

METHODOLOGY

AP VoteCast – 2021 Virginia Governor Election

FEBRUARY 2022

PRESENTED BY:
NORC
at the
University of Chicago

Table of Contents

Study Methodology.....1

Using Weights.....6

About The Associated Press-NORC Center for Public Affairs Research.....7

Appendix A – Likely Voter Models.....8

Study Methodology

AP VoteCast is a survey of the American electorate conducted by NORC at the University of Chicago for Fox News and The Associated Press. The survey is funded by AP. AP VoteCast was conducted in Virginia for their 2021 governor election. The survey of 3,493 registered voters was conducted October 27 to November 2, 2021, concluding as polls closed on Election Day. Interviews were conducted via phone and web, with 169 completing by phone and 3,324 completing by web.

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 were conducted in English and Spanish. Respondents may have received a small monetary incentive for completing the survey. Participants selected from state voter files were contacted by phone, text, and mail, and had the opportunity to take the survey by phone or online.

For those who screened into the survey as eligible based on being a registered voter in the state, the interview completion rates were 94.14% for the probability sample drawn from the state voter files and 94.5% for the nonprobability sample. Quality assurance checks were conducted to ensure data quality. In total, four interviews (0.11%) were removed for nonresponse to at least 50% of the key demographics asked of them – party identification, education, race/ethnicity, gender, and income. These interviews were excluded from the data file prior to weighting.

In Virginia, VoteCast is based on 1,381 probability-based interviews conducted online and via phone, and 2,112 nonprobability interviews conducted online. The margin of sampling error is estimated to be plus or minus 2.5 percentage points for voters (n=2,655) and 4.4 percentage points for nonvoters (n=838). The unweighted response rate for the probability sample drawn from the Virginia state voter file was 3.9%.¹

Sampling error is only one of many potential sources of error and there may be other unmeasured error in this or any other survey.

Sampling Details

Probability-based Registered Voter 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

¹ The design weighted response rate for the probability sample is 4.5 %.

by the four-level partisanship category and five-level predicted response propensity 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 85% of sampled records, including cell phone numbers for 79% 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 “Virginia 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. A single text message was sent to all sampled registered voters with a cell phone number. The text message included the link to the online survey. 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. 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 sample. 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 registered voters in Virginia along with a general population of Virginians age 18 and older were recruited. Among the 1,469 panelists who touched the pre-screener instrument, 1,048 went on to complete the full survey. Panelists confirmed voter registration status in Virginia. Dynata’s system uses 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.

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 voter status. Among the 2,273 panelists who touched the pre-screener instrument, 1,064 went on to complete the full survey. Panelists confirmed voter registration

status in Virginia. In order to ensure the final sample does not include any respondents who completed the survey more than once, Lucid deduplicates by IP address, participant ID, and cookies.

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 Virginia.

First, weights are constructed separately for the probability sample and the nonprobability sample. These weights are adjusted to population totals to correct for demographic imbalances of the responding sample compared to the population of registered voters. The adjustment targets are 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 2019 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)
- County grouping using AP’s party grouping (variable “AP_PARTY_REGION”)
- Probability sample only: Catalist partisanship score quintile (1st quintile, 2nd quintile, 3rd quintile, 4th quintile, 5th quintile).
- Probability sample only: Proportion of total probability sample completed interviews stemming from outbound dials.
- Prior to adjusting to population totals, the probability-based registered voter list sample weights were adjusted for differential non-response by combinations of the following variables:
 - Partisanship category based on Catalist partisanship score (Strong R, Lean R, Lean D, Strong D)
 - Quintile of predicted response propensity
 - Indicator of whether the case was sent a text message or not
 - Indicator of whether the case was offered an incentive or not
 - Indicator of whether the case had a cell phone number available

- Indicator of whether the case had any phone number available

The nonresponse adjustment was done separately for screening stage and the interviewing stage. At each stage, the adjustment cells were collapsed if there were fewer than 10 respondents in the adjustment cell (i.e., 10 completed screeners for the screening stage, 10 completed interviews for the interview stage).

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)

The calibration benchmarks are based on estimates from a linear regression model that incorporates all probability cases to make predictions for registered voters at the state-level for Party ID (Democrat, Independent, Republican). We derived two sets of partisanship benchmarks - one to calibrate those who voted in the 2020 Presidential election and one to calibrate those who voted in the 2018 Senate election. Respondents who voted in both elections were included in the calibration dimension for both 2018 voters and 2020 voters, and those who voted in neither election were weighted to the average party ID distribution of 2018 and 2020 voters. The models for the calibration variables were run using an instrumental calibration approach and 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, flag for voting for Biden in 2020 Presidential election
- Flag for voting for Republican candidate, flag for voting for Democratic candidate in 2018 Senate election

Third, all respondents are weighted to improve estimates for substate geographic regions. This weight combines the weighted probability sample and the non-probability sample, and then uses a small area model to improve the estimate within subregions of the state. We created 21 regions (county groupings) for Virginia 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 governor race.

The following variables were used as potential covariates in the model: 2020 Presidential election results, 2018 House 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, 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 governor election vote results were used as benchmarks for weighting respondents who were voters. This weighting is done in 21 sub-state regions within Virginia.

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

- The FINALVOTE_STATE_WEIGHT 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_STATE_WEIGHT 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 statewide, limit analysis to LIKELYVOTER=1 and cases where vote choice in GOVVOTEVA is not missing. The FINALVOTE_STATE_WEIGHT variable should be used for weights.

About The Associated Press-NORC Center for Public Affairs Research

Celebrating its 10th anniversary this year, 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. In its 10 years, The AP-NORC Center has conducted more than 250 studies exploring the critical issues facing the public, covering topics like health care, the economy, COVID-19, trust in media, and more.

Learn more at www.apnorc.org

Appendix A – Likely Voter Models

Respondents are classified as voters based on the following criteria:

- If respondent says they will vote by mail to WVA or WVB:
 - The respondent says they will definitely vote to LVB, they are certain they will vote to LV, they are extremely or very interested in the election to LVA, and they voted in either the 2018 midterm election or they voted in the 2020 presidential election.
- If the respondent says they will vote in person before or on Election Day to WVA or WVB:
 - 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, and they voted in the 2018 midterm election or they voted in the 2020 presidential election.
- The respondent says they already voted to LVB or LV.