

Methodology

Nashville Poll 2017

Prepared by Princeton Survey Research Associates International
for Vanderbilt University

March 2017

The Nashville Poll 2017 obtained telephone interviews with a representative sample of 1,106 adults living in Nashville-Davidson County, Tennessee. Telephone interviews were conducted by landline (303) and cell phone (803, including 568 without a landline phone). The survey was conducted by Princeton Survey Research Associates International (PSRAI). Interviews were done in English and Spanish by Princeton Data Source and Issues & Answers from February 27 to March 19, 2017. Statistical results are weighted to correct known demographic discrepancies. The margin of sampling error for the complete set of weighted data is ± 3.5 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 Davidson County who have access to either a landline or cellular telephone. Both samples were provided by Survey Sampling International, LLC (SSI) according to PSRAI specifications.

Numbers for the landline sample were drawn with equal probabilities from active blocks (area code + exchange + two-digit block number) containing one or more residential directory listings.

There were two types of cellular sample used. The first cellular sample was RDD, but not list-assisted. The sample was drawn through a systematic sampling from dedicated wireless 100-blocks and shared service 100-blocks with no directory-listed landline numbers. Active cell phone numbers were targeted by using an activity flag appended by SSI which identifies numbers that are likely to be active. Inactive numbers were not dialed thus boosting the cell phone sample working rate and productivity. In order to help increase the number of Hispanic respondents, a second targeted listed cellular sample was used. A total of 303 interviews were

completed from the landline RDD sample, 642 interviews were completed from the cell RDD sample and 161 interviews were completed from the targeted cell sample.

Contact Procedures

Interviews were conducted from February 27 to March 19, 2017. As many as five 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. Each phone number received at least one daytime call when necessary.

For the landline sample, interviewers asked to speak with the adult currently at home with the most recent birthday. Prior to dialing, the landline sample was scrubbed of numbers that have been ported to wireless service by comparing the sample file to the most recently available Intermodal Ported Telephone Number Identification Service database. 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.

Once a potential respondent was on the phone, interviewers confirmed that they lived in Nashville-Davidson County in Tennessee before conducting the full interview.

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 weighting ensures that the demographic profile of the sample matches the profile of the target population. The data was weighted to match Nashville parameters.

A two-stage weighting procedure was used to weight this 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 sample frames and the relative sizes of each frame and each sample.

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

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

$$wt_i = [P1_i + P2_i + P3_i - (P1_i \times P2_i) - (P1_i \times P3_i) - (P2_i \times P3_i) + (P1_i \times P2_i \times P3_i)]^{-1}$$

Where $P1_i$ is the probability that the respondent was sampled in the RDD landline frame, $P2_i$ is the probability that the respondent was sampled in the RDD cell frame and $P3_i$ is the probability that the respondent was sampled in the listed cell frame. Here are the formulas for the individual frame selection probabilities for the i^{th} respondent.

$$P1_i = \frac{LL_i \times S1}{F1 \times AD_i}$$

Where $LL_i=1$ if respondent i has a landline phone and $LL_i=0$ if respondent i has no landline phone. $S1$ is the size of the RDD landline sample and $F1$ is the size of the RDD landline frame. AD_i is the number of adults in household i , capped at three.

$$P2_i = \frac{CP_i \times S2}{F2}$$

Where $CP_i=1$ if respondent i has a cell phone and $CP=0$ if respondent i has no cell phone. $S2$ is the size of the RDD cell sample and $F2$ is the size of the RDD cell frame.

$$P3_i = \frac{CP_i \times HISPF_i \times S3}{F3}$$

Where $HISPF_i=1$ if respondent i is Hispanic and has a cell phone and could therefore be in this frame. Otherwise $HISPF_i=0$. $S3$ is the size of the targeted sample and $F3$ is the potential size of the targeted cell sample frame. $F3$ was estimated as the product of $F2$ and the proportion of Davidson County that is Hispanic (9.9%).

The second stage of weighting balanced total sample demographics to population parameters. The sample is balanced to match Davidson County parameters for sex by age, sex by education, race, and telephone usage. The basic weighting parameters came from the U.S. Census Bureau's 2011-2015 5-Year American Community Survey (ACS) data. The telephone

usage parameter was extrapolated from PSRAI's phone use estimate which is derived from figures from the National Health Interview Survey released by the CDC.²

Weighting was accomplished using SPSSINC RAKE, an SPSS extension module that simultaneously balances the distributions of all variables using the GENLOG procedure. 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. Table 1 compares weighted and unweighted total sample distributions to population parameters.

Two additional weight variables were computed. One for African-American respondents and one for Hispanic respondents (AAWEIGHT and HISPWEIGHT) and should be used for analyses of these groups. These weights balanced those groups' demographics to their population parameters for sex, age and education.

² <https://www.cdc.gov/nchs/nhis/releases.htm>

Table 1: Total Sample Demographics

	<u>Parameter</u>	<u>Unweighted</u>	<u>Weighted</u>	
<u>Gender</u>				
	Male	47.6	45.9	48.2
	Female	52.4	54.1	51.8
<u>Age</u>				
	18-24	13.6	8.2	13.6
	25-34	24.1	13.8	23.8
	35-44	17.5	15.3	18.4
	45-64	30.9	35.0	30.0
	65+	13.9	27.6	14.3
<u>Education</u>				
	HS Grad or less	37.0	35.5	35.8
	Some College/Assoc. Degree	28.5	25.2	28.4
	College Graduate	34.5	39.3	35.9
<u>Race/Ethnicity</u>				
	White/not Hispanic	56.8	55.5	55.9
	Black/not Hispanic	27.4	20.6	27.5
	Hispanic	9.9	19.1	10.6
	Other/not Hispanic	5.9	4.8	6.1
			2.7	2.8
<u>Household Phone Use</u>				
	LLO	2.4	1.9	2.4
	Dual	41.2	46.7	41.0
	CPO	56.4	51.4	56.6

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 a disproportionate sample design and systematic non-response. The total sample design effect for this survey is 1.44.

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 ± 3.5 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 3.5 percentage points away from their true values in the population. Margins of error for subgroups will be larger. The margin of error for African-American respondents is ± 8.1 percentage points and the margin of error for Hispanic respondents is ± 7.0 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.³ Thus the response rate for the landline samples was 9 percent. The response rate for the cellular samples was 12 percent.

³ 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>Total Cell</u>	<u>Cell RDD</u>	<u>Cell Targeted</u>	
1,187	568	553	15	Non-residential/Business (4.500)
813	----	----	----	Ported numbers identified before dialing (4.420)
5	----	----	----	Cell in landline frame (4.420)
2,005	568	553	15	OF = Out of Frame
14,221	5,995	5,786	209	Not working (4.300)
564	18	18	0	Computer/fax/modem (4.200)
14,785	6,013	5,804	209	NWC = Not working/computer
1,559	730	687	43	NA/Busy all attempts (3.120, 3.130)
----	3,159	2,877	282	VM not set up/caller out of range (3.100)
15	20	19	1	On DNC list - not dialed (3.90)
1,574	3,909	3,583	326	UHUO_{NC} = Non-contact, unknown if household/unknown other
2,063	11,387	10,309	1,078	Voice mail (3.140)
0	145	145	0	Blocked call
16	22	17	5	Other non-contact (deaf/disabled/deceased) (3.211)
2,079	11,554	10,471	1,083	UO_{NC} = Non-contact, unknown eligibility
1,291	6,851	6,028	823	Refusals (3.211)
88	1,621	1,347	274	Callbacks (INCLUDE Spanish CBs) (3.211)
1,379	8,472	7,375	1,097	UO_R = Refusal, unknown if eligible
6	54	46	8	O = Other (language) (3.211)
----	687	555	132	Child's cell phone (4.700)
61	2,238	2,148	90	Other ineligible - Not a resident (4.700)
61	2,925	2,703	222	SO = Screen out
57	141	87	54	R = Refusal, known eligible (breakoffs and qualified CBs) (2.100)
303	803	642	161	I = Completed interviews (1.0)
22,249	34,439	31,264	3,175	T = Total numbers sampled
18.8%	78.4%	77.0%	92.1%	$e1 = (I+R+SO+O+UO_R+UO_{NC}) / (I+R+SO+O+UO_R+UO_{NC}+OF+NWC)$ - Est. frame eligibility of non-contacts
85.5%	24.4%	21.2%	49.2%	$e2 = (I+R) / (I+R+SO)$ - Est. screening eligibility of unscreened contacts
43.2%	45.9%	45.1%	52.7%	$CON = [I + R + (e2*[O + UO_R])] / [I + R + (e2*[O + UO_R + UO_{NC}]) + (e1*e2*UHUO_{NC})]$
19.6%	26.6%	27.8%	21.2%	$COOP = I / [I + R + (e2*[O + UO_R])]$
8.5%	12.2%	12.5%	11.2%	AAPOR RR3 = $I / [I+R+[e2*(UO_R+UO_{NC}+O)]+[e1*e2*UHUO_{NC}]] = CON*COOP$