

# Uncertainty Intervals to Quantify and Communicate Polling Estimates

Brandon Steelman (✉ [steelman@cleargenes.com](mailto:steelman@cleargenes.com))

Clear Gene

---

## Article

**Keywords:** election polls, predictive models, confidence intervals, uncertainty

**Posted Date:** November 13th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-99741/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The accuracy of predictive models is constrained by the underlying data. It can be challenging for those who did not develop a predictive model to objectively determine whether the uncertainty of the underlying data is consistent with the claimed probability estimates. Here we distinguish two complementary sources of uncertainty: the resolution of the measurement method and the statistical variation of repeated measurements. Although election polls were widely criticized following the 2016 US Presidential Election, state-level polls correctly predicted the outcome for individual states, but only when the statistical confidence interval lay outside the method-specific uncertainty interval. When confidence intervals overlapped with the uncertainty interval, the results were uncertain: both candidates won some of the uncertain state-level contests. Some elections will be too close to call. Estimating the amount of uncertainty in the data alerts data modelers to overly confident predictions and helps improve the way we explain uncertainty.

## Introduction

Uncertainty can be challenging to quantify scientifically and communicate to general audiences. The probability of a prediction is constrained by the inherent uncertainty of the underlying data. Yet, in the era of big data, complex predictive models can obscure the uncertainty of the underlying data, even to the scientists developing those models. For example, during the 2016 U.S. Presidential Election, some predictions claimed a greater certainty than the underlying data could support. Strategies that quantify uncertainty in underlying polling data can help to identify overly confident predictions, irrespective of the uncertainty in predictive models that may be built from those data. There will continue to be cases where the underlying data does not support a confident prediction. The general public's confidence in social sciences is bolstered when scientists can recognize and communicate simply that an election result is too close to call.

A number of debates about polling predictions have been misguided. Widespread proclamations if incorrect polls fall short in two ways: first, they do not distinguish the underlying polling data ('the polls') from the predictive models that are built from meta-analyses of polling data. Models sometimes erroneously claim more certain predictions than permitted by the data, a modeling error that should not be blamed on the data. Second, general proclamations misframe the central question by implying that uncertainty is a flaw. All scientific measurements are uncertain. Measurements obtained by traditional light microscopy can have different resolutions than those obtained by electron microscopy. Election polling is no different, and, like microscopy, some polling methods have more resolution and produce more reliable measurements. The central question of certainty is whether the data obtained by a given method can answer the relevant research question.

We can approach the question of certainty by distinguishing between two complementary sources of uncertainty. Polling is a class of scientific measurements with a given resolution. Here we implement uncertainty intervals to address method-specific uncertainty. In addition, meta-analyses necessarily

combine polling studies are across time and location. Here we use confidence intervals to address the mathematical uncertainty from repeated measurements. Notably, arithmetic measures of variation in a meta-analysis do not indicate whether the measurand has been resolved to a degree that exceeds the resolution of the measurement method. We therefore classify data as uncertain if the method-specific uncertainty interval overlaps the statistical confidence interval. Together, the combination of method-specific uncertainty intervals and statistical confidence intervals identify whether an election is too close to call based on the available data.

There is an additional level of uncertainty when election results are not based on the popular vote and are instead determined by multiple sub-decisions. This notably occurs in parliamentary elections and the US Presidential Election. Uncertainty in these cases presents a particularly pernicious problem because polling errors can include systematic biases that violate assumptions of independence for each sub-decision, and because well powered meta-analyses require multiple polls for each sub-decision. Both of these additional challenges can be addressed by combining method-specific uncertainty intervals and statistical confidence intervals. The data for a location, for example a US state, are uncertain if the statistical confidence intervals overlap the method-specific uncertainty interval.

## Results

### 2016 US Presidential Election

State-level polls correctly predicted the outcome for individual states—but only when the statistical confidence interval lay outside the method-specific uncertainty interval. Fig.1 and Table 1 show that state-level results were uncertain when confidence intervals overlapped the uncertainty interval: both candidates won some of the uncertain state contests. Leading up to the 2016 election, most forecasts predicted a 90% likelihood that Mrs. Clinton would win, with some claiming a 99% probability. The unappealing statistical reality was that 12 states were too close to call. These states carried 165 electoral votes, and neither candidate had enough certain electoral votes to reach the 270 votes that are required to win the electoral college. Notably, these analyses were performed and disseminated privately prior to the election.

### 2020 US Presidential Election

Polling results prior to the 2020 US Presidential election indicate a different scenario. There are enough certain electoral votes to support confident electoral predictions.

Fig. 2 and Table 2 show that 23 of the states or districts that favor Mr. Biden have confidence intervals outside of the method-specific uncertainty interval of 3.5%. These states provide at least 270 electoral votes that Mr. Biden requires to win the US electoral college. In contrast, the same cannot be said for Mr. Trump. Table 3 lists the states that favor Trump and that have confidence intervals outside of the method-specific uncertainty interval of -3.5%. The electoral votes from these states are not sufficient to conclude that Mr. Trump will claim at least 270 electoral votes. Mr. Trump would not have enough votes

to win the electoral college, even if he were to claim all of the certain states that favor him in Table 3 (126 electoral votes) and all of the states in Table 4 that are too close to call (139 electoral votes).

The number of state-level polls increases in the weeks leading up to an election. The quality and method of polling will be variable. There are not currently enough polls in every state to exclude poorly ranked polls. However, in the states that are classified as uncertain, there are enough polls to repeat the analysis using only well-regarded polls. Limiting the analysis to polls that were graded A, B, or B/C does impact the mean spread between candidates. Table 5 shows that Ohio leans more heavily towards Mr. Trump and Florida and Nevada lean more heavily towards Mr. Biden when poorly ranked polls are excluded from Table 4. However, limiting the dataset to well-regarded polls does not ultimately impact the certainty of the prediction: all states that are categorized as uncertain in Table 4 are still categorized as uncertain in Table 5.

## Discussion

There are uncertainties inherent to a scientific measurement. And there are statistical uncertainties from meta-analysis of repeated studies. The resolution of a scientific method becomes negligible when the resolution exceeds that required to answer the scientific question. But for somewhat blunt measuring devices like opinion polling, the resolution cannot be treated as a negligible variable. Polling uncertainty therefore contains two components: the uncertainty of the polling method itself, and the statistical uncertainty of repeated measurements.

Polls were widely criticized following the 2016 US General Election. A report from the American Association for Public Opinion Research (AAPOR, 2017) noted that “[t]he day after the election, there was a palpable mix of surprise and outrage directed towards the polling community, as many felt that the industry had seriously misled the country about who would win.” However, state-level polls had correctly predicted the outcome for individual states—but only when the statistical confidence interval lay outside the method-specific uncertainty interval. When confidence intervals overlapped with the uncertainty interval, the results were uncertain: both candidates won some of the uncertain state-level contests.

Prior to the 2016 election, we selected  $\pm 3.5\%$  to reflect recent polling errors. For example, Obama beat his polls by 3 points in 2012, and Republicans beat their polls by 4% in the 2014 midterms. Although this was informed by historical datapoints, one limitation of this study was that selecting the method-specific uncertainty interval involved a subjective element. Jennings and Wlezien (2018) subsequently performed a more sophisticated historical analysis of polling errors. Based on 175 presidential elections, they estimated a mean absolute polling error of 2.7 percentage points. This analysis focused on national-level polls and are not directly applicable to sub-national polls that we refer to as state-level polls. Since the analysis by Jennings and Wlezien (2018) spanned multiple countries, additional research may be required to answer whether the uncertainty interval should be tailored to a specific country or system of election.

Polls attempt to infer population-level attributes from a subset of the population. One concern is whether the error rates of modern polls are stable. Jennings and Wlezien (2018) found that the election year had a trivial, non-significant effect ( $P = 0.85$ ) when modeling the absolute error as a dependent variable. This indicates that there has been no discernible decline in the accuracy of polls over time and that it is feasible to set a generalizable method-specific uncertainty interval. However, the selection of uncertainty intervals may still vary by country, polling frequency, quality or polling method, and they do not necessarily need to equal the mean absolute polling error.

The goal of this study was to illustrate that a method-specific uncertainty interval is distinct from and complementary to statistical measures of uncertainty. The uncertainty intervals provide a strategy to assess the reliability of the data that is independent of the predictive models that are then applied to those data. Predictions themselves can be improved in a number of ways. This study did not involve sophisticated predictive models. For simplicity, this analysis treated all state-level contests as digital, even though some states do not award all electoral votes to the candidate who wins the majority of votes in that state. Predictions could also be improved by selecting measures of central tendency other than the mean, and by estimating statistical confidence intervals using non-parametric methods like the bootstrap. For simplicity, this study extracted 95% confidence intervals from a function for a two-sided t-test. These are parametric estimates that assume normally distributed data, an assumption that is easily violated.

This study illustrates a broader limitation of the polling field: the paucity of reliable polling data for every state. Although there was enough data to limit the 2016 analysis to high-quality polls, the 2020 analysis was forced to include a number of polls that were graded D or D-, including polls conducted over the internet. Note that specifically excluding polls rated D or D- did not remove ungraded polls. This was intentional because the timeline for posting a grade and the criteria for an ungraded poll remained unclear to us. We decided to include all polls because the following states or districts did not have enough reliable polls to eliminate poor quality polls: Arkansas, Connecticut, District of Columbia, Delaware, Hawaii, Idaho, Illinois, Indiana, Louisiana, Nebraska, North Dakota, Oregon, Rhode Island, South Dakota, Tennessee, Utah, Vermont, Washington, West Virginia, Wyoming. This finding highlights a need for additional, quality polling data so that the certainty can be more reliably addressed. It is notable that poor-quality polls were noted in the American Association for Public Opinion Research report on the 2016 election polling (AAPOR, 2017). AAPOR proposed several solutions, and yet the number of high-quality state-level polls has subsequently decreased to the point that we were forced to include poor-quality polls in some states in 2020.

Many pollsters use predictive models to estimate probabilities. Predictive models often include their own probability estimates that are based on permutations of the existing data. Permutations and models are reliable, however, only when projections capture the uncertainty of both the underlying data and the uncertainty of the model itself. Modelling can be used to determine an error rate, which in turn can be used to calculate probabilistic forecasts. But modelling cannot generate a forecast that has more certainty than the aggregate uncertainty of the underlying data. As data modeling plays a more central role in science and society, we need to emphasize the distinction between the underlying data and the

models that are derived from them. Polling data should not necessarily be blamed for the incorrect, or overly confident, modeling predictions. In 2016, some models claimed a certainty of 99% (AAPOR, 2017) even though this analysis showed that 30.7% of the electoral-college vote was too close to call. More importantly, no candidate could claim with certainty to reach 270 electoral college votes. The 2016 results were uncertain; it would have been more accurate to report these as too close to call. We cannot claim to make highly certain predictions from highly uncertain data.

In contrast to 2016, the 2020 polling data provide confidence that one candidate will reach the minimum of 270 votes required to win the US electoral college. Although 139 (23.4%) of the 538 available electoral votes are classified as uncertain, the data themselves support more confident predictions because one candidate has reached the minimum requirement of 270 votes in the electoral college based on state-level results that are outside the uncertainty interval.

The goal here is to identify when the underlying data are too uncertain to claim highly certain predictive models, something that is often obscured even from those who develop the models.

Conversely, confidence in the data is not an endorsement of the predictive models themselves. The fact that the data support a certain prediction does not imply that every prediction made using those data will be valid. Variability between models can be attributed to different modeling methods, training methods, decisions used to select training and input data, and the methods used to determine the probability of those results. All of these differences may be scientifically valid, and they may lead to differing predictions.

We need to improve the way we explain uncertainty: uncertain data are not wrong, only uncertain. The unappealing statistical reality is that polls are sometimes too close to call. While general and sophisticated consumers alike often bristle when data scientists present an uncertain conclusion, it is a disservice to make overly confident predictions on inherently uncertain data. Leading up to the 2016 election, many pollsters understood that the election was close, and that there was a high degree of uncertainty. But most did a poor job of articulating that uncertainty. Simply reporting margins “ $\pm 3\%$ ” does not adequately convey the underlying uncertainty. In cases where the underlying data does not support a confident prediction, the general public’s confidence in social sciences is bolstered when scientists can recognize and communicate simply that an election result is too close to call. As the complexity of data models increase, data scientists can help non-specialists become better consumers of data models. When cases of uncertainty arise, we have an opportunity to educate consumers about the limitations of data measurements and data modeling.

## Methods

### Data Source and Preprocessing

Five Thirty Eight compiled polling data for the 2016 and 2020 US Presidential Elections. We retrieved data from the 2016 US Presidential Election from <http://projects.fivethirtyeight.com/general->

model/president\_general\_polls\_2016.csv. Data from the 2020 US Presidential Election were retrieved from [https://projects.fivethirtyeight.com/polls-page/president\\_polls.csv](https://projects.fivethirtyeight.com/polls-page/president_polls.csv).

We performed data analysis in *R*, most recently *R* version 3.6.2 (R Core Team, 2019). We used tools from the *dplyr* (Wickham, François, Henry and Müller, 2020) and *tidyr* (Wickham and Henry, 2020) packages for data manipulation. We used the *knitr* (Xie, 2014; Xie, 2015; Xie, 2020) package to generate reports.

### Inclusion/Exclusion Criteria

Our selection criteria for these meta-analyses limited our analysis to polls that contained state-level data. The reliability of each poll was evaluated by Five Thirty Eight. For baseline analyses of the 2016 data, we selected polls that with grades A+, A, A-, B+, B, B-, C+, C, C-, D and polls without a grade. The grading scheme changed slightly for the 2020 data; we therefore selected polls with grades A+, A, A-, A/B, B+, B, B-, B/C, C+, C, C-, C/D, D+, D, D-, and polls without a grade.

For our meta-analysis of the 2016 US Presidential Election, we selected the 20 most recent polls for each state. For our meta-analysis of the 2020 US Presidential Election, we selected polls that ended before May 1, 2020 because the Democratic race coalesced around Mr. Biden after Mr. Sanders withdrew on April 8.

### Setting the Method-specific Uncertainty Interval

Data collection and analysis can be divided into three phases: data collection, combining repeated measurements, and building a predictive model. The goal of the method-specific uncertainty interval is to explicitly acknowledge that a given technology or method of data measurement has an inherent resolution, and that the imprecision of the data collection method is distinct than both the statistical variations that are quantified for repeated measurements and the probability estimates associated with a given data model.

Training and setting cut-off thresholds are often the least scrutinized aspect of a model. Thresholds can be set based on existing data, but they should be prospectively tested on independent data. Prior to the 2016 elections, we selected a method-specific uncertainty interval for polling data of  $\pm 3.5\%$ . We selected this threshold to reflect recent US polling errors. For example, in the 2012 Presidential Election Obama beat his polls by 3 points. And in the 2014 midterm election Republicans beat their polls by 4 points. After the 2016 US Presidential Election, we prospectively validated the uncertainty interval by labeling each US state based on the outcome of the election.

### Calculating Statistical Confidence Intervals

The goal of the statistical confidence interval is to estimate the variation of the second phase of data collection and analysis: combining repeated data measurements. Here we treat the mathematical uncertainty from meta-analyses of multiple studies as a distinct source of variation to underscore. We are attempting to underscore that this variation is unique from the resolution of the data collection method.

There are robust options for quantifying variation on this level: from traditional parametric tests to non-parametric resampling estimates like  $n$ -fold cross-validation and the bootstrap.

For these analyses, we used parametric methods to estimate confidence intervals. We estimated the arithmetic mean,  $p$ -values and 95% confidence intervals by performing a two-sided Student's  $t$ -test on a vector of polling results using the *t.test* function in the base *R stats* package (R Core Team, 2019).

Combining method-specific uncertainty intervals and statistical confidence intervals

We classified the data for each state or district as certain or uncertain based on whether the statistical confidence intervals for the mean polls in that state were entirely outside the method-specific uncertainty interval. For example, in the 2016 US Presidential Election, both Nevada and Florida favored Mrs. Clinton by a tiny margin (Table 1), but the 95% Confidence Intervals overlapped the 3.5% uncertainty intervals. Prior to the election, both states were classified as uncertain; indeed Mrs. Clinton won Nevada and Mr. Trump won Florida. The same method-specific confidence interval that classified 12 states were classified as uncertain in 2016 only classified 8 states as uncertain in 2020 (Table 1, Table 4).

Tables and Figures

We used the *ggplot2* (Wickham, 2016) and the *ggthemes* (Arnold, 2019) packages to visualize data. We used the *xtable* (Dahl, Scott, Roosen, Magnusson and Swinton, 2019) package to generate tables.

## References

*An Evaluation of 2016 Election Polls in the US* (AAPOR, 2017); [http://www.](http://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx)

[aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-](http://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx)

[Polls-in-the-U-S.aspx](http://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx)

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

Hadley Wickham and Lionel Henry (2020). *tidyr: Tidy Messy Data*. R package version 1.1.1. <https://CRAN.R-project.org/package=tidyr>

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). *dplyr*:

A Grammar of Data Manipulation. R package version 1.0.2.

<https://CRAN.R-project.org/package=dplyr>

Jennings and Wlezien (2018). Election polling errors across time and space.

*Nature Human Behavior*. <https://doi.org/10.1038/s41562-018-0315-6>

R Core Team (2019). R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. URL

<https://www.R-project.org>

Yihui Xie (2020). knitr: A General-Purpose Package for Dynamic Report

Generation in R. R package version 1.29.

Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition. Chapman and

Hall/CRC. ISBN 978-1498716963

Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible Research in R.

In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, Implementing

Reproducible Computational Research. Chapman and Hall/CRC. ISBN 978-

1466561595

David B. Dahl, David Scott, Charles Roosen, Arni Magnusson and Jonathan

Swinton (2019). xtable: Export Tables to LaTeX or HTML. R package version

1.8-4. <https://CRAN.R-project.org/package=xtable>

Jeffrey B. Arnold (2019). ggthemes: Extra Themes, Scales and Geoms for

'ggplot2'. R package version 4.2.0. <https://CRAN.R-project.org/package=ggthemes>

## Tables

Please see the supplementary files section to view the tables.

## Figures

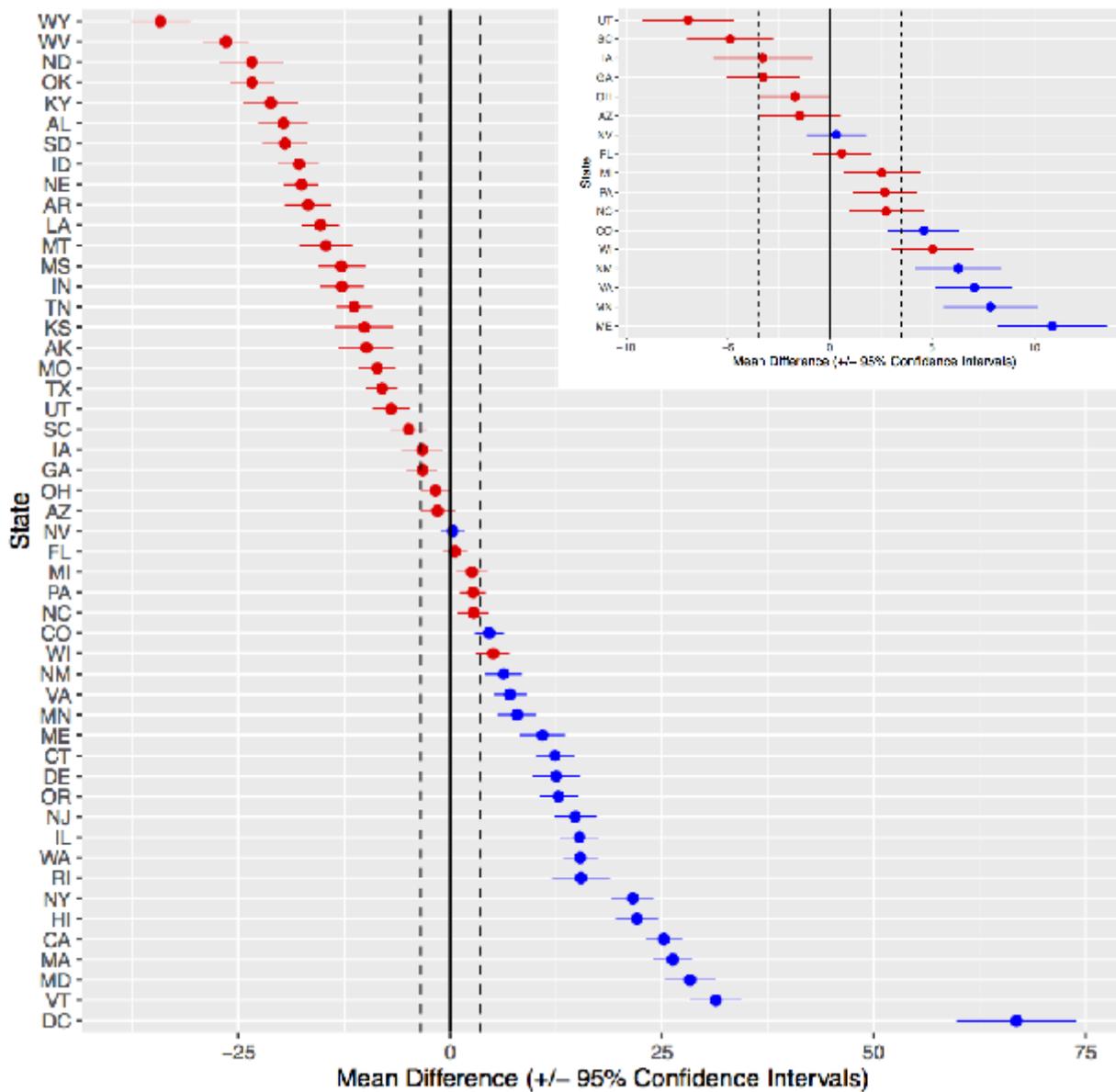


Figure 1

State-level polling data collected and analyzed prior to the 2016 US Presidential Election. For each state a dot represents a measure of central tendency, in this case the mean difference between candidates is plotted along the abscissa. A mean difference less than 0 was arbitrarily chosen to favor Trump and a difference greater than zero favored Clinton. The plot presents two levels of uncertainty. The colored horizontal bars that flank the mean of each state represent the 95% confidence intervals derived from parametric estimates. Method-specific uncertainty intervals are represented as dashed lines. In this case, prior to the 2016 election, we selected general uncertainty intervals of  $\pm 3.5\%$  to reflect polling errors of 3-4% in recent US elections. After the election, states were colored red or blue based on the outcome. States in red were won by Trump. States in blue were won by Clinton. The inset shows states where the statistical confidence interval was close to the method-specific uncertainty interval. State-level polls correctly predicted that outcome for individual states, but only in cases where the statistical confidence interval was outside of the method-specific uncertainty interval. Conversely, results were uncertain when statistical confidence intervals overlapped the method-specific uncertainty interval. Both candidates won some of the uncertain state-level contests.

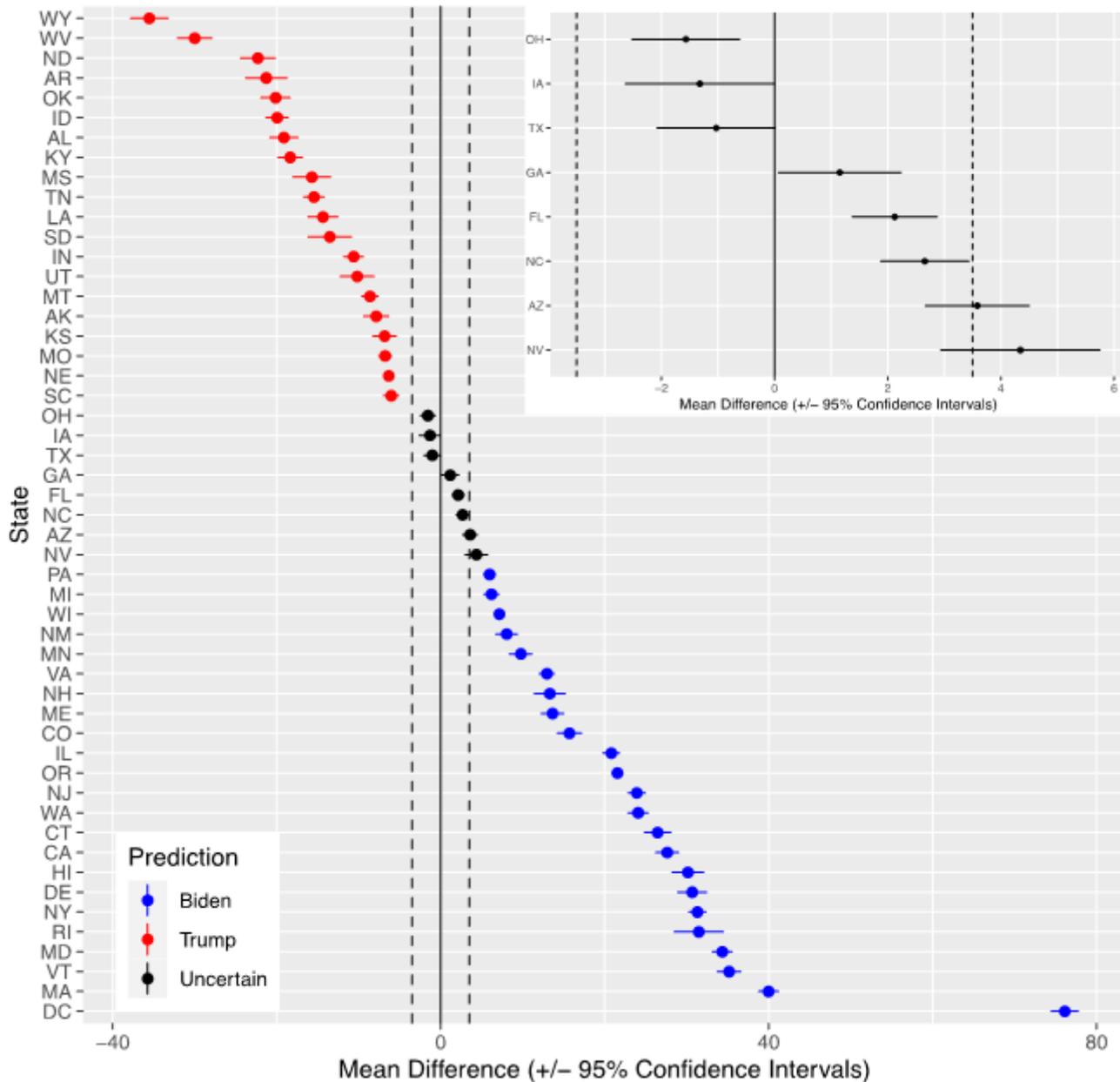


Figure 2

State-level polling data collected and analyzed prior to the 2020 US Presidential Election. The mean difference between candidates is represented as a filled dot. A mean difference less than 0 was arbitrarily chosen to favor Mr. Trump and a difference greater than zero favors Mr. Biden. The colored horizontal bars that flank the mean of each state represent the 95% confidence intervals derived by parametric estimates. Method-specific uncertainty intervals are represented as dashed lines. In this case, we used the general uncertainty intervals of  $\pm 3.5\%$  that we selected prior to the 2016 election. States are colored black if the statistical confidence interval overlaps the method-specific uncertainty interval. States with more certain predictions are colored red if they favor Trump or blue if they favor Biden. The inset shows states where the statistical confidence interval overlaps the method-specific uncertainty interval.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.docx](#)