

Robo-Polls: Taking Cues from Traditional Sources?

May 3, 2012

Joshua D. Clinton¹
Steven Rogers²

After the 2012 Republican New Hampshire primary, there were 159 poll results released prior to the subsequent nomination contests. More than two-thirds of these polls relied on “Interactive Voice Recognition” (IVR) software. In this research note, we evaluate the ability of polls to predict the vote-share for the Republican candidates Romney, Santorum and Gingrich. We find no difference in the average accuracy of IVR and traditional human polls, but IVR polls conducted prior to human polls do significantly worse than traditional human polls even after controlling for characteristics of the states, polls, and electoral environment. These findings provide suggestive 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 therefore be misleadingly precise.

¹ Associate Professor of Political Science and Co-Director Center for the Study of Democratic Institutions, Vanderbilt University, Nashville, TN 37205. Josh.clinton@vanderbilt.edu

² Ph.D. Candidate, Princeton University Department of Politics. rogerssm@princeton.edu

In the Republican primary campaign, polls tracked the surge and decline of nearly every candidate. Even after Iowa and New Hampshire winnowed the field, over a hundred new polls were conducted by pollsters unaffiliated with the candidates and reported on websites such as *RealClearPolitics.com*. Consistent with previous research by political scientists (Bartels 1986; Patterson 2005), coverage of the nomination campaign was driven by the near-continuous release of new polling results (Rosnenstiel et. al. 2012).

There is a reasonable demand for knowing public opinion, but not every poll is equally informative. Interpreting the many reported poll results requires understanding the possible errors that may occur when interviewing respondents using the various 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 can necessitate reevaluations after every election cycle (e.g., Goidel 2011). Comparing the accuracy of various polls (and modes of interviewing) is absolutely essential for interpreting poll results and placing the results in their proper context (see, for example, work by Kiesler and Sproul 1986, Chang and Krosnick 2010).

The validity of polls using employing “Interactive Voice Response” (IVR) technology – so-called robo-polls – is a subject of continuing debate. While the cost-effectiveness of IVR polls combined with diminishing newsroom budgets means that IVR polls are probably here to stay (Asher 2012), some prominent news organizations (e.g., NBC News, the Associated Press, and the *Washington Post*) refuse to run stories based on IVR polls because of methodological concerns (Blumenthal 2011).

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, over two-thirds of the publically reported state-level polls employed IVR technology. Given the preponderance of IVR polls, it is important to objectively evaluate their performance relative to traditional modes of survey interviewing.

IVR and human polls differ in several respects besides the fact of 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.³ 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 4 weeks of a primary election.⁴

³ 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 after the New Hampshire primary and before the suspension of Rick 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 2012NevadaCaucus.com to acquire information about their polling methods without hearing back from them, so we excluded their polls from the analysis.

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

⁴ 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 VA ballot, and Gingrich was not on the Missouri ballot, we divide by the number of major contenders in these two instances (i.e., 1 in VA and 2 in MO).

Table 1 compares IVR and human polls along several dimensions that are potentially relevant for the accuracy of polls (Groves et. al. 2004).

[Insert Table 1 Here]

While IVR polls have larger sample sizes (presumably due to cheaper costs), the average field period of 1.57 days means that most were in the field for only one or two days. This limited field period necessarily limits the ability to conduct callback attempts and reach initially unreachable respondents. Such limits are commonly thought to affect the accuracy of polls, and they led former 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 simply compare the average performance of human and IVR polls Table 1 reveals that there is no difference -- 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 perform worse than human polls, the similar performance of IVR and human polls is consistent with the conclusion that AAPOR reached when 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, pg. 7).

Who Leads Whom?

Digging deeper, however, reveals a pattern worth exploring further. Given the skepticism directed towards IVR polls, perhaps such polls take cues from existing human polls to ensure that their results are not too far off. Pollsters – especially those using a technology

that has yet to gain widespread acceptance – may have an incentive to alter their methodologies to ensure that their reported results are not implausible given prior beliefs about the state of public opinion. While IVR polls are inexpensive relative to traditional polls, the costs of making egregious mistakes -- and potential discrediting the credibility of a firm as a consequence -- are extremely high in the crowded marketplace for polling data. Perhaps as a result, cueing behavior occurs. As one consultant suggestively 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).

While 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 that the poll contains much (if any) additional information. 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 we will therefore know less about public opinion than the number of polls would lead us to believe.

To illustrate the potential implications of pollsters cue-taking, suppose that there are 100 polls, but 99 of them are adjusted to ensure that their results are roughly consistent with the 1 poll thought to be the "gold-standard." If cue-taking occurs, the fact that we have 100 polls with similar results is misleading because we actually only have information from a single

poll -- the results of 99 of the polls are designed to replicate the results of a single poll. If this behavior occurs, methods such as the increasingly common “poll-of-polls” (Hillygus 2011) will report misleadingly precise results, and we will have more uncertainty about public opinion than the number of aggregated polls would suggest.⁵

To better understand the accuracy of reported IVR polls and explore whether there is evidence consistent with cueing behavior, we examine if IVR polls are more accurate when the results of human polls are already known. Because IVR polls were fielded for nearly every Republican nomination contest and human polls were not, we can 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.⁶ As an initial exploration of the cueing hypothesis, we examine if the performance of robo-polls employing IVR technology depends on whether a human poll was also conducted in the state.

[Figure 1 About Here]

Figure 1 plots the average absolute error for 159 reported polls controlling for how many days in advance of the election each poll was conducted for each of the three types of polls.⁷ 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 towards the true election outcome. However, the difference in the dotted and dashed lines reveals that the performance of IVR polls appears to depend on

⁵ To be clear, this is a statement about the variance of the polling error, not a statement about the bias of the results.

⁶ States only with IVR polls were CO, ME, MO, MN, MS, and WA. No public polling was found for AK, KS, ND, and WY.

⁷ To simplify presentation, Figure 1 omits the Jan. 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.

whether not 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), indicating IVR polls conducted in states without human polls do much worse. Meanwhile, in states where both IVR and human polls exist, there is no difference in the average absolute error. The fact that the accuracy of IVR polls seems to depend on whether human polls are present and the fact that the performance of human polls and IVR polls are indistinguishable when both are present, 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 only suggestive and speculative. In particular, we may worry whether the difference in IVR poll performance is due to the difficulty of polling in states where human polls were not conducted. 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. We identify both IVR polls fielded in states without human polls and those conducted *before* the public human polls. These IVR polls cannot take cues from their traditional counterparts and any differences relative to human polls cannot be due to differences in the polling environment.

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 controlling for both poll and state characteristics.⁸ Table 2 reports the results of several specifications and reveals that the findings of Figure 1 persist.

⁸ For state characteristics we control for whether the state was a “red” state or a swing state in the 2008 election, 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,

[Table 2 About Here]

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 a primary or not) 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 not statistically distinguishable from zero.

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 such IVR polls have an average absolute error that is 3.229 more than those conducted after human polls. This difference is statistically distinguishable from zero. Moreover, the fact that the effect of an IVR poll is not statistically distinguishable from zero (an effect of $-.656$) 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 do significantly worse 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 themselves.

Two aspects of the result are worth noting. First, 12 out of the 17 IVR polls conducted before human polls are from the same firm. It is therefore impossible to determine whether

whether the poll included cellphone 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 (i.e., RoboPoll).

the effects are due to a single pollster or not. Second, most of the effect is also attributable to the elections in MO, CO and MN held on February 7th where only IVR polling occurred. We control for the campaign environment (e.g., type of election, number of candidate visits, order of the primary), but this finding is particularly important given the shift in campaign coverage that resulted from “Santorum’s surprising sweep” (Gardner & Heldnerman 2012). Inaccurate IVR polling likely contributed to misguided expectations. Exceeding these gave the Santorum campaign new life with its first victories since Iowa and enabled him to extend the fight for the Republican nomination (Shear 2012).

The important point to emphasize in either case is the fact that not all polls are equal and that the results are consistent with the possibility of pollsters (or at least one pollster) taking cues from one another. This is particularly important insofar as the expectations and campaign narratives used by the reporters covering the campaign are heavily influenced by the polls being released,

Conclusion and Implications:

The apparent equivalence of IVR polls and human polls in the 2012 Republican primary appears to depend on human polls being conducted prior to the IVR polls. IVR polls conducted when there are no human polls do worse than IVR polls fielded after the results of human polls have been made public, and there is no difference between IVR polls and human polls once the results of a human poll have been reported. This suggests, but certainly does not prove, that at least 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 know 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 being told in the press if they want to garner public attention. Pollsters also have further financial incentives to get it right which may make them leery of ignoring the information contained in other polls. The results we find 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 worse when there are no other polls to cue off of. However, the nature of the investigations means that our results are necessarily suggestive rather than definitive. Beyond the implications for interpreting IVR polls, the larger point here 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 would 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 certainly possible that traditional human counterparts employ similar strategies.⁹ Public opinion polls for the 2008 election converged as the general election approached (Moore 2008) and voters may or may not have been coming to their fundamental "enlightened preferences" (Gelman & King 1993), but part of the convergence may also be due to decisions by pollsters rather than voters.

Political polls can be extremely valuable and insightful, but it is important to exercise some care in their interpretation. While taking cues from other polls may improve the

⁹ It is certainly possible that human polls may also take cues from existing polls, but it is difficult to investigate this possibility because of the lack of available data.

performance of the pollster that cues -- and if the goal is to produce an estimate that accounts for all available data and prior beliefs it may even be preferred – such a practice makes it difficult for the objective observer to determine the current state of public opinion. While it is obviously difficult to prove that cueing exists, our results suggest that it is a possibility that consumers of public polls should consider when attempting to gauge what we think we know about the state of public opinion.

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

Table 2: Dependent Variable: Absolute Average Error of Polls from Republican Nomination Contests

Variable	RoboPoll Error	Control for Cues	Pollster Controls
RoboPoll	-0.256 (0.503)	-0.656 (0.497)	-0.173 (1.033)
RoboPoll Conducted before any Human Poll		3.229* (0.913)	3.679* (0.978)
Days Until Election	0.205* (0.036)	0.21* (0.035)	0.217* (0.038)
Total Visits by 3 Major Candidates (Logged)	-0.059 (0.292)	0.092 (0.284)	-0.008 (0.339)
Primary Order	-0.002 (0.074)	-0.019 (0.072)	-0.038 (0.076)
Caucus State	-2.574* (1.21)	-4.512* (1.287)	-4.833* (1.323)
Open Primary	-2.088* (0.76)	-2.495* (0.741)	-2.548* (0.767)
2008 Pres. Red State	-1.305 (0.901)	-1.495 (0.869)	-1.586 (0.904)
Swing State	-3.015* (0.826)	-3.137* (0.797)	-2.772* (0.847)
Southern Dummy	0.851 (0.758)	0.949 (0.73)	1.16 (0.763)
Population Density	-0.007* (0.003)	-0.009* (0.003)	-0.011* (0.004)
Cell Phones			-1.428 (0.82)
Likely Voters Sample			0.89 (1.157)
Sample Size (Logged)			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* (2.007)	8.932* (1.937)	8.603 (4.956)
R-Squared	0.275	0.332	0.365
N	159	159	159

Loess Curves for Error Across Different Types of Polls

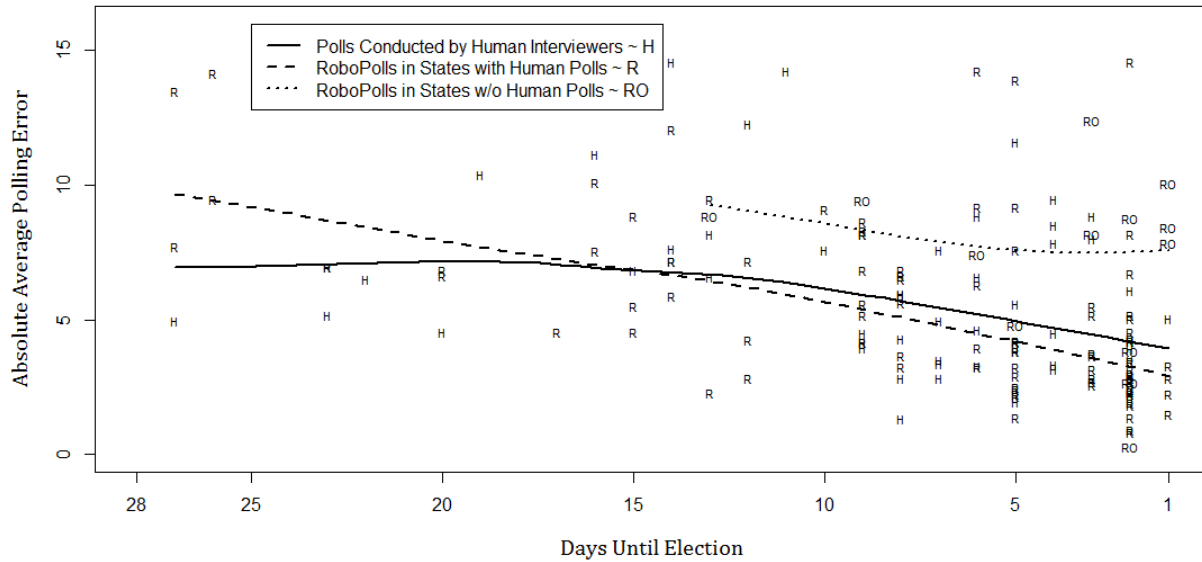


Figure 1: The dotted line plotting the average absolute error for polls conducted in states without human polls is higher than the dashed line plotting the average absolute error for IVR polls in states containing human polls, suggesting IVR polls conducted in states without human polls do much worse.

References

- Asher, Herbert. 2012. *Polling the Public: What Every Citizen Should Know*. CQ Press: Washington DC.
- Bartels, Larry M. 1986. *Presidential Primaries and the Dynamics of Public Choice*. Princeton, NJ: Princeton University Press, 1988.
- Blumenthal, Mark. 2005. "Toward an Open-Source Methodology: What We Can Learn From The Blogosphere," *Public Opinion Quarterly* 69(5): 655-669.
- Blumenthal, Mark. 2009 "The Case For Robo-Pollsters: Automated Interviewers Have Their Drawbacks, But Fewer Than Their Critics Suggest" *National Journal* (<http://www.nationaljournal.com/njonline/the-case-for-robo-pollsters-20090914>). Accessed April 25, 2012.
- Chang, Linchiat, Krosnick, John A. 2009. "National Surveys Via Rdd Telephone Interviewing Versus the Internet, Comparing Sample Representativeness and Response Quality," *Public Opinion Quarterly*, 73(4): 641-678.
- Cohen, Jon, 2012. "Covering Automated Surveys in 2012," *Washington Post*. (http://www.washingtonpost.com/blogs/behind-the-numbers/post/covering-automated-surveys-in-2012/2012/02/01/gIQANvmJIR_blog.html). Accessed April 25, 2012.
- Gardner, Amy and Rosalind S. Helderma2012. "Mitt Romney's presidential campaign stuck in lukewarm," *Washington Post*. Feb. 9. 2012. (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): 409-451.
- Goidel, Kirby. 2011. *Political Polling in the Digital Age: The Challenge of Measuring and Understanding Public Opinion*. Baton Rouge: Louisiana State University.
- [Groves, Robert M.](#), F.J. Fowler, [Mick P. Couper](#), [James M. Lepkowski](#), [Eleanor Singer](#), and R. Tourangeau. 2004. *Survey Methodology*. New York: Wiley.
- Hillygus, D. Sunshine. 2011. "The Evolution of Election Polling in the United States," *Public Opinion Quarterly* 75(5): 962-981.
- Kiesler, Sara, Sproull, Lee S. 1986. "Response Effects in the Electronic Survey," *Public Opinion Quarterly*, 50(3): 402-413.

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*. (http://www.journalism.org/analysis_report/romney_report). Accessed April 25, 2012.

Lavrakas, Paul J. 2008. "The Experts Reply on the Poll Convergence Mystery," *Survey Practice* December 19. (<http://surveypractice.wordpress.com/2008/12/19/the-experts-reply-convergence-mystery/>). Accessed April 25, 2012.

Moore, David. 2008. "The Fluctuating Convergence Mystery," *Pollster.com* December 8. (http://www.pollster.com/blogs/the_fluctuating_convergence_my.php?nr=1). Accessed April 25, 2012.

Patterson, Thomas E. 1985. "Of Polls, Mountains: U.S. Journalists and Their Use of Election Surveys," *Public Opinion Quarterly* 69(5): 716-724.

Schaeffer, Nora Cate and Jennifer Dykema. 2011. "Questions for Surveys: Current Trends and Future Directions," *Public Opinion Quarterly* 75(5): 909-961.

Shear, Michael D. "In Santorum's Sweep, Sign of G.O.P. Unease with Romney." *The New York Times*. February 8, 2012. (<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*, 2009, AAPOR Ad Hoc Committee on the 2008 Presidential Primary Polling.

van Lohuizen, Jan and Robert Wayne Samohyl. 2011. "Method Effects and Robo-polls" *Survey Practice*, February: www.surveypractice.org

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