

# Sampling Methods for Political Polling

It's impractical to poll an entire population—say, all 145 million registered voters in the United States. That is why pollsters select a sample of individuals that represents the whole population. Understanding how respondents come to be selected to be in a poll is a big step toward determining how well their views and opinions mirror those of the voting population.

To sample individuals, polling organizations can choose from a wide variety of options. Pollsters generally divide them into two types: those that are based on probability sampling methods and those based on non-probability sampling techniques.

For more than five decades probability sampling was the standard method for polls. But in recent years, as fewer people respond to polls and the costs of polls have gone up, researchers have turned to non-probability based sampling methods. For example, they may collect data on-line from volunteers who have joined an Internet panel. In a number of instances, these non-probability samples have produced results that were comparable or, in some cases, [more accurate in predicting election outcomes than probability-based surveys](#).

Now, more than ever, journalists and the public need to understand the strengths and weaknesses of both sampling techniques to effectively evaluate the quality of a survey, particularly election polls.

## Probability and Non-probability Samples

In a probability sample, all persons in the target population have a chance of being selected for the survey sample and we know what that chance is. For example, in a telephone survey based on random digit dialing (RDD) sampling, researchers know the chance or probability that a particular telephone number will be selected. (A description of RDD sampling and other techniques commonly used in election surveys appears at the end of this brief.)

The major advantage of a probability-based sampling is that we can calculate how well the findings from the sample represent the total population. That is, we can calculate the margin of sampling error, which measures how much our estimates vary based on the fact we're only measuring a sample of the population and not every member of the population. This ability to estimate, within a specified range, the accuracy of survey findings has made probability-based sampling the cornerstone of modern survey research.

Non-probability sampling methods do not share this feature that everyone in a population has a chance of being selected and researchers know exactly what that chance is. Participants are typically not selected at random to be included in the sample but rather come to be included by other means, for instance because they volunteer, a person's chance of being in the sample is unknown. For example, in an opt-in sample a person accepts an invitation to complete a survey that is offered to all visitors to a website. The chance of that person visiting that website and then choosing to participate in the survey cannot be known. One serious consequence is that only certain types of people may choose to opt into the survey and they may be different than those who do not in ways that could potentially bias the final results.

With non-probability samples there is no simple way to calculate the "margin of error;" instead, estimates of the likely error must be based on a statistical models. As a result, [AAPOR has](#)

cautioned that it may be misleading to report a margin of sampling error for surveys based on non-probability samples.

Nonresponse to polls is a big factor affecting the accuracy of poll results. In a probability sample, the respondents can be thought of as “self-selecting” into the sample. To the extent that the respondents and non-respondents differ systematically on the survey variables—for example, which candidate they support in an upcoming election--nonresponse can bias the poll results, and that is true even if the initial sample was a probability sample. In a similar way, the accuracy of non-probability samples, such as opt-in samples, can be affected by self-selection. In both types of sampling, if the people who participate in the poll are different from those who do not, results can be biased because of these differences.

In addition to sampling method, there are a number of other features of polls that affect the accuracy of the results. For example, how questions are worded or the sequence of questions presented to respondents have been shown to affect poll results and whether they reflect what people in total population really think.

For such reasons, AAPOR’s Code of Professional Ethics calls for transparency in the reporting of sample design, response rates, and the wording of the questions so that these elements can be assessed along with poll results.

## **Types of Sampling Techniques**

### **Probability Samples**

- **Random-Digit Dialing (RDD)**

Samples of telephone area codes and exchanges are selected, and then random digits are added to the end to create 10-digit phone numbers. The first step ensures phone numbers are distributed properly by geography. The second step, adding the random numbers, makes sure that even unlisted numbers are included. This has traditionally been the standard practiced by almost all public pollsters. The major advantage of RDD is the coverage of the population: Everyone with a telephone is eligible to be sampled. The major disadvantage is that it is expensive, since many of the landline telephone numbers generated are non-working numbers and cellphone numbers need to be manually dialed by interviewers.

- **Within Household Sample Selection**

In households in which more than one eligible respondent resides—in the case of election polls, more than one registered voter--further sampling among the members of the household should be done to produce a random sample of voters. Journalists should ask how respondents were selected. Simply taking the person who answers the telephone will not necessarily result in a representative sample.

- **Registration-Based Sampling (RBS)**

This begins with a sample of individuals drawn from lists of registered voters, to which phone numbers are then matched (or sometimes available from the voter list). This is less costly and more efficient, as almost all calls result in reaching a working phone number, which is not true of an RDD sample. One disadvantage of an RBS sample is that voter lists often do not include unlisted telephone numbers or full coverage of cellphone numbers; additionally they may not include voters who have just moved or registered to vote.

### **Non-probability Samples**

- **Self-Selected Samples (SSS)**

In self-selected or opt-in samples, respondents have selected themselves, and this means their answers may not be representative of the larger population. Types of self-selected samples include dial-in polls popular with the media and many Internet-based polls. The American Association for Public Opinion Research (AAPOR) cautions that results of surveys based on respondents who self-select may not be reliable. The characteristics of people who choose to participate in this type of survey may be different than those who do not in ways that bias the final results. These polls may sometimes be accurate, but it is very hard to evaluate whether they are accurate simply because of good luck or because they were able to capture good information about the population they were trying to represent. AAPOR has not yet made a final judgment about the reliability of opt-in samples, but warns that this type of sample is not based on the full target population.

- **Samples from Internet Panels**

One variation of the self-selected sample is the random sample selected from among people who have signed up to be members of an Internet panel. While the sample itself is random, the population from which the sample is drawn is made up of people who have signed up to be members of the panel, which may potential lead to selection bias.