METHODOLOGY

AP VoteCast 2020

MAY 2021

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 Fox News, NPR, PBS NewsHour, Univision News, USA Today Network, The Wall Street Journal and The Associated Press. The survey is funded by AP. The survey of 133,103 registered voters was conducted October 26 to November 3, 2020, concluding as polls closed on Election Day. Interviews were conducted via phone and web, with 4,865 completing by phone and 128,238 completing by web.

AP VoteCast combines interviews with a random sample of registered voters drawn from state voter files; with self-identified registered voters conducted using NORC's probability-based AmeriSpeak® panel, which is designed to be representative of the U.S. population; and with self-identified registered voters selected from nonprobability online panels. Interviews were conducted in English and Spanish. Respondents may receive 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 National Survey

The VoteCast survey of voters and nonvoters nationwide is compiled from results of the 50 state-based surveys and a nationally representative survey of 4,141 registered voters conducted on the probability-based AmeriSpeak panel (3,909 completed online and 232 via phone). It includes 41,776 probability interviews completed online (37,143) and via telephone (4,633), and 87,186 nonprobability interviews completed online. The margin of sampling error[1] is plus or minus 0.4 percentage points for voters (n=110,484) and 0.9 percentage points for nonvoters (n=22,619) including the design effect. Registered voters in the District of Columbia were not included. The overall weighted response rate for the probability sample drawn from the state voter files was 3.4%.[2] The overall weighted response rate for the AmeriSpeak panel sample was 5.0%.

[1] Although there is no statistically agreed upon approach for calculating margins of error for nonprobability samples, these margins of error are estimated using a measure of uncertainty that incorporates the variability associated with the poll estimates, as well as the variability associated with the survey weights as a result of calibration. After calibration, the nonprobability sample yields approximately unbiased estimates. As with all surveys, VoteCast is subject to multiple sources of error, including from sampling, question wording and order, and nonresponse.

[2] The unweighted response rate for the probability sample was also 3.4%.

For those who screened into the survey as eligible based on being a registered voter in the state, the interview completion rates were 96.7% for the probability sample drawn from the state voter files, 98.3% for the AmeriSpeak sample, and 92.0% for the nonprobability sample. Due to quality control checks, 0.10% of respondents were removed from the final sample of completed interviews.

VoteCast State Surveys

In 20 states, VoteCast is based on roughly 1,000 probability-based interviews conducted online and via phone, and roughly 3,000 nonprobability interviews conducted online. In these states, the margin of sampling error including the design effect is estimated to be plus or minus 2.1 percentage points for voters and 4.6 percentage points for nonvoters.

In 20 additional states, VoteCast is based on roughly 500 probability-based interviews conducted online and via phone, and roughly 2,000 nonprobability interviews conducted online. In these states, the margin of sampling error including the design effect is estimated to be plus or minus 2.8 percentage points for voters and 5.6 percentage points for nonvoters.

In 10 additional states, VoteCast is based on between 200 and 1,000 nonprobability interviews conducted online. In these states, the margin of error is estimated to be plus or minus 6.6 percentage points for voters and 12.6 percentage points for nonvoters.

State	Probability Interviews	Non-Probability Interviews	Number of Voters	Margin of Sampling Error for Voters (+/- pp)	Number of Non-Voters	Margin of Sampling Error for Non-Voters (+/- pp)
National	45,917[3]	87,186	110,484	0.4	22,619	0.9
Alaska	623	121	689	6.1	55	17.4
Alabama	652	1,779	1,905	2.7	526	5.3
Arkansas	-	878	599	4.7	279	7.6
Arizona	1,340	3,260	3,772	2.1	828	4.2
California	1,361	3,162	3,683	2.0	840	4.0
Colorado	1,486	2,315	2,995	2.3	806	4.3
Connecticut	781	1,558	1,941	2.8	398	5.9
Delaware	706	490	1,055	4.3	141	11.8
Florida	1,372	3,118	3,698	2.0	792	4.3
Georgia	1,114	3,098	3,291	2.2	921	3.9
Hawaii	-	516	381	7.1	135	12.1
Iowa	1,179	1,521	2,401	2.6	299	7.0
Idaho	-	650	478	5.0	172	9.0
Illinois	715	2,829	2,878	2.2	666	4.6

[3] The national probability total includes both the AmeriSpeak sample and the probability state survey sample.

Indiana	724	2,215	2,367	2.4	572	4.9
Kansas	570	1,157	1,445	3.3	282	7.2
Kentucky	1,273	2,316	2,948	2.3	641	4.9
Louisiana	556	1,780	1,786	3.0	550	5.6
Massachusetts	823	2,127	2,543	2.4	407	5.7
Maryland	794	1,913	2,211	2.6	496	5.5
Maine	1,301	771	1,888	3.2	184	9.6
Michigan	1,231	3,295	3,571	2.0	955	3.9
Minnesota	1,426	2,636	3,613	2.1	449	5.7
Missouri	1,546	2,189	3,119	2.2	616	5.0
Mississippi	541	1,000	1,249	3.6	292	7.9
Montana	812	460	1,135	4.2	137	10.5
North Carolina	1,362	3,070	3,731	1.9	701	4.5
North Dakota	-	268	220	8.1	48	17.5
Nebraska	868	796	1,445	3.2	219	8.5
New Hampshire	1,623	737	2,230	3.1	130	11.7
New Jersey	750	2,278	2,477	2.4	551	4.8
New Mexico	930	909	1,654	3.6	185	9.9
Nevada	1,561	1,572	2,757	2.4	376	6.4
New York	673	2,165	2,222	2.8	616	4.9
Ohio	1,477	3,147	3,902	1.9	722	4.5
Oklahoma	-	965	673	4.5	292	6.9
Oregon	1,611	2,110	3,226	2.3	495	5.3
Pennsylvania	1,218	3,788	4,134	1.8	872	4.0
Rhode Island	-	463	374	6.2	89	13.0
South Carolina	1,137	2,481	2,905	2.4	713	4.8
South Dakota	-	301	238	7.4	63	14.3
Tennessee	768	2,010	2,252	2.5	526	5.2
Texas	1,317[4]	3,362	3,887	1.9	792	4.2
Utah	888	1,167	1,696	3.2	359	6.7
Virginia	727	2,012	2,288	2.6	451	5.5
Vermont	-	245	202	9.6	43	18.7
Washington	776	2,203	2,398	2.5	581	4.8
Wisconsin	1,164	2,935	3,506	2.0	593	5.0
West Virginia	-	847	635	4.5	212	8.1
Wyoming	-	201	154	9.5	47	17.3

Sampling Details

Probability-based Registered Voter Sample

[4] One case was removed from the Texas probability sample following the election at the participant's request.

In each of the 40 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 demographic information, as well as addresses and phone numbers for registered voters, allowing potential respondents to be contacted via mail and telephone. The sample was stratified by state, four-level partisanship category, and five-level predicted response propensity to the postcard category. In addition, NORC attempted to match sampled records to a registered voter database maintained by L2, which provided additional phone numbers and demographic information. After the matching, NORC had phone numbers for 81% of sampled records, including cell phone numbers for 69% of records with a phone number. Prior to dialing, all probability sample records were 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 "[STATE] Registered Voter" in all other cases. Not all sampled records with a telephone number were dialed; instead, dialing was reserved for sampled records in the two lowest predicted response propensity quintiles that had not already responded online. Telephone interviews were 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 and Lucid, including members of their third-party panels. NORC also sampled registered voters in select states who were not included in the probability sample for matching to email addresses. V12 provided the email matching service and emailed these participants a recruitment email to complete the survey online. Digital fingerprint software and panel-level ID validation is used to prevent respondents from completing the VoteCast survey multiple times. Nonprobability respondents provided confirmation of registered voter status in the state. A response rate cannot be calculated for nonprobability samples. While there is no way to quantify the size of the non-covered population for an opt-in panel, the primary population least likely to be included is those without internet access. Interviews were conducted in English and Spanish.

Dynata used router technology to recruit participants, and all available panelists in each state were recruited. Among the 61,951 panelists who touched the pre-screener instrument, 41,808 went on to complete the full survey. Panelists recruited for a specific state were only allowed to complete the survey if they were registered to vote in that state. Dynata's system requires respondents to enter a unique ID to enter the survey, and this built-in technology prevents respondents from completing the survey more than once.

Lucid's suppliers invited respondents to the survey using email invites and panelist recruitment. Before sending them into the survey, Lucid targeted and pre-screened respondents age 18 and older on the basis of state location with zip code and registered voters. Among the 60,340 panelists who touched the pre-screener instrument, 41,737 went on to complete the full survey. Respondents recruited for a specific state were only allowed to complete the survey if they were registered to vote in that state. Lucid assigned respondents a unique ID each time they take a survey, which locked them out if they attempted to complete the survey more than once.

AmeriSpeak Sample

During the initial recruitment phase of the AmeriSpeak panel, randomly selected U.S. households were sampled with a known, non-zero probability of selection from the NORC National Sample Frame and then contacted by U.S. mail, email, telephone and field interviewers (face-to-face). The panel provides sample coverage of approximately 97% of the U.S. household population. Those excluded from the sample include people with P.O. Box-only addresses, some addresses not listed in the USPS Delivery Sequence File and some newly constructed dwellings. AmeriSpeak panelists provided confirmation of registered voter status in the state.

A sample of registered voters was selected from the AmeriSpeak Panel using sampling strata based on age, race/Hispanic ethnicity, education, gender, and whether the panelist completed 2018 AP VoteCast (96 sampling strata in total). The size of the selected sample per sampling stratum was determined by the population distribution for each stratum. In addition, sample selection takes into account expected differential survey completion rate by demographic group so that the set of panel members with a completed interview is a representative sample of the target population of registered voters. If panel household has one more than one active adult panel member, only one adult in the household was eligible for selection (random within-household sampling).

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. For national estimates, the 50 state surveys and the AmeriSpeak survey are weighted separately and then combined into a survey representative of voters in all 50 states.

State Surveys

First, weights are constructed separately for the probability sample (when available) and the nonprobability sample 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 in each state. 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 and the Census Bureau's 2018 American Community Survey. The variables used were:

- Sex * Age (male, female * 18-29, 30-44, 45-64, 65+)
- Race/ethnicity (Hispanic, NH-White, NH-Black, All Other)
- Education (less than high school/high school grad, some college, 4-year college grad, post-graduate)
- Age * race/ethnicity (18-29, 30-44, 45-64, 65+ * NH-White, All Other)
- Education * race/ethnicity (less than HS/HS grad, some college, 4-year college grad+ * NH-White, All Other)
- Probability sample only: Partisanship model score category * Predicted response propensity quintile (strong Republican, lean Republican, lean Democrat, strong Democrat * 1st quintile, 2nd quintile, 3rd quintile, 4th quintile, 5th quintile).
- Probability sample only: Proportion of total completed interviews obtained via outbound dials * Predicted response propensity quintile (Proportion outbound completes * 1st quintile, 2nd quintile, 3rd quintile, 4th quintile, 5th quintile).
- County grouping using AP's party grouping (variable "AP_PARTY_REGION")
- Prior to adjusting to population totals, the probability-based registered voter list sample weights were adjusted for differential non-response by four-level partisanship model score category, five-level predicted response propensity category, and incentive amount offered.

Second, calibration variables were included in weighting for both the probability and non-probability samples 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 demographic adjustments.

- Calibration variables
    - Party ID (Democrat, Independent, Republican)
    - Country on Right/Wrong Track

The calibration benchmarks are based on county level estimates from a linear regression model that incorporates all probability and non-probability cases nationwide. A national level linear regression model was fitted using data from all states (both probability and non-probability samples) to make

predictions for registered voters at the state-level for Party ID (Democrat, Independent, Republican) and Country on Right/Wrong Track. These state-level predicted estimates are used as calibration benchmarks for all states. In states with probability sample, the probability and non-probability samples were separately adjusted to the state-level calibration benchmarks, and then the combined sample was adjusted to regional level benchmarks for the calibration variables. For Party ID, separate models were fitted for predicting the proportion of Democrats and proportion of Republicans. In addition, five separate models were fitted based on how the county voted in the 2016 Presidential election (i.e., based on % Trump vote for county/town). Models included the following individual level variables:

- Flag for interaction between sex (male, female), age (18-29, 30-44, 45-64, 65+) and race/ethnicity (non-Hispanic White, All Other)
- Flag for interaction between sex (male, female) and education (less than high school/high school grad, some college, 4-year college grad, post-graduate)
- Flag for race/ethnicity (non-Hispanic White, non-Hispanic Black, Hispanic, non-Hispanic Other)
- Flag for voting for Trump in 2016 Presidential election

Third, all respondents in each state are weighted to improve estimates for substate geographic regions. This weight combines the weighted probability sample (if available) and the non-probability sample, and then uses a small area model to improve the estimate within subregions of a state. We created between 3 and 35 regions (county groupings) for each state based on vote choice in previous elections and the number of expected survey completes in each county. We then used these groupings to generate model-based estimates of vote choice among likely voters. The small domain model was applied to the Presidential election.

There were two models: 1) predicting percent of vote share that goes for either of the two major parties' candidates, 2) predicting percent of major party vote share that goes for the Democratic/Republican candidate. The following variables were used as potential covariates in the model: 2012 Presidential election results, 2016 Presidential election results, 2018 House election results, population density, median income, percent below poverty line, percent non-Hispanic White below poverty line, percent unemployed, percent college degree, percent non-Hispanic White with college degree, portion on public assistance, percent insurance coverage, percent non-Hispanic White with insurance coverage, percent non-Hispanic White, percent non-Hispanic Black, percent Hispanic, percent citizen, percent 18-29 years old, percent 30-44 years old, percent 65 and older, percent non-Hispanic White 18-29 years old, percent non-Hispanic White 30-44 years old, percent non-Hispanic White 65 and older, percent in rural area, and percent who have not moved in last year. We included in the models at least one variable from each of the

following sets of variables: 1) past vote choice, 2) measure of socioeconomic status, 3) demographic or geographic measure.

Fourth, the survey results are weighted to the actual vote count following the completion of the election. The presidential vote results and, when available, the senate or governor vote results were used as benchmarks for weighting respondents who were voters. This weighting is done in 3-35 sub-state regions within each state.

National Survey

The national survey is weighted to combine the 50 state surveys with the nationwide AmeriSpeak survey. Each of the state surveys is weighted as described. The AmeriSpeak survey receives a nonresponse-adjusted weight that is then adjusted to national totals for registered voters derived from the U.S. Census Bureau's November 2018 Current Population Survey Voting and Registration Supplement, the Catalist voter file and the Census Bureau's 2018 American Community Survey. The state surveys are further adjusted to represent their appropriate proportion of the registered voter population for the country and combined with the AmeriSpeak survey. After all votes are counted, the national data file is adjusted to match the vote for President 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.

- To produce estimates at the state level (e.g., percent of Californians who approve of President Trump), the state weights should be used.
- To produce estimates at the national level (e.g., the percent of registered voters nationwide who think the country is on the right track), the national-level weights should be used.

Additionally, 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 using all interviews collected through poll close, but prior to adjusting the survey outcome to match the final vote count.

To reproduce estimates in AP's publically-available VoteCast crosstabs of voters and estimates of voter demographics nationwide, limit analysis to LIKELYVOTER=1 and cases where vote choice in the race of interest (PRESVOTE or HOUSEVOTE) is not missing. The FINALVOTE_NATIONAL_WEIGHT variable should be used for weights.

To reproduce estimates in each state, limit analysis to LIKELYVOTER=1, the state of interest (using either P_STATE or STATENUM), and cases where vote choice in the race of interest (PRESVOTE, SENVOTE, SENSPVOTE, GOVVOTE, HOUSEVOTE) is not missing. 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 (AP) 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. Online: www.norc.org

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.

Learn more at www.apnorc.org

Appendix A – Likely Voter Models

In the following states – Alaska, California, Illinois, Iowa, Kansas, Kentucky, Massachusetts, Maryland, Minnesota, Mississippi, North Carolina, North Dakota, New Jersey, Nevada, New York, Ohio, Oregon, Texas, Utah, Virginia, Washington, West Virginia – ballots could be received after Election Day as long as it was postmarked by Election Day or the day before. Respondents in these states are classified as voters based on the following criteria:

- If the respondent says they voted by mail to WVA or WVB and:
    - The respondent says they will definitely vote to LVB, and they are certain they will vote to LV, and they are very or extremely interested in the election to LVA, and they voted in the 2018 midterm election or they voted in the 2016 presidential election; or
- If the respondent voted in person before Election Day or on Election Day to WVA or WVB and:
    - 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, and they score an 8 or higher on likelihood to vote to LV, and they voted in the 2018 midterm election or they voted in the 2016 presidential election; or
- The respondent says they already voted to LVB or LV.

The following states – Alabama, Arizona, Arkansas, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Louisiana, Maine, Michigan, Missouri, Montana, Nebraska, New Hampshire, New Mexico, Oklahoma, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Vermont, Wisconsin, Wyoming – required ballots to be received by Election Day or the day before. Those who told us they would vote early but had not yet when interviewed on October 31 or later were classified as nonvoters. Respondents in these states are classified as voters based on the following criteria:

- If the respondent says they already voted to LVB or LV and:
    - The respondent says they voted by mail before Election Day to WVA; or
    - The respondent says they voted in person before or on Election Day to WVA; or
    - If the respondent says they voted by mail on Election Day to WVA and they live in a state where vote by mail can be received after Election Day as long as it was postmarked by Election Day or the day before.
- If the respondent says they definitely will vote to LVB and they are certain they will vote to LV and:

  - The respondent says they will vote in person before Election Day to WVB; or
  - The respondent says they will vote in person at a polling place on Election Day to WVB; or
  - The respondent says they will vote by mail to WVB and they live in a state where vote by mail can be received after Election Day as long as it was postmarked by Election Day or the day before; or
  - The respondent lives in a state that requires vote by mail to be received by Election Day or the day before and says they will vote by mail to WVB and the interview was conducted between October 26 and October 30.
- The respondent says they will probably or definitely vote to LVB, and they score an 8 or higher on likelihood to vote to LV, and they voted in the 2018 midterm election or they voted in the 2020 presidential election and:
  - The respondent says they will vote in person before Election Day to WVB; or
  - The respondent says they will vote in person at a polling place on Election Day to WVB; or
  - The respondent says they will vote by mail to WVB and they live in a state where vote by mail can be received after Election Day as long as it was postmarked by Election Day or the day before; or
  - The respondent lives in a state that requires vote by mail to be received by Election Day or the day before and says they will vote by mail to WVB and the interview was conducted between October 26 and October 30.