Peter K. Enns*, Julius Lagodny and Jonathon P. Schuldt
Understanding the 2016 US Presidential Polls: The Importance of Hidden Trump Supporters

https://doi.org/10.1515/spp-2017-0003

Abstract: Following Donald Trump’s unexpected victory in the 2016 US presidential election, the American Association for Public Opinion Research announced that “the polls clearly got it wrong” and noted that talk of a “crisis in polling” was already emerging. Although the national polls ended up being accurate, surveys just weeks before the election substantially over-stated Clinton’s lead and state polls showed systematic bias in favor of Clinton. Different explanations have been offered for these results, including non-response bias and late deciders. We argue, however, that these explanations cannot fully account for Trump’s underperformance in October surveys. Utilizing data from two national polls that we conducted in October of 2016 ($n > 2100$ total) as well as 14 state-level polls from October, we find consistent evidence for the existence of “hidden” Trump supporters who were included in the surveys but did not openly express their intention to vote for Trump. Most notably, when we account for these hidden Trump supporters in our October survey data, both national and state-level analyses foreshadow Trump’s Election Day support. These results suggest that late-breaking campaign events may have had less influence than previously thought and the findings hold important implications for how scholars, media, and campaigns analyze future election surveys.

1 Introduction

10:30pm: Trump Showing Unexpected Strength in Battleground States
11:40pm: Trump Takes Florida, Closing In on a Stunning Upset
2:30am: TRUMP IS ON THE VERGE OF A STUNNING UPSET
2:50am: TRUMP TRIUMPHS: Shocking Upset as Outsider Harnesses Voters’ Anger
(The New York Times, Nov. 8–9)

*Corresponding author: Peter K. Enns, Associate Professor, Department of Government, Executive Director, Roper Center for Public Opinion Research, Cornell University, Ithaca, USA, e-mail: peterenns@cornell.edu
Julius Lagodny: Department of Government, Cornell University, Ithaca, USA
Jonathon P. Schuldt: Department of Communication, Cornell University, Ithaca, USA
These changing Election Night headlines from the New York Times website chronicle the unfolding surprise of Donald J. Trump’s presidential victory. The American Association for Public Opinion Research reacted to Trump’s unexpected election by announcing, “The polls clearly got it wrong… and already the chorus of concerns about a ‘crisis in polling’ have emerged.” Even Mr. Trump expressed surprise at the outcome (Jacobs and House 2016). In hindsight, however, Trump’s victory should not have seemed so implausible given the polls. Although many forecasts based on the polls indicated a very high probability of a Clinton victory (Katz 2016), the national polls ended up being quite accurate. And already 10 days before the election, ABC News reported that the presidential race had tightened, with Clinton ahead of Trump by just two percentage points (Langer 2016).

Looking back, a bigger puzzle seems to be why surveys conducted in mid-October, just weeks before the election, substantially underestimated Trump’s electoral support. Typically, presidential polls already reflect the final vote margin three to 4 weeks before the election. As Erikson and Wlezien (2012: p. 66) explain, “during the heat of the campaign in October into early November, very little happens to change the national verdicts.” Shirani-Mehr et al. (2016: p. 8) similarly note, “Average [polling] error … appears to stabilize in the final weeks, with little difference in RMSE [root mean square error] one month before the election versus one week before the election.” The deviation in 2016 from these historical trends begs the question: why did Trump perform so poorly in the October polls?

The worst-case scenario for pollsters is non-response bias, which occurs when those who participated in the survey have systematically different attitudes and opinions than those who were not contacted or who refused to participate (Groves 2006). Non-response bias means the sample is not representative of the public, and thus by definition, the survey will produce biased estimates of public opinion. Although weighting survey responses can help alleviate concerns associated with non-response bias, weighting assumes that those with particular demographic characteristics who answered the survey share the same preferences (on average) as those with the same characteristics who did not answer the survey. Concerns about non-response bias have risen as industry-wide response rates have declined, in many cases below 10 percent (Pew Research Center 2012). Many scholars and prominent analysts have pointed to non-response bias as a

3 Some poll-based forecast also reflected a close race. Real Clear Politics only projected Clinton to win the Electoral College by 6 votes (Trende 2016).
possible explanation for why the polls (especially state polls) underestimated Trump support in the 2016 US presidential election (Linzer 2016; Mercer et al. 2016; Silver 2016b; Shepard 2017). Yet, despite declining response rates, research suggests that polls continue to “provide accurate data on most political, social, and economic measures” (Pew Research Center 2012: p. 1) and errors in presidential polling have not increased as response rates have dropped (Franklin 2016). Furthermore, non-response bias seems inconsistent with the accuracy of the final national polls and with evidence that the state polls ended up performing as well or better in 2016 than 2012 (Trende 2016).

The recent AAPOR report on the 2016 Election Polls in the US offers an alternative explanation for why October polls underestimated Trump support: “late deciding” voters (AAPOR 2017). The careful analyses in the report make clear that multiple factors influenced state and national election polls. Yet, the report’s focus on late deciding voters that broke for Trump is particularly relevant for understanding Trump’s underperformance in October polls. The report concludes, “There is evidence of real late change in voter preferences in Trump’s favor in the last week or so of the campaign, especially in the states where Trump won narrowly” (p. 52).

Although some individuals may certainly have waited until the last minute to decide to vote and for whom, we offer a different explanation that we believe more fully accounts for why the October polls substantially underestimated support for Donald Trump: the existence of “hidden” Trump supporters. We define hidden Trump supporters as survey respondents who did not directly express intent to vote for Trump in the survey, but who nevertheless appeared as if they would support Trump on the basis of their responses to other survey questions. Of course, it is impossible to pinpoint exactly when these respondents decided to vote for Trump, and thus we cannot rule out the possibility that these respondents were simply late deciders. We can, however, show that hidden Trump supporters were detectable in October surveys, which holds important implications for forecasting based on polling data as well as for campaign strategy during the final weeks of an election contest. Furthermore, when we consider these hidden supporters, polls in early October come within one percentage point of the actual final vote share and we can correctly predict the winner of 5 of the seven swing

---

4 Many final state polls systematically underestimated support for Trump (see, for example, Kohn et al. (2016) and Drew Linzer: https://twitter.com/DrewLinzer/status/804728243035119620), in part, because some polls failed to weight by education level (AAPOR 2017). Of course, state polls often show greater variability than national polls and a greater proportion of state polls rely on IVR (Interactive Voice Response) and internet polling methods (Hillygus 2011: p. 970), so some difference between state polls and national polls is to be expected.
states with 10 or more electoral college votes. Thus, if the hidden Trump supporters were simply late deciders, they were not deciding equally between Trump and Clinton. Instead, almost a month before the election, those who did not indicate a vote intention for either major party candidate already leaned disproportionately toward Trump.

Our article proceeds as follows. First, we introduce alternative factors beyond non-response bias and late deciders that could explain systematic underestimation of Trump support: social desirability bias and “top of the head” considerations (e.g. Taylor and Fiske 1978; Zaller 1992). We then explain how the nature of the 2016 campaign could lead both of these factors to produce “hidden” Trump supporters who completed the surveys but did not openly declare their support for Trump. We test our expectations with three separate analyses based on 14 state polls and two national surveys conducted during October 2016. Supporting our expectations, the results suggest that more Trump supporters were in the data than standard vote intention questions revealed. In addition to tempering concerns about non-response bias and a “crisis” in polling, as we discuss in the conclusion, these results hold important implications for future analyses of election surveys.

2 The Case for Hidden Trump Supporters

During the 2016 election, surveys consistently found that the percentage of respondents who indicated they were undecided or choosing a third-party candidate was much larger than in previous presidential elections (Linzer 2016; Silver 2016a). Consistent with this finding, our own surveys conducted in October (described below) found that around 20 percent of respondents did not report a voting intention for either Clinton or Trump. The unusually large proportion of respondents who did not express support for either of the two major-party candidates led us to wonder on the eve of the election whether hidden Trump supporters might exist among this group (Enns and Schuldt 2016). Considering Trump’s racist (O’Connor and Marans 2016), sexist (Bahadur 2015), and bigoted comments (Moreno 2015), it seemed especially plausible that some respondents may have felt hesitant to declare their support of him out of a fear of being judged negatively – a phenomenon known to psychologists and pollsters as social desirability bias (e.g. Nederhof 1985). Indeed, survey respondents sometimes “hide” behind the label “independent” to avoid associating with either of the two parties – particularly when those parties are stigmatized (Connors et al. 2016; Klar and Krupnikov 2016). The same process may have occurred with Trump supporters.
Some have argued that a lack of difference between online and phone surveys suggests that social desirability bias was not a factor for Trump supporters (e.g. Dropp 2016). We must remember, however, that although web-based surveys may reduce social desirability bias, there is no evidence that it eliminates such bias entirely (Kreuter et al. 2008); both modes may be sensitive to social desirability considerations. Furthermore, news coverage of Trump’s offensive comments could have affected survey responses in another way. Enns and Richman (2013) show that expressed vote intentions in surveys often reflect different considerations than the final vote choice. Furthermore, they show that the difference between a vote intention expressed in a survey and the actual vote choice does not necessarily result because voters learn new information about the candidates. Instead, some respondents treat the survey question differently than the vote choice, bringing different information to bear.\footnote{Enns and Richman’s (2013) findings are also consistent with research that shows that even among those who report that they are undecided, it can be possible to detect the ultimate vote choice (e.g. Arcuri et al. 2008; Galdi et al. 2008).} In the present case, we might expect that the extremely negative news associated with Trump [e.g. his mocking a disabled reporter or boasting that he groped women against their will (Carmon 2016; Graham 2016)] would create strong “top of the head” (Taylor and Fiske 1978; Zaller 1992) considerations among Trump supporters. If so, these considerations may have led some who would eventually vote for Trump to tell pollsters that they were undecided or voting for a third party.

The key point is that either social desirability bias or top of the head considerations, or some combination thereof, could have led a subset of Trump supporters to refrain from indicating that they would vote for Trump. If so, we would expect Trump to have underperformed other Republican candidates in state polls. We would also expect that, if we could observe the true preferences of those who did not express a vote intention for either Trump or Clinton, the October polls would more closely reflect the final vote share on Election Day. We test these predictions below.

### 3 State Polls Underestimated Trump Support in October

We have argued that Trump supporters participated in the surveys, but negative news at the time kept some of these supporters from directly expressing their intention to vote for Trump. If our argument is correct, in states with a Senate
election, we would expect state polls to systematically underrepresent support for Trump relative to the Republican candidate for Senate. By contrast, if non-response bias explains why Trump’s vote share was underestimated, we would expect support for both Trump and the Republican Senate candidate to be underestimated by roughly equivalent amounts. To test this prediction, we consider October polls from 14 states that had Senate elections. 6 We calculate the percent indicating support for Trump and the percent indicating support for the Republican candidate for Senate and then compare these values to the actual vote share each received in the state. 7

For example, in the October poll in Arizona that we analyzed, Trump received 48.8 percent of the two-party vote intentions. Trump won Arizona with 51.9 percent of the two-party vote, indicating the poll underestimated Trump’s vote by 3.1 percent. We cannot, however, take this alone as evidence of hidden Trump supporters. Enten (2016) has shown that state polls underestimated Republican support in general. Thus, we must consider Trump’s performance relative to the performance of Republican Senate candidates in the same poll. In this case, the Republican candidate for Senate (John McCain) received 59.1 percent of two-party vote intentions in the poll and he won Arizona with 56.9 percent of the two party vote. In other words, McCain’s October vote support exceeded his final vote share by 2.2 percent. Subtracting McCain’s over-performance in the poll (2.2%) from Trump’s underperformance (−3.1%) indicates that relative to the Republican Senate candidate, Trump underperformed by 5.3%. If Trump systematically underperformed Republican candidates for Senate in the October polls, we take that as evidence that some of those indicating they would vote Republican in the Senate race chose not to indicate support for Trump.

Figure 1 presents Trump’s performance relative to the Republican candidate for Senate in the 14 states with relevant poll data. Negative bars indicate Trump underperformance and positive bars indicate Trump over-performance in the state polls. The figure indicates that Trump systematically underperformed. Ten of the states have the expected negative values, whereas just four states show that Trump exceeded expectations. Furthermore, Trump’s average underperformance is more than 2.5 times greater than the average of the four states where he over-performs. In other words, not only were state polls more than twice as likely to underestimate support for Trump, but the magnitude of the difference was much

---

6 State polls were selected based on the following criteria: a sample of likely voters, live telephone interviews, in the field in early October, and included vote intentions for the presidential and Senate candidate. If more than one survey in a single state met these criteria, we selected the survey conducted closest to October 1. Full survey details are provided in the appendix.

7 All actual vote share data come from Leip (2016).
greater in states that underestimated his support. Even after accounting for differences in actual vote share, Trump’s poll numbers underperformed those of other Republican candidates.8

We should note that the AAPOR report on the 2016 Election Polls conducted a nearly identical analysis; however, the polls analyzed were quite different. The

---

8 Enten (2016) notes that Trump’s vote share outperformed state polls most in red states where Trump was most popular. He suggests that this pattern runs contrary to the hidden Trump supporter hypothesis because he proposes that in these states, respondents should have felt the least social desirability bias when expressing support for Trump. We think three factors complicate this interpretation. First, even where Trump was popular, some respondents may have felt concerned about directly expressing support for someone who had mocked a disabled reporter or who crudely described grabbing a woman. Second, as we noted above, social desirability bias is not the only mechanism that could produce hidden Trump supporters. Finally, and perhaps most importantly, because these states had the highest proportion of Trump supporters, by definition, they had the highest potential for hidden Trump supporters. Even if the tendency to not express support for Trump was lower in red states, the aggregate implications could be larger because Trump had the highest proportion of supporters in these states.
AAPOR report only considered surveys from the final 2 weeks of the campaign, and if a survey firm conducted multiple state-level surveys in the last 2 weeks, only the final poll was analyzed. Based on these data, the report finds no evidence that Trump systematically underperformed in state polls, which the report takes as evidence against what the report calls the “shy Trump hypothesis.” The timing of the AAPOR analysis holds important implications. The accuracy of the final national polls indicates that if hidden Trump supporters existed, they had emerged by the end of the campaign. Indeed, as we saw in the ABC report above, some national polls had aligned with the final vote outcome almost 2 weeks before the election. Based on these results, we would not expect hidden Trump supporters in the surveys the AAPOR report analyzed. By contrast, the evidence in Figure 1 that Trump underperformed in early October state polls is consistent with our argument that hidden Trump supporters help account for Clinton’s substantial polling lead at the time.

4 More Evidence of Hidden Trump Supporters

Relative to Republican Senate candidates, Trump underperformed in early October state polls. Our next analysis relies on two nationally representative surveys that we conducted in October 2016 – an online survey fielded by GfK (n = 1541) and a phone survey (cell and landline, n = 625) conducted with Cornell’s Survey Research Institute (SRI). To get a sense of how our surveys compared with other surveys at the time, we begin by identifying all survey questions about presidential vote intentions in the Roper Center’s iPOLL database that were in the field during the dates of our surveys.

As noted above, we focus on October because it is during this period when presidential polls typically begin to reflect the final vote margin (Erikson and Wlezien 2012: p. 66; Shirani-Mehr et al. 2016). Figure 2 illustrates that most polls at this time showed Trump trailing Clinton by a substantial margin, with Trump’s two-party vote share ranging from 41 to 48 percent. Since polls typically converge on the final outcome during this period, these numbers seemed to offer a strong basis for those predicting a Clinton victory. The two national surveys we conducted at this time (light gray bars in Figure 2) closely correspond with other surveys at the time, with the GfK survey aligning with the mean across all surveys (45%
Hidden Trump Supporters

Trump support) and the SRI survey aligning with the modal survey at the time (46% Trump support). Although national-level estimates shifted in November – eventually nearly matching Trump’s actual vote share – our October results are consistent with the standard story that the polls underestimated Trump support.

To further test our argument, we would like to know more about those who did not express support for Clinton or Trump in October surveys. If these individuals leaned toward Trump and we could observe their true preferences, we should see results that are more aligned with the actual election results. By contrast, if these individuals were undecided, we would expect an even split between Clinton and Trump. Or, if the barrage of negative stories about Trump was swaying these individuals, we might expect this group to lean toward Clinton.

To evaluate these competing hypotheses, we analyze a question that we included in our surveys that allows us to estimate which candidate these respondents actually supported: a forced-choice measure of candidate truthfulness that read, “If you HAD to choose, which presidential candidate do you find to be more truthful: Donald Trump or Hillary Clinton?” This question was useful for a variety of reasons. First, because this question was phrased as a forced choice (“if you HAD to choose”), we expected it to alleviate respondents’ concerns about being

Figure 2: Trump support from October 5–25, based on all national surveys in the Roper Center’s iPOLL database and our two national surveys. The vertical bars represent the number of survey questions with that vote share (indicated on the x-axis).
judged for their response – indeed, nearly every survey respondent answered it. Second, automatic associations can predict vote choice – even for those who report that they are undecided (Galdi et al. 2008; Wang and Gold 2008). In this case, we expect that the evaluative judgment of which candidate is more truthful will similarly elicit latent support for that particular candidate. Thus, for those Trump supporters who did not express their vote intention directly (because of social desirability bias or negative “top of the head” considerations), responses to the truthful question should still reflect underlying candidate preferences.

Consistent with the expectation that responses about which candidate is more truthful would reflect latent support for that candidate, for those who answered both this question and the standard vote intension question, the truthful question predicted stated vote intention with very high accuracy (97.5% in our GfK survey and 94.1% in our SRI survey). Thus, it appears that truthfulness is an excellent proxy for vote choice. Recall, however, that our goal is to identify the vote preference of the approximately 20 percent of respondents who did not express a vote intention for Trump or Clinton. Because the truthful question strongly predicts vote preference (empirically and theoretically) and because

![Figure 3: The percent indicating support for Trump (among Trump and Clinton supporters) in two national surveys (GfK and SRI) in October, based on expressed vote intentions and the truthful question.](image-url)
almost all respondents answered the truthful question, it is ideally suited for this purpose. For these reasons, we expected the truthful question would offer a more accurate measure of vote choice than the vote intention question.

Figure 3 tests this prediction with our online GfK survey and our SRI phone survey, both of which were conducted well in advance of Election Day (October 5 through 25). In both, analyzing the full sample (using the truthful question) produces an estimate of Trump’s vote share that is substantially closer to – and within one percentage point of – the actual vote share he received (dashed horizontal line). The GfK result is particularly noteworthy because more than 90 percent of respondents completed the survey before October 15, three-and-a-half weeks prior to the election. Although these results cannot directly confirm our hypothesized mechanisms, if the polls suffered from non-response bias and excluded Trump supporters, analyzing the full sample should not improve the estimates. Additionally, if those who did not express a vote intention were truly undecided (with an equal probability of voting for Clinton or Trump), or if they were leaning toward Clinton based on the barrage of negative Trump coverage, this result should not emerge.

5 Solving Correlated Error in State-Level Estimates: An MRP Analysis

The results above are consistent with hidden Trump supporters in both state and national surveys in October. As a final test of our argument, we return to our “truthful” question and analyze responses for the state level. The results in Figure 3 suggested that this question allowed us to identify hidden Trump supporters (also see Enns and Schuldt 2016). The results in Figure 1 also showed that the state polls systematically underestimated Trump support, a result that has been referred to as “correlated error.” If some of this correlated error resulted because a subset of Trump voters in the sample did not express a vote intention, we should no longer systematically underestimate support for Trump across states when we analyze responses to our truthful question at the state level.

Because our data are based on a national survey, we generate state-level estimates of Trump support (based on the truthful question) with a common approach called multilevel regression and poststratification (MRP).11 MRP proceeds in three steps. First, we used a multilevel model to estimate the relationship between whether the respondent indicated Clinton or Trump is more truthful

and the respondent’s demographic characteristics, state, and region.\textsuperscript{12} Based on these results, we then predict the probability of indicating that Trump was more truthful for each demographic-geographic “type” (e.g. Latino females, age 18–29, with a college degree in New York). Finally, we use census data to poststratify (i.e. weight) the responses to match actual state population values, which allows us to estimate the proportion supporting Trump in each state.\textsuperscript{13} Importantly, our statistical model only includes demographic variables, state, and region. Thus, beyond the census data used for weighting responses, we are not adding additional information to our data. We report the results for the seven swing states that have 10 or more electoral college votes. We focus on states with 10 or more electoral college votes because these states matter most for the election outcome, and because states with fewer electoral college votes have smaller populations and are more sparsely represented in our survey.\textsuperscript{14}

Since we are using a single national-level survey to estimate state-level support, the results are bound to include substantial uncertainty (horizontal bars in Figure 4). Nevertheless, two important patterns stand out. First, consistent with expectations, the errors do not appear to be correlated: our estimates (solid dots) overstate Trump’s actual vote share (hollow diamonds) in four states and underestimate Trump support in three states. We would not expect this pattern if non-response bias was the source of correlated error in state polls. Second, our estimates predict the outcome in 5 of these seven key battleground states, including Wisconsin and Pennsylvania, where state polls made Trump victories appear extremely unlikely.\textsuperscript{15}

Of course, as a post-hoc analysis, any interpretations must be made with caution. At the same time, it is clear that our data would not predict a Clinton landslide in the Rust Belt. Considering that Clinton was the first major-party

\textsuperscript{12} We estimated the multilevel model with glmer in R. The specific model included gender (female or male), race (black, white, Latino, or other), age (18–29, 30–44, 45–64, or 65+), education (less than high school, high school, some college, college graduate or more), state, and region (Northeast, Midwest, South, West, or DC).
\textsuperscript{13} Post-stratification data come from the 2015 American Community Survey (https://usa.ipums.org/usa-action/variables/group).
\textsuperscript{14} We report the results for all 11 swing states in the Appendix. As expected, our results are much less accurate for the four states with fewer than 10 electoral college votes, but the results continue to show that our MRP analysis avoids the correlated error problem. Swing states are based on Politico’s battleground states (http://www.politico.com/2016-election/swing-states).
\textsuperscript{15} Some of the reason our results differ from state polls at the time is our focus on the truthfulness question. Our use of MRP with national data also accounts for some of the difference. The key point, however, is these methods could have been employed in October, and substantially different conclusions would have been reached.
nominee since 1972 not to set foot in Wisconsin, a state where Trump ads outnumbered Clinton ads (DeFour 2016), our evidence that Clinton’s lead was not as secure as it appeared is particularly striking.

6 Implications for the 2016 Election Polls and Beyond

Although the conversation about the 2016 election polls will continue, our results reinforce recent findings that cast doubt on non-response bias as a primary reason that the polls underestimated Trump support (e.g. AAPOR 2017). Across two separate national survey samples, when we analyze the standard voting intention question, our results align with the many other October polls that showed Trump’s support well below Clinton’s. However, when we instead analyze an alternative candidate preference question that nearly all respondents answered – whether respondents viewed Trump or Clinton as more truthful –
our results paint a much different picture. Both surveys are within one percentage point of the actual vote share and a state-level analysis of this question not only eliminates the correlated error problem found in many state surveys but also suggests much improved chances of a Trump victory in several key swing states. We would not expect this pattern if the polls systematically missed a subset of Trump voters. Our analysis of 14 state polls offered further evidence that some likely Trump voters did not directly indicate support for Trump in surveys. Based on these analyses, it appears that there was an important group of respondents in October polls who were surveyed but did not directly voice their intention to vote for Trump. Although our data cannot speak directly to the psychological motives and processes of survey respondents, either social desirability bias or “top of head” considerations (or both) could have led to the pattern of hidden Trump supporters we observe.

These results also hold implications for whether late-breaking news, such as FBI Director Comey’s announcement on October 28 to reopen an investigation into Hillary Clinton’s private email server, influenced the election outcome. In a close election – such as this, in which Clinton lost Michigan, Wisconsin, and Pennsylvania by fewer than 80,000 total votes (Leip 2016) – it is plausible that such news could indeed sway the outcome. Additionally, we cannot completely rule out the possibility that these hidden Trump supporters made their voting decision in the final week or two of the campaign (Blake 2016; AAPOR 2017). Yet, our analyses do not support the idea that these individuals were late deciders that were equally likely to go toward Clinton or Trump. Much of Trump’s Election Day support was detectable in early October. To the extent these individuals were influenced by late-breaking news that favored Trump (or harmed Clinton), our results suggest they were already predisposed to this outcome in early October.

Our findings also suggest important implications for future survey research. Of course, it would be unreasonable to conclude that the polls should have correctly predicted the election outcome. Although polls can help improve forecasts, they are designed to measure preferences at the time the poll was conducted, not to forecast election outcomes. Furthermore, extremely close state races, unanticipated events, and the challenge of predicting voter turnout all complicate the task of using polls to forecast elections. Our analyses do, however, suggest strategies for evaluating the certainty (or lack thereof) of election forecasts. We were able to use an alternate measure (i.e. candidate truthfulness) to approximate vote preferences for the unusually high proportion of respondents who did not indicate an intention to vote for either the Democratic or Republican candidate. It is important to note that the truthfulness of the candidate may not be such a salient consideration in future elections as it was in 2016. Thus, questions about candidate truthfulness may not always correspond with vote intentions. However,
in future elections, researchers can identify other relevant questions that theoretically (based on the campaign issues) and empirically predict vote intentions.\textsuperscript{16} The current findings might also suggest that less direct measures, such as web browsing habits (Wang 2016) or implicit measures (Galdi et al. 2008), could also prove useful for imputing preferences of nominally undecided voters.

Finally, our MRP analysis suggests that Clinton’s lead was much less certain in key swing states than state polls suggested. In future elections, researchers should evaluate whether state-level MRP estimates based on national surveys match the conclusions of state-level surveys. Different results would not reveal which outcome was more likely, but would provide important evidence that the outcome was less certain than expected on the basis of state-level polls alone. Considering how much Trump’s victory came as a surprise, improving our understanding of the uncertainty around election surveys would be a major benefit to researchers, media, and the public.

Appendix 1: State Polls Analyzed in Figure 1

As explained in the text, the state polls analyzed in Figure 1 were selected based on the following criteria: a sample of likely voters, live telephone interviews, in the field around early October, and included vote intentions for the presidential and Senate candidate. If more than one survey in a single state met these criteria, we selected the survey conducted closest to October 1. Additional survey details appear below in Table A.1.

Appendix 2: National Polls Analyzed in Figure 2

The data in Figure 2 come from the Roper Center for Public Opinion Research iPOLL database. The survey questions reported in Figure 2 typically followed the format, “If the (2016) presidential election were being held today and the candidates were Hillary Clinton and Tim Kaine, the Democrats, Donald Trump and Mike Pence, the Republicans, Gary Johnson and Bill Weld of the Libertarian Party, and Jill Stein and Ajamu Baraka of the Green Party, for whom would you vote? Would you lean toward Clinton and Kaine, Trump and Pence, Johnson and Weld,

\textsuperscript{16} Recall that across our two surveys, among the respondents who answered both the vote intention question and the “truthful” question, more than 95 percent of responses agreed across the two questions.
<table>
<thead>
<tr>
<th>State</th>
<th>Polling organization</th>
<th>Dates in the field</th>
<th>Sample size</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indiana</td>
<td>WTHR/Howey politics</td>
<td>10/3–10/5</td>
<td>600</td>
<td><a href="http://www.wthr.com/article/-exclusive-wthrpi-poll-clinton-trump-presidential-race-tightens">http://www.wthr.com/article/-exclusive-wthrpi-poll-clinton-trump-presidential-race-tightens</a></td>
</tr>
<tr>
<td>Missouri</td>
<td>Monmouth</td>
<td>10/9–10/11</td>
<td>406</td>
<td><a href="http://www.monmouth.edu/polling-institute/reports/MonmouthPoll_MO_101216/">http://www.monmouth.edu/polling-institute/reports/MonmouthPoll_MO_101216/</a></td>
</tr>
<tr>
<td>Ohio</td>
<td>Monmouth</td>
<td>10/1–10/4</td>
<td>405</td>
<td><a href="http://www.monmouth.edu/polling-institute/reports/MonmouthPoll_OH_100516/">http://www.monmouth.edu/polling-institute/reports/MonmouthPoll_OH_100516/</a></td>
</tr>
</tbody>
</table>
or Stein and Baraka?” After asking this initial vote intention question, several surveys followed up with respondents who indicated a third party candidate, do not know, or no answer, by asking, “If the only candidates were Hillary Clinton and Tim Kaine the Democrats and Donald Trump and Mike Pence the Republicans, for whom would you vote?” To ensure the results in Figure 2 are as representative as possible, we include the results from both formats when they are available. Additional survey details follow.


Appendix 3: MRP Results for All 11 Swing States

Figure 4 in the text reported our MRP estimates and the actual vote share for the seven swing states with 10 or more electoral college votes. Figure A1 presents the results for all 11 swing states. Not surprisingly, our MRP estimates are much less...
accurate with the four states with fewer than 10 electoral college votes (CO, IA, NH, and NV). These states, after all, have smaller populations and are the least well represented in our surveys. Nevertheless, the key result holds. Unlike state surveys, which consistently underestimated Trump support, even when we consider all 11 swing states, our estimates do not reflect “correlated errors.” Our estimates (hollow dots) over-state Trump’s actual vote share (solid dots) in five states and underestimate Trump support in six states. Although more data would certainly improve our estimates, by employing MRP with one national-level survey from October, we do not systematically underestimate Trump support.

**References**


Trende, Sean (2016) “It Wasn’t the Polls That Missed, It Was the Pundits,” Real Clear Politics. http://www.realclearpolitics.com/articles/2016/11/12/it_wasnt_the_polls_that_missed_it_was_the_pundits_132333.html.


**Article note:** A previous version of this paper was presented at the 2017 Annual Conference of the American Association for Public Opinion Research in New Orleans, LA. We would like to thank Claudia Deane, Gary Langer, Sam Wang, Kathleen Weldon, and Chris Wlezien for helpful comments. We also thank GfK and Cornell’s Survey Research Institute for their support and Alex Rauter for his contribution to the surveys.