Smackdown in Maryland

From the Field

RBS versus RDD

By Richard Morin

Two weeks before Election Day 2002, two rival sampling methodologies squared off in Maryland for a face-to-face showdown designed to determine, once and for all, the best way to sample likely voters in pre-election polls.

In one corner stood the reigning champion, Random Digit Dialing (RDD); in the other, the promising young challenger, Registration-Based Sampling (RBS).

When the final bell sounded on Election Day, neither technique had scored a clean knockout. However, Registration-Based Sampling clearly had the champ on the ropes. The RBS poll was far less expensive and far more efficient and, arguably, at least as accurate as the RDD survey.

But there also were some troubling questions: why were there so few blacks and young people in the Registration-Based Sampling poll? Or, alternatively, why were there so many African Americans and younger people in the Random Digit Dialing sample?

The answers to those questions could decide how the next generation of pre-election polls are done, says Yale University political science professor Donald Green, who with Christopher Mann, a Yale doctoral student and former campaign consultant, conducted the Maryland surveys in collaboration with *The Washington Post*.

Richard Morin is director of polling, The Washington Post.

What follows is a description of a simultaneous test of two ways of capturing samples in election polls—one old and flawed, the other new and untested. While the experiment did not find one method indisputably superior to the other, the test did confirm the potential of Registration-Based Sampling, while also identifying at least one potentially grave failing.

Since the mid-1970s, Random Digit Dialing has been the sampling tool of choice in telephone polling. But as every political pollster knows, the problem with RDD is that all the textbook-pure sampling theory used to create a random sample of telephone numbers is largely abandoned when it comes to identifying "likely voters."

Most pollsters define likely voters based on respondents' answers to as few as one or as many as seven or more questions that attempt to measure the probability that they will vote. The problem is that people say they are registered to vote when they aren't. They swear they're "certain" to vote when they really won't. And they report that they voted in the last election when they really didn't.

In Registration-Based Sampling, the sample is randomly drawn from available lists of registered voters. A pollster automatically knows that everyone in the sample is eligible to vote. As an added bonus, computerized voter rolls often contain prior voting history, so pollsters can produce estimates of the probabilities that someone will vote, based on how often they've voted in the past.

But there are two problems, and they're big ones. Although you have the names, addresses and voting histories of voters, you still must get their home telephone numbers. That critical piece of information usually is not collected when people register to vote. To obtain those numbers, the address on the voter roll file is matched to commercially available files of addresses and their corresponding telephone numbers—a time-consuming and costly task.

And there are real-world examples of bad polls where the final sample—the original RBS sample less those with missing or bad phone numbers—did not mirror the electorate.

The other problem is that in some states with same-day registration, between one in ten and one in five voters registers on Election Day. Consequently, with RBS, *all* of those new voters, those who are registering and voting for the first time, are left out of the sample.

But enough theory. What had been lacking was a real-world test of the two methodologies. And that's exactly what happened in the 2002 election, when the *Washington Post*, CBS News and Quinnipiac University's Polling Institute independently agreed to work with Green.

CBS conducted an RBS and RDD test in South Dakota while Quinnipiac did tests in New York and Pennsylvania. As of this writing, two weeks after the November 5 election, CBS and Quinnipiac had not finished analyzing the data from those surveys.

The *Post* contacted Green to do a paired polls test in Maryland's gubernatorial race featuring Democrat Lieutenant Governor Kathleen Kennedy Townsend and Republican Robert L. Ehrlich, Jr.

The surveys would be conducted simultaneously two weeks before the election, by the same research firm, asking the same horse race and demographic questions. It would be as rigorous and clean a test as we could possibly make it.

"And let the better methodology win," Green said.

The first step was to obtain the names of registered voters in Maryland. Green and Mann purchased 40,000 randomly selected names of Maryland registered voters from Voter Contact Services of Sunnyvale, California, one of a half-dozen commercial firms that sell voter lists to political campaigns and mass marketers.

Two-thirds of these records included a telephone number. Green and Mann randomly selected 10,000 names of voters whose files contained telephone numbers for the sample available to be called. The remaining 30,000 became the control group.

The fact that a third of all voters could not be matched with phone numbers raises the obvious question: were there any differences between the voters with telephone numbers and those without? If there were, that could skew the results. Mann found there were not.

At this point, political consultants may sniff, so what's new about RBS? Campaign pollsters have been buying voter lists for years and using them for their own surveys, canvassing and phone bank operations.

That's true—up to a point, according to Green. In his conversations with political pollsters, he found that they consistently missed a golden opportunity to improve the accuracy of their polls. For example, these pollsters would randomly select to be interviewed only those individuals who had voted in the last election or the last few elections. That means individuals who hadn't done so had no chance of being interviewed. And that's a problem, since these "sometimes" voters often think and vote differently than those who faithfully go to the polls every time.

Green's innovation was to group voters based on their past voting history, then sample from each of these strata in proportion to their probable share of the electorate. That produced a sample representative of all voters who were expected to cast ballots, including those who voted in every election, those who sometimes voted, and even those who rarely went to the polls.

Here's how it worked in Maryland. Weeks before the election, Mann and Green analyzed the Maryland voter rolls to determine who had cast ballots, and how often, in elections back to 1994. Then they grouped voters with similar voting histories into groups, or strata.

They found that 57% of all registered voters in 1998 had voted in two previous statewide elections. About 10% had voted in the previous presidential but not the previous midterm election. Six percent had voted in the previous midterm but not the previous presidential election. Three percent had voted in the midterm but had not voted in the previous two statewide elections. And about 26% were newly registered.

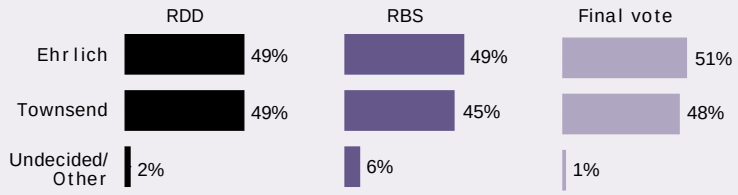
Then they selected voters from each stratum so that each group was correctly represented in the target sample, meaning that 57% of the final sample was composed of voters who had cast ballots in the past two elections, 10% had voted in the previous midterm but not the presidential, and so forth.

The questionnaires were alike in key ways. The first seven questions, which included the likelihood-to-vote question and the candidate-preference questions, were identical. So were the demographic questions at the end of the survey. Interviewing began on October 20 for both surveys. It ended four days later, on schedule. The calls were made by TNS Intersearch of Horsham, Pennsylvania, the firm that conducts election surveys for the *Post*.

The advantages of RBS over RDD were instantly apparent. RBS was spectacularly more

Figure 1
Close Calls

Question:
As you know, the candidates in November's election for governor include Kathleen Kennedy Townsend, the Democrat, and Bob Ehrlich, the Republican. Suppose the election were held today—for whom would you vote?



Note: Asked of likely Maryland voters.
Source: Surveys by *The Washington Post*, October 20-24, 2002.

efficient. "We had to make four [RDD] calls for every one RBS to get a complete," Mann said. That's understandable, since RDD samples typically contain many non-working telephone numbers.

The cooperation rate also was significantly higher for the Registration-Based Sampling survey. One common way to express the cooperation rate is to figure the percentage of completed interviews from households where someone answered the telephone. Using this standard, the cooperation rate for the RBS sample was 48.2% while that for the RDD sample was 34.8%.

The RBS sample also yielded a higher percentage of those who said they were absolutely certain to vote, one commonly used definition of a likely voter. Nearly nine in ten—86%—of those in the list sample declared that they were absolutely certain to vote, compared to 70% of those in the RDD sample.

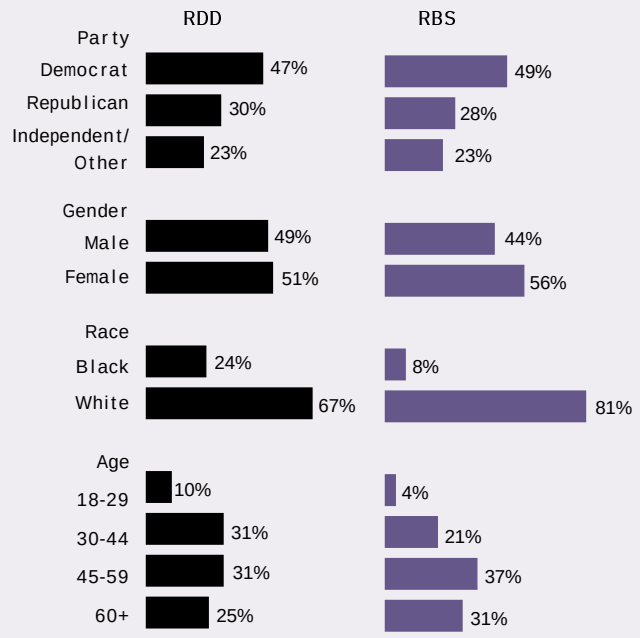
So, it seems, the big winner on election night in Maryland was Registration-Based Sampling. On issues of particular interest to survey practitioners—cost, response rate and efficiency—RBS clearly was superior.

But not so fast. On the question of particular interest to *Post* readers—who's ahead in the Maryland governor's race?—the surveys produced slightly different estimates, and for a very perturbing reason.

To determine the pool of likely voters in the RDD sample, the *Post* created a scale based on respondents' answers to questions measuring their stated likelihood to vote, their past voting history, interest in the gubernatorial election and whether the respondent knew the location of his or her polling place. Those that scored high on this scale became part of the sample of likely voters.

A total of 725 respondents qualified as likely voters, or about 38% of all adults interviewed in the *Post* poll, which corresponded to the estimated turnout.

Figure 2
Selected Sample Characteristics



Note: Results based on registered voters in each sample who said they were "absolutely certain to vote."
Source: Surveys by *The Washington Post*, October 20-24, 2002.

Among likely voters, the two candidates were deadlocked in the Random Digit Dial sample: each received 49% of the vote, with the remainder undecided (see Figure 1). The Registration-Based Sampling survey of 709 registered voters put Ehrlich ahead of Townsend by 49% to 45%.

The difference in the two results was modest and not statistically significant.

What was of greater concern was the reason for the difference: a huge disparity in the proportion of blacks in the two samples. [The RDD sample also contained far more younger people; 41% were between 18 to 44 years old, compared to 25% of all RBS respondents (see Figure 2).]

African Americans comprised 24% of all likely voters in the RDD sample but only 8% in the RBS sample. That's a problem because nine out of ten blacks supported Townsend, so getting the black proportion wrong by even just a few points would mean getting her share of the vote wrong by a similar number of points. Past exit polls only compounded the dilemma. Four years ago, the electorate was 21% black. But in 1994, blacks comprised 12% of all residents who cast ballots.

The scarcity of black voters in the RBS sample clearly troubled Green and Mann. Among their worst fears: perhaps because blacks (and also young people) move more often than whites and older individuals, their telephone numbers are less likely to show up in databases of residential telephone numbers. Thus, the voter files containing phone numbers that come from Voter Contact Services or other vendors inherently might produce samples that are "too white" and "too old."

Green decided to make a change. He had initially planned not to weight the RBS results. In theory, the whole point of Registration-Based Sampling is to do away with the need for such interventions.

But in the end, Green decided to weight

Continued on page 41

Continued from page 9

the RBS results to the average of both surveys, or 16%. In the adjusted sample, Townsend received 45% of the likely vote and Ehrlich got 44%. To produce its "best estimate" of the horse race, the *Post* averaged the results of the two polls, which gave each candidate 47% of the vote.

On November 5, nine days after the *Post* poll was published, Ehrlich beat Townsend 51% to 48%, or about the same margin as the RBS estimate. So the first, unweighted result of the Registration-Based Sampling poll showing Ehrlich up by four points was the more accurate estimate, after all.

Who knows? Perhaps the race changed in the final ten days; races often do, at least a little, but sometimes by a lot. "You're not sure if you've missed because of sampling error or because of some other source of bias, or because the world has changed," Green said.

What was learned from the 2002 Sampling Smackdown in Maryland? For one thing, it was found that Registration-Based Sampling is far less expensive than Random Digit Dialing because it is far more efficient. David Lambert, senior vice president at TNS Intersearch, estimated that it cost $44.91 for each completed RDD interview versus $22.19 for an RBS complete, holding other factors constant.

But more work needs to be done to improve turnout estimates. Mann and Green suspect that both surveys probably misrepresented the black turnout, at least a little.

"We need to examine turnout from African American precincts to see how much of the electorate was black and then determine whether RDD overestimated the proportion, RBS underestimated, or some combination," Mann said.

On most questions, the results obtained by the two polls were roughly similar—at least close enough for newspaper work. "It seems, on their face, the two polls don't really differ," Green said. "The firm conclusion you can draw from this test is that, even if there were no increase in quality, there is such an immense cost savings that this type of research is going to grow."

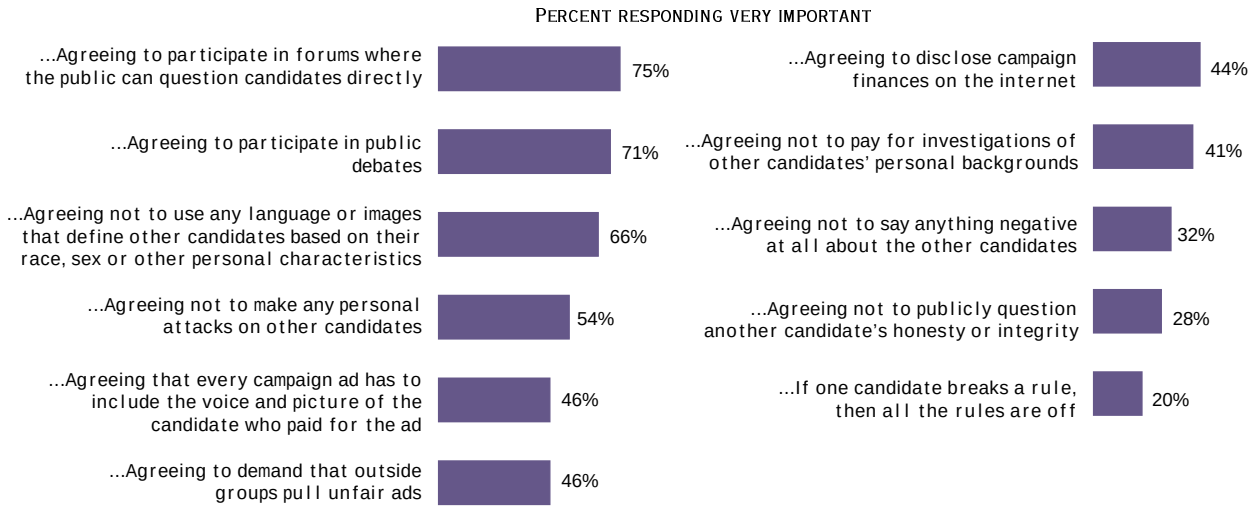
And remember, Green said, "this is the first round. The accuracy will improve."

Correction

On page 21 of the November/December 2002 issue of Public Perspective, labels were missing from some of the response categories for a question battery on election campaign conduct. The question battery is reprinted in its entirety below.

Questions:

I'm going to read a list of different rules that might be included in a code of campaign conduct. For each one, please tell me how important you think it would be to include in a code of conduct for election campaigns—very important, somewhat important, not very important, or not at all important...



Note: Asked of likely voters.
Source: Survey by Lake Snell Perry & Associates and Deardourff/The Media Company, for the Institute for Global Ethics, June 6-11, 2002.