

## Methodology

### **2017 Typology Callback Survey**

Prepared by Princeton Survey Research Associates International  
for the Pew Research Center for the People and the Press

August 2017

#### **SUMMARY**

The 2017 Typology Callback Survey, sponsored by the Pew Research Center for the People and the Press, obtained telephone interviews with 1,893 adults in the United States. The interviews were conducted in English and Spanish by Princeton Data Source, LLC from August 15-21, 2017. Statistical results are weighted to correct known demographic discrepancies. The margin of sampling error for the complete set of weighted data is  $\pm 2.9$  percentage points.

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

#### **DESIGN AND DATA COLLECTION PROCEDURES**

##### **Sample Design**

This project used callback sample from Pew's 2017 Political Landscape survey which fielded in two waves from June 8 to 18 and June 27 to July 9. Sample for the original surveys was drawn using standard *list-assisted random digit dialing* (RDD) methodologies. All respondents from the original surveys were recontacted for this survey.

## Contact Procedures

Interviews were conducted from August 15-21, 2017. As many as 7 attempts were made to contact every sampled telephone number. Sample was released for interviewing in replicates, which are subsamples of the larger sample.<sup>1</sup> Using replicates to control the release of sample ensures that complete call procedures are followed for the entire sample. Interviews were conducted with the person who completed the original interview. Calls were staggered over times of day and days of the week to maximize contact with respondents.

## WEIGHTING AND ANALYSIS

Weighting is generally used in survey analysis to compensate for patterns of nonresponse that might bias results. Two stages of weighting were performed on the data. The weight from the original sample datasets was used as a first-stage weight for this project. This first-stage weight corrects for different probabilities of selection and differential non-response associated with the original sampling and interviewing. The final sample of completes was then raked to match parameters for sex by age, sex by education, age by education, region, race/ethnicity, population density, and phone use. The non-Hispanic, white subgroup was also raked to age, education and region. The basic weighting parameters came from the U.S. Census Bureau's 2015 American Community Survey (ACS) data.<sup>2</sup> The population density parameter was derived from Census 2010 data. The telephone usage parameter came from an analysis of the July-December 2016 National Health Interview Survey.<sup>3</sup> Table 1 compares weighted and unweighted total sample distributions to population parameters.

Weighting was accomplished using Sample Balancing, a special iterative sample weighting program that simultaneously balances the distributions of all variables using a statistical technique called the *Deming Algorithm*. 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 demographic characteristics of the target population.

---

<sup>1</sup> Replicates were assigned to load the lower number replicates with certain typology groups that we felt could be more difficult to conduct the callback interview.

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

<sup>3</sup> Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December, 2016. National Center for Health Statistics. May 2017.

**Table 1. Sample Demographics**

	<u>Parameter</u>	<u>Unweighted</u>	<u>Weighted</u>
<u>Gender</u>			
	Male	48.3	54.1
	Female	51.7	45.9
<u>Age</u>			
	18-24	12.7	7.5
	25-34	17.6	11.8
	35-44	16.6	11.8
	45-54	17.5	16.7
	55-64	16.6	21.1
	65+	19.0	31.1
<u>Education</u>			
	HS Grad or less	40.3	22.2
	Some College/Assoc Degree	31.3	25.4
	College Graduate	28.4	52.4
<u>Race/Ethnicity</u>			
	White/not Hispanic	64.6	76.8
	Black/not Hispanic	11.7	9.0
	Hisp - US born	8.0	5.4
	Hisp - born outside	7.5	2.9
	Other/not Hispanic	8.2	5.9
<u>Region</u>			
	Northeast	17.9	16.1
	Midwest	21.1	23.9
	South	37.4	37.6
	West	23.6	22.4
<u>County Pop. Density</u>			
	1 - Lowest	19.9	21.8
	2	20.0	20.1
	3	20.1	20.8
	4	20.0	21.1
	5 - Highest	20.0	16.2
<u>Household Phone Use</u>			
	LLO	5.3	2.5
	Dual	41.2	54.0
	CPO	53.5	43.5

## Effects of Sample Design on Statistical Inference

Post-data collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. PSRAI calculates the effects of these design features so that an appropriate adjustment can be incorporated into tests of statistical significance when using these data. The so-called "design effect" or *deff* represents the loss in statistical efficiency that results from systematic non-response. The total sample design effect is 1.61.

PSRAI calculates the composite design effect for a sample of size  $n$ , with each case having a weight,  $w_i$  as:

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

In a wide range of situations, the adjusted *standard error* of a statistic should be 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:

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

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  $\pm 2.9$  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 2.9 percentage points away from their true values in that population. 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 callback samples. The response rate estimates the fraction of all eligible respondents in the sample that were ultimately interviewed. Response rates are computed according to American Association for Public Opinion Research standards.<sup>4</sup> The response rate for the landline sample was 50 percent and the response rate for the cell sample was 38 percent.<sup>5</sup>

---

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

<sup>5</sup> The original survey response rates for the August survey were 10 percent for the landline sample and 9 percent for the cell sample. The response rates for the October survey were 8 percent for the landline sample and 8 percent for the cell sample.

**Table 2. Sample Disposition**

<u>Landline</u>	<u>Cell</u>	
16	30	Non-residential/Business (4.500)
0	0	Cell in landline frame (4.420)
16	30	OF = Out of Frame
55	59	Not working (4.300)
7	0	Computer/fax/modem (4.200)
62	59	NWC = Not working/computer
5	1	NA/Busy all attempts (3.120, 3.130)
0	187	VM not set up/caller out of range (3.100)
5	188	UHUO <sub>NC</sub> = Non-contact, unknown if household/unknown other
178	842	Voice mail (3.140)
3	3	Other non-contact (deaf/disabled/deceased) (3.211)
181	845	UO <sub>NC</sub> = Non-contact, unknown eligibility
305	854	Refusals (3.211)
67	231	Callbacks (INCLUDE Spanish CBs) (3.211)
372	1,085	UO <sub>R</sub> = Refusal, unknown if eligible
1	3	O = Other (language) (3.211)
70	0	Target resp does not exist (4.700)
0	21	R did not do original survey (4.700)
70	21	SO = Screen out
31	140	R = Refusal, known eligible (breakoffs and qualified CBs) (2.100)
517	1,376	I = Completed interviews (1.0)
1,255	3,747	T = Total numbers sampled
93.8%	97.5%	$e1 = (I+R+SO+O+UO_R+UO_{NC}) / (I+R+SO+O+UO_R+UO_{NC}+OF+NWC)$ - Est. frame eligibility of non-contacts
88.7%	98.6%	$e2 = (I+R) / (I+R+SO)$ - Est. screening eligibility of unscreened contacts
84.2%	71.9%	$CON = [I + R + (e2*[O + UO_R])] / [I + R + (e2*[O + UO_R + UO_{NC}]) + (e1*e2*UHUO_{NC})]$
58.8%	53.1%	$COOP = I / [I + R + (e2*[O + UO_R])]$
49.5%	38.2%	$AAPOR\ RR3 = I / [I+R+(e2*(UO_R+UO_{NC}+O))+(e1*e2*UHUO_{NC})] = CON*COOP$

