IF THE POLLS AIN'T FIXED:
DON'T BREAK THEM

What 2020-21 misses tell us about the future of pre-election polls

Patrick Murray

May 2022

MONMOUTH
UNIVERSITY

POLLING INSTITUTE

- Presidential polling in 2020 was considered a miss across the board, while 2021 pre-election polls offered a mixed bag. The AAPOR post-mortem of 2020 offered some possible explanations, but no definitive prescription. Despite these concerns, polling in the 2018 midterm and the 2021 Georgia Senate runoffs indicate that "horse race" polls can still accurately sample and model a population of future voters. What is problematic is whether they can do so (a) with enough consistency to stave off declining public trust in polling and (b) without resorting to arbitrary data manipulation to achieve a better "guess" of what the electorate will do.

- The analysis presented here uses validated voter turnout of respondents in pre-election polls drawn from list-based samples to assess whether, in hindsight, polls could have created a more accurate sample of actual voters in 2020. If so, are there indicators – either demographic or opinion-based – that pollsters can use *a priori* to correct for that bias? And if the underlying registered voter samples are accurate, can models be improved to identify only those who will actually vote in an upcoming election. Such fixes can include adjusting weights based on self-reports of motivation and enthusiasm and allocation of undecided voters.

- This study use validated voter information to evaluate sample accuracy in polls conducted by Monmouth University in eight different states during the final weeks before an election. This includes 2020 presidential polls in Georgia, Florida, Arizona, North Carolina, Iowa, and Pennsylvania and 2021 gubernatorial polls in New Jersey and Virginia.

- For the 2020 surveys, Monmouth matched post-election turnout records with the full poll sample in each state, including both respondents and non-respondents. The purpose of this match is to explore demographic factors that could have contributed to partisan bias in non-response. As such, each data set numbers tens of thousands of cases.

- For the 2021 surveys, Monmouth only matched the turnout record with respondents who completed the surveys in NJ and VA.

- For this analysis, "likely voter" results refer to the pre-election modeling of potential turnout. "Validated" results are vote choice results only for those respondents who actually turned out to vote.

2020 Pre-Election Poll Performance

State	Field dates	**Biden lead**				
		All registered voters (poll report)	Likely voter model range (poll report)	Validated voters (post-election)	Actual result	*Validated voters if all undecideds allocated to Trump*
PA	10/28-11/01	+ 5	+ 5-7	+ 6.4	+ 1.2	*+ 2.7*
FL	10/24-28	+ 5	+ 4-6	+ 5.0	− 3.4	*+ 2.9*
GA	10/23-27	+ 5	+ 2-4	+ 6.0	+ 0.3	*+ 5.2*
IA	10/15-19	− 1	+ 3-5	− 0.3	− 8.2	*− 2.5*
AZ	10/09-13	+ 6	+ 2-7	+ 2.9	+ 0.3	*+ 0.1*
NC	10/08-11	+ 3	+ 1-3	+ 2.2	− 1.3	*0.0*

Source: Monmouth University Poll

MONMOUTH UNIVERSITY | POLLING INSTITUTE

2020 Pre-Election Poll Performance

Findings:

Is this pollster's likely voter model range a representative depiction of eventual turnout among the samples of registered voters surveyed in each state?

- **Yes in 4 states, but no in GA (2 points above upper end of LV range) and especially IA (3 points below lower end of LV range)**

Is the sample of validated voters an accurate depiction of the actual electorate?

- **It depends on your error tolerance. Biden margin was off by as little as 2.6 points (AZ) and as much as 8.4 points (FL).**

Does allocating all undecideds to Trump eliminate error in candidate margin?

- **Again, it depends on error tolerance. This estimate would have been reasonably close in PA, AZ, and NC, but still wide of the mark in FL, GA, and IA.**

2020 Pre-Election Poll Performance

Findings:

"Winning margin" is not the most precise method of assessing survey error, but it is the most commonly used metric in media reporting – and thus it is the metric on which pollsters are called to account.

By this metric, there was a small, but systematic Democratic skew in each state based on the vote intentions of validated (i.e. actual) voters. In other words, even if likely voters models were 100% accurate, a partisan bias in the underlying sample was the main contributing factor to these polls being "off."

An arbitrary allocation of all undecided voters to Trump, puts the validated voter sample in 3 states (PA, AZ, NC) within what the media would portray as reasonable variance from the actual result. However, the basis for making such an allocation is not supported by the available data (e.g. there is no demographic or attitudinal evidence among undecided voters in these polls that would warrant making such a blanket allocation).

2020 Pre-Election Poll Regional Variations

Overstatement of Biden Lead based on 2016 vote by county				
	Entire state	Counties that voted for Trump by 10+ pts	Counties that voted for Clinton by 10+ pts	Swing counties
PA	+ 5.2	**+ 10.5**	+ 5.6	− 6.2
FL	+ 8.4	**+ 12.1**	+ 4.0	**+ 15.1**
GA	+ 5.7	+ 6.6	+ 0.4	+ 8.5
IA	+ 7.9	+ 6.5	**+ 15.8**	**+ 10.4**
AZ	+ 2.6	**+ 15.3**	+ 5.8	− 0.4
NC	+ 3.5	+ 4.2	+ 0.7	**+ 10.1**

MONMOUTH UNIVERSITY | POLLING INSTITUTE

2020 Pre-Election Poll Regional Variations

Findings:

There is no consistent pattern in terms of Trump voters being under-represented in the sample based on the dominant partisan tilt of county where they live (e.g. the skew was more prevalent in Trump supporting areas of PA, FL, and AZ, but not the other states).

This analysis also looked at early/mail versus Election Day voters in three of these states and found no consistent skew.

In effect, if we assume we could have accurately predicted the partisan sample skew by region and voting mode in these states, it still would not have eliminated the error.

	Unadjusted Biden lead	*"Corrected" for region/ vote mode*	Actual vote
PA	+ 6.4	*+ 4.7*	+ 1.2
GA	+ 6.0	*− 2.1*	+ 0.3
IA	− 0.3	*+ 0.4*	− 8.2

2020 Pre-Election Poll Error Recap

- **The 2020 presidential pre-election polls showed an across-the-board Democratic skew. In statistical terms, the skew was actually not all that large, but it was in the same direction in each sample.**

- **This skew was due to underlying voter sample bias (e.g. non-response) rather than flaws in likely voter modeling (among the polls examined here).**

- **Analysis of these six state polls (conducted by one organization) did not uncover a uniform correlate, or group of correlates, responsible for the skew.**

Note: In addition to the information presented here, the evaluation looked at other potential factors, including self-reported enthusiasm, ideology and demographic information from the voter file (e.g. gun ownership) – none of which showed a consistent correlation with survey non-response. However, the author acknowledges a more complex analysis design could point to such patterns.

MONMOUTH UNIVERSITY | POLLING INSTITUTE

Turning to 2021 gubernatorial elections in New Jersey and Virginia

Context: VA pre-election polls were seen as largely on-target while NJ polls "missed a Republican surge"

Was partisan non-response bias in the underlying sample to blame or were other factors at play?

MONMOUTH UNIVERSITY | POLLING INSTITUTE

2021 Pre-Election Poll Performance

Democrat lead for Governor					
	Weight A all reg. voters	Weight A validated voters	Weight B all reg. voters	Weight B validated voters	Actual result
NJ	+ 11.0	+ 5.6	+ 9.1	+ 2.9	+ 3.2
VA	− 0.2	+ 0.2	− 3.6	− 3.4	− 2.0

Source: Monmouth University Poll

Note: Weight A (the weight applied for the public pre-election releases of these polls) includes both standard demographic parameters and partisan adjustments based on 2020 regional turnout. Weight B relies on standard demographics only. Weight B is used for the analysis here.

MONMOUTH UNIVERSITY | POLLING INSTITUTE

2021 Pre-Election Poll Performance

Findings:

- **Candidate preference among actual (i.e. validated) voters in Virginia was similar to preference among all registered voters in the poll sample.**

- **Candidate preference among actual voters in New Jersey skewed 6 points more Republican than among all registered voters in the poll sample.**

- **Candidate margin among validated voters reflected actual election results in both states. Thus, underlying voter samples in both states were representative of the population (i.e. no systematic non-response bias).**

- **Perceived polling "miss" in New Jersey was a consequence of likely voter modelling not capturing partisan turnout differential.**

Factors in 2021 NJ/VA Turnout Differential

Is it possible to develop a data-driven likely voter model to "catch" the Republican turnout skew in the New Jersey poll without negatively impacting the accuracy of the Virginia poll – by identifying correlates that predict the differential turnout?

2021 NJ/VA Turnout Differential			
NJ turnout relative to VA (by group)		**Group share among reg. voter base**	
.79	TOTAL	NJ	VA
.55	Soft support for Dem candidate	10.5%	7.0%
.56	Infrequent/low vote history	16.9%	12.5%
.60	Less enthusiastic than usual	10.4%	9.6%
.72	Not motivated to vote	10.2%	6.0%

Factors in 2021 NJ/VA Turnout Differential

Findings:

- **NJ registered voter turnout (40%) was lower than VA (55%). The Monmouth polls showed a similar differential in verified turnout among its registered voter samples in the two states (proportional differential = .79).**

- **Soft Democrats – i.e. respondents who indicated they supported the Democratic candidate for governor but might change their mind, or only leaned toward that candidate – were significantly less likely to turnout in NJ than VA, relative to all other voters (e.g. firm supporters of the Democrat, all supporters of the Republican and independent candidates, and all undecided voters).**

- **For the registered voter sample, a low past voting <u>history</u> and expressing less <u>enthusiasm</u> about the 2021 election were better predictors than a question on voter <u>motivation</u> of the overall differential in NJ/VA turnout.**
  However…

2021 Turnout Factors by Vote Choice

Soft Democrat	NJ	VA	Δ
Low vote history	20.3%	21.5%	-1.2
Less enthusiastic	14.2%	13.6%	+0.6
Not motivated	26.8%	8.2%	**+18.6**

All other voters	NJ	VA	Δ
Low vote history	16.5%	11.9%	+4.6
Less enthusiastic	10.0%	9.3%	+0.7
Not motivated	8.2%	6.0%	+2.2

2021 Turnout Factors by Vote Choice

Findings:

- <u>Motivation</u> was more highly correlated with being a <u>soft Democrat</u> in the NJ poll (27% of soft Dems were "not motivated) than it was in VA (8%). There was no significant relationship between being a soft Democratic supporter and either low vote history or low enthusiasm in the two state samples. Furthermore, there was no difference between these factors (history, enthusiasm, motivation) and any other type of voter in the two states.

- Unmotivated voters made up 10% of NJ sample and 6% of VA sample overall. Entirely eliminating unmotivated voters from the likely voter model only improved the NJ margin by 1.8 points, at the same time increasing the VA error by 0.9 points.

- Adjusting motivation for the interaction of demographic factors correlated with turnout (e.g. race, education) does not significantly improve LV model.

MONMOUTH UNIVERSITY | POLLING INSTITUTE

Summary of Analysis

In line with post-election analysis conducted by **AAPOR** and others, the 2020 presidential polls analyzed here showed a consistent skew toward the Democratic candidate, which was due to an unexplained bias in the underlying voter sample.

This partisan bias did not appear to be an issue for publicly reported polling in elections immediately before 2020 (i.e. 2018 midterms) or after (i.e. 2021 Georgia runoffs). Furthermore, it was not evident in the 2021 NJ and VA gubernatorial polls examined here.

The apparent miss in the 2021 NJ poll was caused by an unusual partisan turnout pattern that was not captured by predictive likely voter models (i.e. it was not due to any flaw in the survey's fundamental sampling methodology).

MONMOUTH UNIVERSITY | POLLING INSTITUTE

Conclusions

Pre-election polling during the 2018 midterms appeared to have corrected for the 2016 presidential election miss. However, a similar partisan skew reappeared in 2020 that raised questions about the validity of election polling as a whole. Poll accuracy in the next cycle suggest this non-response bias surfaces when Donald Trump appears on the ballot – but that is not an adequate causal explanation.

The 2020 polling error fed an already growing public skepticism about polling accuracy. Election polling by design has always experienced, and will always experience, occasional errors. However, in a political environment where facts are denied, it has become even more difficult than in the past for pollsters to convey the underlying uncertainty inherent in polling and get the public's "benefit of the doubt" in return.

Survey research is at its best when it measures the here and now. Likely voter modelling attempts to push past these methodological underpinnings to predict future behavior. But the public does not distinguish between when polls go astray due to the tenuous nature of turnout modelling versus a miss due to systematic sampling error. A miss is a miss in the public's eye.

Conclusions (continued)

On the whole, survey research continues to be sound (e.g. polling on Covid vaccinations and other verifiable behaviors). Unfortunately, the dominant image of polling is the "horse race." And those polls are perceived as getting it wrong more and more often.

Pollsters may be tempted to apply ad hoc or arbitrary adjustments based on errors identified in the last election cycle. They may get the next cycle right and claim to have found a magic formula, despite a lack of evidence that their modifications were based on data-driven assumptions. When these pollsters eventually get it wrong – as they inevitably will – it further feeds the image that pollsters are putting their thumb on the scales.

We need to avoid *breaking* the polling industry by appearing to *fix* the polls. The analysis presented here does not provide a methodological solution for this dilemma. But if the current trend continues, pollsters will be called on more often to defend their discipline. The problem is that these election misses not only undermine public faith in election polls, but in the entire survey research enterprise.

For More Information

The datasets used in this study are available to researchers interested in undertaking further exploration of these issues.

Each data set is appended with information from the relevant state voter file along with inferred demographics (such as race), geocoding, vote history (including post-election turnout verification for the 2020 or 2021 election), and limited data on other political correlates where available (such as gun ownership, political contributions, etc.).

The 2020 datasets also include voter file information on tens of thousands of potential respondents who ether refused or were unable to be contacted for that particular poll.

Contact: _polling@monmouth.edu_