Methods Report for the Center for Public Opinion at UMass-Lowell's:

2018 Survey of Massachusetts Registered Voters

October 9, 2018

Prepared by:

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

@ssrs_research | 484.840.4300 | info@ssrs.com

CONTENTS

Note: SSRS is proud to be a Charter Member of the American Association for Public Opinion Research's (AAPOR) *Transparency Initiative*. We support and encourage transparency in the reporting of research results and are committed to AAPOR's Code of Professional Ethics and Practices.

OVERVIEW

The Center for Public Opinion at UMass-Lowell commissioned SSRS of Glen Mills, PA to conduct the Center's 2018 Survey of Massachusetts Registered Voters before the midterms elections.

The 2018 Poll obtained telephone interviews with a representative sample of 952 adults, age 18 or older, who live in the state of Massachusetts, including 791 Massachusetts registered voters. Telephone interviews were conducted by landline (433) and cell phone (519, including 275 without a landline phone).

Interviews were done in English and Spanish from October 1-7, 2018. Statistical results are weighted to correct known demographic discrepancies. The margin of sampling error for the complete set of weighted data is ± 3.9 percentage points for results based on Total and ± 4.4 percentage points for results based on Registered voters.

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

SAMPLING METHODS

The target population for this study was adults age 18+ who are registered voters in the state of Massachusetts. SSRS used an overlapping dual frame random digit dial (RDD) sample design for this study, with 50% of completions allocated to the landline frame and 50% allocated to the cellular frame. Both samples were provided by Marketing Systems Group (MSG) according to SSRS specifications.

Phone numbers used for this study were randomly generated from landline and cell phone sampling frames, with an overlapping frame design. The RDD landline sample was generated through MSG's GENESYS sampling system. The standard GENESYS RDD methodology produces a strict single-stage, Equal Probability Selection Method (EPSEM) sample of residential telephone numbers. In other words, a GENESYS RDD sample ensures an equal and known probability of selection for every residential telephone number in the sample frame. The sample was generated shortly before the beginning of data collection to provide the most up-to-date sample possible, maximizing the number of valid telephone extensions. Following generation, the RDD sample was prepared using MSG's proprietary GENESYS IDplus procedure, which identifies and eliminates a large percentage of all non-working and business numbers. Numbers flagged as cellular were removed from the landline sample frame.

Using a procedure similar to that used for the landline sample, MSG generated an RDD sample of cellular phone numbers, stratified by state region. To improve the efficiency of the cellular sample, inactive cell phone numbers were flagged utilizing MSG's Cell-Wins procedure and removed prior to dialing. Dialing inactive cellular numbers is largely unproductive and drains interviewer resources. Suppressing inactive cell numbers helps increase the productivity of interviewing hours by limiting telephone dialing to active cell numbers.

The sample size for this poll was n=791 MA registered voters. Overall, approximately 50% of respondents were reached via landline and 50% of respondents were reached via cellular telephone. Under the sample design described above, 397 interviews came from RDD landline sample and 394 interviews came from RDD cellular sample. An additional 36 landline interviews and 125 cellular interviews were conducted among Massachusetts adults who are not registered to vote.

Table 1: Interviews by Sample Frame

Sample	Interviews
RDD Landline, Registered voters	397
RDD Cell, Registered voters	394
RDD LL, Not registered to vote	36
RDD Cell, Not registered to vote	125
Total	952

QUESTIONNAIRE DEVELOPMENT

The questionnaire was developed by the Principal Investigators at the Center for Public Opinion at UMass-Lowell. The SSRS project team provided feedback. 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.

After the first night of interviewing, bases for selected questions were modified from all registered voters to a random half sample of registered voters in order to meet a maximum average length of interview.

Massachusetts adults who were not registered to vote were asked basic demographic questions for weighting purposes.

CONTACT PROCEDURES

For respondent selection for landline samples, interviews were conducted with the youngest adult male/female, age 18 or older, 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.

DATA COLLECTION, PROCESSING, AND INTEGRATION

Survey Administration

The field period for this study was October 1-7, 2018. Interviews were completed in English or Spanish 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 or names, potential obstacles to overcome in order to get good answers to questions, and respondent problems that could be anticipated ahead of time, as well as strategies for addressing the other 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.
- 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.
- 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

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 overlapping dual-frame 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. Weighting involved several stages.

Base weight

The first step of the weighting was to compute a base weight. A base weight adjustment was made for probability of selection based on:

a. A phone number's probability of being sampled from the landline (PN_{LL}) or cell phone (PN_{CELL}) sampling frame. This is calculated as:

$$PN_{LL} = {}^{S_{LL}}/_{F_{LL}}$$
$$PN_{CELL} = {}^{S_{CELL}}/_{F_{CELL}}$$

Where:
S_{LL} = the size of the landline sample
F_{LL} = the size of the landline sample frame
S_{CELL} = the size of the cell sample
F_{CELL} = the size of the cell sample frame

b. The probability that a respondent will be selected when reached on a landline is based on the number of adults in their household. This probability is ${}^{1}/_{AD}$ where AD is the number of adults in the household, capped at 3.

c. Each respondent's access to a landline (LL) and/or cell phone (CP). Respondents reachable by landline were assigned LL=1 and those who could not be reached by landline were assigned LL=0. Similarly, respondents who had a cell phone were assigned CP=1 and those without a cell phone were assigned CP=0.

A respondent's probability (PR) of being reached by each phone type was calculated as:

$$\text{Landline: } PR_{LL} = {LL \times PN_{LL}}/{AD}$$
$$\text{Cell Phone: } PR_{CELL} = CP \times PN_{CELL}$$

And the overall probability of selection from either was calculated as:

$$P_{(LL \cup CELL)} = PR_{LL} + PR_{CELL} - (PR_{LL} \times PR_{CELL})$$

The base weight adjustment (BW) for probability of selection was calculated as:

$$BW = {1}/{P_{(LL \cup CELL)}}$$

Post-stratification

Post-stratification weighting (Raking): With the base weight applied, the sample was weighted by form split to reflect the distribution of the state adult population along known population parameters. Post-stratification was done using Iterative Proportional Fitting (or 'raking'), a procedure in which the data are repeatedly weighted to the parameters until the difference between the weighted data and the population benchmarks is near zero.

The combined landline and cell phone sample was weighted to balance the sample demographics to match estimates for the state population using data from the Census Bureau's 2018 Current Population Survey on sex by age, sex by education, age by education, race, and state region. The sample was also weighted to match current patterns of telephone use using state-level data from the 2012-2016 National Health Interview Survey.[1]

[1] Wireless substitution: Early release of state-level estimates from the National Health Interview Survey, 2012-2016. National Center for Health Statistics. December 2017. Available from: https://www.cdc.gov/nchs/data/nhis/earlyrelease/Wireless_state_201712.pdf

Table 2: Raking Dimensions

Dimension	Value Label
1: Gender by Age	Male, 18-24 Male, 25-34 Male, 35-44 Male, 45-54 Male, 55-64 Male, 65+ Female, 18-24 Female, 25-34 Female, 35-44 Female, 45-54 Female, 55-64 Female, 65+
2: Gender by Education	Male, HS graduate or less Male, Some college/Associate's degree Male, College graduate or more Female, HS graduate or less Female, Some college/Associate's degree Female, College graduate or more
3: Age by Education	18-34, HS graduate or less 18-34, Some college/Associate's degree 18-34, College graduate or more 35-54, HS graduate or less 35-54, Some college/Associate's degree 35-54, College graduate or more 55+, HS graduate or less 55+, Some college/Associate's degree 55+, College graduate or more
4: Race/Ethnicity	White, non-Hispanic Black, non-Hispanic Hispanic Other/Mixed, non-Hispanic
5: State Region	Middlesex County Suffolk County Essex County West East
6: Household Telephone Service	Cell phone only Dual Landline only

Finally, an adjustment was made to control variance among weights ('trimming') by truncating weights at the 2nd and 98th percentiles.

Table 3: Population Parameters and Weighted and Unweighted Total Sample Distributions

Characteristic	Value Label	Parameter	Unweighted	Weighted
Gender	Male	47.7%	52.4%	48.6%
	Female	52.3%	47.6%	51.4%
Age	18-24	11.9%	7.6%	11.7%
	25-34	19.1%	9.8%	18.5%
	35-44	14.9%	10.8%	15.2%
	45-54	17.6%	16.4%	17.7%
	55-64	18.0%	20.3%	18.2%
	65+	18.5%	35.2%	18.7%
Education	HS grad or less	35.0%	23.1%	34.7%
	Some college/Assoc.	22.5%	26.1%	22.9%
	College grad or more	42.5%	50.8%	42.4%
Race/Ethnicity	White, non-Hispanic	73.1%	79.6%	73.5%
	Black, non-Hispanic	6.8%	4.5%	6.7%
	Hispanic	11.0%	6.7%	10.9%
	Other/Mixed, non-Hispanic	9.1%	9.1%	8.8%
State Region	Middlesex County	23.4%	22.8%	23.7%
	Suffolk County	11.9%	10.2%	11.7%
	Essex County	11.3%	12.6%	11.4%
	West	24.0%	23.3%	24.0%
	East	29.4%	31.1%	29.2%
Phone use	Cell phone only	40.2%	28.9%	39.7%
	Dual use	52.6%	67.6%	53.2%
	Landline only	7.1%	3.5%	7.1%

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 1.55 overall and 1.58 based on registered voters.

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 ± 3.9 percentage points, and the margin of error for registered voters is ± 4.4 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 3.9 percentage points away from their true values in the population and estimated proportions based on registered voters will be no more than 4.4 percentage points away from their true values in the population. Margins of error for subgroups will be larger. 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 4 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 4.3 percent.
- The response rate for the RDD cellular samples was 4.8 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 4: Sample Disposition

RDD LL	RDD CELL	TOTAL	
1,565	347	1,912	Non-residential/Business
1	16	17	Cell in landline frame / LL in Cell frame
3	9	12	Duplicate number
1,569	372	1,941	OF = Out of Frame
8,278	1,931	10,209	Not working
745	48	793	Computer/fax/modem
9,023	1,979	11,002	NWC = Not working/computer
3,999	2,578	6,577	NA/Busy all attempts
3,999	2,578	6,577	$UHUO_{NC}$ = Non-contact, unknown if household/unknown other
2,912	6,689	9,601	Voice mail
2,043	752	2,795	Privacy Manager
33	36	69	Other non-contact
4,988	7,477	12,465	UO_{NC} = Non-contact, unknown eligibility
1,450	2,543	3,993	Refusals
1,855	3,336	5,191	Callbacks
3,305	5,879	9,184	UO_R = Refusal, unknown if eligible
59	138	197	O = Other (language)
0	137	137	Child's cell phone
26	155	181	Other ineligible
26	292	318	SO = Screen out (ineligible)
47	46	93	R = Refusal, known eligible (breakoffs and qualified CBs)
433	519	952	I = Completed interviews
23,449	19,280	42,729	T = Total numbers sampled

continued…

Table 4: Sample Disposition (continued…)

RDD LL	RDD CELL	TOTAL	
45.5%	85.9%	64.2%	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
94.9%	65.9%	76.7%	e2 = (I+R)/(I+R+SO) - Est. screening eligibility of unscreened contacts
36.2%	41.5%	39.2%	CON = [I + R + (e2*[O + UO_R])]/[I + R + (e2*[O + UO_R + UO_{NC}]) + (e1*e2*$UHUO_{NC}$)]
11.8%	11.5%	11.6%	COOP = I/[I + R + (e2*[O + UO_R])]
4.3%	**4.8%**	**4.5%**	**AAPOR RR3=I/[I+R+[e2*(UO_R+UO_{NC}+O)]+[e1*e2*$UHUO_{NC}$]] = CON*COOP**

DELIVERABLES

SSRS delivered to the Center for Public Opinion at UMass-Lowell:

- Final questionnaire instrument;
- Weighted dataset in SPSS;
- 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 rigor. 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.