Methodology

Higher Education/Housing Survey

Prepared by Princeton Survey Research Associates International  
for the Pew Social Trends & Demographics Project

March 2011

**SUMMARY**

The Higher Education/Housing Survey, sponsored by the Pew Social & Demographic Trends Project, obtained telephone interviews with a nationally representative sample of 2,142 adults living in the continental United States. The survey was conducted by Princeton Survey Research Associates International. Interviews were done in English and Spanish by Princeton Data Source from March 15-29, 2011. Statistical results are weighted to correct known demographic discrepancies. The margin of sampling error for the complete set of weighted data is ±2.7 percentage points.

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

# Design AND Data Collection Procedures

#### Sample Design

Interviews were conducted using a combination of RDD landline, RDD cell and callback samples. A total of 1,052 interviews were completed with respondents contacted by landline telephone and 1,090 with those contacted on their cellular phone. Five separate sample segments were used for data collection in order to obtain a representative sample while oversampling 18-34 year-olds.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1. Sample Segments** | | | |
| Segment | Sample Type | Population | n= |
| 1 | Landline RDD | All adults | 1,001 |
| 2 | Cell RDD | All adults | 805 |
| 3 | Cell RDD screened for 18-34 | 18-34 | 204 |
| 4 | Landline callback for 18-34 | 18-34 | 51 |
| 5 | Cell callback for 18-34 | 18-34 | 81 |

Sample segment 1 consists of landline random-digit dialing (RDD) samples drawn using standard list-assisted methods, where telephone numbers were drawn with equal probabilities from all active blocks in the continental US. The cell samples used for segments 2 and 3 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. These RDD samples, both landline and cell, were disproportionately-stratified by county based on estimated incidences of African-American and Hispanic respondents.

The landline and cell callback samples used for segments 4-5 were drawn from recent PSRAI surveys and included people who were ages 18-34 at the time of the initial interview.

#### Questionnaire Development and Testing

The questionnaire was developed by the Pew Social Trends & Demographics Project. In order to improve the quality of the data, the questionnaire was pretested with a small number of respondents using RDD cell sample. The monitored pretest interviews were conducted using experienced interviewers who could best judge the quality of the answers given and the degree to which respondents understood the questions. Some final changes were made to the questionnaire based on the monitored pretest interviews.

#### Contact Procedures

Interviews were conducted from March 15-29, 2011. As many as 7 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.

The introduction and screening procedures differed depending on the sample segment. For each contacted household in sample segment 1, interviewers asked to speak with either the youngest adult male or youngest adult female currently at home based on a random rotation. If no male/female was available at the time of the call, interviewers asked to speak with the youngest adult of the opposite sex. This systematic respondent selection technique has been shown to produce samples that closely mirror the population in terms of age and gender when combined with cell sample.

Sample segment 2 included interviews with all adults on cell phones. This segment was administered a standard cell phone screener which simply confirmed that the person was an adult and in a safe place to talk before continuing with the main interview. Sample segment 3 respondents were administered an age screener. Those who were between the ages of 18 and 34 were taken through the main interview while all others were screener out as age-ineligible.

For callback sample segments 4, interviewers started by asking to talk with the person in the household who had previously completed a telephone interview. The person was identified by age and gender. After the target respondent was on the phone, they were administered the age screener before continuing with the main interview. Callback cell sample segments 5 was administered the age screener before continuing with the main interview.

All cell phone respondents were offered $5 for their participation.

# Weighting and analysis

Weighting is generally used in survey analysis to adjust for effects of the sample design and to compensate for patterns of nonresponse that might bias results. The weighting was accomplished in multiple stages to account for the different sample segments as well as the oversampling of 18-34 year-olds. Weighting also balanced sample demographic distributions to match known population parameters.

The first-stage weight is the product of three different adjustments that were made to the different sample segments. These adjustments are summarized in the following table.

|  |  |
| --- | --- |
| **Table 2. Adjustments made in the First-Stage Weight** | |
| Variable name | Description |
| SAMPWT | RDD Sample design weight. Corrects for disproportionately-stratified RDD samples in segments 1-3 |
| OSADJUST | Adjustment to account for oversampling of 18-34 year-olds in segments 2,4 and 5 |
| PUPSA | Phone Use/Probability of Selection Adjustment. Corrects for [a] overlapping cell and landline sample frames and [b] different probabilities of within household selection based on the number of eligible household members |

This first-stage weight (WT2) was used as an input weight for the demographic raking. The data was raked to current population parameters for: sex by age; sex by education; age by education; race/ethnicity; census region; population density and household telephone usage.

The telephone usage parameter was derived from an analysis of recently available National Health Interview Survey data[[1]](#footnote-1). The population density parameter is county-based and was derived from Census 2000 data. All other weighting parameters were derived from the Census Bureau’s 2010 Annual Social and Economic Supplement (ASEC).

This stage of weighting, which incorporated each respondent's first-stage weight, 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*. The raking corrects for differential non-response that is related to particular demographic characteristics of the sample. This weight ensures that the demographic characteristics of the sample closely approximate the demographic characteristics of the population. Table 2 compares weighted and unweighted sample demographics to population parameters.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 3. Sample Demographics** | |  |  |
|  | Parameter | Unweighted | Weighted |
| Gender |  |  |  |
| Male | 48.5 | 47.8 | 48.6 |
| Female | 51.5 | 52.2 | 51.4 |
|  |  |  |  |
| Age |  |  |  |
| 18-24 | 12.8 | 16.6 | 13.2 |
| 25-34 | 17.9 | 20.4 | 18.4 |
| 35-44 | 17.6 | 12.0 | 16.7 |
| 45-54 | 19.4 | 15.9 | 19.3 |
| 55-64 | 15.4 | 15.6 | 15.1 |
| 65+ | 16.8 | 19.6 | 17.3 |
|  |  |  |  |
| Education |  |  |  |
| Less than HS Grad. | 13.7 | 10.9 | 13.2 |
| HS Grad., no PSE | 31.0 | 23.3 | 30.0 |
| Tech/Voc | 4.0 | 3.0 | 3.8 |
| Some College, no degree | 19.3 | 19.2 | 19.7 |
| Assoc degree | 4.6 | 8.2 | 4.9 |
| College Graduate | 18.0 | 21.8 | 18.5 |
| Post-grad training | 9.3 | 13.7 | 9.9 |
|  |  |  |  |
| Race/Ethnicity |  |  |  |
| White/not Hispanic | 68.3 | 61.8 | 67.3 |
| Black/not Hispanic | 11.5 | 14.2 | 12.0 |
| Hispanic | 14.0 | 15.0 | 14.2 |
| Other/not Hispanic | 6.2 | 9.1 | 6.5 |
|  |  |  |  |
| Region |  |  |  |
| Northeast | 18.5 | 14.4 | 18.0 |
| Midwest | 21.9 | 18.0 | 21.6 |
| South | 36.9 | 45.2 | 37.5 |
| West | 22.6 | 22.4 | 23.0 |
|  |  |  |  |
| County Pop. Density |  |  |  |
| 1 - Lowest | 20.1 | 16.2 | 18.8 |
| 2 | 20.0 | 20.3 | 20.0 |
| 3 | 20.1 | 19.7 | 20.4 |
| 4 | 20.2 | 21.0 | 20.5 |
| 5 - Highest | 19.6 | 22.8 | 20.2 |
|  |  |  |  |
| Household Phone Use |  |  |  |
| LLO | 9.4 | 6.0 | 8.4 |
| Dual - few, some cell | 44.4 | 48.6 | 45.0 |
| Dual - most cell | 18.8 | 19.7 | 18.9 |
| CPO | 27.4 | 25.4 | 27.7 |

# 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 1.65.

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



*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 (√*deff* ). Thus, the formula for computing the 95% confidence interval around a percentage is:



*formula 2*

where  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 ±2.7 percentage points. This means that in 95 out of every 100 samples drawn using the same methodology, estimated proportions based on the entire sample will be no more than 2.7 percentage points away from their true values in the population. 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, question wording and reporting inaccuracy may contribute additional error of greater or lesser magnitude. Table 4 shows design effects and margins of error for key subgroups.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 4: Design Effects and Margins of Sampling Error** | | | |
|  | Sample Size | Design Effect | Margin of Error |
| Total Sample | 2,142 | 1.65 | 2.7 percentage points |
|  |  |  |  |
| 18-34 | 781 | 1.62 | 4.5 percentage points |
|  |  |  |  |
| White, not Hispanic | 1,305 | 1.65 | 3.5 percentage points |
| Black, not Hispanic | 299 | 1.47 | 6.9 percentage points |
| Hispanic | 316 | 1.53 | 6.8 percentage points |

# Response Rate

Table 3 reports 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.[[2]](#footnote-2)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 5. Sample Disposition** | | |  |  |  |
| Seg 1 - LL RDD 18+ | Seg 2 - Cell RDD 18+ | Seg 3 - Cell RDD 18-34 | Seg 4 - LL callback | Seg 5 - cell callback |  |
| 1001 | 805 | 204 | 51 | 81 | I=Completes |
| 4496 | 5800 | 4238 | 145 | 126 | R=Refusal and breakoff |
| 1528 | 2303 | 2832 | 39 | 184 | NC=Non contact |
| 57 | 58 | 31 | 0 | 1 | O=Other |
| 16718 | 6711 | 5730 | 93 | 44 | OF=Business/computer/not working/child's cell phone |
| 1982 | 252 | 325 | 11 | 3 | UH/UO=Unknown household/Unknown other |
|  |  |  |  |  |  |
| 0.30 | 0.57 | 0.56 | 0.72 | 0.90 | AAPOR's e=(I+R+NC+O)/(I+R+NC+O+OF) |
|  |  |  |  |  |  |
| **13.0%** | **8.8%** | **2.7%** | **21.0%** | **20.5%** | **AAPOR RR3=I/[I+R+NC+O+(e\*UH/UO)]** |

1. Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June, 2010. National Center for Health Statistics. December 2010. [↑](#footnote-ref-1)
2. PSRAI’s disposition codes and reporting are consistent with the American Association for Public Opinion Research standards. [↑](#footnote-ref-2)