



Methodology

March 2018 Political Survey

Prepared by ICF
for the Pew Research Center

March 2018

SUMMARY

The March 2018 Political Survey, sponsored by the Pew Research Center, obtained telephone interviews with a nationally representative sample of 1,466 adults, age 18 or older, living in the United States. Interviews were conducted via landline ($n_{LL}=384$) and cell phone ($n_C=1,082$; including 653 without a landline phone). ICF administered all interviews in English and Spanish from March 7 to 14, 2018. Statistical results are weighted to correct known demographic discrepancies. The margin of sampling error for the complete set of weighted data is ± 3.2 percentage points.

Details on the design, execution and analysis of the survey are discussed below.

DESIGN AND DATA COLLECTION PROCEDURES

Sample Design

A combination of landline and cellular random digit dial (RDD) samples was used to represent all adults in the United States who have access to either a landline or cellular telephone. Both samples were provided by Marketing Systems Group (MSG) according to ICF specifications.

Numbers for the landline sample were drawn with equal probabilities from active blocks (area code + exchange + two-digit block number) that contained one or residential telephone assignments. The cellular sample was not list-assisted, but was drawn through a systematic sampling from dedicated wireless 100-blocks.

Contact Procedures

Interviews were conducted from March 7 to 14, 2018. As many as seven attempts were made to contact every sampled telephone number. Sample was released for interviewing in replicates, which are representative subsamples of the larger sample. Using replicates to control the release of sample ensures that complete call procedures are followed for the entire sample. Calls were staggered over times of day and days of the week to maximize the chance of making contact with potential respondents. Interviewing was spread as evenly as possible across the days in field. When necessary, each telephone number was called at least one time during the day in an attempt to complete an interview.

For the landline sample, interviewers asked to speak with the youngest male or female currently at home based on a random rotation. If no male/female was available, interviewers asked to speak with the youngest adult of the other gender. This systematic respondent selection technique has been shown to produce samples that closely mirror the population in terms of age and gender when combined with cell interviewing.

For the cellular sample, interviews were conducted with the person who answered the phone. Interviewers verified that the person was an adult and in a safe place before administering the survey. Cellular respondents were offered a post-paid cash reimbursement for their participation.

DATA PROCESSING

Surveys were administered using a computer-assisted telephone interviewing (CATI) system. With CATI the need for data cleaning is greatly reduced, since skips and logic are built into the system. Responses that violated these programmed rules were not allowed. Data cleaning consisted of reviewing and back coding open-ended survey responses to the race question. In addition to the survey response data, we also appended sample frame data, metadata, and constructed variables. We redacted any survey variables that would specifically identify a respondent (e.g., name, address, and telephone number). We then produced a preliminary SPSS file and a final weighted SPSS file.

WEIGHTING AND ANALYSIS

Weighting is generally used in survey analysis to compensate for sample designs and patterns of non-response that might bias results. The sample was weighted to match national adult general population parameters. A two-stage weighting procedure was used to weight this dual-frame sample.

The first stage of weighting corrected for different probabilities of selection associated with the number of adults in each household and each respondent's telephone usage patterns.¹ This weighting also adjusts for the overlapping landline and cell sample frames and the relative sizes of each frame and each sample.

The first-stage weight for the i^{th} case can be expressed as:

$$WT_i = \left[\left(\frac{S_{LL}}{F_{LL}} \times \frac{1}{AD_i} \times LL_i \right) + \left(\frac{S_{CP}}{F_{CP}} \times CP_i \right) - \left(\frac{S_{LL}}{F_{LL}} \times \frac{1}{AD_i} \times LL_i \times \frac{S_{CP}}{F_{CP}} \times CP_i \right) \right]^{-1}$$

Where S_{LL} = the size of the landline sample

F_{LL} = the size of the landline sample frame

S_{CP} = the size of the cell sample

F_{CP} = the size of the cell sample frame

AD_i = Number of adults in household i

$LL_i=1$ if respondent has a landline phone, otherwise $LL=0$.

$CP_i=1$ if respondent has a cell phone, otherwise $CP=0$.

The second stage of weighting balances sample demographics to population parameters. The sample is balanced by form to match national population parameters for sex, age, education, race, Hispanic origin, region (U.S. Census definitions), population density, and telephone usage. The Hispanic origin was split out based on nativity; U.S. born and non-U.S. born. The White, non-Hispanic subgroup was also balanced on age, education and region.

The basic weighting parameters came from the U.S. Census Bureau's 2015 American Community Survey (ACS) data.² The population density parameter was derived from Census 2010 data. The telephone usage parameter came from an analysis of the June-December 2016 National Health Interview Survey.³

Weighting was accomplished using raking, also called sample balancing, an algorithm that iteratively adjusts the sample distributions to population distributions. Weights were trimmed to prevent individual interviews from having too much influence on the final results. The use of these weights in statistical analysis ensures that the demographic characteristics of the sample closely approximate the

¹ i.e., whether respondents have only a landline telephone, only a cell phone, or both kinds of telephone.

² ACS analysis was based on all adults excluding those living in institutional group quarters.

³ Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June, 2017. National Center for Health Statistics. Dec 2017.

demographic characteristics of the national population. Table 1 compares weighted and unweighted sample distributions to population parameters.

Table 1: Sample Demographics

	<u>Parameter</u>	<u>Unweighted</u>	<u>Weighted</u>
<u>Gender</u>			
	Male	48.3%	58.3%
	Female	51.6%	41.7%
<u>Age</u>			
	18-24	12.5%	6.3%
	25-34	17.7%	12.9%
	35-44	16.3%	12.1%
	45-54	17.2%	17.1%
	55-64	16.7%	22.4%
	65+	19.5%	29.2%
<u>Education</u>			
	HS Graduate or Less	39.7%	24.0%
	Some College/Assoc Degree	31.1%	25.9%
	College Graduate	29.1%	50.1%
<u>Race/Ethnicity</u>			
	White/not Hispanic	64.3%	73.3%
	Black/not Hispanic	11.7%	11.1%
	Hisp - US born	8.2%	5.2%
	Hisp - born outside	7.5%	3.1%
	Other/not Hispanic	8.3%	7.3%
<u>Region</u>			
	Northeast	17.9%	18.3%
	Midwest	21.0%	21.9%
	South	37.6%	37.0%
	West	23.6%	22.8%
<u>County Pop. Density</u>			
	1 - Lowest	19.9%	21.1%
	200.0%	20.0%	20.0%
	300.0%	20.1%	22.3%
	400.0%	20.0%	20.0%
	5 - Highest	20.0%	16.6%
<u>Household Phone Use</u>			
	LLO	5.3%	2.7%
	Dual	41.2%	52.8%
	CPO	53.4%	44.5%

Effects of Sample Design on Statistical Inference

Post-data collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. The "design effect" or *deff* represents the loss in statistical efficiency that results from the sample design and operations, including the dual-frame design and weighting adjustments for non-response. The total sample design effect for this survey is 1.54.

The composite design effect for a sample of size n , with each case having a weight, w_i is:

$$deff = \frac{n \sum_{i=1}^n w_i^2}{\left(\sum_{i=1}^n w_i \right)^2} \quad formula\ 1$$

The estimated *standard error* of a survey statistic is calculated by multiplying the usual formula by the square root of the design effect (\sqrt{deff}). Thus, the formula for computing the 95% confidence interval around a percentage is:

$$\hat{p} \pm \left(\sqrt{deff} \times 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right) \quad formula\ 2$$

where \hat{p} is the sample estimate and n is the unweighted number of sample cases in the group being considered.

The survey's *margin of error* is the largest 95% confidence interval for any estimated proportion based on the total sample—the one around 50%. For example, the margin of error for the entire sample is ± 3.2 percentage points. This means that in 95 out every 100 samples drawn using the same methodology, estimated proportions based on the entire sample will be no more than 3.2 percentage points away from their true values in the population. The margin of error for estimates based on form 1 or form 2 respondents is ± 4.5 percentage points. It is important to remember that sampling fluctuations are only one possible source of error in a survey estimate. Other sources, such as respondent selection bias, questionnaire wording and reporting inaccuracy, may contribute additional error of greater or lesser magnitude.

RESPONSE RATE

Table 2 reports the disposition of all sampled telephone numbers ever dialed from the original telephone number samples. The response rate estimates the fraction of all eligible sample that was ultimately interviewed. Response rates are computed according to American Association for Public Opinion Research standards.⁴ Thus the response rate for the landline samples was 4.3 percent. The response rate for the cellular samples was 5.6 percent.

⁴ The American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR.

Table 2. Sample Disposition

<u>Landline</u>	<u>Cell</u>	
5140	513	Non-residential/Business (4.500)
0	0	Ported numbers identified before dialing (4.420)
0	0	Cell in landline frame (4.420)
5140	513	OF = Out of Frame
37195	11657	Not working (4.300)
830	6	Computer/fax/modem (4.200)
38025	11663	NWC = Not working/computer
3828	5551	NA/Busy all attempts (3.120, 3.130)
27	401	VM not set up/caller out of range (3.100)
5	0	On DNC list - not dialed (3.90)
3860	5952	UHUONC = Non-contact, unknown if household/unknown other
4534	10367	Voice mail (3.140)
44	20	Other non-contact (deaf/disabled/deceased) (3.211)
4578	10387	UONC = Non-contact, unknown eligibility
2975	8579	Refusals (3.211)
166	422	Callbacks (INCLUDE Spanish CBs) (3.211)
3141	9001	UOR = Refusal, unknown if eligible
61	430	O = Other (language) (3.211)
0	692	Child's cell phone (4.700)
0	692	SO = Screen out
236	855	R = Refusal, known eligible (breakoffs and qualified CBs) (2.100)
384	1082	I = Completed interviews (1.0)
55425	40575	T = Total numbers sampled
16.3%	64.8%	$e1 = (I+R+SO+O+UOR+UONC)/(I+R+SO+O+UOR+UONC+OF+NWC)$ - Est. frame eligibility of non-contacts
100.0%	73.7%	$e2 = (I+R)/(I+R+SO)$ - Est. screening eligibility of unscreened contacts
42.3%	45.8%	$CON = [I + R + (e2*[O + UOR])]/[I + R + (e2*[O + UOR + UONC]) + (e1*e2*UHUONC)]$
10.0%	12.2%	$COOP = I/[I + R + (e2*[O + UOR])]$
4.3%	5.6%	AAPOR RR3 = $I/[I+R+(e2*(UOR+UONC+O))+(e1*e2*UHUONC)] = CON*COOP$