Methods Report for Vanderbilt University's:

2018 Nashville/ Davidson County Poll

Survey of Adult Residents of Davidson County, Tennessee

February 26, 2018

Prepared by:

Chintan Turakhia, Jonathan Best, and Jennifer Su
1 Braxton Way
Suite 125
Glen Mills, PA 19342

@ssrs_solutions | 484.840.4300 | info@ssrs.com

CONTENTS

OVERVIEW

In 2015 and 2017, the Center for the Study of Democratic Institutions (CSDI) at Vanderbilt University conducted public opinion polls of Nashville/Davidson County residents to help inform community stakeholders, government officials, academics, the general public, etc. about important issues facing the Nashville Metro area. In 2018, Vanderbilt commissioned SSRS of Glen Mills, PA to conduct the University's third poll of the Nashville Metro area.

The 2018 Nashville/Davidson County Poll obtained telephone interviews with a representative sample of 800 adults, age 18 or older, living in Nashville-Davidson County, Tennessee. Telephone interviews were conducted by landline (282) and cell phone (518, including 341 without a landline phone). Interviews were done in English from February 8-19, 2018. Statistical results are weighted to correct known demographic discrepancies. The margin of sampling error for the complete set of weighted data is ± 4.6 percentage points.

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

QUESTIONNAIRE DEVELOPMENT

The questionnaire was developed by the Principal Investigators at the Center for the Study of Democratic Institutions (CSDI) at Vanderbilt in consultation with the SSRS project team. Prior to the field period, SSRS programmed the study into CfMC 8.6 Computer Assisted Telephone Interviewing (CATI) software. Extensive checking of the program was conducted to ensure that skip patterns and sample splits followed the design of the questionnaire.

SAMPLING METHODS

The target population for this poll was adults age 18 or older who are residents of Nashville/Davidson County in Tennessee. SSRS used a hybrid sampling approach via three sources. All samples were provided according to SSRS specifications.

1. Registration-based sample (RBS);
2. Overlapping dual-frame landline and cellular random digit dial (RDD) sample; and
3. Re-contact sample from Vanderbilt's Nashville/Davidson County study conducted in 2017.

RBS sample was procured from L2, one of the major providers of voter list samples. RDD landline and cellular samples were procured from Survey Sampling International (SSI). Re-contact sample was comprised of respondents who completed the 2017 Nashville/Davidson County survey in English.

To improve efficiency and to obtain between 160-200 African-American interviews, a multi-strata design was implemented for this study:

1. Registration-based landline and cellular sample
    a. RBS Likely African-American, Age 18-35
    b. RBS Likely African-American, Age 36 or older
    c. RBS Likely Other race, Age 18-35
    d. RBS Likely Other race, Age 36 or older
2. Dual-frame RDD sample

a. RDD landline sample
b. RDD cellular sample, no billing zip match
c. RDD cellular sample, with Nashville/Davidson county billing zip

3. Re-contact sample from the 2017 study
    a. Re-contact African American
    b. Re-contact Other

To make geographic screening for cellular phones more efficient for the RDD cellular sample, SSRS leveraged cellular billing zip code available in the database and sampled those records that have a Nashville/Davidson county billing zip code at a higher rate. Consistent with prior waves, inactive cellular phone numbers in the RDD samples were suppressed.

Vanderbilt expressed interest in obtaining a robust sample of African-American interviews. To help achieve this goal without screening for race, likely African-Americans were oversampled in the RBS and Re-contact samples.

In order for the frames to be mutually exclusive, the RBS frame was de-duped against the RDD and Re-contact sample frames, and all duplicate records removed from the RDD and Re-contact frames. Similarly, the RDD sample frame was de-duped against the Re-contact sample frame, and all duplicate records removed from the RDD frame. This step was performed prior to configuring the replicates for this study.

The sample size for this poll was n=800 interviews. Overall, approximately 35% of respondents were reached via landline and 65% of respondents were reached via cellular telephone. Under the sample design described above, 338 interviews came from the RBS sample, 259 interviews came from the dual-frame RDD sample, and 203 interviews came from Re-contact sample.

Table 1: Interviews by Sample Frame

Sample	Interviews
Registration-based sample	338
Dual-frame RDD sample	259
Re-contact sample	203
Total	800

CONTACT PROCEDURES

The respondent selection procedures for RBS and RDD samples were identical. For respondent selection for landline samples, interviews were conducted with the youngest adult 18+ male/female at home based on a random rotation. 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. For both landline and cell samples, after an adult was on the phone, geographic eligibility was determined prior to accepting the respondent into the survey.

For Re-contact landline sample, interviewers asked for respondents by gender and age. If the target respondent did not exist, then interviewers asked for the youngest adult 18+ male/female at home based on a random rotation. For Re-contact cell sample, interviewers were conducted with the person who answered the phone after verifying that the person was an adult. If the person who answered the phone

indicated they did not participate in the 2017 study, interviewers attempted to conduct the interview with the person on the phone. For both landline and cell samples, after confirming an adult was on the phone, geographic eligibility was determined prior to accepting the respondent into the survey.

DATA COLLECTION, PROCESSING, AND INTEGRATION

Survey Administration

The field period for this study was February 8-19, 2018. All interviews were completed in English using the CATI system. The CATI system ensured that questions followed logical skip patterns and that complete dispositions of all call attempts were recorded.

CATI interviewers received written materials about the survey instrument and received formal training for this particular project. The written materials were provided prior to commencement of data collection and included an annotated questionnaire that contained information about the goals of the study, detailed explanations about why questions were being asked, the meaning and pronunciation of key terms, potential obstacles to be overcome in getting good answers to questions, and respondent problems that could be anticipated ahead of time, as well as strategies for addressing the potential problems.

Interviewer training was conducted before the study was launched. Interviewers were given instructions to help them maximize response rates and ensure accurate data collection.

In order to maximize survey response, SSRS enacted the following procedures during the field period:

- As many as five (5) 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. At least one daytime call was conducted if necessary.
- Interviewers explained the purpose of the study and its importance.
- Respondents were offered the option of scheduling a call-back at their convenience.
- For Re-contact sample, specially trained interviewers contacted households where the initial call resulted in respondents hanging up the phone or where an earlier call resulted in a suspended interview with a qualified respondent.

Data Processing and Integration

Prior to running cross-tabulations, data were cleaned and checked using standard procedures. This program establishes editing parameters in order to locate any errors. No coding was done for open-end responses.

WEIGHTING AND ANALYSIS

The survey data were weighted to account for the complex sample design and to correct for systematic nonresponse along known parameters. The sample was weighted in two stages. The first-stage of the weighting was the application of a base weight to account for different selection probabilities. In the second stage of weighting, sample demographics were match to population parameters.

Base weight

Sample was drawn from multiple overlapping sample frames. The first stage of weighting assigned a base weight to account for different probabilities of selection within each frame and also to correct for the overlapping frames. The five sample frames are outlined in Table 2.

Table 2: Sample Frames

Frame	Value Label
f1	RBS
f2	RDD landline
f3	RDD cell
f4	Re-contact landline
f5	Re-contact cell

The probability that a respondent was sampled in any sample frame is denoted by p_i and $h_{f,i}$ is an indicator variable that equals 1 if the respondent is in the frame, f, and equals 0 if the respondent is not in the frame. For example, if respondent R1 is a registered voter who has a landline but no cell phone, then we assume $h_{1,R1} = h_{2,R1} = h_{4,R1} = 1$ and $h_{3,R1} = h_{5,R1} = 0$. The overall probability that a respondent was sampled is the sum of the individual frame selection probabilities minus the product of all the two-way frame overlaps.[1] P_i is:

$$P_i = (h_1p_1 + h_2p_2 + h_3p_3 + h_4p_4 + h_5p_5) - (h_1p_1 \times h_2p_2) - (h_1p_1 \times h_3p_3) - (h_1p_1 \times h_4p_4) - (h_1p_1 \times h_5p_5) - (h_2p_2 \times h_3p_3) - (h_2p_2 \times h_4p_4) - (h_2p_2 \times h_5p_5) - (h_3p_3 \times h_4p_4) - (h_3p_3 \times h_5p_5) - (h_4p_4 \times h_5p_5)$$

The base weight is simply the inverse of the overall selection probability:

$$BW = 1/P$$

Post-stratification

The second stage of weighting balanced total sample demographics to population parameters. The sample was balanced to match Davidson County, Tennessee parameters for sex, age, education, and race/ethnicity. For the sex, age, and education parameters, African-Americans were balanced separately to ensure their proper representation. The parameters were derived from the U.S. Census Bureau's 2012-2016 5-year American Community Survey (ACS) data.

The following table lists the dimensions used in the raking.

[1] Selection probabilities for 3-frame, 4-frame and 5-frame overlaps all approach zero and were omitted from the equation.

Table 3: Raking Dimensions

Dimension	Value Label
1: Gender by Race	Male, African-American Female, African-American Male, not African-American Female, not African-American
2: Age by Race	18-34, African-American 35-44, African-American 45-64, African-American 65+, African-American 18-24, not African-American 25-34, not African-American 35-44, not African-American 45-64, not African-American 65+, not African-American
3: Education by Race	HS grad or less, African-American Some college/Assoc., African-American College grad, African-American HS grad or less, not African-American Some college/Assoc., not African-American College grad, not African-American
4: Race/Ethnicity	White, not Hispanic Black or African-American, not Hispanic Hispanic Other, not Hispanic

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 4 compares weighted and unweighted total sample distributions to population parameters.

Table 4: Population Parameters and Weighted and Unweighted Total Sample Distributions

Characteristic	Value Label	Parameter	Unweighted	Weighted
Sex	Male	47.3	45.6	46.7
	Female	52.7	54.4	53.3
Age	18-24	10.5	2.8	7.7
	25-34	25.9	9.1	24.2
	35-44	18	10.0	17.8
	45-64	30.9	41.1	33.9
	65+	14.8	37.0	16.4
Education	HS Grad or less	34.4	20.4	30.1
	Some college/Assoc. degree	26.8	28.9	27.9
	College graduate	38.9	50.8	42.0
Race/Ethnicity	White, not Hispanic	60.7	68.4	63.6
	Black, not Hispanic	25.7	23.3	24.0
	Hispanic	8	3.9	6.4
	Other, not Hispanic	5.6	4.5	6.0

Effects of Sample Design on Statistical Analysis

Post-data collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. SSRS 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 1.77.

SSRS calculates the composite design effect for a sample of size n, with each case having a weight, w as:

$$deff = \frac{n \sum w^2}{(\sum w)^2}$$

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:

$$\hat{p} \pm \sqrt{deff} \times 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

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 ± 4.6 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 4.6 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 ± 9.9 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 RATES

Table 5 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.[2]

- The response rate for the RDD landline samples was 6 percent.
- The response rate for the RDD cellular samples was 8 percent.
- The response rate for RBS landline was 8 percent.
- The response rate for RBS cell was 5 percent.
- The response rate for Re-contact sample was 24 percent.

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

Table 5: Sample Disposition

RDD LL	RDD Cell	RBS LL	RBS Cell	Re-contact	
64	48	69	79	6	Non-residential/Business
2	4	2	0	0	Cell in landline frame / LL in Cell frame
1	1	0	3	1	Duplicate number
67	53	71	82	7	OF = Out of Frame
98	144	974	302	23	Not working
70	4	244	5	1	Computer/fax/modem
168	148	1,218	307	24	NWC = Not working/computer
878	1,076	925	3,382	155	NA/Busy all attempts
878	1,076	925	3,382	155	$UHUO_{NC}$ = Non-contact, unknown if household/unknown other
189	1,592	581	3,508	246	Voice mail
40	14	67	32	2	Privacy Manager
4	2	1	9	1	Other non-contact (deaf/disabled)
233	1,608	649	3,549	249	UO_{NC} = Non-contact, unknown eligibility
181	818	381	1,145	187	Refusals
58	271	118	463	102	Callbacks
239	1,089	499	1,608	289	UO_R = Refusal, unknown if eligible
4	65	6	68	4	O = Other (language)
0	81	0	85	9	Child's cell phone
6	61	4	138	15	Other ineligible
6	142	4	223	24	SO = Screen out (ineligible)
18	28	25	19	15	R = Refusal, known eligible (breakoffs and qualified CBs)
68	191	140	198	203	I = Completed interviews
1,681	4,400	3,537	9,436	970	T = Total numbers sampled

Table 5: Sample Disposition (continued)

					continued…
70.7%	94.0%	50.7%	93.6%	96.2%	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
93.5%	60.7%	97.6%	49.3%	90.1%	e2 = (I+R)/(I+R+SO) - Est. screening eligibility of unscreened contacts
28.2%	36.6%	37.6%	24.0%	57.3%	CON = $[I + R + (e2*[O + UO_R])]/[I + R + (e2*[O + UO_R + UO_{NC}]) + (e1*e2*UHUO_{NC})]$
21.7%	20.8%	21.3%	19.0%	42.1%	COOP = $I/[I + R + (e2*[O + UO_R])]$
6.1%	**7.6%**	**8.0%**	**4.5%**	**24.2%**	**AAPOR RR3=$I/[I+R+[e2*(UO_R+UO_{NC}+O)]+[e1*e2*UHUO_{NC}]]$ = CON*COOP**

DELIVERABLES

SSRS delivered to Vanderbilt University:

- Final questionnaire instrument;
- Weighted dataset in SPSS;
- Weighted banners in Microsoft Word;
- A detailed methodology report; and
- A contact file for reporter follow-up.

ABOUT SSRS

SSRS is a full-service market and survey research firm managed by a core of dedicated professionals with advanced degrees in the social sciences. Service offerings include the Omnibus Survey, Probability Panel and other Online Solutions as well as custom research programs – all driven by a central commitment to methodological vigor. The SSRS team is renowned for its multimodal approach, as well as its sophisticated and proprietary sample designs. Typical projects for the company include complex strategic, tactical and public opinion initiatives in the U.S. and in more than 40 countries worldwide. SSRS is research, refined. Visit www.ssrs.com for more information.