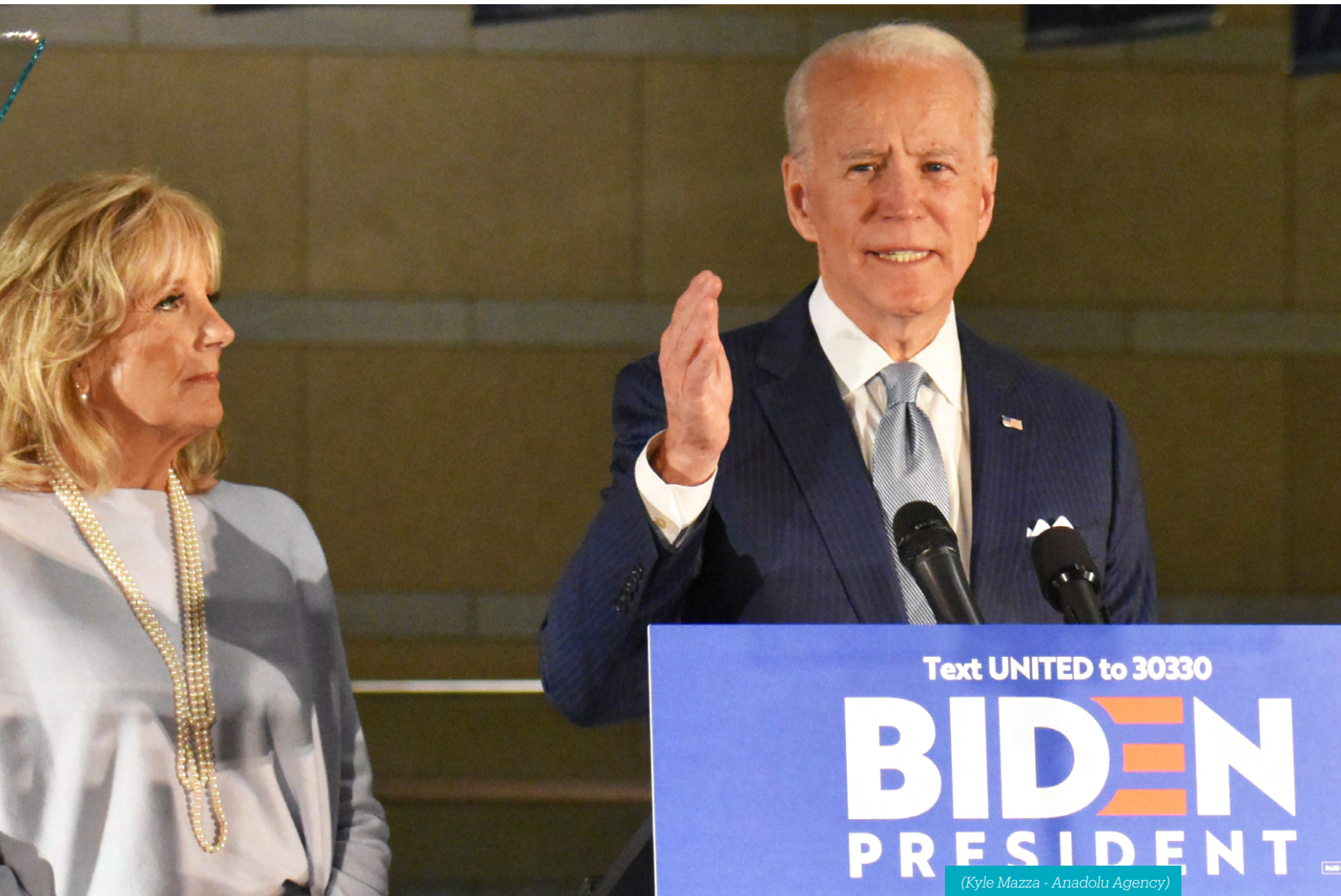


The Politics of Polling

Why the winner of the 2020 US Presidential Election is a 'known unknown'

Ravale Mohydin



History of Polling in the US

Applying the [Johari window model](#) to polling science, there are the 'known knowns', defined as things that we know we know. Then there is the concept of a 'known unknown', which is defined as something we know we do not know. Predicting election outcomes is a process of carefully navigating between the known knowns and the known unknowns, and this process is, at times, affected by the 'unknown unknowns' - or the things we do not know we do not know. This is perhaps what gives polling research its lasting appeal.

Attempts to predict election outcomes have been made since as long as elections have been held. In the United States (US), where the first presidential election to be decided by popular vote was held in 1824, unscientific polls in the form of trial heat tallies¹ took place in mostly informal settings such as taverns and public meetings. This continued until the 1936 Presidential election when George Gallup conducted the first ever quota-selected polls. The results were published in the January 1937 edition of *Public Opinion Quarterly* and explained how "George Gallup's [quota-controlled survey](#) of a few thousand triumphed over the Literary Digest's straw poll of millions in correctly predicting the election outcome". Many people, however, were unconvinced of the ability to forecast an election based on a sample. As one expert mused: "Is it possible to [sample public opinion sufficiently](#) accurately to forecast an election, particularly a close one?" i.e. how can a small sample of survey respondents reflect the opinions of millions? People only became convinced of the power of statistical theory to predict an entire population's voting patterns based on a representative sample when quota-based surveys became [part of election campaigns](#) in the 1960s. Since then, there has been an explosion in the number of US election polls. According to research, there was an estimated [900 per cent increase](#) in trial heat polls between 1984 and 2000.

The nature of election polling itself has evolved. From conducting polls via telephone on behalf of politicians or media organisations to conducting internet surveys and disseminating information directly online, not only have preferred methods changed but also the [initiative](#) is no longer restricted to media organisations or political campaigns and "journalists are no longer the [formal gatekeepers](#) determining if a given poll is of sufficient quality and interest to warrant the public's attention". Technology has changed the way not only of how polls are conducted but also how they are disseminated and used.

In terms of functionality, polls are being used for three different purposes in the US: to forecast election outcomes, understand voter behaviour and plan campaign strategy.

Media coverage, candidate behaviour and voter behaviour are all [shaped by polling](#), and while there has been remarkable growth in the polling industry based on [significant success](#), there have also been some noteworthy misses. For example, polls predicted that incumbent Democratic president Harry Truman would lose to Republican Thomas Dewey in the 1948 US election, but he won instead². Even though CBS News correctly predicted Bill Clinton as the winner in 1996, they incorrectly forecast the winning margin, predicting an 18-per cent margin of victory compared to the actual 8 per cent. Gallup predicted the [wrong winner](#) of the popular vote in 2000 by overestimating support for minor-party candidates, while most 2008 polls predicted wrongly that Barack Obama would beat Hillary Clinton in the New Hampshire Democratic primary. All these polling results impacted both campaign strategy as well as campaign and voter behaviour, with significant and far-reaching socio-political consequences, making this a [high stakes business](#). Tellingly, polling for the next presidential election now begins the day after the previous one ends.

Achieving Accuracy

Why do some polls get it right while others fail disastrously? To start, considering that many entrepreneurs looking to self-promote initiate and fund election polls, motives can be questionable, a fact highlighted by disgruntled pundits who may not agree with findings from such polls. Given that internet surveys - the most commonly used these days - are typically opt-in, which results in nonprobability samples³, the quality of predictions can be compromised. Besides that, given that a sample is polled rather than the entire population, a random sampling error can arise, which has to be mitigated by [providing margins of error](#) that may invite criticism of imprecision or calls to increase the sample size, which may be too costly. Besides random sampling errors, there are other types of errors also, including [systematic ones](#) such as not polling a group that exists within the population but not the sample, which can introduce bias into the polling results and thereby skew the forecast. Systematically excluding households that do not have a telephone landline, typically households with a lower socioeconomic status is an example of coverage bias⁴. Such systematic errors, potentially reflecting systemic biases and discrimination, are of greater concern.

Research design is an important factor when it comes to election polling. Crucially, 'no amount of analysis can fix what was [bungled by design](#)'. Methodological design decisions including the timing of the study, weighting certain factors such as race and education, [question-wording or order](#) can all play a role in the accuracy of results. Having a [survey on a weekday versus a weekend](#) can lead to a difference in how many and also the type of person who responds versus who does not, leading to a nonresponse

¹ Asking a [respondent how he or she is going to vote](#) in a given race at the time.

² Most experts believed Truman would lose but [voter sentiment and turnout changed in the final weeks before the election](#), ensuring Truman's win over Dewey.

³ Samples that may not be truly reflective of the population.

⁴ Research has found such studies to be biased against Democratic candidates.

error. Data quality assurance is hard in a polling activity, particularly in the case of pre-election polling when “[every poll is...a sample of a sample](#)” as the sample is taken from an unknown population as there is no way to accurately predict who will actually [turn up to vote](#) on election day itself. Not every election is the same with different cross-sections of voters making an impact in different elections, which makes the selection of who is [likely to vote](#) in a particular election very significant. In 2008, for example, an unusually large number of young and minority voters turned out to vote for Barack Obama. Likelihood of voting is hard to predict in the US, particularly if using the same model for all 50 states that vary in voter registration requirements and early voting rules, among other factors. Besides the likelihood of voting, undecided voters are another unknown and contribute to polling error. As per research, the accuracy of polling is not improved by reducing the proportion of [undecided voters](#) in the sample.

Pre-election polls can also feature respondent sourced errors such as not providing honest answers to both questions of intent to vote as well as whom to vote for, resulting in reduced accuracy as far as outcome predictions are concerned. Extensive research has documented [overreporting of turnout](#) (and turnout intention), primarily the result of social desirability bias. Voting intention is not said to be particularly stable either. Research shows that more than [40 per cent of voters](#) change their mind at least once during the election campaign. While some experts believe that these shifts are predictable in the sense that they would happen every election in the same way as voters gravitate toward their [fundamental](#) political positions or can be attributable to [sampling error](#), others claim that [certain events or ‘campaign shocks’](#) can also contribute, particularly during the late stages of a campaign. As evidenced, there are many explanations as to which data can vary as well as why it does when it comes to polls, but it can be difficult to decipher this data variance as many survey firms do not practice methodological transparency.

To reduce volatility in polling numbers, news organisations tend to aggregate individual poll results. Though this can be effective when ‘house effects’⁵ are accounted for carefully, how it is [done](#) can also lend to imprecision. Ignoring house effects and simply taking the average of all available data is considered naive as ‘the various sources of bias in the individual poll numbers will cancel each other out, but [if biases work in the same direction](#), aggregation will not reduce bias’.

There are alternatives to poll-based predictions. One way is looking at macroeconomic, or market indicators, such as economic performance, grounded in the theory that incumbents are re-elected if such indicators are [positive](#) and are not if they are negative. [Experts](#) contend that this methodology, being based on voting behaviour theory, is more reliable than poll-based predictions. However, in the

US, they only “typically predict the two-party popular vote rather than the [Electoral College](#)⁶ outcome”, which is arguably not enough. It is also claimed that these theory-based models are not as [powerful](#) in terms of predictions because they do not take into account the current political candidates’ [messaging](#) about the economy and their attention to other issues.

Prediction markets are another election-forecasting alternative. These predictions are based on expectations about the winner from “those willing to put [money](#) behind their judgments”. These are essentially betting markets where traders buy and sell candidate futures based on whom they think will win the election. For example, in September 2008, Obama shares were selling for 0.60 USD which meant he had been given a [60 per cent](#) chance of winning in November 2008. Research shows that these can be [more accurate](#) than both poll and market-based predictions. However, traders are [not representative](#) of all voters and are only assumed to make informed judgements based on, among other factors, the same poll numbers that suffer from all the inconsistencies highlighted above.

From the point of view of voter behaviour, a significant source of measurement error comes from [turnout over-reporting](#), which studies show [exaggerates](#) the impact of factors such as education and religiosity on turnout. [Social desirability bias](#) is another potential source of measurement error, with survey participants reporting being more politically engaged and informed than they actually are.

Associated with accuracy when it comes to prediction, establishing a causal relationship using survey research alone is very hard. For example, while there may be a correlation, there is no way to know whether being a Democrat causes you to support a certain issue. In order to manage the measurement errors and to establish causality as described above, some experts and political parties have considered [panel surveys](#)⁷, experimental studies in the form of field experiments as well as using information technology to develop enormous databases to not have to generalise from a sample that can be cross-referenced with other governmental data to understand voter behaviour and plan accordingly.

Lessons from the 2016 Presidential Election

If there are many ways to achieve an accuracy as described above, why were the results of the 2016 US Presidential Election so unexpected? Almost all pollsters forecast Hillary Clinton to win decisively over Donald Trump. In fact, [two renowned forecasters](#) predicted that Clinton had a 99 per cent chance to win in 2016, based on polling research leading up to the election. In terms of how

⁵ All methodological decisions made by any survey firm.

⁶ In the Electoral College system, [each state gets a certain number of electors based on its total number of representatives in Congress](#). Each elector casts one electoral vote following the general election; there [are a total of 538 electoral votes](#). The candidate that [gets more than half \(270\) wins the election](#).

⁷ Surveys that track the same individual over time.

the US Presidential Election system works, each state gets a certain number of electors based on its total number of representatives in Congress (as per the US Electoral College system). Clinton lost electors in key states that went to Trump instead, resulting in her losing the election even though she won the popular vote. This makes state polls critical when it comes to predicting election outcomes. Predictably, there was a significant backlash against the same polling companies and experts, with most media outlets blaming them for their incompetence and [‘failed algorithms’](#). In fact, only [one notable major public poll](#) was able to correctly predict a Donald Trump win.

Post-election analysis showed that education - a decisive component of presidential vote preference in 2016 - was not accounted for by many pollsters, [particularly in state polls](#). This proved to be disastrous when it came to predicting the winner based on the Electoral College as described above. College-educated voters are more likely to respond to surveys about voting preferences than non-college-educated voters, and college-educated voters were more likely to indicate they would vote for Clinton than Trump. Some experts also believed that Trump voters were not as likely to respond to pre-election surveys as Clinton supporters due to potentially lower levels of [civic engagement](#). Both of these factors may have led to over-representation of voters who indicated they would vote for Clinton. The fact that Clinton only had a [4 point lead](#) over Barack Obama among college-educated voters in pre-election polls in 2012 may have been why education was not considered that important a factor in the 2016 election polling strategy.

Most state polls were not weighted by education

	Weighted by education		Unknown	Did not weight by education
National polls	49%		28	23
Battleground states	32	20	48	
Battleground non-online	14	19	67	

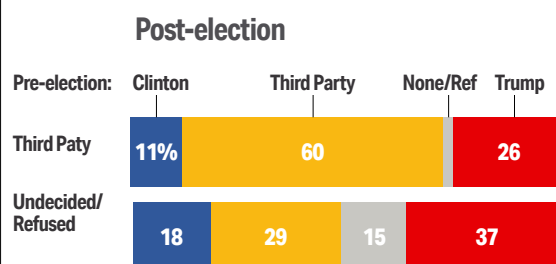
Source: Upshot analysis of polls used by the AAPOR post-election report.

Retrospectively speaking, many experts believe that assuming a large number of undecided voters were likely to vote for Clinton did not [reflect reality](#), as many chose to vote for Trump instead. Multiple survey companies re-contacted people they had surveyed after the election and found that a larger percentage than the [usual number](#) of undecided and minor-party voters voted for Trump. Apparently, “the [systematic failure](#) of dozens of surveys across several states suggest that...(this) was due to a real shift in the closing days of the race, rather than an earlier, more fundamental error”. The ‘real shift’ was attributed by

many experts to a letter by [FBI Director James Comey](#) to Congress on 28 October 2016 about the existence of Clinton’s official emails sent from a private server that needed to be investigated. According to Nate Silver, a prominent pollster, it may have ‘shifted the race by [3 or 4 percentage points](#) toward Donald Trump, swinging Michigan, Pennsylvania, Wisconsin and Florida to him, perhaps along with North Carolina and Arizona’.

Post-election surveys say Trump won undecided voters

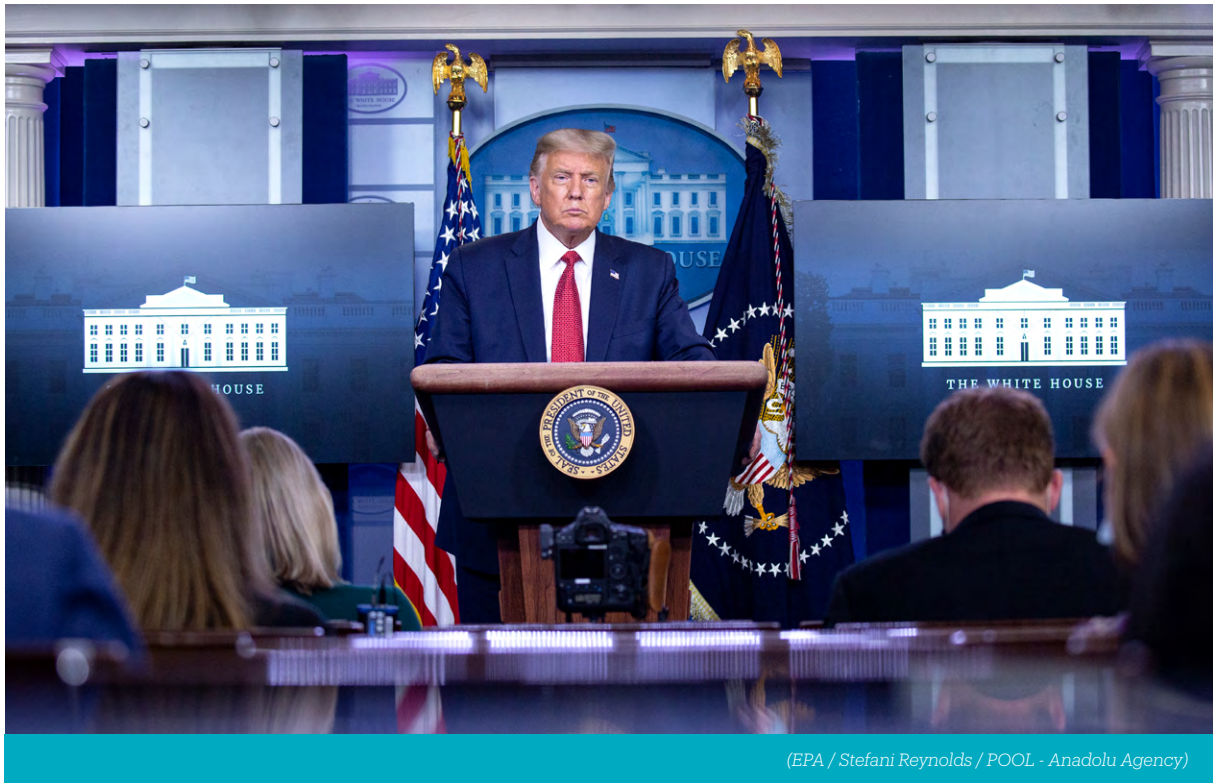
Voters who said they were undecided or supported a third-party candidate in pre-election surveys were likelier to say they voted for Mr. Trump.



Source: Pew research, AAPOR report on 2016 polling; Global Strategy Group, presented at NY AAPOR. For concision, the chart is an average of the two findings.

The postelection shift was also attributed to the [‘shy Trump’](#) effect. As delineated above, self-report measures including [social desirability bias](#) are important to consider as survey participants may have been embarrassed to divulge their support for a socially undesirable candidate a.k.a. Trump: perhaps [the undecided voters were not undecided at all](#), they just did not want to admit their choice to an unknown stranger working for a polling company until their chosen candidate won. Moreover, turnout for Trump was underestimated, particularly in rural counties in traditionally Republican states. On the other hand, as mentioned above, a significant source of measurement error comes from turnout overreporting. Clinton’s supporters were more likely than Trump’s supporters to [stay home](#) after indicating their intention to vote.

As highlighted above, there was only one major public poll, the USC/Los Angeles Times Daybreak poll, which accurately predicted Trump’s victory in 2016. As indicated above, to manage the measurement errors highlighted when reflecting upon 2016 polling inaccuracies, some experts and political parties used panel surveys. The USC/Los Angeles Times Daybreak also used panel surveys, tracking the same people over time. The poll also correctly detected, unlike other major polls, Trump’s appeal to a [key bloc of voters](#) who strongly favoured him: conservative whites who had sat out the 2012 election but intended to vote in 2016. It also did not ask survey respondents to choose between candidates but instead asked them to



(EPA / Stefani Reynolds / POOL - Anadolu Agency)

rate, on a 0-100 scale, the [probability](#) of them voting for any candidate, thereby avoiding forcing voters to decide before they had decided, resulting in false positives for the Clinton campaign, and capturing this ambiguity that other polls did not. Interestingly, the creator of the USC/Los Angeles Times Daybreak poll was a Danish scientist, uninterested in political science or activism, whose election prediction results were an [outgrowth](#) of his main research interest on how people make decisions on their finances". This example is a testament to polling research's ability to accurately predict election outcomes no matter how implausible it may have seemed at the time, provided the research methodology is well-thought-out and preferably void of sentiment.

Alternatively, macroeconomic, or market, indicators such as economic performance, grounded in the theory that incumbents are [re-elected](#) if such indicators are positive and are not elected if they are negative, were more accurate. Studies have shown that the strength of the S&P 500 Index has accurately predicted [every election](#) outcome since 1984 and 87 per cent since 1928: specifically speaking, when stocks are going up, the [incumbent tends to win](#) the White House. Tellingly, the S&P 500 was down since July 2016 onwards, meaning the Democrat party - the incumbent party at the time led by President Barack Obama - would lose and Donald Trump from the Republican Party, according to this metric, would win in 2016, which he did. Prediction markets such as the Iowa Electronic Markets (IEM), another election-forecasting alternative, also predicted a much tighter race than expected: Clinton's chances of winning the election were around 90 per cent till the FBI news broke. After the FBI letter, "those chances

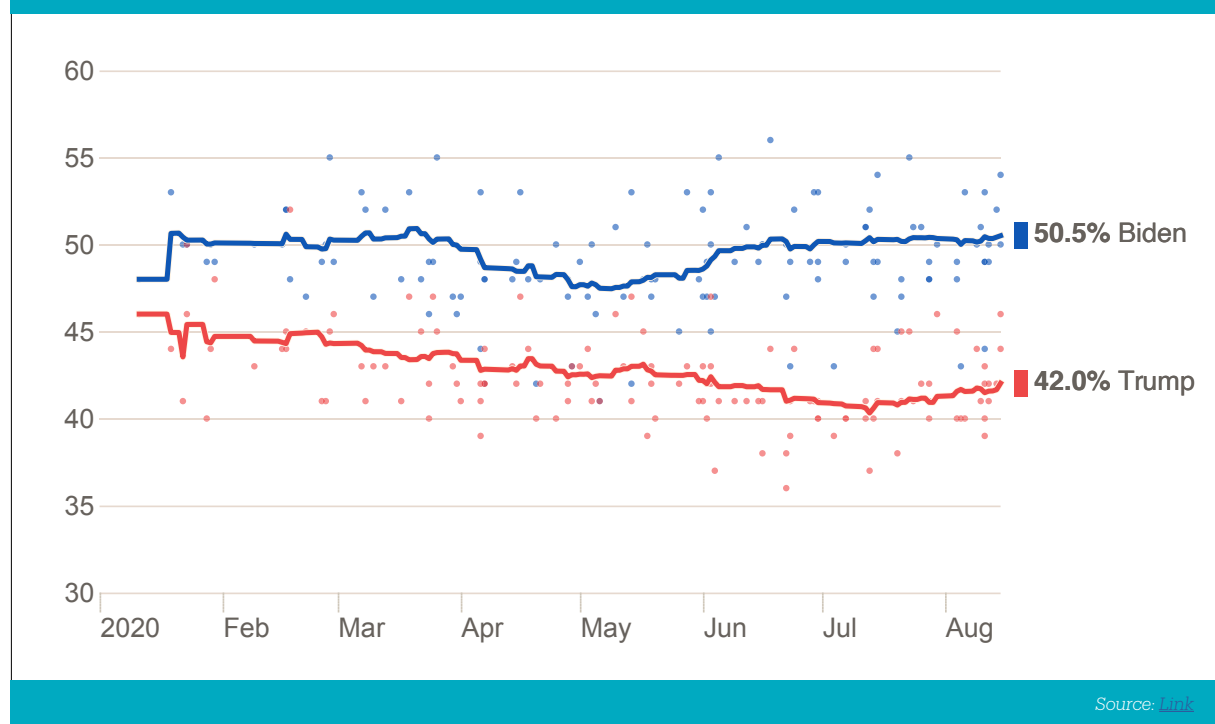
[plunged](#) to around 68 per cent. GOP candidate Donald Trump's chances rose correspondingly, from 10 to around 32 per cent".

Predicting the 2020 US Election winner?

The year 2020 is not a normal election year by most standards. The novel coronavirus pandemic has swept the world with the US being the worst affected country on the planet with more than 5 million cases at the time of writing. The impact of the coronavirus pandemic and the resulting economic and social fall out cannot be underestimated. In [Florida](#), where Covid-19 cases have surged in recent weeks, Joe Biden, the presumptive Democratic nominee, leads Trump by 6 percentage points. In typically Republican states such as Texas, polling results show that the race is tightening and promises to be a close one in November 2020. White senior citizens, a group that was very supportive of Trump in 2016, have seemingly disapproved of the Trump Administration's handling of the pandemic.

This is already being reflected in national polls where Biden is taking an ever-growing lead over Trump. However, after Trump's surprise win in 2016, it would be reasonable for Americans as well as members of the international community to take these polling results with a grain of salt. There are still more self-identified Democrats than Republicans in most surveys, at times by a wide margin, similar to 2016. Furthermore, the education gap among Trump and Biden supporters persists, the effects of which may not be

COMPARING BIDEN AND TRUMP IN THE NATIONAL POLLS



sufficiently captured in the polling results, particularly in states with a large number of white working-class voters have a big effect when it comes to the Electoral College⁸ system.

However, strong signs are pointing to a Biden victory in November. Firstly, Trump has never polled at more than 46 per cent. This entails a loss of the popular vote, which may result in a loss in the Presidential Election. Only [five](#) US Presidents thus far have lost the popular vote and won the election in US history. Trump is one of them, but there is no guarantee that he will win the same way twice. Secondly, [minor party voters](#) that helped divide the Democrat vote as well as helped Trump in the Electoral College are likely to be less than they were in 2016 (6 per cent of voters voted minor party then) with third-party candidates being much less prominent this election cycle than the last one. Thirdly, considering that there was a fierce backlash against Clinton's Presidential bid in 2016 that Trump benefited from, voters do not report [disliking Biden](#) to the same extent. However, even though Biden reportedly works closely with Senator Bernie Sanders⁹, there appears to be a significant proportion of Democrat voters who believe that the [Biden-Harris](#)¹⁰ the ticket is [not appealing](#) to the Democratic party's progressive base and may not vote at all. A divided Democrat vote could help Trump immensely.

Additionally, even though Biden has an impressive lead over Trump in the polls, voting patterns may [change](#) from

now until November 2020, particularly with regards to the coronavirus and its impact on the economy. As noted above, the strength of the S&P 500 Index has accurately predicted election results: when the S&P 500 is [trending higher](#) as of three months before the election, the incumbent party usually wins. As the economy recovers from the coronavirus lockdowns in the US, the S&P 500 is bound to trend higher. In August 2020, three months out from the US election, [the S&P 500 is reaching an all-time high](#). When it comes to the predictive markets, the IEM showed Biden had an approximately 60 per cent chance of winning by early July 2020, while Trump had a 40 per cent chance. In early August 2020 at the time of writing, Trump's chances have risen to [43 per cent](#). Whether Biden can maintain his considerable lead over Trump is yet to be seen.

It is clear from the analysis above that the polling results are not clear. While there is merit in the concept of well thought out research methodology accurately predicting results, the fact that voters' behaviour is going to be impacted by the unprecedented nature of the issues surrounding the 2020 US Presidential Election – making this an unknown unknown – is likely to make forecasting difficult. As the year unfolds in terms of how the novel coronavirus affects the US economy and society, the trajectory of the election is also likely to change. At this point, the winner of the 2020 US Presidential Election is truly a known unknown.

⁸ In the Electoral College system, [each state gets a certain number of electors based on its total number of representatives in Congress](#). Each elector casts one electoral vote following the general election; there [are a total of 538 electoral votes](#). The candidate that [gets more than half \(270\) wins the election](#).

⁹ Senator Bernie Sanders is a democratic socialist who made a second run for the White House in 2020. He withdrew after a [series of primary losses to Biden](#).

¹⁰ California Senator Kamala Harris was chosen by Biden to be his running mate for the 2020 Presidential Election.