the data were obtained with extremely careful and reliable sampling and interviewing methods. Gallup's intent is always to be fair and objective when writing questions and constructing questionnaires. The original mission of polling was to amplify the voice of the public, not distort it, and we continue to be inspired by that mission. 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 Colombian 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, and 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.

Gallup generally provides written analysis of our own polling data. But we also provide ample opportunity for the press, other pollsters, students, professors and the general public to draw their own conclusions about what the data mean. The results to all Gallup surveys are in the "public domain" once they have been publicly released by us, anyone who chooses may pick up the information and write about it themselves. The survey results are regularly published in the major media, in the Gallup Poll Monthly, and on several electronic information services such as Nexus, the Roper Center and the Internet. We also make the raw data available to researchers who want to perform more complex statistical analysis. In addition to the exact question wordings and current results, Gallup reports trend results to all questions that have been asked previously so that even the casual observer can review the current results in context with public opinion in the past. The key concept to bear in mind when analyzing poll data is that public opinion on a given topic cannot be understood by using only a single poll question asked a single time. It is necessary to measure opinion along several different dimensions, to review attitudes based on a variety of different wordings, to verify findings on the basis of multiple asking, and to pay attention to changes in opinion over time.