Polls Gone Wild

(From *Probabilities: The Little Numbers that Rule Our Lives*, Chapter 8.)

Before the presidential election in 1936, the magazine *Literary Digest* published an opinion poll that predicted an easy win for Republican candidate Alf Landon over President Franklin D. Roosevelt: 57% for Landon and 43% for Roosevelt. The *Digest* had a solid reputation after having correctly predicted the winner in each presidential election since 1916, and the poll was based on responses from 2.3 million people. Yes, you read correctly, two million three hundred thousand people! The election result? Roosevelt got 62% and Landon 38% (of those who voted for either of the two), which is one of the largest margins of victory in any presidential election. The *Digest* poll has gone down in history as the worst opinion poll, ever and the magazine went out of business shortly thereafter.

How on earth could this happen? With 2.3 million people, the rule of thumb from above gives a margin of error of only 0.07%, so the predicted numbers ought to be almost certain. Did something happen that made people suddenly change their minds? No, the error is in the methods of the *Digest*. The reliability of the estimate as measured by the margin of error is only valid if we have a *random sample*, meaning that everybody is equally likely to be selected. In theory, if there had been a list of all voters and 2.3 million people had been chosen from this list, the prediction would have been very accurate. But this is not how it was done. The *Digest* made two errors that resulted in heavily biased results.

Their first error was selection bias. When they selected the people to be included in the poll, they used various available address lists such as their own subscription lists, telephone directories, automobile registration lists, and club membership lists. Now, this was during the Great Depression and you would not show up on any such list unless you had disposable income. A young man who had just enrolled in the Civilian Conservation Corps would most likely not spend his daily dollar on a subscription to the *Literary Digest*, nor would a recently laidoff steel worker decide to join the local country club. Cars and telephones were also far less common than today; for example, only 25% of households had a telephone. The selection of individuals favored the rich, and in 1936, they were less likely to support Roosevelt's New Deal than Landon's more restrictive financial policies. This might have been the first election year when there was this kind of divide in the electorate that

mattered to the *Digest*'s polls. After all, they had managed to get it right before.

As if the selection bias was not bad enough, the second error was *nonresponse bias*. The *Digest* mailed postcards to 10 million people and based their poll on the 2.3 million cards that were returned. One can imagine that even if the recently laidoff steel worker was to receive a postcard from the *Digest*, he would be far more concerned with feeding his family than filling out the card and mailing it back to the magazine. The bias that was introduced by selection was further enforced by nonresponse, and whereas 2.3 million may seem an impressive number, a response rate of 23% is not. It may be a bit more speculative that nonresponse bias would favor Landon, but a special poll done in Chicago showed that over half of those who responded favored Landon but Roosevelt still got two thirds of the vote in the city. The Chicago poll had a 20% response rate and did not suffer from selection bias because the individuals were chosen from lists of registered voters.

In the 1936 election, the *Digest* faced competition from some new kids on the block. Archibald Crossley, George Gallup, and Elmo Roper were three bright young fellows who had realized that samples must be random in order for results to be reliable. Each of them predicted a win for Roosevelt, and Gallup also managed to correctly predict the *Digest*'s erroneous numbers, a feat that established George Gallup more than anybody else as Mr. Opinion Poll. He also gained this reputation in Europe where he later correctly managed to predict Winston Churchill's defeat in the U.K. elections in 1945 when almost everybody else predicted a Churchill victory.

In the 1936 U.S. presidential election, Gallup predicted 56% for Roosevelt (still a bit off the actual 62%) based on a random sample of 50,000 people. Moreover, based on a random sample of 3,000 people from the lists used by the *Digest*, Gallup predicted that the magazine would predict 44% for Roosevelt. Gallup realized that a sample of 3,000 would give a good idea of how the 10 million on those lists would vote and he realized that the huge sample size of 2.3 million would do no good because the selection procedure was skewed from the beginning. See Table 1 for predicted and actual numbers in the 1936 election.

The unemployed editors of the *Digest* would however get a small revenge on the pesky newcomers 12 years later. The 1948 American presidential election is the time of the second famous erroneous poll, and this time it was Gallup & Co. who got it wrong. You have probably seen the famous photo of Harry Truman holding a copy of the *Chicago Daily Tribune* with the headline

Table 1: Polls and election numbers for Roosevelt in the 1936 election

Source	Roosevelt's %
Election result	62
Gallup	56
Literary Digest	43
Gallup's prediction of *Digest*	44

"Dewey Defeats Truman" just after Truman had won the election. Crossley, Gallup, and Roper had predicted a victory for Republican candidate Thomas Dewey by a handsome 5–7% margin, and in reality, the results turned out the opposite. What went wrong this time?

The reason for the failure of Gallup & Co. was that the three pollsters had still not managed to get rid of all bias in their sampling. They realized correctly that their polls would be more accurate if they made sure that their samples reflected the population composition. Thus, they would try to get half men and half women and various other population traits such as race, age, and income in their correct proportions in the samples. The polls were conducted by interviewers who visited the selected individuals and asked for their opinions. However, once an interviewer was informed that he had to interview, for example, five white men over the age of 40 in suburban Chicago, he was free to choose whomever he wanted, and there enters the potential bias. For whatever reasons you can imagine (nicer neighborhoods, shinier cars in the driveways to draw attention, housewives who are at home to open the door, etc.), the interviewers tended to interview disproportionally many Dewey supporters. And this was no coincidence because the pollsters consistently overestimated the Republican vote in the elections 1936–1948. Republicans were simply slightly easier to interview and that biased the results somewhat, but only in 1948 was the difference between the parties small enough that the bias made the pollsters actually predict a Republican victory. See Table 2 for some numbers regarding the 1948 election results and predictions.

To avoid selection bias as much as possible, the individuals included in the sample must be identified when the poll starts. If interviews are done over the phone, interviewers must talk to the person that is selected, not whoever

Table 2: The 1948 election: Predictions by the three pollsters Crossley, Gallup, and Roper and the actual election result

Candidate	Crossley	Gallup	Roper	Election result
Truman	45	44	38	50
Dewey	50	50	53	45
Others	5	6	9	5

happens to pick up the phone. If interviews are done in person and nobody is at home at the moment, the interviewer should not ask the neighbor or the mailman for their opinions instead. A modern type of selection bias stems from the fact that polling is often done from phone directories, but more and more people, especially the young, have only cell phones and no land line and they are thus excluded from the samples. It is unclear if and how this affects the outcome of political opinion polls, but it would probably have great impact if the question was about support for a ban on cell phones in public places.

It is harder to avoid nonresponse bias, but the polling companies usually try to contact people several times before they give up. If nonresponse occurs randomly and is not too large, it is not too problematic but if it is believed to skew the results, it could be. What if a poll is done by mail and asks about whether people read their junk mail or just throw it away? Nonresponse is a reason that the number of people reported in opinion polls is often not a nice round number but something like 1,014, in which case, there were probably 486 out of 1,500 people who did not respond.

A special form of nonresponse is when a poll asks potentially embarrassing or otherwise loaded questions; in which case, people may not want to answer or may simply not tell the truth. This could, for instance, be questions about drug use or illegal behavior. A clever trick in this situation is to ask everybody to roll a die before answering the question. If the die shows 6, the person answers "yes" and otherwise the person answers truthfully. In this way the interviewer never knows if an affirmative answer is truthful or due to rolling a 6. How is the true proportion then estimated? For example, suppose that 6,000 people are polled and 3,000 answer "yes." As we expect 1,000 people to roll a 6 with the die, we expect that 1,000 of the 3,000 "yes"

answers are from these rolls and the remaining 2,000 are truthful. We thus count 2,000 out of 5,000 "yes" answers, and our estimate becomes 40%. The number of 6s will of course rarely be exactly 1,000, and the effect of the randomness in the die rolls will be reflected in a wider margin of error than an ordinary poll. When this special type of poll is done, it has been decided already from the beginning that one sixth of the sample be wasted.

The political parties also do their own polling, and somehow magically they always seem to get results that support their own candidates. Other than introducing selection bias and disregarding nonresponse bias, it is also possible to introduce bias in the results by the phrasing of the questions, on purpose or unconsciously. A poll in 2005 regarding the tragic Terri Schiavo case might have been an example of unclear phrasing. In this poll, 55% sided with Terri's husband and 53% sided with her parents—an overlap even though the two parties held diametrically opposite positions. All in all, the polls done by the major polling companies are well planned and executed and give accurate results. After all, there is competition going on between the polling companies and nobody wants to face the fate of the *Literary Digest*.

Serious polls that are done by random sampling are often called "scientific." In contrast, "unscientific" polls are, for example, when people are asked to call a TV show or vote on a website. Such polls have little value other than entertainment because they suffer from *self-selection* bias. I remember one particularly unscientific poll in Sweden in 1990. It was time for the census, and unlike the United States, Sweden does not have the census mandated in its constitution and there can be quite a bit of resistance against the government intruding in people's lives by requiring them to fill out the census form (Swedes are otherwise fairly tolerant of governmental intrusion). On an evening talk show with the charismatic and entertaining host Robert Aschberg (hard to describe to a non-Swedish audience), the census was discussed and Aschberg pulled out his form and set it on fire. At the same time, people were asked to call in for or against the census and, lo and behold, 95% of the callers were against! Certainly no "delicate handling" of statistics, but it was fun to watch. By the way, the response rate for the census ended up being 97.5%, so if this had been a serious opinion poll, it would by far have beaten the *Literary Digest* for the all time low.