



PewResearchCenter

MENU

RESEARCH AREAS

U.S. SURVEY RESEARCH

Our survey methodology in detail

Sampling

The typical Pew Research Center national survey selects a random digit sample of both landline and cellphone numbers in all 50 U.S. states and the District of Columbia. As the proportion of Americans who rely solely or mostly on cellphones for their telephone service continues to grow, sampling both landline and cellphone numbers helps to ensure that our surveys represent all adults who have access to either (only about 3% of households in the U.S. do not have access to any phone). We sample landline and cellphone numbers to yield a combined sample with approximately 25% of the interviews conducted by landline and 75% by cellphone. This ratio is based on an analysis that attempts to balance cost and fieldwork considerations as well as to improve the overall demographic composition of the sample (in terms of age, race/ethnicity and education). This ratio also ensures an adequate number of cell-only respondents in each survey.

The design of the landline sample ensures representation of both listed and unlisted numbers (including those not yet listed) by using random digit dialing. This method uses random generation of the last two digits of telephone numbers selected on the basis of the area code, telephone exchange and bank number. A bank is defined as 100 contiguous telephone numbers, for example 800-555-1200 to 800-555-1299. The telephone exchanges are selected to be proportionally stratified by county and by telephone exchange within the county. That is, the number of telephone numbers randomly sampled from within a given county is proportional to that county's share of telephone numbers in the U.S. Only banks of telephone numbers containing one or more listed residential numbers are selected.

The cellphone sample is drawn through systematic sampling from dedicated wireless banks of 100 contiguous numbers and shared service banks with no directory-listed landline numbers (to ensure that the cellphone sample does not include banks that are also included in the landline sample). The sample is designed to be representative both geographically and by large and small wireless carriers (also see cellphones (<http://www.pewresearch.org/methodology/u-s-survey-research/sampling/#cellphones>) for more information).

Both the landline and cell samples are released for interviewing in replicates, which are small random samples of each larger sample. Using replicates to control the release of telephone numbers ensures that the complete call procedures are followed for all numbers dialed. The use of replicates also improves the overall representativeness of the survey by helping to ensure that the regional distribution of numbers called is appropriate.

When interviewers reach someone on a landline phone, they randomly ask half the sample if they could speak with “the youngest male, 18 years of age or older, who is now at home” and the other half of the sample to speak with “the youngest female, 18 years of age or older, who is now at home.” If there is no eligible person of the requested gender currently at home, interviewers ask to speak with the youngest adult of the opposite gender, who is now at home. This method of selecting respondents within each household improves participation among young people who are often more difficult to interview than older people because of their lifestyles.

Unlike a landline phone, a cellphone is assumed in Pew Research Center polls to be a personal device. Interviewers ask if the person who answers the cellphone is 18 years of age or older to determine if the person is eligible to complete the survey (also see cellphone surveys (<http://www.pewresearch.org/methodology/u-s-survey-research/collecting-survey-data/#cell-phone-surveys>) for more information). This means that, for those in the cell sample, no effort is made to give other household members a chance to be interviewed. Although some people share cellphones, it is still uncertain whether the benefits of sampling among the users of a shared cellphone outweigh the disadvantages.

Sampling error results from collecting data from some, rather than all, members of the population. For each of our surveys, we report a margin of sampling error for the total sample and usually for key subgroups analyzed in the report (e.g., registered voters, Democrats, Republicans, etc.). For example, the sampling error for a typical Pew Research Center national survey of 1,500 completed interviews is plus or minus 2.9 percentage points with a 95% confidence interval. This means that in 95 out of every 100 samples of the same size and type, the results we obtain would vary by no more than plus or minus 2.9 percentage points from the result we would get if we could interview every member of the population. Thus, the chances are very high (95 out of 100) that any sample we draw will be within 3 points of the true population value. The sampling errors we report also take into account the effect of weighting. (Also see probability and non-probability sampling (<http://www.pewresearch.org/methodology/u-s-survey-research/sampling/#probability-and-non-probability-sampling>) for more information.)

Nonresponse

At least seven attempts are made to complete an interview at every sampled telephone number. The calls are staggered over times of day and days of the week (including at least one daytime call) to maximize the chances of making contact with a potential respondent. Interviewing is also spread as evenly as possible across the field period. An effort is made to recontact most interview breakoffs and refusals to attempt to convert them to completed interviews.

Response rates for Pew Research polls typically range from 5% to 15%; these response rates are comparable to those for other major opinion polls. The response rate is the percentage of known or

assumed residential households for which a completed interview was obtained. The response rate we report is computed using the American Association for Public Opinion Research's (AAPOR) Response Rate 3 (RR3) method. (For a full discussion of response rates, see AAPOR's Standard Definitions (<http://www.aapor.org/AAPORKentico/Communications/AAPOR-Journals/Standard-Definitions.aspx>) .) Fortunately, low response rates are not necessarily an indication of nonresponse bias, as we discuss in the problem of declining response rates (<http://www.pewresearch.org/methodology/u-s-survey-research/collecting-survey-data/#the-problem-of-declining-response-rates>) .

In addition to the response rate, we sometimes report the contact rate, cooperation rate or the completion rate for a survey. The contact rate is the proportion of working numbers where a request for an interview was made. The cooperation rate is the proportion of contacted numbers where someone gave initial consent to be interviewed. The completion rate is the proportion of initially cooperating and eligible households where someone completed the interview.

Data weighting

Nonresponse in telephone interview surveys can produce biases in survey-derived estimates. Survey participation tends to vary for different subgroups of the population, and these subgroups are likely to also vary on questions of substantive interest. To compensate for these known biases, the sample data are weighted for analysis.

The landline sample is first weighted by household size to account for the fact that people in larger households have a lower probability of being selected. In addition, the combined landline and cellphone sample is weighted to account for the fact that respondents with both a landline and cellphone have a greater probability of being included in the sample.

The sample is then weighted using population parameters from the U.S. Census Bureau for adults 18 years of age or older. The population parameters used for weighting are: gender by age; gender by education; age by education; region; race and Hispanic origin, which includes a break for Hispanics based on whether they were born in the U.S. or not; population density; and among non-Hispanic whites – age, education and region. The parameters for these variables are from the Census Bureau's 2014 American Community Survey (excluding those in institutionalized group quarters), except for the parameter for population density, which is from the 2010 census. These population parameters are compared with the sample characteristics to construct the weights. In addition to the demographic parameters, the sample is also weighted to match current patterns of telephone status (landline only, cellphone only or both landline and cellphone), based on extrapolations from the 2015 National Health Interview Survey. The final weights are derived using an iterative technique that simultaneously balances the distributions of all weighting parameters. These (<http://www.pewresearch.org/files/2015/03/Demographic-Questions-Phone-English-3-20-2015.pdf>) are the standard demographics we ask on our telephone polls in order to weight and analyze the data. You can also access the version of these demographics that we use on Web surveys (<http://www.pewresearch.org/files/2015/03/Demographic-Questions-Web-and-Mail-English-3-20-2015.pdf>) , as well as the Spanish phone (<http://www.pewresearch.org/files/2015/03/Demographic-Questions-Phone-Spanish-3-20-2015.pdf>) and Spanish Web

(<http://www.pewresearch.org/files/2015/03/Demographic-Question-Web-and-Mail-Spanish-3-20-2015.pdf>) versions.

Weighting cannot eliminate every source of nonresponse bias. Nonetheless, properly conducted public opinion polls have a good record in achieving unbiased samples. In particular, election polling – where a comparison of the polls with the actual election results provides an opportunity to validate the survey results – has been very accurate over the years (see the National Council on Public Polls’ evaluations of the 2012 (<http://www.ncpp.org/?q=node/137>) and 2010 (<http://www.ncpp.org/?q=node/134>) elections).

Data analysis

Each Pew Research survey report includes a “topline questionnaire” with all of the questions from that survey with the exact question wording and response options as they were read to respondents. This topline provides the results from the current survey for each question, as well as results from previous surveys in which the same or similar questions were asked.

For discussion of the results in reports and commentaries, differences among groups are reported when we have determined that the relationship is statistically significant and therefore is unlikely to occur by chance. Statistical tests of significance take into account the effect of weighting. In addition, to support any causal relationships discussed, more advanced multivariate statistical modeling techniques are often employed to test whether these connections exist, although the results of these models may or may not be shown in the actual report.

For most studies, it is our policy to release datasets from Pew Research surveys five months after the data were collected and archive them on our website as quickly as possible. Please visit our datasets (<http://www.pewresearch.org/data/download-datasets/>) page for further information.