Routledge
Taylor & Francis Group

Journal of Elections, Public Opinion and Parties

ISSN: 1745-7289 (Print) 1745-7297 (Online) Journal homepage: http://www.tandfonline.com/loi/fbep20

Preelection poll accuracy and bias in the 2016 U.S. general elections

Costas Panagopoulos, Kyle Endres & Aaron C. Weinschenk

To cite this article: Costas Panagopoulos, Kyle Endres & Aaron C. Weinschenk (2018) Preelection poll accuracy and bias in the 2016 U.S. general elections, Journal of Elections, Public Opinion and Parties, 28:2, 157-172, DOI: 10.1080/17457289.2018.1441850

To link to this article: https://doi.org/10.1080/17457289.2018.1441850

Published online: 14 Mar 2018.

JOURNAL OF ELECTIONS, PUBLIC OPINION AND PARTIES, 2018
VOL. 28, NO. 2, 157–172
https://doi.org/10.1080/17457289.2018.1441850

Routledge
Taylor & Francis Group

Preelection poll accuracy and bias in the 2016 U.S. general elections

Costas Panagopoulos[a], Kyle Endres[b] and Aaron C. Weinschenk[c]

[a]Department of Political Science, Northeastern University, Boston, MA, USA; [b]Duke Initiative on Survey Methodology, Duke University, Durham, NC, USA; [c]Political Science, University of Wisconsin, Green Bay, WI, USA

ABSTRACT

This report examines accuracy and bias in national- and state-level preelection polls conducted during the 2016 U.S. general election cycle. Overall, national polls in 2016 were somewhat more accurate than in 2012, but statewide polls were less accurate. Patterns across the board suggest polls underestimated Republican support in the presidential, U.S. Senate and gubernatorial races. Nevertheless, these biases were generally statistically insignificant, suggesting significant bias in preelection polls was scarce in 2016.

Introduction

Polls matter. Not only can preelection polls help voters, analysts, and campaign operatives alike to assess campaign dynamics and the evolution of candidate preferences (Erikson and Wlezien 2012; Panagopoulos 2009, 2013), but projections about outcomes on Election Day that are based on preelection polls can have important consequences for voter decision-making and behavior. For better or worse, studies have revealed that polling information can provide important cues about candidate viability (Bartels 1988), affect strategic considerations based on expectations about election outcomes and result in contagion (or bandwagon) effects, by which voters come to evaluate candidates or parties more favorably if their prospects for victory appear strong and more negatively if their chances are slim (Ansolabehere and Iyengar 1994; Blais, Gidengil, and Nevitte 2006). Preelection poll information can even influence turnout in elections (Vannette and Westwood 2013). These effects can be quite potent, despite the fact that most voters reject the notion that poll information affects them (Price and Stroud 2005).

Given how influential poll information can be in elections, scholars have turned their attention to examining the quality of preelection polls, focusing, in part, on technical or methodological aspects (Crespi 1988) but also on polls'

CONTACT Costas Panagopoulos costas@post.harvard.edu Northeastern University, Boston, MA 02115, USA

capacity to accurately forecast actual outcomes on Election Day (Martin, Traugott, and Kennedy 2005; Panagopoulos 2009; Panagopoulos and Farrer 2014; Traugott 2001, 2005). Actual election outcomes are a useful benchmark against which to evaluate the accuracy of preelection polls, especially those conducted just before Election Day, in part because verifying the accuracy of daily snapshots in the absence of objective measures is virtually impossible. Sizable deviations from the vote on Election Day or failures to accurately project results can potentially call poll quality into question and, perhaps more importantly, mislead voters who rely on this information to make judgments about parties and candidates. Such analyses can also reveal whether aspects of poll methodology or other poll attributes systematically affect accuracy and can help pollsters to adopt refinements to improve accuracy.

In this report, we assess poll accuracy in the 2016 elections in the United States. We use a number of different metrics to examine accuracy – Mosteller et al.'s (1949) M3 and M5, and the *A* measure proposed by Martin, Traugott, and Kennedy (2005) – in order to provide a more comprehensive look at the accuracy of publicly available preelection polls. First, we assess the accuracy of the final, national, presidential preelection polls. While doing so, we compare our findings to similar analyses released by a committee of polling professionals commisioned by the American Association for Public Opinion Research (AAPOR). Our analyses and findings overlap in some regards, but they are distinct in others. One primary difference is the set of polls examined; we evaluate polls conducted in the week prior to the election while AAPOR sets 13 days prior to the election as their cutoff point. Our inclusion criteria are consistent with those used in prior election years which allow for direct comparisons to the extant literature. Second, we evaluate the accuracy of the final, national, generic Congressional vote polls. Third, we assess the accuracy of the final statewide presidential, gubernatorial, and U.S. Senate preelection polls. Using a dataset containing state-level preelection polls conducted in the final week of the election cycle, we also conduct a series of multivariate regression analyses to explain overall levels of poll accuracy and bias. Finally, we place these findings in historical context, comparing the results for 2016 to similar analyses conducted in previous presidential election cycles.

The 2016 presidential election pitted the Democratic nominee Hillary Clinton against the Republican party's Donald Trump. Many political analysts, and most preelection poll projections expected Clinton to triumph. But as the votes were tabulated and each state declared a winner, it became clear that Trump would prevail. The polls, and the polling industry more broadly, quickly became a scapegoat for perpetuating the erroneous narrative that Trump would not (and could not) win (Wells 2016).

Trump's win was a surprise to many, but the polls depicted a close race. Trump consistently trailed Clinton, albeit narrowly, in the national polls throughout 2016. The prominent poll aggregators, on the other hand,

projected an almost-certain Clinton victory. Poll aggregators, such as *HuffPost* Pollster, *FiveThirtyEight*, and *RealClearPolitics*, predicted a Clinton win with estimates ranging from a 71% (*FiveThirtyEight)* to a 98% chance of winning by the *HuffPost* (Katz 2016). These lopsided projections did not coincide with the snapshot provided by any, individual national preelection poll taken in the days prior to the election, nor did they appear to reflect the larger, and still relatively close, race depicted in the aggregations. For example, on Election Day, *FiveThirtyEight* had Clinton beating Trump by a margin of 3.9 percentage points (45.7–41.8%). Similarly, *HuffPost Pollster*'s final estimate showed Clinton beating Trump by 5.3 points (47.3–42%). The final projection from *RealClearPolitics* showed that Clinton would win 46.8% of the vote and Trump would win 43.6% of the vote, a margin of 3.2 points.

In the end, Clinton did win the national popular vote, earning 48.2% nationally to Trump's 46.1% (and exceeding Trump by more than 2 million votes), but the Electoral College victory went to Trump. Ultimately, Trump's win stimulated intense national debates about poll reliability and sobering retrospectives and critiques about how preelection preferences were evaluated, reported, aggregated, and interpreted by both the media (e.g. Jackson 2016) and industry leaders (Kennedy et al. 2017).

National, presidential preelection polls in 2016

Preelection polls were abundant in 2016. The poll aggregation website *Realclearpolitics.com* tracked 221 national polls conducted between 1 January 2016 and Election Day (8 November). Interestingly, Trump led Clinton in only 26 (11.8%) of these polls. Fourteen national polls were conducted during the final week before the 2016 election (1–8 November 2016). We use the final estimates from these 14 national preelection polls to assess the accuracy of the 2016 polls relative to the popular vote using their final estimates in a two-way race between Clinton and Trump.

We gauge the accuracy using Mosteller's (1949) M3 and M5, and Martin, Traugott, and Kennedy's (2005) *A* measure. All three metrics are commonly used to evaluate the accuracy of electoral polls. We use these standard measures, in part, to be comparable to similar analyses conducted in previous election cycles (Panagopoulos 2009; Panagopoulos and Farrer 2014). Table 1 displays all three metrics for each poll and is arranged from most to least accurate based on the *A* measure. The *A* metric is calculated using the natural logarithm of the odds ratio of the outcome in a poll and the popular vote (see Martin, Traugott, and Kennedy 2005 for more details). M3 is computed by taking the average of the absolute difference between the poll estimate and the final election result for each candidate. Using the McClatchy/Marist poll as an example, M3 is equal to 2.15. This poll put Trump at 44% (2.1% below his share of the popular vote) and Clinton at 46% (2.2% below her

Table 1. Final, national presidential and U.S. House Preelection Poll Accuracy, 2016.

Rank/Firm	Trump/ Rep	Clinton/ Dem	Sample Size (*N*)	Mosteller Measure 3	Mosteller Measure 5	Predictive accuracy
Election result (*President*)	*46.1*	*48.2*				
1. McClatchy/ Marist	44	46	940	2.15	0.1	0.0001
2. ABC/Wash Post Tracking	46	49	2220	0.45	0.9	−0.0186
3. IBD/TIPP Tracking	42	43	1107	4.65	1.1	0.0210
4. CBS News/*NY Times*	44	47	1333	1.65	0.9	−0.0214
5. Bloomberg	43	46	799	2.65	0.9	−0.0229
6. *Economist*/ YouGov	45	49	3669	0.95	1.9	−0.0406
7. FOX News	44	48	1295	1.15	1.9	−0.0425
8. CBS News	43	47	1426	2.15	1.9	−0.0444
9. Gravis Marketing	50	50	5360	2.85	2.1	0.0445
10. NBC News/*WSJ*	43	48	1282	1.65	2.9	−0.0655
11. Reuters/Ipsos	39	44	2195	5.65	2.9	−0.0761
12. Monmouth	44	50	748	1.95	3.9	−0.0833
13. NBC News/ SurveyMonkey	44	51	30145	2.45	4.9	−0.1031
14. *LA Times*/USC Tracking	47	44	2935	2.55	5.1	0.1105
Average				*2.35*	*2.24*	*−0.0244*
election result (*U.S. House*)	*49.1*	*48.0*				
1. FOX News	46	45	1295	3.05	0.1	−0.0004
2. Bloomberg	48	45	799	2.05	1.9	0.0422
3. McClatchy/ Marist	47	48	940	1.05	2.1	−0.0434
4. Reuters/Ipsos	41	42	1858	7.05	2.1	−0.0465
5. CBS News/*NY Times*	46	49	1333	2.05	4.1	−0.0856
6. NBC News/*WSJ*	44	47	1282	3.05	4.1	−0.0883
7. Economist/ YouGov	42	45	3677	5.05	4.1	−0.0914
Average				*3.34*	*2.64*	*−0.0448*

Note: To be consistent with previous years' analyses of poll accuracy, we include poll estimates produced within the final week of the election.

share of the popular vote); the difference between each candidate's actual vote share and their poll estimate is averaged to calculate M3. M5 compares the polled margin between the two leading candidates to the official electoral margin between the same candidates and is estimated by taking the absolute value of the difference between these margins. In the case of the McClatchy/ Marist poll, M5 would be 0.1 since Clinton ended up winning the popular vote by 2.1% compared to her lead of 2% in the final McClatchy/Marist poll.

All three metrics provide insight about the accuracy of each poll's final estimate. With respect to *A*, its absolute value provides an estimate of accuracy, that is to say, deviation from the actual election result without regard to the partisan direction of the deviation. In the case of our most accurate

poll, the McClatchy/Marist, *A* is 0.0001. The *A* measure has the added benefit of signaling the direction of any potential partisan bias since the metric results in a signed statistic instead of an absolute value. When *A* is equal to zero, the poll is not biased at all. Positive values of *A* represent pro-Republican bias whereas negative values indicate a pro-Democratic bias (Traugott 2005).

The values for each metric are displayed in Table 1 for all 14 polls conducted in the week prior to the election. Table 1 also includes the average for each metric, across all polls. Overall, the average value for *A* is −0.0244, indicating a modest Democratic bias. Assuming a tied election, our estimate of bias implies polls favored Clinton by about 0.6 percentage points on average. However, the standard error associated with the mean value for *A* we report for the full sample of national polls is .051, indicating the bias overall was not statistically significant. In fact, in all but one case (*Los Angeles Times*/USC), individual polls were not significantly biased in 2016. In the case of the *Los Angeles Times*/USC tracking poll, however, *A* is positive and statistically distinguishable from zero, implying the final poll projection significantly overestimated Trump support by 2.75 percentage points, assuming a tied election.

The 2016 average for Mosteller's Measure 3 is 2.35, and the average value for Mosteller's Measure 5 is 2.24. To help put the 2016 polling in the context, we present the averages for M3 and M5, the number of candidates, and the number of polls during the final week before the election from 1956 to the present in Table 2 (see Panagopoulos 2009; Traugott 2005). Mosteller's measures present a mixed picture. Based on M3, the 2016 polls performed somewhat worse than the historical average (M3 = 1.85, 1956–2012), while

Table 2. Average errors in presidential polls, 1948–2016.

Year	# of Polls	# of Candidates	M3	M5	*A*
1956	1	2	1.8	3.5	
1960	1	2	1.0	1.9	
1964	2	2	2.7	5.3	
1968	2	3	1.3	2.5	
1972	3	2	2.0	2.6	
1976	3	3	1.5	2.0	
1980	4	3	3.0	6.1	
1984	6	2	2.4	4.4	
1988	5	2	1.5	2.8	
1992	6	3	2.2	2.7	
1996	9	3	1.7	3.6	−0.084
2000	19	3	1.7	3.5	+ 0.063
2004	19	2	1.7	2.1	−0.024
2008	20	2	1.5	1.5	−0.013
2012	21	2	1.8	2.8	+ 0.054
2016	14	2	2.4	2.2	−0.024
Average (1956–2016)			1.9	3.1	−0.005

Notes: Data for 1956–2004 period obtained from Traugott (2005, 649); 2008 from Panagopoulos (2009), 2012 from Panagopoulos and Farrer (2014), and 2016 update compiled by authors. The table displays only polls conducted in the last week before the election.

M5 indicates that the 2016 polls were somewhat more accurate than average (M5 = 3.15, 1956–2012).

We do not have as lengthy of a time series for the *A* measure. Table 2 displays the *A* measure for the last 20 years, beginning with the 1996 presidential election. The polls in 1996, 2004, 2008, and 2016 all had slight biases toward Democratic candidates while the 2000 and 2012 polls had slight biases toward Republicans. Based on *A*, the 2016 polls ($A = -0.024$) were more accurate than in 2012 ($A = 0.054$), but less accurate than in 2008 ($A = -0.013$) and comparable to the 2004 ($A = -0.024$) presidential polls.

Our results are fairly comparable to those produced by the committee convened by AAPOR to evaluate presidential polls in 2016. The committee concluded the national polls were historically quite accurate and more accurate than in the previous presidential election (Kennedy et al. 2017), whereas we find, based on M3, that the 2016 polls were worse (but not statistically significantly so) than the historical average, but, based on M5, somewhat better than average (see Table 2). It is important to note we approached the task somewhat differently in terms of our methodological and measurement decisions. We focus our analysis on the final polls conducted in the last week before the election, whereas the AAPOR committee included all polls fielded within 13 days of the election. Much can happen in the final weeks of an election, and 2016 was no exception. For example, on 25 October 2016, the incumbent administration announced that Obamacare premiums would be increasing by 25%, and, on 28 October 2016, FBI director James Comey sent a letter to congressional leaders informing them about the potential for new evidence related to Clinton's email investigation. Thus, the period of the second week before the election was especially challenging for Clinton. AAPOR's evaluation that polls were more accurate than the historical average is based on the absolute value of the difference between the vote margin between Clinton and Trump in the final polls and the vote margin in the national popular vote. This method is comparable to our M5 measure, by which we also conclude the 2016 national polls were more accurate on average than in prior cycles. AAPOR's analysis from the last two weeks found the average error to be 2.2 percentage points (Kennedy et al. 2017), whereas our M5 calculation for polls conducted in the last week of the election cycle is 2.24 percentage points.

National generic congressional vote preelection polls in 2016

In 2016, seven polling organizations reported final poll results projecting the national, generic Congressional vote in the week before Election Day. These estimates, displayed in the lower part of Table 2, project the overall share of the U.S. House vote nationally for each party, rather than the share of the vote for candidates within each congressional district. Nevertheless, we

assess the accuracy of these polls relative to the actual outcome on Election Day. Overall, Republican U.S. House candidates secured 49.1% of the vote nationally (63,173,815 votes) compared to 48.0% for Democratic candidates (61,776,554). Only two of the seven (Fox News and Bloomberg) polls accurately forecasted the Republican advantage in the national U.S. House vote, and all polls except Bloomberg overestimated the Democratic vote (based on *A*). The mean value of *A* was −0.0448, confirming the pro-Democratic bias overall. Nevertheless, only the *Economist*/YouGov poll significantly overestimated the Democratic bias, so we conclude the vast majority (86%) of the final, national generic U.S. House vote projections were not significantly biased in 2016.[1] We also calculated mean values of M3 (3.34) and M5 (2.64) for this set of polls.

Statewide preelection polls in 2016

In addition to assessing presidential and congressional preferences nationally, pollsters also probe state-level preferences for president, as well as U.S. Senate and gubernatorial candidates.[2] We use a parallel dataset of state-level polls conducted in the final week of the election cycle (1–7 November 2016) to examine accuracy in these statewide races. These observations were obtained from the compendium provided by *HuffPost* Pollster, which means we were subjected to the inclusion restrictions adopted by Pollster. To be consistent with previous examinations, we restricted our sample to polls fielded by organizations that conducted at least 10 statewide polls over the course of the 2016 campaign and that polled in multiple (at least three) states.[3] For presidential contests, we also limited our sample to polls that probed respondents about the two-way match-up between Trump and Clinton. Accordingly, our analyses are restricted to a sample of 183 statewide polls that examined preferences for president, U.S. Senate or governor in each state, respectively. Most of the polls in our sample (93%) were conducted on the Internet, while the remainder used Interactive Voice Response (IVR) online.[4]

Using each poll as a single observation, we present the frequency distribution of *A* in Figure 1. In the absence of overall bias, we would expect the distribution to be centered on zero. The mean value of *A* in the complete sample of polls is −0.11, suggesting the pattern of pro-Democratic bias detected in the national presidential and congressional generic vote polls

[1]Comparable analyses are not available for previous election cycles.
[2]We assess the accuracy of the statewide polls as a group.
[3]This is consistent with Martin, Traugott, and Kennedy (2005, 362), Panagopoulos (2009) and Panagopoulos and Farrer (2014).
[4]IVR surveys in the United States are only allowed to call landline phones due to current legal requirements. Thus, some polling firms will supplement their IVR polls with online polls designed to target cell-phone only individuals.

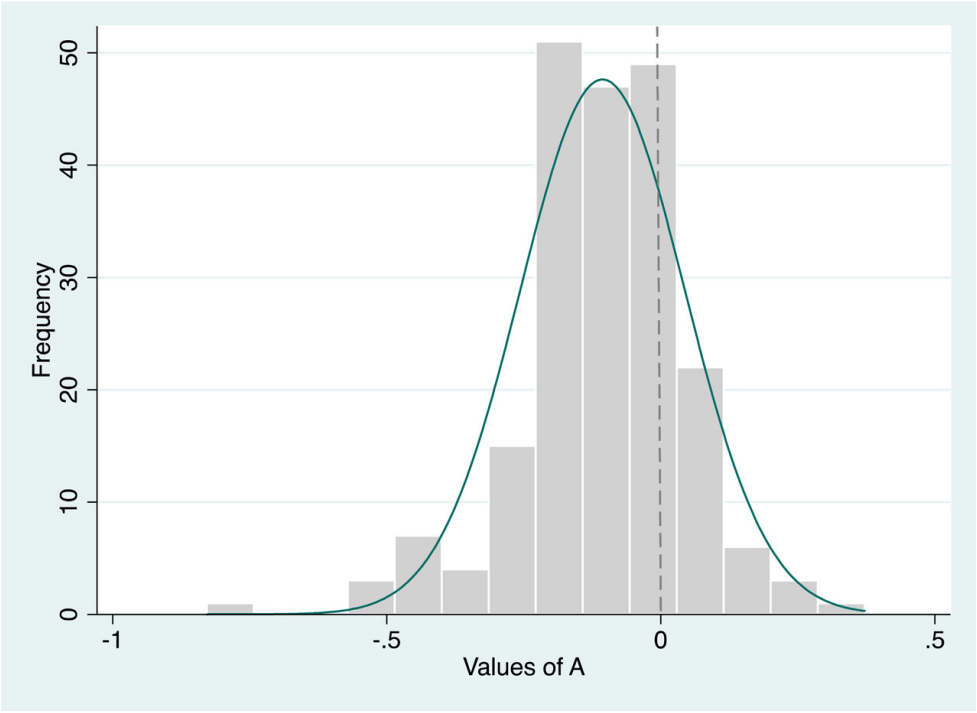

Figure 1. Frequency distribution of *A* in Statewide Preelection Polls, 2016.

also characterizes statewide polls, but the bias is not statistically significant at conventional levels (standard error = .085). Nevertheless, assuming all races were perfectly tied, this would translate into a difference (or Democratic overstatement) of 2.75 percentage points (see Panagopoulos and Farrer 2014).

We further use the *A* measure to investigate the poll accuracy by common poll characteristics, including election type, survey mode, sample type, interviewing period, and sponsor. In recent years, polling organizations have adopted methodological refinements to improve poll performance by adjusting their weighting schemes to make polls more representative, and updating their approaches to modeling the electorate (likely voters) (see Erikson, Panagopoulos, and Wlezien 2004). Despite these efforts, error persists, and analyses of poll accuracy can help to reveal approaches and features that are potentially problematic – or at least to confirm that procedures generally produce reasonable (that is to say, accurate or unbiased) results. The analyses below can shed light on these issues. We present the mean and standard error of *A*, in Table 3, by each characteristic. All of the polling organizations, as shown in Table 3, had small, statistically insignificant biases toward the Democratic Party. This is comparable to our findings for the final, national presidential polls. There seems to have been small Democratic biases across the board in 2016. Most of the biases, however, were statistically insignificant. We proceed by evaluating the accuracy based on *A* for each characteristic.

Table 3. Mean predictive accuracy (*A*) by poll characteristics, 2016 statewide polls.

Poll characteristics (type/sponsor)	Number of polls	Mean predictive accuracy (*A*)	Standard error
Democratic	12	−0.097	0.068
Nonpartisan	171	−0.111	0.087
Presidential	126	−0.092	0.094
U.S. Senate	43	−0.154	0.064
Governor	14	−0.138	0.074
IVR/online	13	−0.090	0.069
Internet	170	−0.112	0.087
Sponsors			
CBS/YouGov	5	−0.052	0.069
Ipsos/Reuters	42	−0.070	0.088
PPP (D)	13	−0.090	0.069
SurveyMonkey	74	−0.155	0.069
UPI/Cvoter	49	−0.087	0.114

Notes: Following Martin, Traugott, and Kennedy (2005; Panagopoulos 2009), only polling organizations that conducted at least 10 statewide polls over the course of the 2016 campaign in at least 3 separate states in 2016 are included in the analysis.

We begin with the partisan affiliation of the polling firm. Based on the inclusion criteria we describe above, we only have polls conducted by Democratic and non-partisan polling firms. Democratic firms conducted 12 polls while nonpartisan firms conducted 171. Both Democratic and non-partisan firms had mean values of *A* that indicate small and statistically insignificant pro-Democratic biases, as shown in the top two rows of Table 3. By contrast, the bias in 2012 pointed in the other direction; comparable analyses revealed statistically insignificant pro-Republican biases for statewide polls conducted by nonpartisan as well as Democratic and Republican polling organizations (Panagopoulos and Farrer 2014).

Next, we turn our attention to election type. Our sample includes 126 presidential polls, 43 U.S. Senate polls and 14 gubernatorial polls. Based on mean values of *A*, polls overall at each level reflected biases that favored Democratic candidates, as shown in Table 3. The Democratic biases were statistically insignificant for both presidential and gubernatorial polls. Statewide U.S. Senate polls, however, were significantly biased in a pro-Democratic direction (mean value of $A = -0.154$, SE = 0.064). Compared to parallel analyses conducted for the 2012 election cycle (Panagopoulos and Farrer 2014), these results suggest accuracy overall was considerably lower across all types of races in 2016.

Poll mode is our next consideration. Our sample includes 13 polls conducted with IVR technology (often with online supplements targeted toward members of the public without landline phones) and 170 Internet polls. Once again, mean values of *A* point to statistically insignificant biases favoring Democratic candidates for both survey modes. The results suggest that Internet-based surveys were slightly less accurate ($A = -0.112$) than IVR polls ($A = -0.090$) on average in 2016. This is consistent with patterns detected

in the 2008 cycle, in which statewide IVR polls were most accurate (Panagopoulos 2009), but contrasts with the 2012 election when IVR polls were the least accurate (Panagopoulos and Farrer 2014).[5]

The extant research connecting the poll timing and accuracy is decidedly mixed. Some research argues that polls' predictive accuracy grows as Election Day approaches (see Crespi 1988; Erikson and Wlezien 2012). Other studies fail to find a statistically significant relationship between the poll timing and accuracy (see Lau 1994; Martin, Traugott, and Kennedy 2005; Panagopoulos 2009). We use longitudinal analysis to assess whether poll accuracy in 2016 improved closer to Election Day. The dashed line presents the smoothed pattern of the absolute value of *A* over the course of the campaign period and suggests accuracy likely improved slightly in the final few days of the election cycle (the absolute value of *A* trended toward zero), but this pattern reversed in polls completed the day before the election. In 2008 (Panagopoulos 2009), accuracy improved most dramatically during the final few days prior to the election. The solid line in Figure 2 plots lowess-smoothed levels of mean predictive accuracy (*A*) over the same duration. The pattern suggests statewide polls reflected a stable, pro-Democratic bias on average for most of the period we study, and that this bias strengthened in the final day of the election cycle. Next, we conduct multivariate regression analyses to determine if any of the poll characteristics systematically affect the accuracy and bias of the polls.

We estimate two regression models with *A*, or the absolute value of *A*, as the dependent variables, respectively. The independent variables include sample size, the number of days to the election, and indicator variables for U.S. Senate races, gubernatorial races, and nonpartisan polling organizations. We also include controls (fixed effects) for "house" (polling organization) and state effects. The dependent variable for the first model is the absolute value of *A*, with values close to zero representing high levels of accuracy and larger values representing greater inaccuracy. The results for Model 1, estimated using ordinary least squares (OLS) regression, are presented in Table 4. The results suggest statewide polls for U.S. Senate and gubernatorial candidates were significantly (at the $p < .10$ level, two-tailed) less accurate, compared to presidential polls (the excluded category) in 2016, all else equal. Comparable analyses from previous presidential election years also documented significantly lower levels of accuracy for U.S. Senate polls compared to state-level presidential polls, both in 2008 (Panagopoulos 2009) and in 2012 (Panagopoulos and Farrer 2014). Statewide gubernatorial polls were less accurate than statewide presidential polls in 2008 (Panagopoulos 2009), but about

[5]Other research has found that IVR polls are as accurate as other polling modes when the results from more traditional modes have been publically released, which suggests that IVR polls may adjust their weighing procedures so their results more closely align with other polls (Clinton and Rogers 2013). This is one possibility why we do not find significant differences between modes.

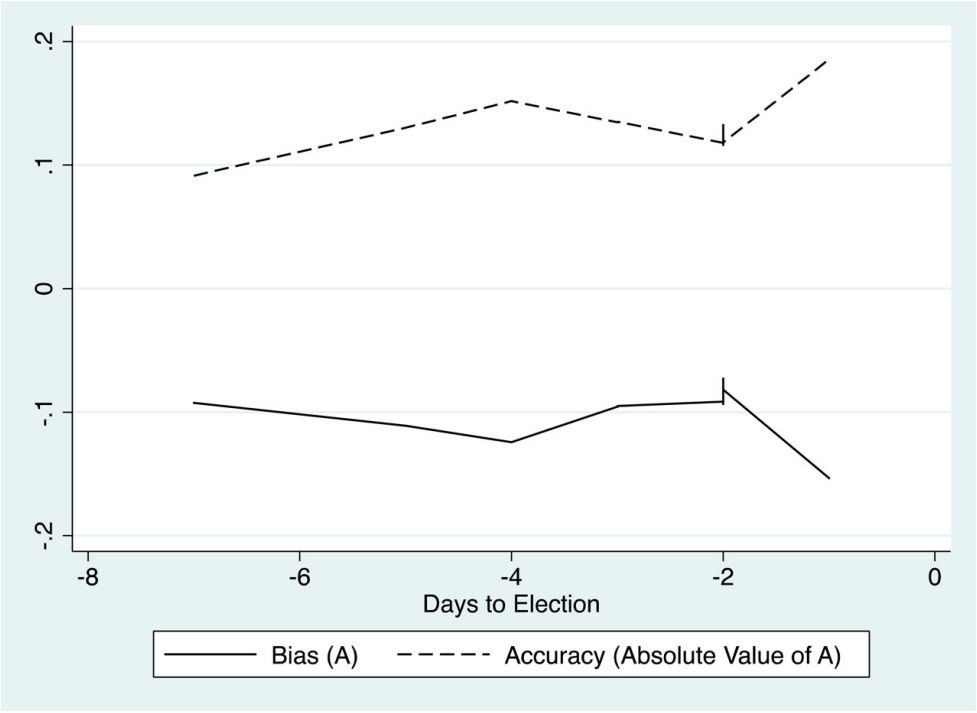

Figure 2. Smoothed (lowess) levels of overall bias (*A*) and accuracy (absolute value *A*) by period to election, 2016.

Table 4. The impact of poll attributes on bias and accuracy in statewide preelection polls, 2016.

Independent variables: (Poll characteristics)	Model 1: Accuracy	Model 2: Bias (Pro-Democratic)
U.S. Senate	0.04*	−2.16
	(*0.02*)	(*1.42*)
Governor	0.05*	−2.20*
	(*0.03*)	(*1.28*)
Sample size	−0.00	0.01***
	(*0.00*)	(*0.00*)
Nonpartisan polling organization	−0.01	–
	(*0.11*)	–
IVR/online	–	–
	–	–
Days to election	−0.01	−5.16
	(*0.02*)	(*1226.47*)
Constant	0.004	−2.36
	(*0.18*)	(*2491.61*)
N	183	67
R^2/pseudo R^2	0.61	0.56
Log likelihood	—	−19.54

Notes: Model 1: OLS. Dependent variable is the absolute value of *A*; Model 2: Probit. Dependent variable = 1 if $A < 0$, and 0 if $A > 0$. Clustered for polling organizations. Observations with covariate patterns that predict outcome perfectly are excluded from the model, resulting in the smaller number of cases. ***Statistical significance at the $p < .01$ level, and ** at the $p < .05$ level, and * at the $p < .10$ level using two-tailed tests. Standard errors in parentheses.

as accurate as presidential polls in 2012 (Panagopoulos and Farrer 2014). Neither sample size, days to the election, nor the partisanship of the polling organization were statistically related to poll accuracy in 2016.

The second model is estimated using probit regression with a dichotomous dependent variable equal to 1 if the poll reflected a pro-Democratic bias ($A < 0$) and 0 if the poll reflected a pro-Republican bias ($A > 0$). The explanatory variables are identical to model 1 and also include controls (fixed effects) for "house" (polling organization) and state effects. Statewide gubernatorial polls in 2016 were, all else equal, significantly (at the $p < .10$ level, two-tailed) less likely to be biased in a Democratic direction in 2016, compared to presidential polls (the excluded category), but the level of bias in statewide U.S. Senate polls was not significantly different from state-level presidential poll in 2016, all else equal. In both the 2008 and 2012 cycles, final, statewide preelection polls in both U.S. Senate and gubernatorial races were significantly biased in the Republican direction, compared to statewide presidential polls (Panagopoulos 2009; Panagopoulos and Farrer 2014). The 2016 analysis also suggests pro-Democratic bias was positively associated with larger samples, but no other factors were significantly related to poll bias in these polls 2016.

The multivariate analyses confirm some of the findings from the last two election cycles. For example, statewide U.S. Senate polls have been significantly less accurate in each of the three, most recent election cycles, and gubernatorial polls in two (2008 and 2016) of the last three cycles, compared to statewide presidential polls. Gubernatorial polls have also been significantly biased in a pro-Republican direction in the last three election cycles, while U.S. Senate polls, which reflected significant pro-Republican biases in 2008 and 2012 were not significantly biased in a GOP direction in 2016 (Panagopoulos 2009; Panagopoulos and Farrer 2014). Such developments, along with the accumulation of comparative information, can foster speculation about explanations for persistent patterns (or aberrations), and eventually improvements in polling methodology as well as poll interpretation.

Conclusion

Polling in 2016 received considerable attention both before and after the election. In the aftermath of an outcome in the presidential race that surprised many experts and contradicted most projections based on preelection polls, interest in investigating poll accuracy has been reinvigorated, leading to numerous, detailed, and comprehensive studies, including the report commissioned by AAPOR mainly to study poll accuracy in the presidential race (Kennedy et al. 2017). We are encouraged that our findings corroborate aspects of this report (and undoubtedly others), but the current study also aims to make other contributions, most notably by extending our analyses beyond the presidential race to encompass congressional and gubernatorial

races. We believe this is important, in part, because poll accuracy is these races typically lags behind presidential contests (Panagopoulos 2009; Panagopoulos and Farrer 2014). We also seek to apply analytic procedures that parallel previous analyses of this sort (Panagopoulos 2009; Panagopoulos and Farrer 2014) in order to document developments in 2016 as well as changes over time. By our estimate, the final preelection polls in the 2016 cycle were fairly accurate. At least by some measures, national polls overall were actually more accurate in 2016, at least compared to 2012, but statewide polls were less accurate, suggesting sources of inaccuracy merit continued scrutiny (Panagopoulos and Farrer 2014).

One intriguing possibility is that poll accuracy fluctuates across election cycles as a function of the size of the late-deciding electorate, for example, which makes it challenging for pollsters to capture such developments in time to be reflected in their final poll projections. Analysis of available data from the American National Election Studies suggests there has been an uptick in the number of voters who reached voting decisions in the final two weeks of the election since 1948, with about one-in-five voters on average reporting deciding during this period in presidential elections since 1992. According to exit polls conducted in 2016, 13% of voters reported making their voting decisions in the final week of the election.[6] We explored the hypothesis that overall poll accuracy in national polls is linked to the proportions of late-deciders (measured as the percent of voters who reported making their vote decisions in the last two weeks of the election) empirically using the available data and the over-time measures of poll accuracy described above. The evidence is inconclusive. For the period 1956–2004 (when the ANES stopped asking the "timing of vote decision" item in identical fashion), measures of M3 and M5 (see Table 2) are positively correlated (Pearson's $r = .14$ and .07, respectively) with the proportion of late-deciders, but the relationships are weak and statistically insignificant ($N = 13$ in both analyses). The correlation between available measures of predictive accuracy (absolute values of A) since 1996 (see Table 2) and the proportion of voters who indicated in exit polls that they decided who to vote for in the final week of the election is positive and considerably stronger, but it is also statistically insignificant (Pearson's $r = .54$, $p = .27$), perhaps due to the limited number of observations ($N = 6$). While fairly inconclusive, these preliminary analyses suggest polls may be less accurate when voters delay (or change) their voting decisions, a line of inquiry that is worthy of further scholarly investigation.[7]

[6]Exit poll data on the timing of vote decision in 2016 can be found at the following website: http://www.cnn.com/election/results/exit-polls.

[7]Kennedy et al. (2017) also examined the impact of late deciding in 2016. The investigators found limited evidence that late-deciding affected accuracy at the national level, but they argued noticeable, late-deciding disproportionately favoring Trump could potentially help to explain at least part of the discrepancy between polls and outcomes in four decisive states (MI, WI, FL, and PA).

Our findings also revealed that preelection polls, more or less across the board in 2016, consistently reflected pro-Democratic bias; even though these biases were insignificant, this pattern suggests systemic contextual factors may account, at least in part, for potential bias in poll results. These hints of partisan bias, reflected in patterns of partisan advantage detected in prior cycles, imply polls may have a tendency to converge on an overall election narrative that may (or may not) be accurate and that can exert important effects on voter decision-making and behavior. Some of this can result from "herding," or the practice of adjusting poll findings to match or approximate the results of other polls to avoid embarrassment, that artificially produces consistent results across polls and polling organizations that may not accurately reflect public attitudes and preferences (AAPOR 2016; Silver 2014). In our assessment, we are heartened by the fact that these biases were generally not statistically significant in 2016, but we do recognize that the pattern is striking. Documenting such findings and patterns in successive election cycles is also crucial to ongoing assessments of poll accuracy in elections.

Notwithstanding our overall conclusions, we recognize that alternative methodological approaches, selection criteria or interpretations have the potential to lead analysts to adopt a more circumspect view of poll accuracy in the 2016 cycle; assessments of poll accuracy are almost invariably subjective and can vary depending on how accuracy is measured or defined. The lack of a consensus about how accuracy is conceptualized and measured makes it difficult for analysts to adopt consistent analytic approaches, and the perspectives and approaches adopted by different analysts can yield dramatically different interpretations. Taken as a whole, we recognize that our conclusions overall may paint a more favorable picture of poll accuracy in the 2016 election cycle, but we do not dismiss or otherwise overlook serious shortcomings in poll reliability. Polls were far from perfect in 2016. The degree of imperfection, however, is perhaps as much a subjective as it is an empirical matter. While there may not be consensus about how accurate polls were in 2016, there is clearly agreement that perfect accuracy remains elusive. And so, the pursuit of further improvements in polling methodology, and the quest to continue to stay apace and responsive to new and intensifying challenges, including, but certainly not limited to, growing nonresponse rates and coverage problems, remain crucial. Speculations about systematic sources of poll error are also critical. Our investigations explore some factors potentially linked to poll error, but others, like sampling-based explanations, are also plausible. The AAPOR report noted, for example, that failure to weight for education in many polls (especially at the state level) may have affected overall accuracy 2016 because, unlike in previous cycles, the preferences of highly educated voters (who tend to be overrepresented in poll samples) did not resemble the views of low-education voters in the election

cycle (Kennedy et al. 2017); adjustments for education would have meaningfully improved accuracy by reducing the over-statement of Clinton support in two states (New Hampshire and Michigan) in which the polling organizations (University of New Hampshire and Michigan State University, respectively) conducted *post hoc* analyses along these lines (Kennedy et al. 2017). Such analyses are instructive, but access to detailed weighting procedures is typically unavailable, and approaches can vary across polling entities. Factors beyond weighting schemes, like choices made by polling organizations to model the likely electorate (Erikson, Panagopoulos, and Wlezien 2004; Kennedy et al. 2017), can also be consequential, and deserve continued scrutiny. In the end, polls will probably always be imperfect, and there may be more reliable ways to forecast election outcomes (e.g. see Beauchamp 2017). But these alternatives are subject to their own shortcomings and imperfections, and election projections made on the basis of these indicators are not always unequivocal. At least for now, we are heartened by the fact that polls seem to hit the mark more often than not, and, in our assessment, they did so in 2016.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Costas Panagopoulos is Professor of Political Science and Director of the Big Data & Quantitative Methods and Networks Initiatives at Northeastern University.

Kyle Endres is a postdoctoral associate at the Duke Initiative on Survey Methodology (DISM) at Duke University.

Aaron C. Weinschenk is an associate professor of Political Science at the University of Wisconsin-Green Bay.

References

AAPOR. 2016. "Herding." Accessed November 2, 2017. https://www.aapor.org/Education-Resources/Election-Polling-Resources/Herding.aspx.

Ansolabehere, S., and S. Iyengar. 1994. "Of Horseshoes and Horseraces: Experimental Studies of the Impact of Poll Results on Electoral Behavior." *Political Communication* 11 (4): 413–430.

Bartels, L. M. 1988. *Presidential Primaries and the Dynamics of Public Choice*. Princeton, NJ: Princeton University Press.

Beauchamp, N. 2017. "Predicting and Interpolating State-level Polls Using Twitter Textual Data." *American Journal of Political Science* 61: 490–503.

Blais, A., E. Gidengil, and N. Nevitte. 2006. "Do Polls Influence the Vote?" In *Capturing Campaign Effects*, edited by H. Brady, and R. Johnston, 263–279. Ann Arbor: University of Michigan Press.

Clinton, J. D., and S. Rogers. 2013. "Robo-Polls: Taking Cues from Traditional Sources?" *PS: Political Science and Politics* 46: 333–337.

Crespi, I. 1988. *Pre-Election Polling: Sources of Accuracy and Error*. New York: Russell Sage Foundation.

Erikson, R. S., C. Panagopoulos, and C. Wlezien. 2004. "Likely (and Unlikely) Voters and the Assessment of Poll Dynamics." *Public Opinion Quarterly* 68 (4): 588–601.

Erikson, R. S., and C. Wlezien. 2012. *The Timeline of Presidential Elections: How Campaigns Do (and Do Not) Matter*. Chicago: University of Chicago Press.

Jackson, N. 2016. "Why HuffPost's Presidential Forecast Didn't See a Donald Trump Win Coming: Here's How We Blew It and What We're Doing to Prevent a Repeat." *The Huffington Post*. Accessed November 10, 2016. http://www.huffingtonpost.com/entry/pollster-forecast-donald-trump-wrong_us_5823e1e5e4b0e80b02ceca15.

Katz, J. 2016. "Hillary Clinton has an 85% Chance to Win." *The New York Times*. Accessed November 8, 2016. https://www.nytimes.com/interactive/2016/upshot/presidential-polls-forecast.html.

Kennedy, C., M. Blumenthal, S. Clement, J. D. Clinton, C. Durand, C. Franklin, K. McGeeney, et al. 2017. "An Evaluation of 2016 Election Polls in the United States." Report commissioned by the American Association of Public Opinion Research.

Lau, R. R. 1994. "An Analysis of the Accuracy of "Trial Heat" Polls During the 1992 Presidential Election." *Public Opinion Quarterly* 58(1): 2–20.

Martin, E., M. Traugott, and C. Kennedy. 2005. "A Review and Proposal for a New Measure of Poll Accuracy." *Public Opinion Quarterly* 69 (3): 342–369.

Mosteller, F., H. Hyman, P. McCarthy, E. Marks, and D. Truman. (1949). *The Preelection Polls of 1948: Report to the Committee on Analysis of Pre- Election Polls and Forecasts*. New York: Social Science Research Council.

Panagopoulos, C. 2009. "*Polls and Elections*: Preelection Poll Accuracy in the 2008 General Elections." *Presidential Studies Quarterly* 39 (4): 896–907.

Panagopoulos, C. 2013. "Campaign Effects and Dynamics in the 2012 Election." *The Forum* 10 (4): 36–39.

Panagopoulos, C., and B. Farrer. 2014. "*Polls and Elections* Preelection Poll Accuracy and Bias in the 2012 General Elections." *Presidential Studies Quarterly* 44 (2): 352–363.

Price, V., and N. J. Stroud. 2005. "Public Attitudes Toward Polls: Evidence from the 2000 U.S. Presidential Election." *International Journal of Public Opinion Research* 18 (4): 393–421.

Silver, N. 2014. "Here's Proof Some Pollsters are Putting a Thumb on the Scale." *Five Thirty Eight Blog*. Accessed November 14, 2016. https://fivethirtyeight.com/features/heres-proof-some-pollsters-are-putting-a-thumb-on-the-scale/.

Traugott, M. 2001. "Assessing Poll Performance in the 2000 Campaign." *Public Opinion Quarterly* 65: 389–419.

Traugott, M. 2005. "The Accuracy of the National Preelection Polls in the 2004 Presidential Election." *Public Opinion Quarterly* 69 (5): 642–654.

Vannette, D. and S. Westwood. (2013). "Voter Mobilization Effects of Poll Reports During the 2012 Presidential Campaign." Paper presented at the 68th Annual AAPOR Conference, May 17.

Wells, N. 2016. "Brexit and Trump Surprise Wins: The Polling Industry's Latest Black Eyes." CTV News. Accessed December 12, 2016. http://www.ctvnews.ca/world/brexit-and-trump-surprise-wins-the-polling-industry-s-latest-black-eyes-1.3200677#.