



From SIAM News, Volume 37, Number 9, November 2004

The Art of Election Polling

A glimpse of the widely different methods used by pollsters helps make sense of puzzlingly divergent poll results.

Sara Robinson

In September, following the Republican convention, a series of U.S. presidential election polls seemed to defy logic. Some, such as those by Gallup and Newsweek, favored President George W. Bush over Senator John Kerry by huge margins, while others, including those by Zogby International and Pew Research, showed the race as essentially a dead heat.

A few weeks later, after a strong performance by Kerry in the first debate, the huge leads for Bush reflected in some polls evaporated, while the results from other polls remained constant. More than a dozen polls conducted in early October showed a difference between the candidates that fell within the margin of error. Why are there so many discrepancies between polls, and why do some polls fluctuate more than others? Are some polling methods just wrong?

Many of the differences can be traced to the varying methods of the polling organizations. Because of time constraints and the practical difficulties in sampling from large populations, pollsters resort to tricks that can introduce dozens of potential biases. There are ways to adjust for some of the biases, but each organization has its own philosophy of how best to do it, making it difficult to compare results across different polls.

Complicating things further, election polling is more than just statistical sampling. Pollsters are trying to predict an outcome based on data collected weeks or months in advance. Part of the art of election polling, then, is guessing which of the people interviewed will actually go to the polls; each organization has its way of doing this, too.

None of these methodological differences is reflected in the official "margin-of-error" for a poll. That refers only to the statistical sampling error, the error that comes from determining, with 95% confidence, the opinions of a large population from those of a much smaller random sample.

Varying Techniques

Most election polls are conducted by news organizations, by independent research organizations, often in conjunction with news organizations, and by partisan groups (whose polling methods are not discussed here). Because opinion polling is expensive, organizations often team up to get the data, although they sometimes analyze it separately. *The Washington Post* and ABC have a polling relationship, for instance, as do *The New York Times* and CBS. The independent polling company Gallup does a few of its polls alone, but usually works in conjunction with its media partners, CNN and *USA Today*, says David Moore, Gallup's senior polling editor.

The Los Angeles Times is unusual in doing most of its polling alone, not because it doesn't want partners, says polling director Susan Pinkus, but because the major television networks are already in partnerships with competing newspapers.

Polling results are greatly influenced by the phrasing and ordering of the questions. For this reason, careful pollsters pretest their questions and order the possible responses

randomly with each interview. But because so many polls are done by media organizations operating under tight time constraints, questions are often tested less thoroughly than they would be in academic settings, says Michael Dimock, director of research at the Pew Research Center for the People and the Press, an independent polling organization funded by a charitable trust.

Dimock says that his organization pretests newly devised questions. The same is true of the *LA Times*, according to Pinkus.

Gallup rarely pretests its questions, Moore explains, in part because it reuses many questions from previous polls. "The issues are ones with which we have a great deal of experience," he says. "Designing a balanced question and one that is unbiased is something we take very seriously."

Polls also vary in how much they pressure undecided people to lean one way or another. Seven or eight percent of the respondents in recent Pew polls were labeled "undecided," as compared with only 3% of Gallup respondents, suggesting that Gallup puts more pressure on undecided voters to pick a candidate.

Once pollsters settle on a list of questions, they obtain a sample of residential phone numbers in the area to be polled, typically from an independent company that specializes in this task. Often, an organization contracts with yet another independent company to do the calls. "Polling is not a minimum-wage job, and you need to get highly trained people to do it well," says Dimock, of Pew Research, which works with two such companies.

The *LA Times* staffs its own 90-station phone bank; some of the poll workers have been there for 20 years or more, Pinkus says. While the newspaper does buy phone number samples for most of its state polls, it generates the phone numbers for its national polls itself. Gallup obtains its phone number samples from an outside firm, but, like the *LA Times*, uses an in-house staff to make the calls.

A polling organization will make a specified number of attempts to call a household for a poll before giving up. Gallup calls each number up to six times over three days; Pew calls up to 10 times over five days. In the case of refusals to cooperate, Pew Research has its callers categorize the firmness of the refusal before deciding whether to call again. "If someone says, 'Don't ever call me again, I hate pollsters,' we don't call them again," Dimock says.

For election polls, organizations usually report responses only for people who say they are registered voters. From this group, most organizations attempt to single out those who are likely to vote in the upcoming election; the responses of this "likely voter" group are reported separately.

Polling groups have to start with far more phone numbers than their target response number: Many of the phone numbers will be non-working, calls will go unanswered, and many people will refuse to participate. To get 1500 responses (and even fewer registered voters and likely voters), Pew Research starts with a sample of 5000 numbers.

The people who respond to a poll may not have representative opinions. To address this and other biases, polling organizations weight their responses, using census information. Gallup and Pew Research both weight their responses by age, gender, education level, geographic region, and race. Pollsters do not weight for urban versus rural areas, one of the big "red state-blue state" divisions, or (in most cases) for ideology.

Gallup, along with most other polling organizations, also asks for the number of working phone lines in each household it reaches, excluding those used for data only. Because multiple phone lines make a household more likely to be selected, the responses from multiline households are given less weight. Some organizations also weight for the number of people of voting age who share the phone line. Pew Research does not weight its results according to either factor, Dimock says: Historically, it hasn't done so, and it

wants to keep a consistent methodology.

The Tricky Task of Sampling Phone Numbers

The most obvious method for sampling phone numbers---generating random 10-digit phone numbers---is impractical because only a tiny fraction (about 2%) of the six billion allowable phone numbers will correspond to working lines. Many of those lines will correspond to businesses or cell phones, which are illegal to dial automatically. In any case, employing people to try all those numbers would be prohibitively expensive.

Working from lists of registered voters isn't feasible either, because the rolls are often maintained only at the county or even the town level. Some voter lists, moreover, are infrequently purged of deceased voters, convicted felons, and people who have left the area.

Because of these hurdles, a specialized industry has sprung up to provide pollsters with phone number samples containing many working residential lines. Linda B. Piekarski, a vice president at Survey Sampling International, which provides samples to Gallup and Pew Research, sent a 50-page paper describing the company's immensely complicated procedure.

SSI's techniques produce phone samples consisting of at least 40% working residential numbers, a huge increase in efficiency. The downside is the introduction of myriad potential biases in the samples, although studies suggest that most of them are small. A larger bias comes from the exclusion of households without landlines, currently about 5% of all households. People who rely on cell phones tend to be young and urban, and those with no phones at all tend to be poor. Demographic weighting may partially offset this bias, and SSI is currently exploring ways to include cell phone numbers in its samples, in case the law is changed.

To pare down the sample set, SSI first restricts its attention to working "exchanges," the three digits following the area code (e.g., 382 is the exchange for SIAM: (215) 382-9800). SSI regularly obtains lists of active area codes and exchanges from Telcordia Technologies, which maintains the North American Numbering Plan for phone numbers.

Restricting the sample set to exchanges in Telcordia's file increases the fraction of working lines from 2% to 8%, still not high enough to be practical. To increase the fraction further, SSI also relies on the fact that, historically, phone numbers have been issued in banks of 100 contiguous numbers that agree on all but the last two digits. This is still true in rural areas that use old switching technology, although in most places, working numbers are now scattered more sparsely across many banks. Still, businesses, government agencies, and cell phone companies are often issued numbers in banks, and many banks are unused.

SSI samples only from banks with n or more directory-listed numbers, which cuts out about half the banks in each working exchange when $n = 1$, bringing the fraction of working residential lines up to 40%. To further increase efficiency, some polling organizations choose a higher value of n . (Pew Research uses $n = 3$.) Of course, choosing only those banks means that residential numbers in recently introduced banks and banks with few listed numbers are never selected, perhaps biasing samples in favor of rural areas.

Nationwide, approximately 30% of residential phone numbers are unlisted. Those numbers are not uniformly distributed throughout the country; in urban Chicago, for instance, 46% of all households are unlisted, as compared with 13% of households in rural West Virginia. Here, SSI cites two studies from the early 1990s suggesting that the associated bias may be slight.

Even maintaining an accurate database of listed numbers is more complex than it seems. Listings are updated at different times, depending on the region, and the recent proliferation of area codes means that two people on the same street could have the same exchange and number, but in different area codes. Some new area codes are assigned as

overlays, with both the old and new numbers valid at first. SSI keeps track of places where new area codes will be introduced and updates its data for listed blocks from a master directory nine times a year. SSI's database lists telephone exchanges by county, assigning the 36% of exchanges that cut across county (or even state) lines to the county to which the majority of the listed numbers belong.

To get a phone sample for an election poll, SSI typically begins by calculating the total number of phone numbers in banks with one or more listed numbers in each county within the area to be polled; it determines a sampling interval by dividing the total number of possible phone numbers by the desired sample size. Choosing a random starting point in the first interval, SSI then takes each number that is a set number of intervals away from that point. This process will select proportionally more numbers in areas in which phone banks are sparsely utilized, but a smaller percentage of those numbers will correspond to working lines.

The Telemarketer Effect

A little understood source of sampling bias in polls comes from non-responders, people who never answer the phone or those who listen just long enough to hear the word "survey" before hanging up. Low response rates could introduce a bias if the opinions of non-responders tended to differ from those of easy-to-reach people.

With the proliferation of telemarketers, the percentage of non-responders has increased markedly over the last few years, although pollsters are hopeful that the new national do-not-call list will turn that trend around. For an August survey that resulted in 1512 interviews, Pew Research started with 5000 phone numbers. Of the people who were contacted, only 41% agreed to be interviewed. In a recent *LA Times* survey, 50% of those contacted declined to participate.

Getting data from non-responders isn't easy, but Pew Research has made valiant efforts. In two studies, in 1997 and 2003, Pew first surveyed just the people who responded over its typical five-day time frame, and then extended the poll for several weeks, making extraordinary efforts to elicit more responses. The organization called people repeatedly, pressuring them to respond to the poll, and sent letters, some including small financial incentives, explaining the study and the importance of responding. Through these extra efforts, Pew was able to boost its cooperation rate from 38% to 59%.

The study then compared the responses of the initial sample with those of the larger "rigorous" sample, as well as with those of the "hardest-to-reach" people (people who agreed to be interviewed only after declining at least twice, or were reached only after more than 20 attempts).

Respondents answered about 90 questions covering their demographic information, social attitudes, financial situation, and voting practices. For the most part, there was surprisingly little difference between the three groups. The numbers were similar for age distribution, education level, and financial status, and even for attitudes toward social issues. Still, a few things stand out.

About 82% of the respondents for the standard five-day sample were white, which is consistent with U.S. Census data, but the hardest-to-reach group was 74% white. More striking, the hardest-to-reach sample had fewer Republicans (32% versus 23%) and more independents (11% versus 19%). These anomalies could just be statistical fluctuations (several large discrepancies would be expected with 90 questions), but they are intriguing.

The Elusive Nature of Party Affiliation

To help counteract sources of sampling bias, virtually all pollsters weight their results by demographics. A few, like Zogby International and some partisan pollsters, also weight the results by party, trying to achieve an appropriate balance of Democrats and Republicans. Because of the elusive nature of party affiliation, however, this type of weight is merely a prediction of how people might view themselves on Election Day.

For this reason, most polling organizations do not weight by party. "There is virtual unanimity among academic and media pollsters that you will misrepresent your poll numbers if you try to weight by party," says Dimock of Pew Research.

The National Council on Public Polls, an association of polling organizations, cites the weighting of likely voters by party as the most common bad weighting practice in political polls. What is missing, the organization explains on its Web site, is good data on the actual fractions of Democrats and Republicans in the population, forcing pollsters to resort to using data from exit polls in previous elections. The actual numbers change constantly and are updated inconsistently, often only at the local level. Voters in some states do not register by party, or even at all.

In general, Dimock says, people's stated party affiliation is an indication of their state of mind at the moment. By way of example, he points out that following the well-received Republican convention, it is likely that more Americans thought of themselves as Republicans; attempts to give less weight to Republican respondents would thus distort results.

John Zogby, Zogby's director, was not available for an interview, but in an article on the organization's Web site, he explained his reasoning. People's party affiliation "is no small consideration," Zogby wrote. "Given the fact that each candidate receives anywhere between eight in ten and nine in ten support from voters in his own party, any change in party affiliation [for the respondents to a poll] trades point for point in the candidate's support."

Indeed, party affiliation alone seems to account for most of the discrepancies in the September polls. A Gallup poll showing Bush well ahead used a sample consisting of 38% Republicans and 31% Democrats. Zogby weights its samples according to the proportions that turned out for the 2000 presidential election: 39% Democrats, 35% Republicans, and 26% independents. When the Gallup sample is weighted according to Zogby's proportions, the two organizations' poll numbers look similar.

The sample for an anomalous *LA Times* poll in June showing Kerry well ahead consisted of 38% Democrats and 25% Republicans. At the time, some pollsters suggested that the poll results were wrong because the sample contained too many Democrats. Pinkus, of the *LA Times*, says that she does not weight national polls by party, although she will weight state polls for states that, like California, require people to register by party and keep central records.

"Likely Voters" and Other Constructs

Whereas only some pollsters try to predict Democratic and Republican turnouts, just about all pollsters will try to predict which of their respondents will vote. This, too, is an art: People often aren't honest about their intentions, and the intention to vote doesn't always translate into a vote. When registered voters are asked whether they intend to vote, "something like 98% will say 'yes,'" Dimock says, and after the election about 78% will say that they voted. In fact, typical voter turnouts are under 60%.

"Likely voter" models involve a lot of guesswork. Gallup's system is to ask a series of seven questions gauging past voting practices, and interest in and knowledge about the election, assigning respondents a score from zero to seven. Gallup then takes the top 55% of scorers, the percentage of eligible voters who voted in 2000.

Pew uses essentially the Gallup system, but other organizations have different methods, with some, according to Dimock, using only one question. The *LA Times* asks six or seven questions and then weights the responses according to a model developed for that election, Pinkus says.

2004: An Especially Challenging Election Year

Because each polling organization uses its own methods, comparing results across polls

from different organizations is like comparing apples and oranges. More useful is to look at trends, the results over time for a single polling organization.

Polls become better predictors as an election approaches. Many polls conducted immediately before the last presidential election were reasonably accurate, correctly forecasting the result---a statistical tie---within their margins of error. Still, almost all the polls gave Bush an edge (the CBS-*New York Times* poll even predicted a five-point difference), although, ultimately, Al Gore won the popular vote by a tiny margin.

This year's election may be even harder to predict. There are now more cell phone-only households and a larger split in opinions between rural and urban voters, factors that could increase the effects of sampling biases. Furthermore, with so much at stake, far more people are registering and planning to vote, which could invalidate current "likely voter" models. Only November will tell.

Sara Robinson is a freelance writer based in Los Angeles, California.

Society for Industrial and Applied Mathematics
3600 University City Science Center
Philadelphia, PA 19104 USA
+1-215-382-9800

webmaster@siam.org

© Society for Industrial and Applied Mathematics



Updated: DJC 11/23/04