METHODOLOGY

AP VoteCast 2022

FEBRUARY 2023

PRESENTED BY:
NORC
at the
University of Chicago

✳NORC at the University of Chicago

Table of Contents

Study Methodology

AP VoteCast was a survey of the American electorate conducted by NORC at the University of Chicago for The Associated Press and Fox News. The survey was funded by AP. The survey of 120,896 registered voters was conducted October 31 to November 8, 2022, concluding as polls closed on Election Day. Interviews were conducted via phone and web, with 4,506 completing by phone and 116,390 completing by web.

AP VoteCast combined interviews from a random sample of registered voters drawn from state voter files; with self-identified registered voters randomly sampled from 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 panels. Interviews were conducted in English and Spanish. Telephone interviews were conducted using live interviewers. Respondents may have received a small monetary incentive up to $10 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. AmeriSpeak participants could be contacted by email, mail, or phone, and had the opportunity to take the survey by phone or online. Participants selected from the nonprobability panels took the survey online.

VoteCast National Survey

The VoteCast survey of voters and nonvoters nationwide was compiled from results of 50 state-based surveys and a nationally representative survey of 4,276 registered voters conducted on the probability-based AmeriSpeak panel (4,169 completed online and 107 via phone). The state surveys included 26,856 probability interviews from state voter files completed online (22,457) and via telephone (4,399), and 89,764 nonprobability interviews completed online. The margin of sampling error[1] was plus or minus 0.5 percentage points for voters (n=94,296) and 1.4 percentage points for nonvoters (n=26,600), including the design effect. Registered voters in the District of Columbia were not included. The overall weighted

---

[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.

response rate for the probability sample drawn from the state voter files was 2.7%.[2] The overall weighted response rate for the AmeriSpeak panel sample was 9.7%.

For those who screened into the survey as eligible based on being a registered voter in the state, the interview completion rates were 96.3% for the probability sample drawn from the state voter files, 98.1% for the AmeriSpeak sample, and 89.8% for the nonprobability sample. Due to quality control checks, fewer than 1% of respondents were removed from the final sample of completed interviews prior to weighting.

VoteCast State Surveys

The target number of completed interviews varied by state. In some states that did not have a statewide election in 2022 (Delaware and West Virginia), interviews contributed to national estimates only.

State	Probability Interviews	Nonprobability Interviews	Number of Voters	Margin of Sampling Error for Voters (+/- pp)	Number of Non-Voters	Margin of Sampling Error for Non-Voters (+/- pp)
National	31,132[3]	89,764	94,296	0.5	26,600	1.4
Alaska	817	328	1,037	4.8	108	13.0
Alabama	438	2,110	1,905	3.5	643	5.6
Arkansas	721	2,046	2,003	2.6	764	4.6
Arizona	1,019	3,464	3,230	2.4	1,253	3.6
California	668	3,437	3,396	1.9	709	4.6
Colorado	676	3,377	2,728	2.2	1,325	3.4
Connecticut	340	2,210	1,887	2.7	663	5.0
Delaware	-	305	246	-	59	-
Florida	1,176	3,121	3,379	2.0	918	3.9
Georgia	851	3,362	3,234	2.1	979	3.9
Hawaii	675	500	881	5.3	294	7.7
Iowa	729	2,014	2,064	2.6	679	5.4
Idaho	523	861	1,049	3.3	335	7.8
Illinois	337	2,160	2,071	2.6	426	6.1
Indiana	367	2,152	1,801	2.8	718	4.8
Kansas	598	1,734	1,773	2.7	559	6.5
Kentucky	414	2,027	1,770	3.0	671	5.5
Louisiana	458	2,046	1,816	2.9	688	5.1
Massachusetts	293	2,246	2,076	2.5	463	5.8
Maryland	354	2,083	1,926	3.4	511	5.9
Maine	785	690	1,265	3.4	210	9.4
Michigan	941	3,125	3,244	2.1	822	4.3

---

[2] The unweighted response rate for the probability sample was 2.4%.

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

Minnesota	697	2,933	2,904	2.1	726	5.0
Missouri	458	2,014	1,845	3.2	627	5.1
Mississippi	562	324	772	4.8	114	13.3
Montana	564	323	749	4.3	138	11.6
North Carolina	762	3,318	3,175	2.0	905	4.1
North Dakota	411	303	622	4.7	92	14.1
Nebraska	-	1,061	700	4.2	361	6.4
New Hampshire	640	823	1,248	3.6	215	10.3
New Jersey	309	318	544	5.3	83	13.9
New Mexico	309	1,171	1,042	4.0	438	6.2
Nevada	633	2,036	2,190	2.7	479	5.8
New York	674	3,279	3,156	2.7	797	4.7
Ohio	764	3,265	3,152	2.7	877	4.7
Oklahoma	-	1,039	697	5.1	342	7.7
Oregon	266	2,356	2,138	3.1	484	5.8
Pennsylvania	771	3,188	3,179	2.5	780	5.0
Rhode Island	271	515	619	5.5	167	11.8
South Carolina	947	2,758	2,830	2.3	875	4.5
South Dakota	346	348	597	5.5	97	12.2
Tennessee	-	1,054	741	4.1	313	6.7
Texas	1,167	3,201	3,420	2.0	948	3.8
Utah	650	1,803	1,904	2.9	549	5.5
Virginia	351	328	602	5.0	77	15.4
Vermont	510	323	682	5.0	151	10.7
Washington	326	2,324	2,169	2.4	481	5.5
Wisconsin	832	3,418	3,290	2.0	960	4.6
West Virginia	-	258	198	-	60	-
Wyoming	456	285	638	5.4	103	14.1

Sampling Details

Probability-based Registered Voter Sample

In each of the 45 states in which VoteCast included a probability-based sample, NORC obtained a sample of registered voters from Catalist LLC's registered voter database. This database included 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, a four-level partisanship variable, and a five-level predicted response propensity variable. In states with a large non-white voter population, race and ethnicity was also a stratifying factor. 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 91% of sampled records, including cell phone numbers for 91% 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 were addressed by name to

the sampled registered voter if that individual was under age 35; postcards were addressed to "[STATE] Registered Voter" in all other cases. Additional outbound dialing was conducted for sampled records in the two lowest predicted response propensity quintiles who had not already responded online. Telephone interviews were conducted with the adult that answered the phone. Both online and telephone respondents provided confirmation of registered voter status in the state.

Nonprobability Sample

Nonprobability participants were provided by Dynata, Cint, and Prodege, including members of their third-party panels. Digital fingerprint software and panel-level ID validation was 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 was those without internet access. Interviews were conducted in English and Spanish.

Dynata used router technology to recruit participants, and all available panelists age 18 and older in each state were recruited. Among the 28,838 panelists who touched the pre-screener instrument, 21,796 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 used built-in technology that uses digital fingerprinting, geolocation clues, and checks at enrollment to confirm identity and to identify suspicious behavior to prevent respondents from completing the survey more than once.

Cint's suppliers invited respondents to the survey using email invites and panelist recruitment. Before sending them into the survey, Cint targeted and pre-screened respondents age 18 and older on the basis of registered voter status and state location using zip codes. Among the 155,138 panelists who touched the pre-screener instrument, 51,095 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. In order to ensure the final sample did not include any respondents who completed the survey more than once, Cint removed duplicates by IP address, participant ID, and cookies.

Prodege invited its members to the survey using invitations sent by email and through the panelist portal. Before sending them into the survey, Prodege targeted and pre-screened respondents age 18 and older on the basis of age, gender, education, income, race/ethnicity, state, and registered voter status. Among the 21,710 panelists who touched the pre-screener instrument, 16,900 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. To prevent respondents from completing more than once, Prodege verified

panelists using email double opt-in verification, physical address verification, device fingerprinting, mobile verifications, and CAPTCHA while continuously monitoring their users to avoid panelists duplication. From there, they managed how many surveys were sent to each user, and only allowed each user to enter the survey 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, supplemented by the USPS Delivery Sequence File, and then contacted by U.S. mail, email, telephone and field interviewers (face-to-face). The panel provided sample coverage of approximately 97% of the U.S. household population. Those excluded from the sample included 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 2020 vote. The size of the selected sample per sampling stratum was determined by the population distribution for each stratum. In addition, sample selection took into account expected differential survey completion rate by group so that the set of panel members with a completed interview was a representative sample of the target population of registered voters. If a panel household had 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 employed a four-step weighting approach that combines the probability sample with the nonprobability sample and refined estimates at a subregional level within each state. For national estimates, the 50 state surveys and the AmeriSpeak survey were weighted separately and then combined into a survey representative of voters in all 50 states.

State Surveys

First, weights were constructed separately for the probability sample (when available) and the nonprobability sample for each state survey. These weights were 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 were derived from a combination of data from the U.S. Census Bureau's November 2020 Current Population Survey Voting and Registration Supplement, Catalist's voter file, and the Census Bureau's 2021 American Community Survey. The categories used for

weighting were collapsed in some states based on the sample sizes and population distributions. 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)

- Housing tenure (owned, rented/occupied without payment)

- Only in AL, AZ, HI, MD, MO, NY, OH, OK, OR, PA: 2022 Presidential primary vote (voted in Democrat or Republican primary, did not vote)

- Probability sample only: Partisanship model score quintile (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 a four-level partisanship model score variable, a five-level predicted response propensity variable, and Catalist voter file race/ethnicity.

Second, calibration variables were included in weighting for both the probability and nonprobability samples to ensure the nonprobability sample was similar to a probability sample in regard to variables that were predictive of vote choice that cannot be fully captured through demographic adjustments.

- <u>Calibration variables</u>

    o Party ID (Democrat, Independent, Republican)

The calibration benchmarks were based on estimates from a combination of national and state linear regression models that made predictions for registered voters at the state-level for Party ID (Democrat, Independent, Republican). The models for the calibration variables were run using an instrumental calibration approach. 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 vote in the 2020 Presidential election (Biden, Trump, other)

Third, all respondents in each state were weighted to improve estimates for substate geographic regions. This weight combined the weighted probability sample (if available) and the nonprobability sample, and then used a small area model to improve the estimate within subregions of a state. We created between 2 and 28 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 Senate or governor election or overall House vote in each state.

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: 2020 Presidential election results, population density, median income, percent below poverty line, percent unemployed, percent college degree, portion on public assistance, percent 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 in rural area, percent in suburban area, percent living in rented households, and percent who had not moved in the 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 were weighted to the certified vote count following the completion of the election. The Senate, governor, or House vote results were used as benchmarks for weighting respondents who were voters. This weighting was done in 2-28 sub-state regions within each state.

National Survey

The national survey was weighted to combine the 50 state surveys with the nationwide AmeriSpeak survey. Each of the state surveys was weighted as described. The AmeriSpeak survey received a nonresponse-adjusted weight that was then adjusted to national totals for registered voters derived from the U.S. Census Bureau's November 2020 Current Population Survey Voting and Registration Supplement, the Catalist voter file, and the Census Bureau's 2021 American Community Survey. The

state surveys were further adjusted to represent their appropriate proportion of the registered voter population for the country and combined with the AmeriSpeak survey. After all votes were counted and certified, the national data file was adjusted to match the vote for Senate, governor, or House within each state.

Contact

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

Using Weights

AP VoteCast was 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 California registered voters who approve of President Biden), 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 publicly 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 (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 (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, Kansas, Massachusetts, Maryland, Mississippi, North Carolina, North Dakota, New Jersey, Nevada, New York, Ohio, Oregon, Texas, Utah, Virginia, Washington, West Virginia – respondents were classified as voters based on the following criteria:

- The respondent said they will definitely vote to LVB; or
- The respondent said they probably will vote to LVB, and they voted in either the 2018 midterm election or they voted in the 2020 presidential election; or
- The respondent said they already voted to LVB

The following states – Alabama, Arizona, Arkansas, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Iowa, Idaho, Indiana, Kentucky, Louisiana, Maine, Michigan, Minnesota, 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 voted when interviewed on November 6 or later were classified as nonvoters. Respondents in these states were classified as voters based on the following criteria:

- If the respondent said they already voted to LVB and:
    - The respondent said they voted by mail before Election Day to WVA; or
    - The respondent said they voted in person before or on Election Day to WVA; or
- If the respondent said they definitely will vote to LVB, or that they probably will vote and voted in either the 2018 midterm election or the 2020 presidential election, and:
    - The respondent said they will vote in person before Election Day to WVB; or
    - The respondent said they will vote in person at a polling place on Election Day to WVB; or
    - The respondent said they will vote by mail to WVB and the interview was conducted between October 31 and November 5.