

**The Big East Two:
A Comparison of RBS and RDD polls in the 2002 elections
in New York and Pennsylvania**

By Doug Schwartz and Clay F. Richards
The Quinnipiac University Polling Institute

Abstract: In its most extensive use to date, RBS did an excellent job of predicting the outcome of the elections for governor in New York and Pennsylvania last year. The Quinnipiac University Polling Institute study interviewed more than 3,300 likely voters in side-by-side RBS/RDD polls in the two states. The final RBS polls in both states mirrored the election outcome. A closer look at the demographics showed similarities in some areas and differences in others. The greatest concern uncovered was an under representation in RBS of the urban areas of both states. But the dramatically greater efficiency of RBS sampling with its resulting savings in polling costs, when combined with the extremely promising accuracy found in the study, bodes strongly for extensive testing of RBS in future election polling.

Paper prepared for presentation at:

APSA 2003

Philadelphia, PA

The Scene:

At first glance, New York and Pennsylvania appear to be two very similar neighboring northeastern states. Each is dominated by a major metropolitan city in the East that dates from the earliest days of America, and each has in the west a second, smaller but sizeable city that once was a major producer of steel, but now is part of the American rustbelt. The urban population of both states is ethnically mixed, with strong representations of the groups that produced the American melting pot, the rural areas are largely white and conservative and both have large suburban populations that have come to represent upper middle class America and the key swing vote in elections.

In 2002 though New York and Pennsylvania, so much alike demographically and geographically, had two very different gubernatorial elections.

In New York, a relatively bland two term Republican Governor, George Pataki, who had managed to oust the increasingly controversial Democratic Governor Mario Cuomo, was seeking a third term. One of his would-be Democratic rivals, Andrew Cuomo, son of Mario Cuomo, and former U.S. Housing Secretary, sealed his defeat in the Democratic primary by calling Pataki merely a “coat carrier” for Mayor Rudy Guiliani after the terrorists attacks of 9/11. The comment, while caustic and seemingly unpatriotic, was in fact an accurate description of the politician who always seemed to be in the background whenever Guiliani held a news conference. The Democratic candidate turned out to be popular but also uncharismatic State Comptroller Carl McCall, the first black to run for governor of New York as a major party candidate. Even though Cuomo withdrew before the primary vote, McCall's candidacy never caught on and Pataki easily won a third term. The Republican incumbent easily won in every poll match up prior to the election.

Thomas Golisano, the independent third party candidate for governor, was thought to be a somewhat complicated factor in polling this election. Golisano, a businessman, spent millions of his own fortune to get about 14 percent of the vote. However, Golisano's presence as a third party candidate made no significant difference in this study, as detailed below.

In Pennsylvania, a state with a history of electing bland Republican governors, the political scene was much different. After 9/11, popular Republican Governor Tom Ridge went to Washington to become Director of Homeland Security, leaving Lieutenant Governor Mark Schweiker, a likeable unknown who had never achieved enough of a name to consider running on his own, in charge of the state. (Ironically, late in his term when a group of Pennsylvania coal miners were rescued after being trapped underground for several days, Schweiker became an instant hero for his leadership role, but it was too late for him to get in the race). Avoiding a divisive primary, Republicans put up Attorney General Mike Fisher, the kind of competent, but unexciting, long serving public official that Pennsylvanians usually elect governor. Meanwhile, the Democrats seemed headed for a bruising primary, pitting state auditor Bob Casey, Jr., the son of a popular former governor, against former Philadelphia Mayor Ed Rendell, who had served as Bill Clinton's Democratic national chairman. Rendell won the primary fairly easily, gaining

strong statewide name recognition, while Fisher, with no primary, remained somewhat unknown despite having served as attorney general.

There was never any real contest between Rendell and Fisher and Pennsylvanians discarded their political traditions and elected the Jewish, former mayor of Philadelphia as their governor.

In New York, where charismatic Democrats usually win, a bland Republican was re-elected governor, while in Pennsylvania, where workhorse Republicans are the norm, a charismatic Democrat won.

The two big-state governors' contests, so much alike and so different, provided the Quinnipiac University Polling Institute with the opportunity to conduct the largest experiment in simultaneous use of Random Digit Dialing (RDD) and Registration-Based Sampling (RBS) conducted anywhere in the country. The two states combined produced 3,320 likely voters who were interviewed for this study.

The Polling Background:

Quinnipiac University in Hamden, Conn., conducts about 40 polls a year in Connecticut, New York, New Jersey, Pennsylvania and nationally. The polls are conducted by student and non-student interviewers in the 39-station polling center located across from the main campus. Standard RDD polling methods are used and the results are released to the news media, usually within 36 hours of completion of the poll. In the 2002 election cycle, Quinnipiac conducted 10 polls in the New York governor's race and 5 in the Pennsylvania governor's race.

In the spring of 2002, Quinnipiac Polling Director Douglas Schwartz attended a conference in Washington, D.C. on pre-election polling sponsored by The Gallup Organization. At the conference, Yale University political scientist Don Green challenged pollsters to try a new way of predicting election outcomes using voter registration lists and using the voter history those lists contained to create a better sample of likely voters than is provided by traditional RDD samples. Schwartz proposed that Quinnipiac conduct parallel surveys during the 2002 elections. Green's graduate assistant, Chris Mann, coordinated the project with Schwartz and directed the RBS study. Yale provided the RBS samples and Quinnipiac underwrote the polling costs.

When Quinnipiac conducted its final RDD poll of 962 Pennsylvania likely voters (registered voters who say they will definitely vote) Oct. 21-27, it simultaneously surveyed 754 Pennsylvania likely voters using an RBS sample. When the final RDD poll of 851 likely voters in New York Oct. 28-Nov 3., 739 New York likely voters were simultaneously interviewed using an RBS sample. The polls were conducted over the same time span by interviewers drawn from the same pool of regular Quinnipiac interviewers. The RBS interviews were conducted in the Law School legal clinic on the main campus, a few hundred yards from the polling center. The results are being reported here for the first time.

Methodology of the RBS Sample:

Readily available lists of registered voters are used to create an RBS sample. This creates the first advantage over RD, because every individual in the sample is eligible to vote. Computerized voter rolls also contain the individual's history of voting, allowing for the creation of a formula to determine estimates of likely voters based on an individual's past history. The pollster does not have to worry about whether the respondent is truthful about being registered, or what the individual's voting behavior was in past elections.

Voter Contact Services (VCS) maintains a list of all registered voters in New York and Pennsylvania and provides home telephone numbers using commercially available lists of addresses and matching phone numbers. VCS updated those lists between May and August of 2002 in each county.

This test was different from the previous use of voters rolls for polling because each registered voter was put into one of five strata based on their past voter history and a sample was drawn with the proportion of voters in each stratum reflecting the likely proportion of actual voters who would fall into that stratum. In New York, for example, each voter was placed in each of the following strata: 1) voted in both the 2000 and 1998 elections; 2) voted in the 1998 general election, but not the 2000 election; 3) voted in the 2000 general election, but not in the 1998 election; 4) registered to vote, but had not voted in either the 2000 or 1998 elections; and 5) newly registered voters since the 2000 election.

A complete explanation of RBS methodology and how the experiment was constructed can be found in other papers by Mann and Green. In effect, Quinnipiac did the fieldwork for the RBS proposal designed by Green, while Green and Mann discussed the system they developed. In reporting Quinnipiac's part of this experiment, for the sake of simplicity, the RDD likely voters were defined as those respondents who said they were registered to vote and would definitely vote in the election for governor. Quinnipiac's numbers vary slightly from those used by Green. It is the purpose of this paper to first, discuss the results in terms of predicting the outcome of the elections in New York and Pennsylvania, and second, to examine the demographics in both surveys to see how those results compare. In such an examination, lie the clues to the usefulness of using RBS in future election polling.

Now on to the results:

New York State (see attached Table 1-1)

RDD vs. Actual Turnout

Quinnipiac's final RDD poll in the New York State gubernatorial election, released the day before the election, predicted George Pataki would be easily re-elected and came very close to the actual margin of victory. Pataki defeated McCall, the Democratic state comptroller, by 15 percentage points, 49-34 percent, with Independence Party candidate Thomas Golisano winning 14 percent. The RDD sample showed Pataki winning by 14 points, 43-29, with Golisano at 16 percent.

Comparing RDD to the actual electorate, RDD slightly over represented New York City, but at the same time underrepresented registered Democrats. RDD mirrored the actual electorate in terms of gender and was younger than the actual electorate.

-- While 33 percent of the RDD sample came from New York City, 30 percent of the actual electorate was from New York City

-- Whereas 18 percent of the RDD sample was under 35, only 11 percent of the actual electorate was under 35.

-- In the RDD sample, registered Democrats outnumbered registered Republicans 42-34 percent, a +8 Democratic advantage. The actual electorate was more Democratic, 47 percent Democrat and 34 percent Republican, a +13 Democratic voter registration advantage.

RBS vs. Actual electorate

Like the RDD poll, RBS accurately forecast the New York election. Indeed RBS and RDD both had the same 14 percentage point margin of victory for Pataki, though RDD came closer to Golisano's 14 percent of the vote. In the RDD sample, Golisano received 16 percent of the vote, while RBS had Golisano at 19 percent. The fact that both polls exceeded the actual Golisano vote is probably proof that some voters tell pollsters they will vote for a third party candidate, but when they get into the booth decide not to "throw away" their vote.

RBS basically mirrored the actual electorate in terms of gender and age. However, like RDD, RBS underestimated the number of registered Democrats. In the RBS sample, Democrats outnumbered Republicans 41 – 35 percent, a +6 Democratic advantage. This compares to a +13 Democratic advantage in the actual electorate, 47-34 percent Democrat.

The biggest difference between RBS and the actual electorate occurred on the regional distribution. RBS significantly underestimated the share of the vote from New York City. RBS had 20 percent of its voters from New York City, yet New York City represented 30 percent of the vote on Election Day.

The under representation of New York City probably does not indicate a major flaw in using RBS, but the presence of a problem historic to polling in New York State, whatever method is used. Traditionally in New York polling there is a big drop off in the New York City numbers when you go from the sample of telephone numbers to the actual completed surveys. In RBS, the drop off in New York City was 9 percentage points. Twenty-nine percent of the phone numbers came from New York City, but only 20 percent of the completed surveys came from New York City. Similarly, there was a 9-percentage point drop in the New York City numbers in the RDD sample. New York City was 42 percent of the phone numbers, but only 33 percent of the total completes.

RBS initially starts out with a lower percentage of phone numbers from New York City than RDD (29 percent compared to 42 percent) because New York City has a higher percentage of unlisted phone numbers, and a lower voter registration rate than the rest of the state.

The RBS sample of phone numbers, with 29 percent coming from New York City, is actually more accurate in terms of representing the voting population than the RDD sample of phone numbers in which 42 percent come from New York City. The RDD sample ends up so closely mirroring the geographic distribution of the vote is because it is simply harder to get completes in New York City. Among other things, there are more language problems (Spanish respondents were interviewed in both samples) and disconnected numbers in New York City.

RDD vs. RBS

The RDD and RBS polls horseshoe measurements were both extremely accurate in New York. They both came within a percentage point of the actual 15-percentage point Republican victory. Consequently, the average candidate error for both polls was only .5 (half a percentage point)

In summary:

--The RBS sample was older, more female, and had fewer blacks than the RDD sample.

--While 18 percent of the voters in the RDD sample were under 35, only 9 percent of the voters in the RBS sample were under 35.

--In the RDD sample 35 percent had college degrees, compared to 45 percent in the RBS sample.

--In the RDD sample, 9 percent were black, compared to 6 percent in the RBS sample.

--The biggest difference between the RDD and RBS numbers in the New York study was on region. There was a 13-percentage point gap on the percentage of voters from New York City. RDD had New York City at 33 percent, while RBS had it at 20 percent. Despite such a big difference on region, the RBS and RDD numbers were surprisingly similar on party identification and political philosophy.

Pennsylvania (See attached Table 1-2)

Analyzing the Pennsylvania RBS and RDD studies is more difficult because the polls were taken about two weeks before Election Day and released a week before the election, therefore making it more difficult to compare the poll results with the actual vote. Keating Holland of CNN is fond of saying that trying to predict the outcome of the election with a poll a week before Election Day is like trying to predict the outcome of a baseball game in the 7th inning.

Quinnipiac's final RDD poll correctly predicted that Ed Rendell would easily defeat Mike Fisher for governor, but overestimated the margin of Rendell's victory. The actual margin of victory was 9-percentage points, 53-44 percent, while the RDD sample had the margin at 22 points, 55-33 percent.

Thus the RDD sample overestimated Rendell's margin of victory by 13-percentage points. The average candidate error was 6.5 percentage points.

It is worth noting that the RDD poll had 9 percent undecided. The rule of thumb is that the vast majority of undecided voters break for the challenger. While Rendell was not an incumbent, he ran like he already had the job (in the eyes of the Philadelphia area voters), and he was much better known than Fisher. An allocation of undecided voters would have placed most of their votes in the Fisher column, making the margin narrower.

It is also worth noting that much happened in the final week of the election. Nationwide, there is much evidence of a swing toward Republican candidates in the closing days, and President Bush's last minute campaigning for Fisher in Pennsylvania probably gave the Republican a bump. There were also stories after the election that Rendell, confident of election, backed off his "get out the vote" effort at the end so that Republicans coming out to vote for the Democratic gubernatorial candidate would not cross back over the vote for Republican House candidates in at least two very close races.

Finally the RDD sample was more Democratic in terms of party registration than the electorate, showing a +9 Democratic edge over Republicans (48 percent Democrat, 39

percent Republican), while the actual party registration figures showed only a +2 Democratic edge (48 percent Democrat, 46 percent Republican).

Interestingly, The Keystone Poll, the oldest and most respected of Pennsylvania polls run by the State University at Millersville, was taken at the same time as the Quinnipiac poll and showed almost the same margin for Rendell.

The RBS sample taken a week before the election was right on target in predicting the Pennsylvania race, giving Rendell the same nine point lead that turned out to be his margin of victory. However, a closer look reveals some internal differences.

-- The RBS sample was more Republicans in terms of party registration than the electorate. The RBS sample was off on party registration more than the RDD sample. The RBS sample composition was 50 percent Republican and 41 percent Democrat, a +9 Republican advantage. Democrats had a +2 advantage in the actual electorate, which meant that RBS was off by 11 points on party registration, while RDD was off by only seven points.

-- RBS was older than the actual electorate and slightly more female. It had slightly fewer males.

-- There were fewer Philadelphians in the RBS sample, though the gap was nowhere near as wide as that found in New York. Philadelphia was accurately represented in the original RBS sample of phone numbers, which had it at 10 percent. The actual vote was 11 percent. The drop-off was a result there was a lower contact rate in Philadelphia (more disconnects for one). By comparison, 12 percent of the original RDD sample of phone numbers was from Philadelphia, and 13 percent of the likely voters in the completed RDD poll were from Philadelphia.

RBS vs RDD

The biggest demographic differences between RBS and RDD were on age and party identification. The RBS sample was older and more Republican than the RBS sample.

-- Whereas 44 percent of the voters from RBS were over 60, only 30 percent from RDD were that age.

-- While 44 percent of voters from RBS identified themselves as Republicans, only 33 percent of voters in the RDD identified themselves as Republicans.

There were other difference between RDD and RBS. The RBS sample was more educated, had fewer Philadelphians, fewer African-Americans, and was more conservative.

-- While 38 percent of the RBS sample had a college degree, only 29 percent of RDD had a college degree. (This does not help account for the Rendell overestimate because higher educated folks were more likely to vote for Rendell).

-- Whereas, 13 percent of the RDD sample came from Philadelphia, only 7 percent of the RBS sample came from Philadelphia.

-- While 10 percent of the RDD sample was black, only 5 percent of the RBS sample was black.

-- While 30 percent of the RDD sample was conservative, 36 percent of the RBS sample was conservative

Concluding Thoughts:

As others have already reported, using an RBS sample is much more efficient than using an RDD sample. There are far more "good numbers" in an RBS sample. In the Pennsylvania and New York tests, the interviewers were twice as likely to find a likely voter with RBS than we were with RDD. The dollar savings cannot be calculated because the interviewers used CATI in the RDD poll and paper with the RBS sample, so the time factor was not the same. Suffice it to say if it could be proved to be as accurate or better than RDD sampling, all election polling would convert to RBS for the cost/time savings alone.

From this study, these conclusions can be drawn:

1. RBS did an excellent job of accurately predicting the outcome of these two elections. As a result, Quinnipiac will continue testing RBS samples in upcoming elections including the presidential election.
2. The greater efficiency of RBS sampling in obtaining likely voters is especially important in primary voting. Only a very small number of voters cast ballots in a primary, making RDD primary polling historically difficult and less accurate than general election polling.
3. The greatest concern about RBS uncovered in this study is that it underrepresented major urban areas, New York City, and to a lesser extent, Philadelphia, but this does not appear to be a major flaw in RBS sampling. This factor can be fixed by starting out with more telephone numbers from the major urban areas to compensate for the lower contact rate.
4. The under representation of voters from major urban areas may explain at least in part the under representation of Democrats and blacks.

There is no compelling finding in this study that would lead to the rejection of RBS sampling as a better way to conduct election polls. To the contrary, the system offers much promise of not only more efficient and less costly polling, but potentially a more accurate system as well. While Quinnipiac has made no decision how to employ RBS in the future, some method of using it for at least part of polling is likely.

Table 1-1 New York State 2002 Election for Governor, Selected Sample Characteristics

	RDD	RBS	Actual Vote
McCall	29%	28%	34%
Pataki	43	42	49
Golisano	16	19	14
Undecided/Other	13	11	3
Upstate	43	58	47
New York City	33	20	30
Suburbs	24	22	23
Male	47	45	46
Female	53	55	54
18-24	6	2	3
25-34	12	7	8
35-44	17	17	17
45-59	31	34	34
60+	30	37	38
Reg Dem	42	41	47
Reg Rep	34	35	34
Reg Ind/Other	24	25	19

Table 1-1 New York State 2002 Election for Governor, Selected Sample Characteristics
(Continued)

	RDD	RBS
Party Identification		
Democrat	37	35
Republican	31	30
Independent/Other	31	35
Liberal	26	23
Moderate	42	44
Conservative	27	27
White	75	85
Black	9	6
Hispanic	8	4
Less than HS	9	4
HS Grad	20	20
Some Coll	36	30
Coll Grad	13	16
Post Grad	22	29

Table 1-2 Pennsylvania 2002 Election for Governor, Selected Sample Characteristics

	RDD	RBS	Actual Vote
Rendell	55%	49%	53%
Fisher	33	40	44
Undecided/Other	9	11	2
Male	48	45	47
Female	52	55	52
18-24	5	1	3
25-34	12	6	8
35-44	18	16	17
45-59	32	30	34
60+	30	44	38
Reg Dem	48	41	48
Reg Rep	39	50	46
Reg Ind/Other	10	10	6
Allegheny	11	11	11
Philadelphia	13	7	11
Northeast	12	10	11
Southeast	22	21	22
Northwest	8	10	10
Southwest	11	11	10
Central	23	30	25

Table 1-2 Pennsylvania 2002 Election for Governor, Selected Sample Characteristics

	RDD	RBS
Party Identification		
Democrat	41	34
Republican	33	44
Independent/Other	27	22
Liberal	20	18
Moderate	44	42
Conservative	30	36
White	87	88
Black	10	5
Hispanic	1	4
Less than HS	7	5
HS Grad	28	29
Some Coll	36	27
Coll Grad	13	13
Post Grad	16	25