

ELECTION 20

AP VOTECAST

Assessing AP VoteCast 2020

MAY 7, 2021

FOR MORE INFORMATION CONTACT:

Trevor Tompson
Senior Vice President
NORC at the University of Chicago
55 East Monroe Street
30th Floor
Chicago, IL 60603
(312) 759-4000

Emily Swanson
Director of Public Opinion Research
The Associated Press
1100 13th Street NW
Washington, DC 20005
(202) 731-1491



Table of Contents

AP VoteCast 2020 Executive Summary	1
Overview of Approach	2
Summary of Results	5
Georgia Runoff Elections.....	11
Modeling Improvements Implemented for the Georgia Runoff Elections	11
Further Research.....	13
Appendix I: Vote Choice Estimates: 2020 General Election and Georgia Runoff Elections.....	14
Appendix II: Likely Voter Models for 2020 and Summary of Results.....	22
Questions used for Likely Voter Model.....	22
Specifications for Likely Voter Models.....	25
Appendix III: Voter Classification Rates by State	26

AP VoteCast 2020 Executive Summary

AP VoteCast is a modern, innovative survey of the American electorate conducted in all 50 states that is designed to meet voters where they are. Because VoteCast is not based on in-person interviews conducted outside of polling locations, its methodology was well suited for the explosion of advance voting in 2020 that resulted from the COVID-19 pandemic. Many voters opted to avoid in-person polling places on Election Day; about 70% of ballots in 2020 were cast either early, absentee, or by mail. VoteCast also captures the opinions of both people who vote and registered voters who decide not to cast a ballot. Taken together, this means that VoteCast delivers a broader portrait of the American electorate than any other election survey.

VoteCast was developed to provide a more accurate and comprehensive approach to understanding elections, particularly as the American electorate increasingly chooses to vote in advance of Election Day. VoteCast [debuted successfully for the 2018 midterm elections](#) after years of research and testing. Using a random, probability-based sample of registered voters to carefully calibrate a very large sample from opt-in sample sources, VoteCast delivers the best of both methods – the accuracy of probability-based surveys combined with the scale provided by an opt-in survey that interviews tens of thousands of people quickly.

AP VoteCast 2020 was 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 2020 election was the first time VoteCast was used for a presidential race. Amid a pandemic, record turnout, and many close races, the 2020 election was challenging for the polling industry; many high-quality surveys underestimated the vote for Republican candidates. VoteCast was not immune to some of the broader problems confronting polls, and AP and NORC are deeply committed to further research to improve VoteCast and build upon its successes. That said, we believe VoteCast delivered on its mandate to facilitate AP's declaration of election winners and provide a real-time portrait of the electorate that enabled journalists around the world to provide readers, viewers, and listeners with a deep understanding of the 2020 election. Specifically, VoteCast:

- Completed interviews with more than 133,000 registered voters in nine days leading up to the election;
- For the first time, provided estimates of the electorate in all 50 states;
- Provided an accurate picture of the demographic makeup of voters participating in an election with record turnout. The final estimates of the composition of the electorate largely matched state voter file data and Census Bureau estimates for age groups, gender, and racial/ethnic groups;
- Correctly projected the winner in 90% of U.S. Senate, governor, and presidential elections including 45 of 50 states in the presidential race before final vote adjustments;
- Projected a national margin of 10.0 percentage points in favor of Joe Biden at 4 p.m. on Election Day; the final vote count had a 4.5 percentage point advantage for Biden;
- On average, at poll close on Election Day, overestimated Democratic candidates' share of the vote by 2.3 percentage points and underestimated the share of the vote for Republican candidates by 3.1 percentage points;

- Assessed multiple likely voter models in real-time with the final model correctly classifying 93% of the probability-based sample respondents identified as voters, according to a voter validation study.¹

News organizations across the United States and around the world used the data to explain the mood of the electorate in their election-night coverage and AP and Fox News used VoteCast to inform race-calling decisions.

We are committed to transparency of VoteCast’s methods and results and the continual improvement of the VoteCast methodology over time. For example, innovations to the calibration modeling were tested following the general election and successfully implemented for the Georgia runoff elections.

This report provides the results of a thorough assessment of VoteCast’s performance in the 2020 general election and the Georgia runoff elections.

Overview of Approach

The VoteCast general election survey of 133,103 registered voters nationwide was conducted between Oct. 26 and Nov. 3, 2020, concluding as polls closed on Election Day. The survey provided estimates of the presidential vote in all 50 states, votes in 35 U.S. Senate elections, 11 gubernatorial elections, and the national House vote, as well as the opinions of both voters and non-voters nationwide.

VoteCast features a large, nationally representative survey of voters and non-voters that allows for in-depth analysis of state-level election attitudes and behaviors. The survey combines three different samples: probability-based state samples drawn from voter files, non-probability samples primarily from online panels, and a probability-based sample from a nationally representative panel.

The survey delivered robust samples within each of the 50 states based on a combination of probability- and non-probability samples. In 40 states, competitive at the presidential level or with high-profile governor or U.S. Senate races, VoteCast combined interviews of registered voters randomly sampled from state voter files with interviews of self-reported registered voters from opt-in online sample sources. Respondents sampled from the voter files were recruited to the survey via postcard and telephone. Respondents were given the option of completing the survey either online or by telephone. Each of these 40 states had between 541 and 1,623 probability interviews and between 121 and 3,295 non-probability interviews. In the other 10 states, VoteCast used opt-in online sample sources to collect between 200 and 1,000 interviews of self-reported registered voters in each state. Across all 50 states, 41,776 interviews came from probability-based sampling of state voter files, and 87,186 interviews came from non-probability sample sources.

In addition to the 50 state surveys, VoteCast used NORC’s AmeriSpeak® panel. The AmeriSpeak® panel is a mixed-mode panel recruited from a probability, address-based sample and is designed to be representative of the U.S. population. There were 4,141 total interviews from AmeriSpeak.

¹ This voter validation study was conducted with data available as of March 15, 2021. At this time data were available in 30 states.

After combining the interviews from AmeriSpeak, the probability-based state samples, and the non-probability cases, the survey included interviews with 110,484 voters and 22,619 non-voters. The state samples ranged from 201 interviews to 5,006 interviews.²

Sampling Summary

Probability-based Registered Voter Sample

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, 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. 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. Before dialing, all probability sample records received 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. Respondents may receive a small monetary incentive for completing the survey. 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 who answered the phone. Both online and telephone respondents provided confirmation of registered voter status in the state.

Non-probability Sample

Non-probability 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 were used to prevent respondents from completing the VoteCast survey multiple times. Non-probability respondents provided confirmation of registered voter status in the state.

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.

A sample of registered voters was selected from the AmeriSpeak panel using sampling strata based on age, race/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 the expected

²The study interview was available in English or Spanish. Additionally, for all registered voter file cases flagged as likely Hispanic households, the postcards had both English and Spanish text introducing the survey and explaining how to complete it.

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 a panel household has more than one active adult panel member, only one adult in the household was eligible for selection (random within-household sampling). AmeriSpeak panelists provided confirmation of registered voter status in the state.

Weighting Summary

VoteCast employs a four-step weighting approach that combines the probability sample with the non-probability sample and refines estimates at a subregional level within each state.

First, weights are constructed separately for the probability sample (when available) and the non-probability 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 were 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 included gender, age, race/ethnicity, age, education, and region.

Before 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 sampling condition.

Second, calibration variables (party identification and country right/wrong track) were included in weighting for both the probability and non-probability samples to ensure the non-probability sample is similar to a probability sample regarding variables that are predictive of vote choice that cannot be fully captured through demographic adjustments.

The two 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 to make predictions for registered voters at the state-level for party identification (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 samples, 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. 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 individual-level variables such as gender, age, race/ethnicity, education, and 2016 presidential vote.

Third, all respondents in each state are weighted to improve estimates for sub-state 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 sub-regions 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 completed surveys in each county. We then used these groupings to generate model-based estimates of vote choice among likely voters. The small area model was applied to the presidential election.

There were two models: 1) predicting the percent of vote share that goes for either of the two major parties' candidates, 2) predicting the percent of major party vote share that goes for the Democratic/Republican candidate. The variables used as potential covariates in the model included county-level variables related to previous election results, population density, income, age, education,

and race/ethnicity. We included the variables most highly correlated with vote among 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 U.S. Senate or governor vote results were used as benchmarks for weighting respondents who were voters. This weighting is done in between 3 and 35 sub-state regions 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.

The AmeriSpeak survey receives a non-response-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.

Summary of Results

We conducted an assessment of VoteCast 2020 to evaluate the sample performance, the composition of the electorate, the accuracy of the vote choice estimates, and voter validation as a way to identify avenues for future methodological improvements.

Sample Performance

With data collection for the state samples, we collected 133,103 interviews. The number of state sample interviews varied from 201 in Wyoming to 5,006 in Pennsylvania. The survey included more than 4,500 interviews in each of the following states: Pennsylvania, Michigan, Ohio, Texas, Arizona, and California.

VoteCast featured 41,776 probability interviews. In particular, it had more than 1,000 interviews in the 20 states deemed most newsworthy and more than 500 in another 20 states. The states with the most complete probability-based interviews were New Hampshire (1,623), Oregon (1,611), and Nevada (1,561). Eighty-nine percent of probability completes came from people logging into the website provided on the postcard. The overall response rate for the probability sample was 3.4% (AAPOR Response Rate 3).

The survey had 87,186 non-probability interviews across all 50 states. There were more than 3,000 non-probability completes in several large battleground states such as Arizona, Florida, Georgia, Michigan, Pennsylvania, and Ohio. However, there were less than 300 non-probability interviews in less populous states such as Alaska, Wyoming, North Dakota, and Vermont.

The AmeriSpeak survey with 4,141 completed interviews was included in the national popular vote estimates. The overall AmeriSpeak response rate was 5.0% (AAPOR Response Rate 3).

Composition of Electorate

VoteCast provided a reliable picture of the demographic makeup of voters. The VoteCast estimates of the composition of the electorate are largely consistent with an initial comparison to publicly available government records and data. In particular, estimates for the electoral composition are similar to the

estimates for voters from updated, post-election state voter files, as well as available data on voting and registration from the U.S. Census Bureau’s Current Population Survey (CPS).

VoteCast estimates of the age, gender, race/ethnicity, and education makeup of voters nationally both prior to and after adjustments to the final vote count are within a few percentage points of the 2020 CPS estimates of voters (see Table 1 below).

Table 1. Comparison of Census Voting Data and VoteCast Estimates for National Composition of the 2020 Electorate

	Current Population Survey estimates	VoteCast estimates PRIOR to adjustments to vote count	VoteCast estimates AFTER adjustments for final vote count
Age			
18 to 29	16	14	13
30 to 44	23	23	23
45 to 64	35	36	36
65+	26	27	28
Gender			
Male	47	46	47
Female	53	53	53
Race/ethnicity			
White	71	73	74
Black	12	12	11
Hispanic	11	10	9
Other	6	6	6
Education			
High school or less	29	27	27
Some college	30	34	34
BA	26	24	25
Postgrad	15	14	15

In addition, VoteCast state-level estimates of the makeup of voters with all interviews collected through poll close on election night, but before any adjustments to the vote count, look similar to the final electorate. Further, VoteCast’s final estimates of the demographic characteristics of the electorate following the final adjustment of the results to vote count closely resemble the state voter files in key states that have updated their voter records.

For example, race, age, and gender - both before and after the vote count - all closely align with the voter file data in states such as Arizona, Pennsylvania, and Colorado.

Table 2. Comparison of Catalist Voter Files and VoteCast Estimates for Composition of the 2020 Electorate in Key States where VoteCast Demographics were Most Accurate

		Catalist Voter file estimates for percent of electorate	VoteCast estimates PRIOR to adjustments to vote count	VoteCast estimates AFTER adjustments for final vote count
Arizona	Age			
	18 to 29	15	14	14
	30 to 44	22	22	22
	45 to 64	33	34	34
	65+	30	30	30
	Gender			
	Male	47	47	48
	Female	53	52	52
	Race/ethnicity			
	White	70	71	71
	Black	5	4	4
	Hispanic	18	18	17
	Other	7	7	8
Pennsylvania	Age			
	18 to 29	15	14	14
	30 to 44	22	23	23
	45 to 64	36	36	36
	65+	27	27	27
	Gender			
	Male	47	46	47
	Female	53	53	53
	Race/ethnicity			
	White	84	84	85
	Black	10	10	9
	Hispanic	4	4	4
	Other	3	3	3
Colorado				
	Age			
	18 to 29	17	17	17
	30 to 44	26	26	26
	45 to 64	33	34	34

	65+	23	23	23
	Gender			
	Male	48	49	49
	Female	52	50	50
	Race/ethnicity			
	White	80	80	81
	Black	4	3	3
	Hispanic	12	11	11
	Other	4	4	4

Even in states in which VoteCast’s estimates of the vote before adjusting to the vote count were less accurate, such as Florida or Ohio, the VoteCast electorate at poll close is very similar to the voter file when it comes to age, gender, and race/ethnicity.

The reliability of VoteCast’s estimates of the electorate after adjusting to the final vote allowed AP and other media partners to report stories on election night and in the days and weeks after the election that accurately reflect the demographics of voters.

Table 3. Comparison of Catalist Voter Files and VoteCast Estimates for Composition of the 2020 Electorate in States where Estimates were Less Accurate

		Catalist Voter file estimates for percent of electorate	VoteCast estimates PRIOR to adjustments to vote count	VoteCast estimates AFTER adjustments for final vote count
Florida	Age			
	18 to 29	13	13	12
	30 to 44	20	21	21
	45 to 64	35	35	35
	65+	31	31	31
	Gender			
	Male	45	46	46
	Female	55	54	53
	Race/ethnicity			
	White	66	66	66
	Black	13	13	12
	Hispanic	17	17	18
	Other	4	4	4
Georgia	Age			
	18 to 29	16	17	16

	30 to 44	24	24	23
	45 to 64	37	36	37
	65+	23	22	23
	Gender			
	Male	44	46	47
	Female	56	53	52
	Race/ethnicity			
	White	62	61	63
	Black	29	31	29
	Hispanic	3	3	3
	Other	5	4	4
North Carolina	Age			
	18 to 29	16	15	15
	30 to 44	22	22	22
	45 to 64	36	36	36
	65+	26	27	27
	Gender			
	Male	45	45	45
	Female	55	55	55
	Race/ethnicity			
	White	72	73	74
	Black	20	20	19
	Hispanic	3	3	3
	Other	4	4	4
Ohio	Age			
	18 to 29	14	13	13
	30 to 44	22	23	22
	45 to 64	37	36	36
	65+	28	28	28
	Gender			
	Male	47	47	47
	Female	53	52	52
	Race/ethnicity			
	White	86	85	86
	Black	10	11	10
	Hispanic	2	1	1
	Other	2	2	2

Accuracy of Vote Choice

In terms of estimates with all cases collected at poll close on Election Day, but prior to any adjustments to the vote count, VoteCast correctly predicted the winner in 90% of the 96 state-wide races for president, U.S. Senate, and governor in 2020. On average, the survey overestimated the Democratic candidates' share of the vote by 2.3 percentage points and underestimated the share of the vote for Republican candidates by 3.1 percentage points. VoteCast's estimates were within a percentage point of Joe Biden's final vote choice in twelve states, including Colorado, Kentucky, Missouri, and Massachusetts. In contrast, there were six states in which the survey over-estimated Biden's vote share by more than 4 percentage points, including Florida and Iowa. See Appendix I for the estimates for each race.

The calibration of the probability and non-probability samples using the model estimates of partisanship and right/wrong track of the country tended to boost Republican vote share, which significantly improved the overall estimates. As a result, NORC will continue to analyze different models that allow for more state-level variation and seek a solution that improves estimates across all states.

The small area modeling adjustments were highly effective in improving accuracy within states. On average, the small area models reduced the average absolute error in regions within states by about 50% (the average error for Biden vote fell from 5.4 percentage points to 2.2 percentage points).

Voter Validation

During data collection, we monitored 20 likely voter models and found little difference in the vote choice estimates for the various models. The initial results of a voter validation study of 31,540 probability cases found that 93% of those identified as likely voters were correctly classified.

The two likely voter models used for the final estimates both accounted for past vote, self-reported intent to vote, and the method/timing of vote. The difference in the two models used was based on whether a respondent lived in a state in which mail ballots had to be received on Election Day or could be received after Election Day. In states in which ballots had to be received by Election Day, we used a model that classified people as unlikely voters if they reported that they planned to vote by mail but had not yet voted when interviewed on Oct. 31, 2020 or later. See Appendix II for detailed specifications for the two models we used and a summary of the results. Overall, 83% of respondents were classified as likely voters, including 95% of probability cases and 77% of non-probability cases.

The initial results of a voter validation study of the probability-based interviews show that 90% of respondents were correctly classified as voters or non-voters in the study.³ Moreover, 93% of those classified as voters did vote according to state voter files. There was no significant difference in the national vote estimates of likely voters who actually voted and those who did not vote (54% Biden/44% Trump vs. 55% Biden/43% Trump).

The likely voter model chosen for each state accurately identified voters across partisanship, gender, race/ethnicity, education, and income groups. The model correctly classified at least 93% of voters from each of the three partisan groups (Democrats, Republicans, and independents). The model worked

³ Due to privacy agreements with our vendors, VoteCast was unable to ask non-probability completes from the opt-in panel for the personally identifiable information needed to validate their voter status.

relatively well in accurately classifying voters of all demographic characteristics including those age 18-29 (91%), age 65 and older (95%), with a college degree (94%), without a college degree (92%), Hispanic (91%), African American (89%), and white respondents (94%).

The model did well across a variety of states. See Appendix III for a list of states and voter classification rates.

Georgia Runoff Elections

The VoteCast Georgia runoff elections survey of 4,565 registered voters was conducted from Dec. 28, 2020 to Jan. 5, 2021, concluding as polls closed on Election Day. Interviews were conducted via phone and web, with 186 completing by phone and 4,379 completing by web. The survey is based on 1,173 probability-based interviews conducted online and via phone, and 3,392 non-probability interviews conducted online. The response rate for the probability sample drawn from the Georgia state voter file was 1.9%.

Similar to AP VoteCast 2020 for the general election, the 2021 Georgia Runoff survey combines interviews with a random sample of registered voters drawn from state voter files with self-identified registered voters selected from Lucid and Dynata's non-probability online panels.

The random sample of registered voters was obtained from Catalist LLC's registered voter database. Interviews were conducted in English. 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. All sampled registered voters from state voter files 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. A subset of records with telephone numbers were dialed.

The survey employed a similar four-step weighting approach that combines the probability sample with the non-probability sample, and refined the estimates at the subregional level within Georgia. The improvements noted below to this approach compared to the November 2020 survey were also implemented.

The VoteCast estimates for the Georgia runoffs were accurate. At poll close, VoteCast estimated it was 50.0-50.0 in the Ossoff/Perdue election (final result was 50.6-49.4) and 50.3-49.7 in the Warnock/Loeffler election (final result was 51.0-49.0).

Modeling Improvements Implemented for the Georgia Runoff Elections

For the Georgia runoff elections, we adjusted the modeling for the calibration variables based on some preliminary analysis from the November 2020 election data. We found our modeling in November improved the reliability of the estimates, and our analysis showed the models would have further improved the results if they were run multiple times until they converged on an estimate (similar to an iterative raking process used in weighting when demographic variables are raked until convergence). With this iterative modeling, the model is run with the data, the data are weighted to the new calibration estimate, the model is re-run with the weighted data, the data are re-weighted with the new calibration estimate and this process continues until there is no change in the estimate of the calibration variable by further running the model.

We implemented the iterative modeling approach for deriving calibration variable benchmarks for partisanship in Georgia, and it improved the reliability of the results. For example, our estimates using

the iterative modeling showed the Purdue/Ossoff race at 50-50 on the Saturday before the election. Using the same data but running the calibration models only once, the estimates were 52-48 for Ossoff.

When testing the iterative approach on the November 2020 election data, it significantly improved the accuracy of the overall results. While the November estimates had an average error of 2.2 percentage points for Biden and 2.9 percentage points for Trump using all cases collected at poll close on Election Day, but prior to any adjustments to the vote count, the iterative version had an average error of 1.5 percentage points for Trump and 1.1 percentage points for Biden. The iterative approach would have improved the results in most states and would have been especially effective in several Sunbelt swing states. The table below highlights several states in which the iterative approach would have improved results the most. The table shows the final spread in the election, our November estimated spread, and the estimated spread if we had used the iterate approach.

Table 4. Comparison of Final Spread, Estimated Spread, and Iterated Spread

State	Final Result (Biden-Trump)	November Estimate at Poll Close (Biden-Trump)	November Estimate with Iteration (Biden-Trump)
United States	4.5	9.8	7.7
Arizona	0.3	6.2	0.5
Georgia	0.2	4.8	-0.9
North Carolina	-1.4	5.5	1.1
South Carolina	-11.7	-3.4	-11.3
Texas	-5.6	1.8	-5.2
Michigan	2.8	6.1	3.7
Indiana	-16.1	-9.6	-17.9
Alabama	-25.5	-19.4	-27.4
New Hampshire	7.4	14.2	4.3
Iowa	-8.2	2.3	-3.4
Ohio	-8.0	0.6	-3.8
Wisconsin	0.6	8.8	5.9
Utah	-20.5	-8.6	-17.8
Virginia	10.2	14.7	10.6

However, the iterative approach would not have significantly improved estimates in several key states such as Florida, Pennsylvania, or Nevada, and it would have slightly reduced accuracy in a couple of blowout states such as California.

Moving forward, NORC believes it is important to conduct further research to determine the broader applicability of the approach, assess under what conditions it is most and least effective, and determine how best to operationalize it for different elections. For example, we plan to further explore different modeling approaches and specifications and what variables we should use in the models (e.g. what past elections).

Further Research

AP VoteCast is committed to continuous assessment and research to improve the methodology and product. Based on the results from the 2020 VoteCast experience, we plan to undertake the following additional research to improve the survey for future election cycles:

- Explore ways to increase non-probability sample availability, especially in small states where online panels cannot support VoteCast's target sample needs.
- Analyze potential refinements to the calibration modeling approach to improve accuracy and reduce variability of the error across states.
- Test recruitment materials to improve overall response and response among low-propensity subgroups such as racial and ethnic minority groups, lower education voters, weaker partisans, and infrequent voters.

AP VoteCast is also committed to full transparency. A complete public-use data file from the 2020 general election and Georgia runoff elections, along with a methodology statement and codebook, is available at: <https://apnorc.org/projects/ap-votecast-2020-general-elections>.

Appendix I: Vote Choice Estimates: 2020 General Election and Georgia Runoff Elections

State	Race	Party	VoteCast estimates PRIOR to adjustments to vote count	Actual Result (via AP Vote Count)	VoteCast Estimates AFTER Final Vote Count ⁴
US	pres	Democrat	53.6	51.4	51.2
US	pres	Republican	43.8	46.9	47.0
US	pres	Other	2.6	1.7	1.8
AK	pres	Democrat	44.9	43.0	43.0
AK	pres	Republican	52.1	54	53.1
AK	pres	Other	3.0	3.9	3.9
AK	sen	Democrat	45.1	41.3	40.6
AK	sen	Republican	50.1	54.0	53.0
AK	sen	Other	4.8	4.7	6.4
AL	pres	Democrat	38.9	36.7	35.5
AL	pres	Republican	58.2	62.2	61.9
AL	pres	Other	2.9	1.1	1.6
AL	sen	Democrat	43.1	39.8	39.7
AL	sen	Republican	54.4	60.2	60.1
AL	sen	Other	2.5	0.0	.26
AR	pres	Democrat	36.9	34.8	34.8
AR	pres	Republican	61.2	62.4	62.4
AR	pres	Other	1.9	2.8	2.8
AR	sen	Democrat	0.0	0.0	0.0
AR	sen	Republican	65.8	66.5	66.7
AR	sen	Other	34.2	33.4	33.3
AZ	pres	Democrat	51.8	49.4	49.3
AZ	pres	Republican	45.7	49.1	49.0
AZ	pres	Other	2.5	1.5	1.6
AZ	sen	Democrat	53.8	51.2	51.0
AZ	sen	Republican	44.7	48.8	48.9
AZ	sen	Other	1.5	0.0	0.1
CA	pres	Democrat	66.9	63.5	63.5
CA	pres	Republican	30.5	34.3	34.3
CA	pres	Other	2.7	2.2	2.2
CO	pres	Democrat	55.4	55.4	55.4
CO	pres	Republican	41.2	41.9	41.9
CO	pres	Other	3.4	2.7	2.7

⁴ All of the survey estimates were adjusted to match vote count data as of January 22, 2021. Some states such as Connecticut and Oklahoma made small adjustments to the final vote estimates after that date, which explains why some of the final survey estimates don't exactly match the final vote numbers shown in the table.

CO	sen	Democrat	54.8	53.5	54.2
CO	sen	Republican	42.4	44.2	44.8
CO	sen	Other	2.8	2.3	1.0
CT	pres	Democrat	60.1	59.3	59.0
CT	pres	Republican	37.5	39.2	39.0
CT	pres	Other	2.4	1.5	2.0
DE	pres	Democrat	58.7	58.8	58.5
DE	pres	Republican	39.1	39.8	39.6
DE	pres	Other	2.2	1.5	1.9
DE	sen	Democrat	61.2	59.4	60.2
DE	sen	Republican	35.7	37.9	38.5
DE	sen	Other	3.1	2.7	1.2
DE	gov	Democrat	61.1	59.5	60.3
DE	gov	Republican	36.4	38.6	38.1
DE	gov	Other	2.5	1.9	1.7
FL	pres	Democrat	52.9	47.9	47.9
FL	pres	Republican	45.5	51.2	51.2
FL	pres	Other	1.6	0.8	0.9
GA	pres	Democrat	51.1	49.5	49.3
GA	pres	Republican	46.3	49.3	49.0
GA	pres	Other	2.7	1.2	1.7
GA	sen	Democrat	49.5	47.9	48.2
GA	sen	Republican	45.4	49.7	49.7
GA	sen	Other	5.0	2.3	2.1
GA	senspecial	Democrat	38.4	32.9	47.2
GA	senspecial	Republican	25.4	25.9	49.2
GA	senspecial	Other	36.3	41.2	3.6
HI	pres	Democrat	65.0	63.7	63.7
HI	pres	Republican	33.1	34.3	34.3
HI	pres	Other	1.9	2.1	2.0
IA	pres	Democrat	49.7	45.0	45.0
IA	pres	Republican	47.4	53.2	53.2
IA	pres	Other	2.9	1.8	1.8
IA	sen	Democrat	50.7	45.2	45.2
IA	sen	Republican	45.6	51.8	52.7
IA	sen	Other	3.7	3.0	2.1
ID	pres	Democrat	35.2	33.1	33.1
ID	pres	Republican	61.9	63.9	63.9
ID	pres	Other	2.8	3.0	3.0
ID	sen	Democrat	35.6	33.2	34.0
ID	sen	Republican	60.8	62.6	64.4

ID	sen	Other	3.6	4.1	1.6
IL	pres	Democrat	59.9	57.5	57.5
IL	pres	Republican	37.0	40.6	40.6
IL	pres	Other	3.2	1.9	1.9
IL	sen	Democrat	59.3	54.9	56.0
IL	sen	Republican	35.8	38.9	39.3
IL	sen	Other	5.0	6.2	4.6
IN	pres	Democrat	43.5	41.0	40.8
IN	pres	Republican	53.1	57.1	56.8
IN	pres	Other	3.4	2.4	2.4
IN	gov	Democrat	32.2	32.1	32.2
IN	gov	Republican	56.2	56.5	57.1
IN	gov	Other	11.5	11.4	10.7
KS	pres	Democrat	43.0	41.6	41.4
KS	pres	Republican	53.2	56.2	55.9
KS	pres	Other	3.8	2.2	2.7
KS	sen	Democrat	44.8	41.8	42.5
KS	sen	Republican	51.2	53.2	53.8
KS	sen	Other	3.9	5	3.7
KY	pres	Democrat	36.3	36.2	36.2
KY	pres	Republican	61.6	62.1	62.1
KY	pres	Other	2.1	1.7	1.7
KY	sen	Democrat	39.0	38.2	38.4
KY	sen	Republican	55.7	57.8	58.3
KY	sen	Other	5.2	4.0	3.3
LA	pres	Democrat	41.0	39.9	39.9
LA	pres	Republican	56.2	58.5	58.5
LA	pres	Other	2.8	1.7	1.7
LA	sen	Democrat	21.0	19.0	35.0
LA	sen	Republican	56.3	59.3	60.8
LA	sen	Other	22.7	21.7	4.2
MA	pres	Democrat	64.9	65.9	65.6
MA	pres	Republican	33.0	32.3	32.1
MA	pres	Other	2.1	1.8	2.3
MA	sen	Democrat	64.5	66.7	66.4
MA	sen	Republican	32.1	33.3	33.3
MA	sen	Other	3.4	0.0	0.3
MD	pres	Democrat	67.2	65.8	65.8
MD	pres	Republican	30.2	32.4	32.4
MD	pres	Other	2.6	1.8	1.8
ME	pres	Democrat	56.0	53.1	53.1

ME	pres	Republican	41.8	44.0	44.0
ME	pres	Other	2.2	2.8	2.9
ME	sen	Democrat	48.2	42.4	42.9
ME	sen	Republican	44.2	51.0	50.9
ME	sen	Other	7.6	6.6	6.2
MI	pres	Democrat	51.7	50.6	50.6
MI	pres	Republican	45.6	47.8	47.9
MI	pres	Other	2.7	1.5	1.5
MI	sen	Democrat	51.7	49.9	50.6
MI	sen	Republican	45.7	48.8	48.7
MI	sen	Other	2.7	1.9	0.7
MN	pres	Democrat	53.6	52.6	52.6
MN	pres	Republican	43.6	45.4	45.4
MN	pres	Other	2.9	2.0	2.0
MN	sen	Democrat	53.9	48.8	50.9
MN	sen	Republican	43.8	43.5	45.4
MN	sen	Other	2.3	7.7	3.8
MO	pres	Democrat	41.9	41.4	41.4
MO	pres	Republican	54.1	56.8	56.8
MO	pres	Other	4.1	1.8	1.8
MO	gov	Democrat	41.5	40.7	40.8
MO	gov	Republican	54.5	57.1	57.6
MO	gov	Other	4.1	2.2	1.6
MS	pres	Democrat	42.5	41.1	41.1
MS	pres	Republican	55.2	57.6	57.6
MS	pres	Other	2.3	1.3	1.3
MS	sen	Democrat	47.4	44.1	45.0
MS	sen	Republican	49.4	54.1	53.9
MS	sen	Other	3.3	1.8	1.0
MT	pres	Democrat	43.6	40.6	40.4
MT	pres	Republican	53.1	56.9	56.6
MT	pres	Other	3.2	2.5	3.0
MT	sen	Democrat	47.2	45.0	44.6
MT	sen	Republican	51.5	55.0	55.2
MT	sen	Other	1.3	0.0	0.3
MT	gov	Democrat	44.6	41.6	42.0
MT	gov	Republican	51.5	54.4	54.7
MT	gov	Other	3.9	4.0	3.4
NC	pres	Democrat	51.8	48.7	48.7
NC	pres	Republican	46.4	50.1	50.1
NC	pres	Other	1.8	1.2	1.2

NC	sen	Democrat	50.3	46.9	47.5
NC	sen	Republican	45.7	48.7	49.4
NC	sen	Other	4.0	4.4	3.1
NC	gov	Democrat	55.9	51.5	53.0
NC	gov	Republican	42.3	47.0	45.4
NC	gov	Other	1.8	1.5	1.7
ND	pres	Democrat	37.3	31.9	31.8
ND	pres	Republican	59.2	65.5	65.1
ND	pres	Other	3.4	2.6	3.1
ND	gov	Democrat	30.3	26.7	26.7
ND	gov	Republican	60.4	69.2	70.0
ND	gov	Other	9.3	4.1	3.3
NE	pres	Democrat	40.2	39.4	39.2
NE	pres	Republican	56.5	58.5	58.2
NE	pres	Other	3.2	2.1	2.6
NE	sen	Democrat	21.3	26.2	26.2
NE	sen	Republican	59.3	67.4	68.2
NE	sen	Other	19.3	6.4	5.6
NH	pres	Democrat	56.0	52.9	52.6
NH	pres	Republican	41.9	45.5	45.3
NH	pres	Other	2.1	1.6	2.1
NH	sen	Democrat	59.0	56.7	56.4
NH	sen	Republican	38.2	41	41.4
NH	sen	Other	2.7	2.3	2.3
NH	gov	Democrat	38.8	33.4	36.7
NH	gov	Republican	59.3	65.2	61.3
NH	gov	Other	1.9	1.4	1.9
NJ	pres	Democrat	60.4	57.3	57.3
NJ	pres	Republican	36.9	41.4	41.4
NJ	pres	Other	2.7	1.3	1.3
NJ	sen	Democrat	59.8	57.2	57.7
NJ	sen	Republican	36.2	40.9	41.4
NJ	sen	Other	3.9	1.9	0.9
NM	pres	Democrat	53.8	54.3	54.3
NM	pres	Republican	43.4	43.5	43.5
NM	pres	Other	2.7	2.3	2.2
NM	sen	Democrat	51.8	51.7	52.5
NM	sen	Republican	46.6	45.6	46.4
NM	sen	Other	1.6	2.6	1.1
NV	pres	Democrat	53.1	50.1	50.1
NV	pres	Republican	44.1	47.7	47.7

NV	pres	Other	2.8	2.3	2.3
NY	pres	Democrat	65.7	60.9	60.9
NY	pres	Republican	31.7	37.8	37.8
NY	pres	Other	2.6	1.4	1.4
OH	pres	Democrat	49.1	45.3	45.0
OH	pres	Republican	48.5	53.3	53.0
OH	pres	Other	2.4	1.4	2.0
OK	pres	Democrat	31.3	32.5	32.5
OK	pres	Republican	66.6	65.9	65.9
OK	pres	Other	2.0	1.6	1.6
OK	sen	Democrat	31.2	32.8	33.0
OK	sen	Republican	64.8	62.9	64.9
OK	sen	Other	4.0	4.3	2.1
OR	pres	Democrat	57.8	57.0	56.9
OR	pres	Republican	38.8	40.7	40.7
OR	pres	Other	3.4	2.5	2.5
OR	sen	Democrat	59.1	57.0	57.9
OR	sen	Republican	37.9	39.4	40.2
OR	sen	Other	3.1	3.7	1.9
PA	pres	Democrat	50.9	50.0	50.0
PA	pres	Republican	46.7	48.8	48.8
PA	pres	Other	2.4	1.1	1.3
RI	pres	Democrat	61.4	59.7	59.7
RI	pres	Republican	37.0	38.8	38.8
RI	pres	Other	1.6	1.5	1.5
RI	sen	Democrat	64.5	66.6	66.0
RI	sen	Republican	32.9	33.4	33.9
RI	sen	Other	2.6	0.0	0.1
SC	pres	Democrat	47.0	43.4	43.4
SC	pres	Republican	50.5	55.1	55.1
SC	pres	Other	2.5	1.5	1.5
SC	sen	Democrat	49.6	44.2	44.3
SC	sen	Republican	47.8	54.5	55.1
SC	sen	Other	2.6	1.3	0.7
SD	pres	Democrat	38.9	35.6	35.6
SD	pres	Republican	57.6	61.8	61.8
SD	pres	Other	3.5	2.6	2.6
SD	sen	Democrat	42.1	34.3	35.1
SD	sen	Republican	57.9	65.7	64.9
SD	sen	Other	0.0	0.0	0.0
TN	pres	Democrat	39.9	37.5	37.5

TN	pres	Republican	57.7	60.7	60.7
TN	pres	Other	2.4	1.8	1.9
TN	sen	Democrat	39.0	35.2	35.8
TN	sen	Republican	57.6	62.2	63.0
TN	sen	Other	3.4	2.6	1.3
TX	pres	Democrat	50.0	46.5	46.5
TX	pres	Republican	48.2	52.1	52.1
TX	pres	Other	1.8	1.4	1.4
TX	sen	Democrat	48.4	43.9	44.9
TX	sen	Republican	49.7	53.5	54.4
TX	sen	Other	2.0	2.6	0.7
UT	pres	Democrat	42.8	37.7	37.7
UT	pres	Republican	51.4	58.2	58.2
UT	pres	Other	5.8	4.2	4.2
UT	gov	Democrat	35.4	30.8	31.2
UT	gov	Republican	56.5	63.9	64.8
UT	gov	Other	8.0	5.4	4.0
VA	pres	Democrat	56.0	54.4	54.1
VA	pres	Republican	41.3	44.2	44.0
VA	pres	Other	2.6	1.5	2.0
VA	sen	Democrat	57.3	56.0	55.9
VA	sen	Republican	40.8	44.0	43.9
VA	sen	Other	1.9	0.0	0.2
VT	pres	Democrat	70.2	66.4	66.4
VT	pres	Republican	29.4	30.8	30.8
VT	pres	Other	0.4	2.8	2.7
VT	gov	Democrat	29.2	27.5	28.0
VT	gov	Republican	66.8	68.8	70.3
VT	gov	Other	4.0	3.7	1.7
WA	pres	Democrat	61.1	58.4	58.4
WA	pres	Republican	35.4	39.0	39.0
WA	pres	Other	3.6	2.7	2.6
WA	gov	Democrat	61.4	56.7	56.7
WA	gov	Republican	37.2	43.3	43.1
WA	gov	Other	1.5	0.0	0.2
WI	pres	Democrat	53.2	49.6	49.6
WI	pres	Republican	44.4	48.9	48.9
WI	pres	Other	2.4	1.5	1.5
WV	pres	Democrat	28.9	29.7	29.6
WV	pres	Republican	69.3	68.6	68.3
WV	pres	Other	1.8	1.6	2.2

WV	sen	Democrat	26.7	27.0	27.3
WV	sen	Republican	71.0	70.3	71.4
WV	sen	Other	2.3	2.7	1.2
WV	gov	Democrat	27.1	30.8	27.5
WV	gov	Republican	67.6	64.8	67.5
WV	gov	Other	5.3	4.4	5.0
WY	pres	Democrat	32.9	26.7	26.7
WY	pres	Republican	62.9	70.4	70.4
WY	pres	Other	4.2	2.9	2.9
WY	sen	Democrat	33.9	26.9	27.7
WY	sen	Republican	61.6	73.1	67.7
WY	sen	Other	4.5	0.0	4.6
GA	senrunoff	Democrat	50.0	50.6	50.6
GA	senrunoff	Republican	50.0	49.4	49.4
GA	senrunoff	Other	0.0	0.0	0.0
GA	sensprunoff	Democrat	50.3	51.0	51.0
GA	sensprunoff	Republican	49.7	49.0	49.0
GA	sensprunoff	Other	0.0	0.0	0.0

Appendix II: Likely Voter Models for 2020 and Summary of Results

Questions used for Likely Voter Model

LVA.

[TEXT IF NOT ELECTIONDAY]

How interested are you in the election on November 3rd in [STATENAM]?

[TEXT IF ELECTIONDAY]

How interested are you in the election taking place today in [STATENAM]?

RESPONSE OPTIONS:

1. Extremely interested
 2. Very interested
 3. Somewhat interested
 4. Only a little interested
 5. Not at all interested
-

LVB.

There are a range of reasons why people do or do not vote. We're interested in hearing from voters and non-voters. How likely are you to vote in the election?

RESPONSE OPTIONS:

1. Definitely will vote
 2. Probably will vote
 3. Probably will not vote
 4. Definitely will not vote
 5. I already voted
-

LV.

On a scale from zero to 10, where 10 means you're certain you will vote and zero means there is no chance you will vote, please indicate how likely it is that you will vote in this election.

RESPONSE OPTIONS:

1. 0-Certain will not vote
2. 1
3. 2
4. 3
5. 4
6. 5
7. 6
8. 7
9. 8
10. 9
11. 10-Certain will vote
12. Already voted

WVA.

[SHOW IF LVB=5 AND STATENAM<>OREGON OR (LVB=5 AND ELECTIONDAY=1 AND NOINEARLYVOT=0)]

[IF LVB=5 AND STATENAM<>OREGON AND ELECTIONDAY=1 AND NOINEARLYVOT=1, THEN FILL WVA=2]

Which best describes how you voted?

RESPONSE OPTIONS:

1. [IF STATE HAS EARLY IN PERSON VOTING] I voted in person at a polling place BEFORE Election Day
2. I voted by mail BEFORE Election Day
3. [IF ELECTION DAY] I voted in person at a polling place TODAY
4. [IF ELECTION DAY] I voted by mail TODAY

WVB.

[SP]

[SHOW IF LVB=1, 2, 3 AND STATENAM<>OREGON]

If you do vote in this election, how will you cast your ballot?

RESPONSE OPTIONS:

1. [IF BEFORE ELECTION DAY] I will vote by mail
 2. [IF BEFORE ELECTION DAY AND STATE HAS EARLY IN PERSON VOTING] I will vote in person before Election Day
 3. [IF BEFORE ELECTION DAY] I will vote in person at a polling place on Election Day
 4. [ELECTION DAY] I will vote by mail today
 5. [ELECTION DAY] I will vote in person at a polling place today
-

QPVVOTE.

In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have time. Which one of the following statements best describes you?

RESPONSE OPTIONS:

1. I did not vote in the 2016 presidential election.
 2. I thought about voting in the 2016 presidential election, but didn't.
 3. I usually vote, but I didn't in the 2016 presidential election.
 4. I'm sure I voted.
-

QPVVOTE18.

In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have time. Which one of the following statements best describes you?

RESPONSE OPTIONS:

1. I did not vote in the 2018 midterm election.
 2. I thought about voting in the 2018 midterm election, but didn't.
 3. I usually vote, but I didn't in the 2018 midterm election.
 4. I'm sure I voted.
-

QPVVOTE16G.

What about voting in the 2016 election for president? Which one of the following statements best describes you?

RESPONSE OPTIONS:

1. I did not vote in the 2016 presidential election.
 2. I thought about voting in the 2016 presidential election, but didn't.
 3. I usually vote, but I didn't in the 2016 presidential election.
 4. I'm sure I voted.
-

Specifications for Likely Voter Models

LV_alt14

- Likely voters: 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
- Likely voters: 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
- Likely voters: The respondent says they already voted to LVB or LV.

LV_alt16

- Likely voters: 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.
- Likely voters: 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.
- Likely voters: If 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 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.

Appendix III: Voter Classification Rates by State

State	Percent of Voters Correctly Classified
National	93
Alaska	94
Alabama	88
Arizona	92
California	94
Colorado	94
Connecticut	92
Delaware	94
Florida	93
Georgia	91
Iowa	96
Indiana	91
Kansas	95
Louisiana	90
Massachusetts	90
Maryland	95
Maine	93
Michigan	93
Missouri	93
Montana	96
North Carolina	94
Nebraska	96
New Mexico	94
Nevada	92
Ohio	94
Oregon	95
Pennsylvania	94
Tennessee	90
Virginia	95
Washington	95
Wisconsin	94