Robo-Polls: Taking Cues from Traditional Sources?

Joshua D. Clinton, *Vanderbilt University*
Steven Rogers, *Princeton University*

ABSTRACT After the 2012 Republican New Hampshire primary, 159 poll results were released prior to the subsequent nomination contests in the Republican presidential primary. More than two-thirds of these polls relied on interactive voice response (IVR) software to conduct the interviews. We evaluate the ability of polls to predict the vote-share for the Republican candidates Romney, Santorum, and Gingrich. We find no overall difference in the average accuracy of IVR and traditional human polls, but IVR polls conducted prior to human polls are significantly poorer predictors of election outcomes than traditional human polls even after controlling for characteristics of the states, polls, and electoral environment. These findings provide suggestive, but not conclusive, evidence that pollsters may take cues from one another given the stakes involved. If so, reported polls should not be assumed to be independent of one another and so-called poll-of-polls will be misleadingly precise.

In the Republican primary campaign, polls tracked the surge and decline of nearly every candidate. Even after Iowa and New Hampshire winnowed the field of candidates, more than 100 new polls were conducted by pollsters unaffiliated with the candidates and reported on websites such as RealClearPolitics.com. Consistent with previous research (Bartels 1986; Patterson 2005), coverage of the nomination campaign was driven by the near-continuous release of new polling results (Rosenstiel, Jurkowitz, and Sartor 2012).

Although a reasonable demand for knowing public opinion exists, not every poll is equally informative. Interpreting the many reported poll results requires understanding the possible errors that occur when interviewing respondents using the existing survey modes (see, for example, the review of Schaeffer and Dykema 2011). Gauging the reliability of various survey methodologies is difficult because the polling environment is affected by rapid technological and demographic changes that require reevaluations after every election cycle (e.g., Goidel 2011). Comparing the accuracy of various polls (and modes of interviewing) is essential for both interpreting poll results and placing their results in the proper context (see, for example, work by Kiesler and Sproull 1986; Chang and Krosnick 2009).

The validity of polls using interactive voice response (IVR) technology—so-called robo-polls—is continually debated. Although the cost-effectiveness of IVR polls combined with diminishing newsroom budgets means that IVR polls probably are here to stay (Asher 2012), some prominent news organizations (e.g., ABC News, NBC News, and the Associated Press) refuse to run stories based on IVR polls because of methodological concerns (Blumenthal 2009).

Even so, more IVR polls were conducted during the 2012 Republican primary than any other type of poll. In fact, following the New Hampshire primary, more than two-thirds of the publically reported state-level polls used IVR technology. Given the preponderance of IVR polls, objectively evaluating their performance relative to traditional modes of survey interviewing is important.

IVR and human polls differ in several respects besides whether the interviewer is a human being or not. To characterize the differences, we examine the sample of polls conducted during the 2012 Republican primary. We take advantage of the "open-source" nature of public opinion polls (Blumenthal 2005) and analyze all reported polls listed on RealClearPolitics.com along with a handful of additional polls.[1] To measure a poll's accuracy, we calculate the average absolute error between the primary outcome and the percentage predicted for Romney, Santorum, and Gingrich for each of the 159 polls conducted within four weeks of a primary election.[2]

Table 1 compares IVR and human polls along several dimensions that are potentially relevant for the accuracy of polls (Groves et al. 2004). Although IVR polls have larger sample sizes (presumably because of much cheaper costs), the average field period of 1.57 days means that most polls were in the field for only one or two days. This limited field period limits the ability to conduct callback attempts and reach initially unreachable respondents. It is also illegal to contact cell phone numbers with IVR polls. Potential problems with nonresponse and noncoverage are commonly

Joshua D. Clinton *is an associate professor of political science and codirector for the Center for the Study of Democratic Institutions at Vanderbilt University. He can be reached at josh.clinton@vanderbilt.edu.*
Steven Rogers *is a PhD candidate in the Department of Politics at Princeton University. He can be reached at rogerssm@princeton.edu.*

doi:10.1017/S1049096513000012
https://doi.org/10.1017/S1049096513000012 Published online by Cambridge University Press

Table 1

Comparison of IVR and Human Poll Characteristics

	IVR POLLS	HUMAN POLLS
Average Absolute Error	5.55	5.87
Field Period	1.57 Days	4.34 Days
Average Sample Size	816.41	528.79
Number	106	53

thought to affect the accuracy of polls, and these led former American Association for Public Opinion Research (AAPOR) president Peter Miller to argue that automated polls "rely too much on assumptions to make estimates based on data from an increasingly unrepresentative part of the population. Heroic assumptions will lead to big, unpredictable errors" (Cohen 2012).

In terms of the relative accuracy of human and IVR polls in the 2012 Republican primary, if we compare the average overall performance of human and IVR polls there is no difference in performance: table 1 reveals that the average absolute error is 5.87% for human polls and 5.55% for IVR polls. Despite conceptual reasons to think that IVR polls may be poorer predictors of election outcomes than human polls, the similar performance of IVR and human polls is consistent with AAPOR's conclusion after examining the performance of primary polls in 2008: "The use of either computerized telephone interviewing (CATI) techniques or interactive voice response (IVR) techniques made no difference to the accuracy of estimates" (Traugott 2009, 7).

WHO LEADS WHOM?

Digging deeper, however, reveals a pattern worth exploring. Given the skepticism directed toward IVR polls, perhaps these polls take cues from existing human polls to ensure that their results are not too different. Pollsters, especially those who use a technology that has not gained widespread acceptance, may have an incentive to ensure that their reported results are not implausible in light of existing beliefs about the state of public opinion at the time (presumably through the use of postestimation weighting). Although IVR polls are inexpensive relative to traditional polls, the costs of making egregious mistakes and, as a consequence, potential discrediting the credibility of a polling firm are extremely high given the large marketplace of polling firms. This risk-aversion may lead pollsters to take cues from existing polls. As one consultant acknowledges, "Taking into account what other polls on the same topic are reporting is one useful and appropriate piece of information in deciding how to 'tweak' the screening and weighting used in one's most recent poll" (Moore 2008).

Although it may make sense for a pollster to use all available information—including the results of other polls—to minimize the "total survey error" of the poll they are conducting (Weisberg 2005), exploring whether there is evidence consistent with pollsters taking cues from one another is important for two reasons. First, if the accuracy of a poll is largely attributable to the pollster's ability to devise appropriate weights to match existing poll results, it is unclear whether the poll contributes much (if any) new information about the state of public opinion. Second, if pollsters routinely take cues from one another to ensure that their poll results are reasonably similar, the poll results will no longer be independent, and therefore we will know less about public opinion than the number of polls would lead us to believe.

To illustrate the potential implications of pollsters taking cues from one another, suppose that there are 100 polls, but 99 are adjusted to ensure that their results are roughly consistent with the one poll thought to be the "gold-standard." If this cue-taking occurs, the fact that we have 100 polls with similar results is misleading; we actually have only information from a single poll because the results of the 99 other polls are weighted to replicate the results of a single poll. If this behavior occurs, averaging these poll results to produce a "poll-of-polls" (Hillygus 2011) will produce estimates with a larger margin of error than the number of aggregated polls would otherwise suggest because the results of the averaged polls cannot be considered independent.[3]

To better understand the accuracy of reported IVR polls and explore whether evidence consistent with cueing behavior exists, we examine if IVR polls are more accurate predictors of Election Day results when the results of human polls are already known. Because IVR polls were fielded for nearly every Republican nomination contest but human polls were not, we compare the average absolute error for three groups of polls: human polls, IVR polls where no human polls were conducted, and IVR polls in states where human polls were reported.[4] As an initial exploration of the cueing hypothesis, we examine if the accuracy of polls using IVR technology depends on whether a human poll was conducted prior to the IVR polls in the state.

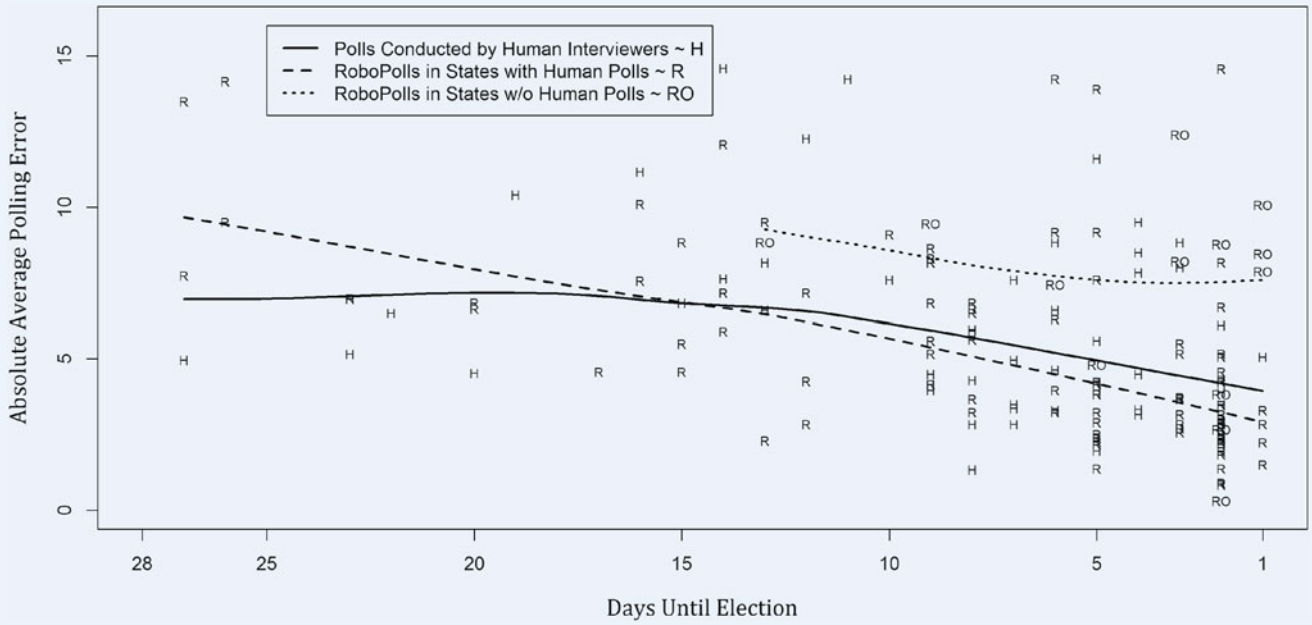
Figure 1 plots the average absolute error for 159 reported polls in the 31 Republican nomination contests in which a poll was conducted within four weeks of the election controlling for how many days in advance of the election each poll was conducted for each of the three types of polls.[5] We summarize the average performance over time for each group of polls using a simple loess regression line. The decreasing slope of the three plotted lines reveals that all three types of polls converge toward the true election outcome as Election Day approached. However, the difference in the dotted and dashed lines reveals that the performance of IVR polls appears to depend on whether a human poll was also conducted in the state. The average absolute error for polls conducted in states without human polls (dotted line) is higher than error for IVR polls in states containing human polls (dashed line). This pattern indicates that IVR polls conducted in states without human polls are poorer predictors of election outcomes. Meanwhile, in states where both IVR and human polls exist, no difference in the average absolute error is shown. Overall, the accuracy of IVR polls seems to depend on whether human polls are present and the performance of human polls and IVR polls are indistinguishable when both are present. These results are both consistent with the hypothesis that IVR polls in the 2012 Republican primary took cues from human polls.

The patterns evident in figure 1 are clearly only suggestive and speculative. In particular, we may worry whether the difference in IVR poll performance is because of the difficulty of polling in states where human polls were not conducted (e.g., the Missouri Republican primary was ignored by telephone polls because no delegates were at stake). To account for this possibility and increase the precision of our comparisons, we collect information on the states that are being polled and the characteristics of the reported polls. Moreover, we identify both IVR polls fielded in states without human polls and those conducted *before* the public human

https://doi.org/10.1017/S1049096513000012 Published online by Cambridge University Press

Figure 1

Loess Curves for Error Across Different Types of Polls



The dotted line plotting the average absolute error for polls conducted in states without human polls is higher than the dashed line plotting the average absolute error for IVR polls in states containing human polls, suggesting IVR polls conducted in states without human polls are poorer predictors of the actual election outcome.

polls. By analyzing the performance of IVR polls in states where telephone polls are conducted shortly thereafter we can ensure that (1) the analyzed IVR polls cannot take cues from their traditional counterparts, and (2) any differences relative to human polls cannot be the result of differences in the polling environment. Descriptively, IVR polls were conducted before human polls in seven states, and human polls were conducted first in 24 states.

To determine if the pattern in figure 1 persists after refining the analysis and accounting for potential confounds, we use a regression model to predict the average absolute error between the predicted support for the leading candidate and the candidate's vote share on Election Day controlling for both poll and state characteristics.[6] Table 2 reports the results of several specifications and reveals that the findings of figure 1 persist. The first column of regression coefficients treats all IVR polls as equivalent and confirms the lack of difference between IVR and human polls. Controlling for characteristics of the states (e.g., whether the election was an open primary, closed primary, or caucus) as well as characteristics of the electoral geography (e.g., how many times candidates visited the state, whether the state voted for McCain in 2008) we find that the difference in average absolute error for IVR and human polls is −.256 and statistically indistinguishable from zero. This confirms the early finding that, on average, no difference in the performance of IVR and telephone polls exists.

The second column controls for whether the IVR poll was conducted prior to any human polls in the state. The results of this regression reveal that IVR polls that are conducted without the benefit of a preceding human poll have an average absolute error that is 3.229 larger than the average absolute error of IVR polls that are conducted after human polls. This difference is statistically distinguishable from zero. Moreover, because the effect of an IVR poll is not statistically distinguishable from zero (an effect of −.656) this implies that IVR polls and human polls perform equally well when the results of a human poll have already been reported. Together, these results confirm the pattern evident in figure 1: there is no difference in the accuracy of IVR polls and human polls when IVR polls occur after a human poll, but IVR polls are significantly poorer predictors if human polls are not conducted first. The third column of table 1 reveals that these conclusions persist if we expand the specification to control for various features of the polls and polling organizations themselves.

To put the results in broader context, two aspects of our results are worth noting. First, 12 of the 17 IVR polls conducted before human polls are from the same firm. Therefore it is impossible to determine whether the effects we document are because of a single pollster or not. Second, most of the effect is also attributable to the elections in Missouri, Colorado, and Minnesota held on February 7 when only IVR polling occurred. We control for the campaign environment (e.g., type of election, number of candidate visits, order of the primary), but it is possible that the campaign environment plays an important role. Even so, our finding is important in light of the shift in campaign coverage that resulted from "Santorum's surprising sweep" (Gardner and Helderman 2012) because inaccurate IVR polling likely contributed to misguided expectations. Exceeding the expectations that were set by IVR polls gave Santorum's campaign new life with his first victories since Iowa and enabled him to extend the fight for the Republican nomination (Shear 2012).

In either case, not all polls are equal, and the results are consistent with the possibility that pollsters (or at least one pollster) take cues from one another. This is particularly important as the expectations and campaign narratives used by the reporters, who cover the campaign, are heavily influenced by the polls being released.

https://doi.org/10.1017/S1049096513000012 Published online by Cambridge University Press

Table 2

Dependent Variable: Absolute Average Error of Polls from Republican Nomination Contests

VARIABLE	ROBO-POLL ERROR	CONTROL FOR CUES	POLLSTER CONTROLS
Robo-Poll	−0.256	−0.656	−0.173
	(0.503)	(0.497)	(1.033)
Robo-Poll Conducted before any Human Poll		3.229*	3.679*
		(0.913)	(0.978)
Days Until Election	0.205*	0.210*	0.217*
	(0.036)	(0.035)	(0.038)
Total Visits by 3 Major Candidates (Logged)	−0.059	0.092	−0.008
	(0.292)	(0.284)	(0.339)
Primary Order	−0.002	−0.019	−0.038
	(0.074)	(0.072)	(0.076)
Caucus State	−2.574*	−4.512*	−4.833*
	(1.210)	(1.287)	(1.323)
Open Primary	−2.088*	−2.495*	−2.548*
	(0.760)	(0.741)	(0.767)
2008 Pres. Red State	−1.305	−1.495	−1.586
	(0.901)	(0.869)	(0.904)
Swing State	−3.015*	−3.137*	−2.772*
	(0.826)	(0.797)	(0.847)
Southern Dummy	0.851	0.949	1.160
	(0.758)	(0.730)	(0.763)
Population Density	−0.007	−0.009*	−0.011*
	(0.003)	(0.003)	(0.004)
Cell Phones			−1.428
			(0.820)
Likely Voters Sample			0.890
			(1.157)
Sample Size			0.017
			(0.704)
Days in Field			−0.022
			(0.142)
Public Policy Polling Dummy			−0.626
			(0.745)
Rasmussen Dummy			−0.717
			(0.689)
Academic Poll Dummy			0.883
			(1.031)
News Organization Poll Dummy			0.825
			(0.776)
Debate During Poll Dummy			−0.596
			(0.621)
Constant	8.489*	8.932*	8.603*
	(2.007)	(1.937)	(4.956)
R-Squared	0.275	0.332	0.365
N	159	159	159

*Denotes two-sided significance at .05 or lower.

CONCLUSION AND IMPLICATIONS

The apparent equivalence of IVR polls and human polls in the 2012 Republican primary appears to depend on human polls conducted prior to the IVR polls. IVR polls conducted when no human polls exist are poorer predictors of election outcomes than IVR polls fielded after the results of human polls have been made public. There is no difference between IVR polls and human polls after the results of a human poll have been reported. This suggests, but certainly does not prove, that some IVR polls may use earlier human polls to adjust their results to ensure that they are not notably different from existing polls and beliefs.

Pollsters understand that their results are being compared to the results of prior polls, and polls created for public consumption have incentives to ensure that their results are roughly consistent with the narrative presented by the press if they want to garner public attention. Also, pollsters have financial incentives to "get it right" that may make them leery of ignoring the information contained in other polls. Our results are consistent with what we would expect if IVR polls took cues from the results of more established methodologies—IVR polls do as well as traditional human polls when both are present, but they do poorly when there are no other polls from which to cue. However, the nature of these investigations means that our results are necessarily suggestive rather than definitive. Beyond the implications for interpreting IVR polls, the larger point is that if polls take cues from one another, then the hundreds of polls being reported are not really as informative as the number of polls imply.

Our results suggest we should closely examine pollsters' methodological decisions as they may have implications for how we interpret the results. For example, although we present suggestive evidence of IVR pollsters cueing in the Republican primary, it is possible that traditional human counterparts use similar strategies.[7] Public opinion polls for the 2008 election converged as the general election approached (Moore 2008), and while voters may have been coming to their fundamental "enlightened preferences" (Gelman and King 1993), part of the observed convergence may also be due to decisions by pollsters rather than voters.

Political polls can be extremely valuable and insightful, but we should exercise care in their interpretation. Although taking cues from other polls may improve the performance of the pollster who cues and if the goal is to produce an estimate that accounts for all available data and prior beliefs it even may be preferred—this practice makes it difficult for the objective observer to determine the current state of public opinion. To

https://doi.org/10.1017/S1049096513000012 Published online by Cambridge University Press

prove that cueing exists is exceptionally difficult but our results suggest that it is a possibility that consumers of public polls should at least consider when gauging what we think we know about public opinion.

ACKNOWLEDGMENT

The authors thank students from the Vanderbilt University undergraduate honors seminar, Predicting Elections, for their assistance in data collection. ■

NOTES

1. To supplement our list, we used WikiPedia (http://en.wikipedia.org/wiki/Statewide_opinion_polling_for_the_Republican_Party_presidential_primaries,_2012#Polling_for_completed_primaries). The analysis relies on polling data from states whose Republican nomination contests were held after the New Hampshire primary and before the suspension of Sen. Santorum's campaign. We examine polls whose field period ended within four weeks of that state's Election Day. To identify states where no human polls took place, we continued data collection until there was a two-week window without a public poll. We contacted American Research Group and 2012 NevadaCaucus.com to acquire information about their polling methods. They did not respond so we excluded their polls from the analysis.
2. To measure poll performance in a multi-candidate election, we use the average distance in the absolute value between the Election Day results and the support for Romney, Gingrich, and Santorum. For each of the three candidates we took the absolute value of the difference between the percentage of votes that the candidate received and the support indicated by the poll. For each poll, we then calculate the average error using Romney, Santorum, and Gingrich. Because Gingrich and Santorum were not on the Virginia ballot, and Gingrich was not on the Missouri ballot, we divide by the number of major contenders in these two instances (i.e., one in Virginia and two in Missouri). We look at polling in states excluding New Hampshire and Iowa because polling started in these states more than a year before the actual primary date.
3. To be clear, this is a statement about the variance of the polling error, not a statement about the bias of the results.
4. The states that only had IVR polls were Colorado, Maine, Missouri, Minnesota, Mississippi, and Washington state. No public polling was found for Alaska, Kansas, North Dakota, and Wyoming.
5. To simplify presentation, figure 1 omits the January 21, 2012, Public Policy Polling poll from Minnesota. This state had no human polls, and the error of this IVR poll was the highest of any IVR poll at 18.1. We additionally omit it from the regression analysis.
6. For state characteristics we control for whether the state was a "red" state or thought to be an important swing state in the 2012 election (coded to be the states of Colorado, Florida, Ohio, Virginia, and Wisconsin following Kuhn (2011)), whether it was a southern state, the type of primary used (caucus, closed primary, open primary), state population density, and the order of the state's primary. For poll characteristics we control for how many days prior to the election the poll was conducted, the length of field period, the sample size of the poll, whether the poll included cell phone respondents or not, whether the sample was a likely voter or registered vote sample, whether a debate occurred during the survey's field period, whether the poll was conducted by an academic institution, news organization, or a potential partisan, and whether the poll was conducted using IVR technology.
7. Although we can compare the performance of early IVR polls to the performance of early human polls, it is difficult to determine if early human polls are also affected by the lack of earlier polls because of the lack of a suitable benchmark. It is impossible to determine if early human polls perform with less accuracy than later human polls because of the early polls have no polls from which they can take cues from or whether it is because polling so many days prior to the election inevitably contains more error.

REFERENCES

Asher, Herbert. 2012. *Polling the Public: What Every Citizen Should Know*. Washington, DC: CQ Press.

Bartels, Larry M. 1986. *Presidential Primaries and the Dynamics of Public Choice*. Princeton, NJ: Princeton University Press.

Blumenthal, Mark. 2005. "Toward an Open-Source Methodology: What We Can Learn from the Blogosphere." *Public Opinion Quarterly* 69 (5): 655–69.

Blumenthal, Mark. 2009. "The Case for Robo-Pollsters: Automated Interviewers Have Their Drawbacks, But Fewer Than Their Critics Suggest." *National Journal*, September 14. http://www.nationaljournal.com/njonline/the-case-for-robo-pollsters-20090914.

Chang, Linchiat, and John A. Krosnick. 2009. "National Surveys via Rdd Telephone Interviewing versus the Internet, Comparing Sample Representativeness and Response Quality." *Public Opinion Quarterly* 73 (4): 641–78.

Cohen, Jon. 2012. "Covering Automated Surveys in 2012." *Washington Post*, February 1. http://www.washingtonpost.com/blogs/behind-the-numbers/post/covering-automated-surveys-in-2012/2012/02/01/gIQANvmJIR_blog.html.

Gardner, Amy, and Rosalind S. Helderman. 2012. "Mitt Romney's Presidential Campaign Stuck in Lukewarm." *Washington Post*, February 9. http://www.washingtonpost.com/politics/mitt-romneys-presidential-campaign-stuck-in-lukewarm/2012/02/09/gIQAhEMh2Q_story.html.

Gelman, Andrew, and Gary King. 1993. "Why Are American Presidential Election Campaign Polls So Variable When Votes Are So Predictable?" *British Journal of Political Science* 23 (4): 409–51.

Goidel, Kirby. 2011. *Political Polling in the Digital Age: The Challenge of Measuring and Understanding Public Opinion*. Baton Rouge: Louisiana State University Press.

Groves, Robert M., F.J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and R. Tourangeau. 2004. *Survey Methodology*. New York: Wiley-Interscience.

Hillygus, D. Sunshine. 2011. "The Evolution of Election Polling in the United States." *Public Opinion Quarterly* 75 (5): 962–81.

Kiesler, Sara, and Lee S. Sproull. 1986. "Response Effects in the Electronic Survey." *Public Opinion Quarterly* 50 (3): 402–13.

Kuhn, David Paul. 2011. "The Critical 2012 Swing States." *RealClearPolitics.com*, March 4. http://www.realclearpolitics.com/articles/2011/03/04/critical_big_five_swing_states_electoral_map_obama_ohio_florida_virginia__wisconsin_colorado_109116.html.

Moore, David. 2008. "The Fluctuating Convergence Mystery," *Pollster.com* December 8. http://www.pollster.com/blogs/the_fluctuating_convergence_my.php?nr=1.

Patterson, Thomas E. 2005. "Of Polls, Mountains: U.S. Journalists and Their Use of Election Surveys." *Public Opinion Quarterly* 69 (5): 716–24.

Rosenstiel, Tom, Mark Jurkowitz, and Tricia Sartor. 2012. "How the Media Covered the 2012 Primary Campaign." *Pew Research Center's Project for Excellence in Journalism*, April 23. http://www.journalism.org/analysis_report/romney_report.

Schaeffer, Nora Cate, and Jennifer Dykema. 2011. "Questions for Surveys: Current Trends and Future Directions." *Public Opinion Quarterly* 75 (5): 909–61.

Shear, Michael D. 2012. "In Santorum's Sweep, Sign of G.O.P. Unease with Romney." *New York Times*, February 8. http://www.nytimes.com/2012/02/09/us/politics/santorum-sweep-sets-stage-for-new-battle-in-republican-race.html?pagewanted=all.

Traugott, M.W. 2009. *An Evaluation of the Methodology of the 2008 Pre-Election Primary Polls*, Unpublished Manuscript. AAPOR Ad Hoc Committee on the 2008 Presidential Primary Polling.

Weisberg, Herbert. 2005. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.

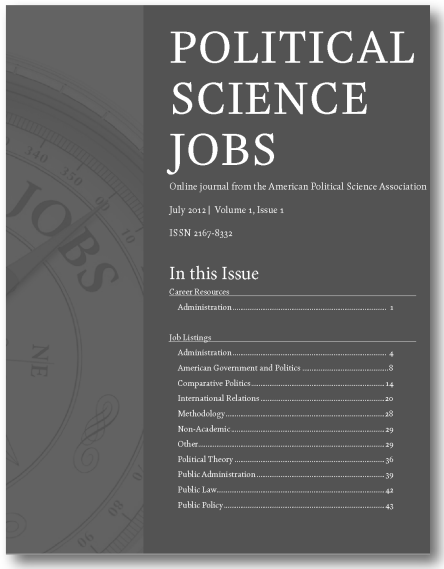
https://doi.org/10.1017/S1049096513000012 Published online by Cambridge University Press

American Political Science Association

New eJournal

Political Science Jobs
New eJournal from APSA

In July 2012, APSA launched a new monthly and online journal focused on jobs in political science. *Political Science Jobs* contains active job listings from APSA's eJobs database. The combination of the new online journal and the daily APSA eJobs database provides a continued platform for a transparent, efficient, and ethically grounded political science labor market. APSA has maintained an open academic job market for many years, and this new journal forms the next phase of these efforts.



Format and Frequency

Political Science Jobs includes all active positions in a monthly downloadable PDF and link to the searchable eJobs system online.

APSA Members

APSA members (individuals and departments) have free access to the online journal as a benefit of membership. All members also receive free access to the daily PDF of eJobs listings.

Replacing *PS Supplement* with *Political Science Jobs*

Political Science Jobs replaces the former *PS Supplement* of job listings. The April 2012 *PS Supplement* was the last published volume since establishing the journal in 2008 to fulfill federal regulatory requirements tied to international hiring requirements. These federal rules recently changed to recognize online or web-based professional journals as an eligible outlet.

www.apsanet.org/jobs

apsa

1527 New Hampshire Avenue NW, Washington, DC 20036 | 202.483.2512 | www.apsanet.org

https://doi.org/10.1017/S1049096513000012 Published online by Cambridge University Press