

Methodology

Library Services Survey

Prepared by Princeton Survey Research Associates International
for the Pew Research Center's Internet & American Life Project

November 2012

SUMMARY

The Library Services Survey obtained telephone interviews with a nationally representative sample of 2,252 people ages 16 and older living in the United States. Interviews were conducted via landline ($n_{LL}=1,127$) and cell phone ($n_C=1,125$, including 543 without a landline phone). The survey was conducted by Princeton Survey Research Associates International. The interviews were administered in English and Spanish by Princeton Data Source from October 15 to November 10, 2012. Statistical results are weighted to correct known demographic discrepancies. The margin of sampling error for results based on the complete set of weighted data is ± 2.3 percentage points. Results based on the 1,945 internet users¹ have a margin of sampling error of ± 2.5 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 the United States 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 probabilities in proportion to their share of listed telephone households from active blocks (area code + exchange + two-digit block number) that contained three 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.

Contact Procedures

Interviews were conducted from October 15 to November 10, 2012. 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

¹ Internet user is defined based on those accessing the internet occasionally, sending or receiving email, and/or accessing the internet on a cell phone, tablet, or other mobile handheld device.

respondents. Interviewing was spread as evenly as possible across the days in field. Each telephone number was called at least one time during the day in an attempt to complete an interview.

For the landline sample, interviewers asked to speak with the youngest male or female ages 16 or older currently at home based on a random rotation. If no male/female was available, interviewers asked to speak with the youngest person age 16 or older of the other gender. 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 interviewing.

For the cellular sample, interviews were conducted with the person who answered the phone. Interviewers verified that the person was age 16 or older and in a safe place before administering the survey. Cellular respondents were offered a post-paid cash reimbursement for their participation.

Weighting and analysis

The first stage of weighting corrected for different probabilities of selection associated with the number of adults in each household and each respondent's telephone usage patterns.² This weighting also adjusts for the overlapping landline and cell sample frames and the relative sizes of each frame and each sample.

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

$$WT_i = \frac{1}{\left(\frac{S_{LL}}{S_{CP}} \times \frac{1}{AD_i}\right)} \text{ if respondent has no cell phone}$$
$$WT_i = \frac{1}{\left(\frac{S_{LL}}{S_{CP}} \times \frac{1}{AD_i}\right) + R} \text{ if respondent has both kinds of phones}$$
$$WT_i = \frac{1}{R} \text{ if respondent has no land line phone}$$

Where S_{LL} = size of the landline sample

S_{CP} = size of the cell phone sample

AD_i = Number of adults in the household

R = Estimated ratio of the land line sample frame to the cell phone sample frame

The equations can be simplified by plugging in the values for $S_{LL} = 1,127$ and $S_{CP} = 1,125$. Additionally, we will estimate of the ratio of the size of landline sample frame to the cell phone sample frame $R = 0.60$.

The final stage of weighting balances sample demographics to population parameters. The sample is balanced by form to match national population parameters for sex, age, education, race, Hispanic origin, region (U.S. Census definitions), population density, and telephone usage. The Hispanic origin was split out based on nativity; U.S born and non-U.S. born. The White, non-Hispanic subgroup is also balanced on age, education and region. The basic weighting parameters came from a special analysis of the

² i.e., whether respondents have only a landline telephone, only a cell phone, or both kinds of telephone.

Census Bureau’s 2011 Annual Social and Economic Supplement (ASEC) that included all households in the United States. The population density parameter was derived from Census data. The cell phone usage parameter came from an analysis of the July-December 2011 National Health Interview Survey.³⁴

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: Sample Demographics

Parameter (16+)		Unweighted	Weight
<u>Gender</u>			
Male	48.7%	47.0%	48.7%
Female	51.3%	53.0%	51.3%
<u>Age</u>			
16-24	16.0%	14.2%	16.5%
25-34	17.3%	13.2%	16.9%
35-44	16.6%	12.3%	15.6%
45-54	18.3%	16.6%	18.0%
55-64	15.4%	18.5%	15.3%
65+	16.3%	23.6%	16.5%
<u>Education</u>			
Less than HS Graduate	16.4%	11.3%	16.0%
HS Graduate	29.4%	27.1%	29.2%
Some College/Assoc Degree	27.5%	25.0%	26.6%
College Graduate	26.8%	36.1%	27.6%
<u>Race/Ethnicity</u>			
White/not Hispanic	67.4%	69.8%	66.4%
Black/not Hispanic	11.6%	10.8%	11.5%
Hisp - US born	7.0%	7.1%	7.1%
Hisp - born outside	7.3%	5.2%	7.0%
Other/not Hispanic	6.7%	5.6%	6.5%
<u>Region</u>			
Northeast	18.3%	16.6%	18.9%
Midwest	21.7%	22.6%	21.6%
South	36.8%	36.5%	36.7%
West	23.2%	24.3%	22.8%

³ Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December, 2011. National Center for Health Statistics. June 2012.

⁴ The phone use parameter used for this 16+ sample is the same as the parameter we use for all 18+ surveys. In other words, no adjustment was made to account for the fact that the target population for this survey is slightly different than a standard 18+ general population survey.

<u>County Pop. Density</u>			
1 - Lowest	19.9%	23.2%	20.2%
2	20.0%	18.8%	19.8%
3	20.1%	21.7%	20.2%
4	20.0%	19.8%	20.2%
5 - Highest	20.0%	16.5%	19.6%
<u>Household Phone Use</u>			
LLO	7.0%	5.6%	6.8%
Dual - few, some cell	39.0%	49.8%	39.5%
Dual - most cell	18.8%	20.3%	18.9%
CPO	35.2%	24.1%	34.6%

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 systematic non-response. The total sample design effect for this survey is 1.24.

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

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

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:

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

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 ± 2.3 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 2.3 percentage points away from their true values in the population. The margin of error for estimates based on form 1 or form 2 respondents is ± 3.3 percentage points. 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.

Response Rate

Table 2 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 respondents in the sample that were ultimately interviewed. At PSRAI it is calculated by taking the product of three component rates:⁵

- Contact rate – the proportion of working numbers where a request for interview was made⁶
- 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 landline sample was 11.4 percent. The response rate for the cellular sample was 11 percent.

⁵ PSRAI's disposition codes and reporting are consistent with the American Association for Public Opinion Research standards.

⁶ PSRAI assumes that 75 percent of cases that result in a constant disposition of "No answer" or "Busy" are actually not working numbers.

Table 2: Sample Disposition

Landline	Cell	
27,813	23,844	Total Numbers Dialed
1,100	404	Non-residential
1,120	45	Computer/Fax
8	----	Cell phone
13,815	9,183	Other not working
1,577	321	Additional projected not working
10,193	13,891	Working numbers
36.6%	58.3%	Working Rate
526	107	No Answer / Busy
3,296	4,073	Voice Mail
27	11	Other Non-Contact
6,344	9,700	Contacted numbers
62.2%	69.8%	Contact Rate
373	1,504	Callback
4,749	6,630	Refusal
1,222	1,566	Cooperating numbers
19.3%	16.1%	Cooperation Rate
40	42	Language Barrier
----	375	Screen out / Child's cell phone
1,182	1,149	Eligible numbers
96.7%	73.4%	Eligibility Rate
55	24	Break-off
1,127	1,125	Completes
95.3%	97.9%	Completion Rate
11.4%	11.0%	Response Rate