Methodology

New York State in Turmoil

Prepared by Princeton Survey Research Associates International
for Hofstra/Newsday/Cablevision

May 2010

SUMMARY

The New York State in Turmoil Survey, sponsored by the Hofstra, Newsday and Cablevision obtained telephone interviews with a representative sample of 1,500 registered voters in New York state. The survey was conducted by Princeton Survey Research International (PSRAI). Interviews were done in English by Princeton Data Source from April 19 to May 9, 2010. Statistical results are weighted to correct known demographic discrepancies. The margin of sampling error for the complete set of weighted data is ±3.3 percentage points.

Details on the design, execution and analysis of the survey are discussed below.

DESIGN AND DATA COLLECTION PROCEDURES

Sample Design

A combination of landline and cellular random digit dial (RDD) samples was used to represent all adults in New York who have access to either a landline or cellular telephone. Both samples were provided by Survey Sampling International, LLC (SSI) according to PSRAI specifications.

Numbers for the landline sample were drawn with equal probabilities from active blocks (area code + exchange + two-digit block number) that contained one or more residential directory listings. The cellular sample was not list-assisted, but was drawn through a systematic sampling from dedicated wireless 100-blocks and shared service 100-blocks with no directory-listed landline numbers.

Additional sample was released in Nassau and Suffolk counties for an oversample of 300 registered voters on Long Island.

Contact Procedures

Interviews were conducted from April 19 to May 9, 2010. As many as seven attempts were made to contact every sampled telephone number. Sample was released for interviewing in replicates, which are representative subsamples of the larger sample. Using replicates to control the release of sample ensures that complete call procedures are followed for the entire sample. Calls were staggered over times of day and days of the week to maximize the chance of making contact with potential respondents. Each phone number received at least one daytime call when necessary.

For the landline sample, interviewers asked to speak with the youngest adult male or female currently at home based on a random rotation. If no male/female was available, interviewers asked to speak with the youngest adult of the other gender. For the cellular sample, interviews were conducted with the person who answered the phone. Interviewers verified that the person was an adult and in a safe place before administering the survey.

Full interviews were conducted with respondents who were registered to vote in New York or who said that they were likely to register in New York State this year. Non-registered voters who said they are not likely to register this year were deemed ineligible and were only asked basic demographic questions for weighting.

WEIGHTING AND ANALYSIS

Weighting is generally used in survey analysis to compensate for sample designs and patterns of non-response that might bias results. The full sample – completes plus those not registered to vote – was weighted to match New York State general population parameters. A three-stage weighting procedure was used to weight this dual-frame sample.

The first stage of weighting corrected for the oversample of Long Island respondents. The second stage weight accounted for different probabilities of selection associated with the number of adults in each household and each respondent's telephone usage patterns.[1] This adjustment also adjusts for the overlapping landline and cell sample frames and the relative sizes of each frame and each sample.

[1] i.e., whether respondents have only a landline telephone(s), only a cell phone, or both kinds of telephone.

This second-stage weight for the i^{th} case can be expressed as:

$$WT_i = \frac{1}{\left(\frac{S_{LL}}{S_{CP}} \times \frac{1}{AD_i}\right)} \; if\ respondent\ has\ no\ cell\ phone$$

$$WT_i = \frac{1}{\left(\frac{S_{LL}}{S_{CP}} \times \frac{1}{AD_i}\right) + \frac{U_{LL}}{U_{CP}}} \; if\ respondent\ has\ both\ kinds\ of\ phones$$

$$WT_i = \frac{1}{U_{LL}/U_{CP}} \; if\ respondent\ has\ no\ land\ line\ phone$$

Where S_{LL} = size of the landline sample

S_{CP} = size of the cell phone sample

U_{LL} = size of the landline sample frame

U_{CP} = size of the cell phone sample frame

LL_i = Number of landline telephones in household

AD_i = Number of adults in the household

CP_i = Number of cell phones that respondent uses

The equations can be simplified by plugging in the values for S_{LL} = 1,109 and S_{CP} = 626. Additionally, we will estimate of the ratio of the size of landline sample frame to the cell phone sample frame U_{LL}/U_{CP} = 0.60.

$$WT_i = \frac{1}{\left(1.77 \times \frac{1}{AD_i}\right)} \; if\ respondent\ has\ no\ cell\ phone$$

$$WT_i = \frac{1}{\left(1.77 \times \frac{1}{AD_i}\right) + 0.60} \; if\ respondent\ has\ both\ kinds\ of\ phones$$

$$WT_i = \frac{1}{0.60} \; if\ respondent\ has\ no\ land\ line\ phone$$

The third and final stage of weighting balanced total sample demographics (completes plus screen-outs) to population parameters. The sample was balanced - by form - to match New York state population parameters for sex, age, education, race, Hispanic origin, region and population density. The parameters were derived from Census 2000 data[2].

[2] Census 2000 Summary File 4 (SF 4).

Weighting was accomplished using Sample Balancing, a special iterative sample weighting program that simultaneously balances the distributions of all variables using a statistical technique called the *Deming Algorithm*. Weights were trimmed to prevent individual interviews from having too much influence on the final results. The use of these weights in statistical analysis ensures that the demographic characteristics of the sample closely approximate the demographic characteristics of the national population. Table 1 compares weighted and unweighted sample distributions to population parameters.

Table 1: Full Sample Demographics – Completes plus Screenouts

	Parameter	Unweighted	Weighted
Gender			
Male	47.1%	47.3%	45.8%
Female	52.9%	52.7%	54.2%
Age			
18-24	12.3%	11.2%	12.1%
25-34	19.1%	11.9%	17.4%
35-44	21.9%	13.1%	20.0%
45-64	29.6%	39.3%	29.6%
65+	17.1%	21.6%	17.7%
Education			
Less than HS Graduate	21.3%	5.6%	16.2%
HS Graduate	27.4%	25.8%	28.6%
Some College	25.8%	25.9%	26.4%
College Graduate	25.5%	41.4%	27.4%
Race/Ethnicity			
White/not Hispanic	64.3%	70.3%	64.2%
Black/not Hispanic	13.7%	11.1%	12.8%
Hispanic	13.8%	9.5%	12.8%
Other/not Hispanic	8.1%	6.5%	7.3%
Region			
New York City	42.5%	30.1%	41.0%
Long Island	14.4%	31.9%	17.2%
Hudson Valley	14.1%	12.4%	13.9%
Urban Upstate	15.1%	13.4%	14.3%
Rural Upstate	14.0%	12.2%	13.6%
County Pop. Density			
1 - Lowest	6.6%	5.9%	6.7%
2	11.8%	12.7%	11.7%
3	9.7%	8.4%	9.8%
4	17.9%	25.0%	18.5%
5 - Highest	54.1%	48.1%	53.4%

Effects of Sample Design on Statistical Inference

Post-data collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. PSRAI calculates the effects of these design features so that an appropriate adjustment can be incorporated into tests of statistical significance when using these data. The so-called "design effect" or *deff* represents the loss in statistical efficiency that results from a disproportionate sample design and systematic non-response. The total sample design effect for this survey is 1.70.

PSRAI calculates the composite design effect for a sample of size n, with each case having a weight, w_i as:

$$deff = \frac{n\sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2} \qquad \textit{formula 1}$$

In a wide range of situations, the adjusted *standard error* of a statistic should be calculated by multiplying the usual formula by the square root of the design effect ($\sqrt{deff}$). Thus, the formula for computing the 95% confidence interval around a percentage is:

$$\hat{p} \pm \left(\sqrt{deff} \times 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) \qquad \textit{formula 2}$$

where $\hat{p}$ is the sample estimate and n is the unweighted number of sample cases in the group being considered.

The survey's *margin of error* is the largest 95% confidence interval for any estimated proportion based on the total sample— the one around 50%. For example, the margin of error for the entire sample is ±3.3 percentage points. This means that in 95 out every 100 samples drawn using the same methodology, estimated proportions based on the entire sample will be no more than 3.3 percentage points away from their true values in the population. Table 2 reports design effect and margins of sampling error for key subgroups. It is important to remember that sampling fluctuations are only one possible source of error in a survey estimate. Other sources, such as respondent selection bias, questionnaire wording and reporting inaccuracy, may contribute additional error of greater or lesser magnitude.

Table 2: Design effects and Margins of Sampling Error

	n	Design Effect	Margin of Error
Total	1500	1.70	3.3%
New York City	435	1.46	5.7%
Long Island	492	1.69	5.7%
Hudson Valley	189	1.52	8.8%
Urban Upstate	203	1.60	8.7%
Rural Upstate	181	1.60	9.2%

RESPONSE RATE

Table 3 report the disposition of all sampled telephone numbers ever dialed from the original telephone number samples. The response rate estimates the fraction of all eligible sample that was ultimately interviewed. At PSRAI it is calculated by taking the product of three component rates:[3]

- Contact rate – the proportion of working numbers where a request for interview was made[4]
- Cooperation rate – the proportion of contacted numbers where a consent for interview was at least initially obtained, versus those refused
- Completion rate – the proportion of initially cooperating and eligible interviews that were completed

Thus the response rate for the land line samples was 16 percent. The response rate for the cellular samples was 19 percent.

[3] PSRAI's disposition codes and reporting are consistent with the American Association for Public Opinion Research standards.

[4] PSRAI assumes that 75 percent of cases that result in a constant disposition of "No answer" or "Busy" are actually not working numbers.

Table 3:Sample Disposition

Landline	Cell	
25321	12498	**T** Total Numbers Dialed
1268	174	**OF** Non-residential
1147	3	**OF** Computer/Fax
4		**OF** Cell phone
10723	3666	**OF** Other not working
1598	309	**UH** Additional projected not working
10581	8346	Working numbers
41.8%	66.8%	Working Rate
533	103	**UH** No Answer / Busy
3250	2775	**UO_{NC}** Voice Mail
23	4	**UO_{NC}** Other Non-Contact
6775	5464	Contacted numbers
64.0%	65.5%	Contact Rate
933	861	**UO_{R}** Callback
4130	2955	**UO_{R}** Refusal
1712	1648	Cooperating numbers
25.3%	30.2%	Cooperation Rate
418	343	**IN1** Language Barrier
255	791	**IN2** Not NY registered voter/Child's cell phone
1039	514	Eligible numbers
60.7%	31.2%	Eligibility Rate
39	14	**R** Break-off
1000	500	**I** Completes
96.2%	97.3%	Completion Rate
15.6%	19.2%	Response Rate