METHODOLOGY

AP VoteCast - 2020 Democratic Primaries

OCTOBER 2020

PRESENTED BY:
NORC at the
University of Chicago

NORC
at the UNIVERSITY *of* CHICAGO

Table of Contents

Study Methodology

AP VoteCast is a survey of the American electorate conducted by NORC at the University of Chicago for The Associated Press and Fox News. The survey is funded by AP. In 2020, AP VoteCast covered the Democratic Presidential Primary Elections in 17 states across six election dates. For all election dates, interviews were conducted via phone and web. The surveys concluded as polls closed on Election Day.

- The Iowa caucus on February 3: This survey was conducted January 28 to February 3, 2020.
- The New Hampshire primary on February 11: This survey was conducted February 5 to February 11, 2020.
- The South Carolina primary on February 29: This survey was conducted February 23 to February 29, 2020.
- The Alabama, California, Colorado, Massachusetts, Minnesota, North Carolina, Texas, and Virginia primaries on March 3: These surveys were conducted February 26 to March 3, 2020.
- The Michigan, Missouri, and Mississippi primaries on March 10: These surveys were conducted March 4 to March 10, 2020.
- The Arizona, Florida, and Illinois primaries on March 17: These surveys were conducted March 11 to March 17, 2020.

Respondents who completed the survey during the first four days of the field period were asked if they would be willing to be recontacted for a follow-up survey. Those who consented were contacted duirn the final three days of the field period and reasked their intent to vote and vote choice. The vote intent and vote choice provided on the re-contact survey was used for the final estimates. Respondents were contacted using their preferred mode with a call from a NORC interviewer or a text or email with a link to complete the short recontact survey online.

AP VoteCast combines interviews with a random sample of registered voters drawn from state voter files with self-identified registered voters selected from nonprobability online panels. Interviews for probability and nonprobability sample were conducted in English and Spanish. Respondents received a small monetary incentive for completing the survey. Participants selected from state voter files were contacted by phone and mail, and had the opportunity to take the survey by phone or online.

VoteCast interviewed both voters and non-voters in the Democratic primaries. Eligibility to complete the survey was based on the type of caucus or primary held in the state. There were three types of caucuses/primaries:

- Open: In open states[1], all registered voters were eligible. Those who said they intended to vote in the Democratic primary completed the full survey. Those who said they intended to vote in the Republican primary/caucus or neither completed a set of demographic questions for weighting purposes. The following states had open primaries: Alabama, Illinois, Michigan, Minnesota, Mississippi, Missouri, South Carolina, Texas, Virginia. Iowa's caucus was also open.

- Semi-closed: In semi-closed states, respondents were first asked if they are registered as a Democrat, Republican, in another party, or if they are not affiliated with a party. Those who said they were not affiliated with a party were then asked if they planned to vote in the Democratic primary, Republican primary, or neither. Those registered as Democrats or those who said they planned to vote in the Democratic primary were eligible and completed the full survey. Unaffiliated respondents who said they would vote in the Republican primary or neither were also eligible to complete and completed a set of demographic questions for weighting purposes. The following states had semi-closed primaries: California, Colorado, Massachusetts, New Hampshire, North Carolina.

- Closed: In closed states, respondents were asked if they are registered as a Democrat, Republican, in another party, or if they are not affiliated with a party. Only those registered as Democrats were eligible to complete and completed the full survey. The following states had closed primaries: Arizona, Florida.

The table below shows the total completed probability interviews among those registered voters classified as eligible based on the type of primary or caucus.

State	Total Probability Interviews	Probability Web interviews	Probability Phone interviews	AAPOR Response Rate 3
Alabama	1,743	1,366	377	2.3
Arizona	2,088	1,859	229	4.2
California	2,884	2,655	229	2.5
Colorado	3,392	3,063	329	4.0
Florida	1,947	1,733	214	3.2
Iowa	4,340	3,695	645	5.9
Illinois	2,310	2,039	271	2.9
Massachusetts	3,304	2,999	305	3.8
Michigan	2,164	1,959	205	3.3
Minnesota	3,235	2,941	294	5.5
Missouri	2,168	1,905	263	3.7
Mississippi	1,282	963	319	1.8

---

[1] For states that cancelled the Republican primary, respondents were asked whether they would have voted in the Republican primary had it not been cancelled. If they said they would have, they were not categorized as a Democratic voter or nonvoter.

North Carolina	2,408	2,087	321	2.9
New Hampshire	3,462	3,133	329	4.3
South Carolina	2,219	1,863	356	2.8
Texas	1,834	1,482	352	2.4
Virginia	2,732	2,495	237	3.8

This dataset includes Democratic Primary voters and nonvoters who were asked the full survey and excludes those who were only asked demographic questions for weighting purposes. Among these voters and nonvoters:

In 9 states, VoteCast is based on between 1,000 and 3,700 probability-based interviews conducted online and via phone. In an additional 8 states, VoteCast is based on between 1,000 and 3,300 probability-based interviews conducted online and via phone, and between 600 and 2,200 nonprobability interviews conducted online.

For Super Tuesday estimates, the results of the re-contact survey were weighted to represent all early respondents, and then were combined with the results of later responders who completed the initial interview within the final two days of the election. In Minnesota, only responders who completed after Senator Amy Klobuchar dropped out on Monday morning or those who told us earlier they had already voted are included in the final sample. Voters were defined using a likely voter model for each state. Appendix A includes the likely voter model definition for each Democratic Primary contest. Due to quality control checks, 0.10% of respondents were removed from the final sample of completed interviews.

State	Interviews with Democratic Primary Voters and Non-Voters	Probability Interviews	Non-Probability Interviews	Phone interviews	Web interviews
Alabama	1,281	1,281	-	260	1,021
Arizona	2,088	2,088	-	230	1,858
California	4,398	2,427	1,971	186	4,212
Colorado	3,083	3,083	-	286	2,797
Florida	3,904	1,947	1,957	215	3,689
Iowa	3,653	3,653	-	617	3,036
Illinois	3,048	2,206	842	257	2,791
Massachusetts	3,360	2,674	686	221	3,139
Michigan	2,687	2,049	638	194	2,493

Minnesota	1,460	1,460	-	112	1,348
Missouri	2,042	2,042	-	245	1,797
Mississippi	1,175	1,175	-	303	872
North Carolina	2,968	2,045	923	259	2,709
New Hampshire	3,237	3,237	-	297	2,940
South Carolina	1,884	1,884	-	312	1,572
Texas	3,795	1,605	2,190	321	3,474
Virginia	3,069	2,293	776	189	2,880

State	Number of Interviews with Democratic Primary Voters	Margin of Sampling Error for Voters (+/- pp)	Number of Interviews with Democratic Primary Non-Voters	Margin of Sampling Error for Non-Voters (+/- pp)
Alabama	1,193	4.2	88	16.4
Arizona	2,015	3.3	73	17.1
California	4,023	2.3	375	9.6
Colorado	3,006	3.0	77	16.1
Florida	3,412	2.2	492	6.7
Iowa	3,036	3.0	617	5.7
Illinois	2,738	2.7	310	8.3
Massachusetts	3,085	3.1	275	12.1
Michigan	2,460	2.8	227	9.2
Minnesota	1,337	4.7	123	13.0
Missouri	1,891	3.5	151	11.7
Mississippi	1,091	4.1	84	14.0
North Carolina	2,706	2.8	262	8.9
New Hampshire	3,111	3.0	126	13.6
South Carolina	1,499	3.7	385	7.4
Texas	3,268	2.4	527	6.6
Virginia	2,604	2.7	465	7.8

Although there is no statistically agreed upon approach for calculating margins of error for nonprobability samples, these margins of error are estimated using a calculation called the root mean squared error. Unlike a standard error, a root mean square error does not easily translate into a traditional statistical confidence interval for the vote count for an individual candidate or the vote differential between two candidates. However, a mean square error is a measure of uncertainty that incorporates the variability associated with the poll estimates, as well as the bias associated with the estimates derived from a nonprobabilty sample.

As with all surveys, VoteCast is subject to multiple sources of error, including from sampling, question wording and order, and nonresponse.

Sampling Details

Probability-based Registered Voter Sample

In each of the 17 states in which VoteCast includes a probability-based sample, NORC obtained a sample of registered voters from Catalist LLC's registered voter database. This database includes listed or modeled demographic information, as well as addresses for registered voters. Additionally some registered voters also have one or more telephone numbers. This allows all registered voters to be contacted either by mail alone or by mail and telephone. Noncoverage is limited to registered voters who register to vote immediately prior to the election (e.g., states with same day registration laws) and are therefore not on the voter rolls at the time of sampling.

The sample was stratified by state, partisanship (either party affiliation on the voter file or Catalist partisanship model), past primary vote participation, age, and race. In addition, NORC attempted to match sampled records to a registered voter database maintained by L2, which provided additional phone numbers. After the matching, NORC had phone numbers for 79 percent of sampled records, including cell phone numbers for 84 percent of records with a phone number. Prior to dialing, all probability sample records are mailed a postcard inviting them to complete the survey either online using a unique PIN or via telephone by calling a toll-free number. Postcards are addressed by name to the sampled registered voter if that individual is under age 35; postcards are addressed to "registered voter" in all other cases. Telephone interviews are conducted with the adult that answers the phone. Both online and telephone respondents provided confirmation of registered voter status in the state.

Nonprobability Sample

Nonprobability participants were provided by Dynata or Lucid, including members of its third-party panels. The non-probability sample providers used routers as well as direct invitations for recruitment. Among the 12,174 panelists who touched the pre-screener instrument, 9,983 went on to complete the full survey. A response rate cannot be calculated for nonprobability sample. There were no firm demographic quotas, however in CA, FL, and TX, targets were provided for gender, age, and race. Digital fingerprint software and panel-level ID validation is used to prevent respondents from completing the VoteCast survey multiple times. Nonprobability respondents provided self-reported confirmation of registered voter status in the state. Interviews were conducted in English and Spanish.

Weighting Details

VoteCast employs a four-step weighting approach that combines the probability sample with the nonprobability sample, and refines estimates at a subregional level within each state. The 17 state surveys are weighted separately.

First, weights are constructed separately for the probability sample and the nonprobability sample (when available) for each state survey. These weights are adjusted to population totals to correct for demographic imbalances of the responding sample compared to the population of registered voters eligible to vote in the primary contest. In open states, all registered voters are eligible to participate while in semi-closed or closed states eligibility depends on party registration. The adjustment targets are derived from a combination of data from the U.S. Census Bureau's November 2018 Current Population Survey Voting and Registration Supplement, Catalist's voter file, the 2018 AP VoteCast Survey, and the Census Bureau's 2019 American Community Survey. The variables used were tailored to each state based on whether the state was conducting an open, semi-closed, or closed primary. Appendix B includes further details on the variables used in each state.

Prior to adjusting to population totals, the probability-based registered voter list sample weights are adjusted for probabilities of selection and differential non-response related to factors such as availability of phone numbers, age, race, and partisanship.

Second, all non-probability sample respondents receive a calibration weight. The calibration weight is designed to ensure the non-probability sample is similar to a probability sample in regard to variables that are predictive of vote choice that cannot be fully captured through the prior demographic adjustments. The calibration benchmarks are based on estimates derived from the probability sample. The three variables used for calibration were political ideology, vote in the 2016 Democratic presidential primary, and confidence about the fairness of the 2020 Democratic primary process. For each of these three variables, a benchmark was calculated based on the probability estimate and the non-probability sample was raked to these benchmarks along with the demographic characteristics list below.

	CA	MA	NC	VA	TX	MS	FL	IL
Political ideology – (Very/somewhat liberal, Closer to liberal/neither/closer to conservative/somewhat/very conservative)	x	x	x	x	x	x	x	x
Vote in the 2016 Democratic presidential primary – (Hillary Clinton, Bernie Sanders, Someone else/nonvoter)	x	x	x	x	x	x	x	x

Confidence in the fairness of the 2020 Democratic primary process (Confident, Not confident)	x	x	x	x	x	x	x	x
Age*sex - (18-34 , 35-44, 45-54, 55-64, 65+) * (male, female)	x	x	x	x	x	x	x	x
Race*age - (white, Hispanic, black, other) * (18-34, 35 plus)	x	x	x	x	x	x	x	x
Neighborhood education - Census Block Group divided into four education categories based on % with college degree in gen pop	x	x	x	x	x	x		
Education - (HS grad or less, some college, 4-year college grad, post-graduate)	x	x	x	x	x	x	x	x
Party registration – Democrat, unaffiliated	x	x	x					
Age*sex within registered Democrats ((18-34, 35-54, 55+) * (male, female))	x	x	x					
Race within registered Democrats (NH-White, NH-Black, Hispanic, NH-All Other)	x	x	x					
Age*race within registered Democrats (18-34, 35-54, 55+) * (NH-White, All Other)	x	x	x					
Self-reported party*education (Democrat * (HS grad or less, some college, 4-year college grad, post-graduate))				x	x	x		x
Self-reported party * age ((Democrat, Independent) * (18-34, 35-54, 55+))				x	x	x		x
Self-reported Democrat*race (Democrat * (white, black, other))				x		x		
Self-reported Democrat*race (Democrat * (white, black, Hispanic, other))					x			x

Self-reported party*gender (Democrat *Gender(male, female), Other Party)						x		

Third, all respondents in each state are weighted to improve estimates for substate geographic regions. This weight combines the weighted probability sample and the calibrated non-probability sample, and then uses a small area model to improve the estimate within subregions of a state. We created between 6and 19 regions (county groupings) for each state based on vote choice in previous Democratic primary election results in each county. We then used these groupings to generate model-based estimates of vote choice among likely voters.

For each state, there was a separate small area model predicting percent of vote share for each main candidate in the race. The combined sample weight was ratio adjusted to agree with the small area estimates by region (within state) and vote choice (for the main candidates in a given state); this ratio adjustment only applied to likely voters with known vote choice. For all other respondents, they retain their original combined sample weight. The following variables were used as potential covariates in the models: 2016 Democratic Presidential primary election results, 2018 Democratic Senate/Governor primary election results, Catalist partisanship score, Catalist ideology score, population density, median income, percent below poverty line, percent unemployed, percent college degree, portion on public assistance, percent insurance coverage, percent nonwhite, percent citizen, percent 18-34 years old, percent 65 and older, and percent who have not moved in last year. For each state, we included in the models at least one: 1) a political measure such as a past election or Catalist ideology/partisanship score, 2) a measure of socioeconomic status, 3) at least one demographic or geographic measure.
Fourth, the survey results are weighted to the actual vote count following the completion of the election. This weighting is done in 6-19 sub-state regions within each state.

Contact

For more information, visit www.apnorc.org or email info@apnorc.org.

Using Weights

AP VoteCast is designed to be analyzed using weighted data. The data file includes different weights for different types of analyses. For this dataset, all analysis should be done at the state level (e.g. percent of California Democratic Primary Voters who voted for Joe Biden). The weights are not designed to be used for overall measures across states.

The data file includes weights that represent results at two different stages of data collection.

- The FINALVOTE weights should be used to produce estimates that are adjusted to reflect the final vote counts in addition to demographic, geographic, and calibration adjustments. Certified vote count data was provided by AP. AP VoteCast recommends using these weights for most analyses.
- The POLLCLOSE weights can be used to produce estimates prior to any adjustments to final vote counts. These weights are provided for transparency of the methodology to permit comparison of the survey's estimates at poll close but prior to adjusting the survey outcome to match the final vote count.

To reproduce estimates in each state, limit analysis to LIKELYVOTER=1, the state of interest (using either P_STATE or STATENUM) and cases that are not missing PRESVOTEPRIM. The FINALVOTE_STATE_WEIGHT variable should be used for weights.

About The Associated Press-NORC Center for Public Affairs Research

The AP-NORC Center for Public Affairs Research taps into the power of social science research and the highest quality journalism to bring key information to people across the nation and throughout the world.

- The Associated Press is an independent global news organization dedicated to factual reporting. Founded in 1846, AP today remains the most trusted source of fast, accurate, unbiased news in all formats and the essential provider of the technology and services vital to the news business. More than half the world's population sees AP journalism every day. Online: www.ap.org.
- NORC at the University of Chicago is one of the oldest and most respected, objective social science research institutions in the world.

The two organizations have established The AP-NORC Center to conduct, analyze, and distribute social science research in the public interest on newsworthy topics, and to use the power of journalism to tell the stories that research reveals.

Appendix A – Likely Voter Models

In Iowa, respondents are classified as voters based on the following criteria:

- The respondent says they will definitely vote to LVB and they are very or extremely interested in the election to LVA; or
- The respondent says they will probably or definitely vote to LVB, they are somewhat, very, or extremely interested in the election to LVA, they voted in the 2018 midterm election, and they voted in the Democratic primary in 2016.

In South Carolina, respondents are classified as voters based on the following criteria:

- The respondent says they will definitely vote to LVB, they are certain they will vote to LV, and they are very or extremely interested in the election to LVA; or
- The respondent says they will probably or definitely vote to LVB, they score an 8 or higher on likelihood to vote to LV, they are somewhat, very, or extremely interested in the election to LVA, and they voted in the Democratic primary in 2016; or
- The respondent says they already voted to LVB or LV.

In New Hampshire, Alabama, California, Colorado, Massachusetts, Minnesota, North Carolina, Texas, Virginia, Michigan, Missouri, Mississippi, Arizona, Florida, and Illinois, respondents are classified as voters based on the following criteria:

- The respondent says they will definitely vote to LVB and they are certain they will vote to LV; or
- The respondent says they will probably or definitely vote to LVB, they score an 8 or higher on likelihood to vote to LV, they voted in the 2018 midterm election, and they voted in the 2016 primary or caucus; or
- The respondent says they already voted to LVB or LV.

Appendix B – Weighting variables by state

Alabama – Open state without 2016 vote information or party registration: age*sex, race*age, education, neighborhood education, self-reported Democrat*race, self-reported party*education, self-reported party*age.

Arizona – Closed state with party registration and 2016 vote information: age*sex, race*age, neighborhood education, party registration*education (geographic level).

California – Semi-closed state with party registration but not 2016 vote information: age*sex, race*age, education, neighborhood education, party registration, party registration*education (geographic level), age*sex (within registered Democrats), race (within registered Democrats), age*race (within registered Democrats).

Colorado – Semi-closed state with party registration but not 2016 vote information: age*sex, race*age, education, neighborhood education, party registration, party registration*education (geographic level), age*sex (within registered Democrats), race (within registered Democrats), age*race (within registered Democrats).

Florida – Closed state with party registration but not 2016 vote information: age*sex, race*age, neighborhood education, party registration*education (geographic level).

Illinois – Open state with 2016 vote information but not party registration: age*sex, race*age, education, neighborhood education, self-reported Democrat*race, self-reported party*education, self-reported party*age.

Iowa – Open state with party registration: age*sex, race*age, education, neighborhood education, party registration, 2016 primary vote choice, party registration*education (geographic level), age*sex (within Democrats), race (within Democrats), age*race (within Democrats).

Massachusetts – Semi-closed state with party registration and 2016 vote information: age*sex, race*age, education, neighborhood education, party registration, party registration*education (geographic level), age*sex (within registered Democrats), race (within registered Democrats), age*race (within registered Democrats).

Michigan – Open state with 2016 vote information but not party registration: age*sex, race*age, education, neighborhood education, self-reported Democrat*race, self-reported party*education, self-reported party*age.

Minnesota – Open state without 2016 vote information or party registration: age*sex, race*age, education, neighborhood education, self-reported Democrat*race, self-reported party*education, self-reported party*age.

Mississippi – Open state with 2016 vote information but not party registration: age*sex, race*age, education, neighborhood education, self-reported Democrat*race, self-reported party*education, self-reported party*age.

Missouri – Open state without 2016 vote information or party registration: age*sex, race*age, education, neighborhood education, self-reported Democrat*race, self-reported party*education, self-reported party*age.

New Hampshire – Semi-closed state with party registration and 2016 vote information: age, sex, race, education, neighborhood education, party registration, 2016 primary vote choice, party registration*education (geographic level), sex (within Democrats), race (within Democrats).

North Carolina – Semi-closed state with party registration and 2016 vote information: age*sex, race*age, education, neighborhood education, party registration, party registration*education (geographic level), age*sex (within registered Democrats), race (within registered Democrats), age*race (within registered Democrats).

South Carolina – Open state with 2016 information but not party registration: age*sex, race*age, education, neighborhood education, self-reported Democrat*race, self-reported Democrat*education (geographic level).

Texas – Open state with 2016 vote information but not party registration: age*sex, race*age, education, neighborhood education, self-reported Democrat*race, self-reported party*education, self-reported party*age.

Virginia – Open state with 2016 vote information but not party registration: age*sex, race*age, education, neighborhood education, self-reported Democrat*race, self-reported party*education, self-reported party*age.