Touchdown!

Online polling scores big in November 2000

*By Humphrey Taylor, John Bremer, Cary Overmeyer,
Jonathan W. Siegel, and George Terhanian*

The past two years have seen many vigorous debates and disagreements between defenders and critics of internet-based polls. The fiercest critics argue that, because most online polls do not use probability samples of the population, they are bound to be seriously inaccurate.

The November 2000 elections provided Harris Interactive with a unique opportunity to test the accuracy of our online survey methods and, specifically, our ability to predict elections. Happily, the accuracy of our efforts in the 73 different races we covered exceeded our most optimistic expectations.

• While almost all the polls did very well in the national vote for president in 2000, only two polls correctly showed the two candidates tied: the Harris Interactive online poll and the Harris Interactive telephone poll.

• Our predictions for the presidential election votes in 38 states were also quite accurate, although not as accurate as the national polls. (We did not cover all 50 states because of difficulties in developing adequate online samples in 12 of the smallest states. Specifically, we felt that some significant subpopulations, such as rural blacks in Mississippi, Alabama, Arkansas and South Carolina, were not adequately represented in our database.)

Humphrey Taylor is chairman of the Harris Poll. John Bremer is research scientist, Cary Overmeyer is senior research associate, Jonathan W. Siegel is vice president of general markets, and George Terhanian is vice president of research and methodology, Harris Interactive.

According to the *Research Business Report*, on average, Harris Interactive's results in these 38 states "were twice as accurate" as all other telephone polls conducted in the same states.

• In the Senate races we covered in 27 states, our forecasts were marginally more accurate. We had the winner ahead in 26 of the 27 races, and our average error was 2.2 percentage points for the two main candidates.

• Predictions for seven governors' elections were even more accurate. There the average error was 1.9 points on the two main candidates.

If we had conducted only one online poll on the national vote for president, it would have been easy to believe this success was just luck. Indeed, with so many national polls coming as close as they did to the final result in this election, having the candidates tied, and being "the best poll," was surely luck in part. But with 73 different elections to compare, it is obvious that our accuracy cannot be explained by statistical accident. We have demonstrated that online polls can be designed and executed to measure voting intentions with high accuracy.

The key elements of the methods used in our final online election surveys included, first, the development and building of a large national online panel of willing respondents. The panel was recruited through multiple sources, including banner advertisements and sweepstakes that have run across the web, the Harris/Excite daily poll, product registrations on Excite and Netscape, and telephone surveys. Since the resulting sample was by no means representative in a statistical sense of either the US internet population or the total US population, a sampling and weighting approach had to be designed to adjust for the different likelihood, or propensity, of different respondents being in our online or telephone samples.

In September and October we conducted two similar, but smaller, surveys as "trial runs" for the main event. Then, between October 31 and November 6, we did interactive online interviews with a total of 240,666 adults who, based on their answers, were categorized as "likely voters."

Finally, the data in the 39 samples (the national sample and the 38 state samples) were weighted, using both demographic weights virtually identical to those used in the nationwide Harris Interactive telephone poll, and a propensity score reflecting respondents' probability of being online.

The propensity score was developed based on questions measuring behavior and attitudes which, as our proprietary research among parallel online and telephone surveys had taught us, were substantially different in samples drawn from our online database, even after demographic weighting. In these election surveys we used questions which measured alienation, readership, participation and investment, which were selected on a strictly empirical basis from hundreds of questions tested. A comparison of our online and telephone surveys showed that weighting by propensity scores using these questions did the most to reduce biases efficiently.

It should be noted that it would not have been possible to produce accurate predictions for these elections without using *both* the demographic weights and the propensity score measure. However, we have not discovered the holy grail. One reason our predictions for the presidential state-by-state votes were less accurate than our national projections may be that we did not have propensity score targets in the states, because we had done no telephone surveys there. We used the same propensity score assumptions everywhere, based on the nationwide research we had done.

We will continue to test and develop new or better questions to be used in propensity score weighting to reduce the biases in our online samples, not just for political surveys but for marketing, social, and other research.

It would be a mistake to assume that all online pre-election surveys will achieve the same level of accuracy in predicting outcomes as did the Harris Interactive polls in November 2000. There are enormous differences in the ways different organizations are using the internet to conduct research—differences much greater than those in the methods used to conduct telephone surveys.

One key to the development of new survey methods is to run scared and keep on running. It would be a mistake to assume, on the basis of our success in the 2000 elections, that any of the following are true:

• The same weighting variables and weights can be used in other countries with equal success.

• The same weighting variables and weights will work equally well with samples drawn from other databases, or by other means, in the US.

• The same weighting variables and weights will work equally well for surveys on other topics, even when using the same database in the US.

Indeed, it is our strong belief that *all of these assumptions will prove to be wrong.*

If so, the ability to conduct reliable online research will depend on a continuing investment in testing and improving different weighting schemes. Research into improving online survey methodology will need to continue for a long time to come.

Table 1

Accuracy of Online Election Polls in 38 States

	Harris Interactive Forecasts		Spread Between Candidates		Error (on spread)
	Bush %	Gore %	Harris Result	Actual Result	
Alaska	54%	31%	23%	31%	8 points
Arizona	52	40	12	6	6
California	43	49	6	11	5
Colorado	48	43	5	9	4
Connecticut	40	51	11	18	7
Florida	46	49	3	0	3[1]
Georgia	51	43	8	12	4
Hawaii	38	53	15	19	4
Idaho	65	27	38	39	1
Illinois	45	51	6	12	6
Indiana	39	56	17	16	1
Iowa	44	50	6	1	5
Kansas	55	38	17	21	4
Kentucky	54	42	12	16	4
Maine	42	50	8	5	3
Maryland	39	56	17	17	0
Massachusetts	32	58	26	27	1
Michigan	45	50	5	5	0
Minnesota	43	46	3	2	1
Missouri	49	46	3	3	0
Nebraska	57	38	19	29	10
Nevada	48	45	3	4	1
New Hampshire	45	47	2	1	3[1]
New Jersey	40	54	14	16	2
New Mexico	44	49	5	0	5
New York	38	53	15	25	10
North Carolina	51	46	5	13	8
Ohio	49	46	3	4	1
Oklahoma	59	37	22	22	0
Oregon	42	49	7	0	7
Pennsylvania	45	50	5	5	0
Tennessee	49	47	2	4	2
Texas	57	39	18	21	3
Utah	65	25	40	41	1
Virginia	51	45	6	8	2
Washington	47	46	1	5	6[1]
West Virginia	50	46	4	6	2
Wisconsin	47	47	0	0	0
				Average Error	**3.4 points**

[1]Because Harris had the loser ahead, the errors on the candidates are added to get the error on the spread.