

Methodology

2017 Hoosier Survey

Prepared by Princeton Survey Research Associates International
for Ball State University

October 2017

SUMMARY

The 2017 Hoosier Survey, sponsored by Ball State University, obtained telephone interviews with a representative sample of 600 adults living in Indiana. Telephone interviews were conducted by landline (360) and cell phone (240, including 156 with adults with no landline phone). The survey was conducted by Princeton Survey Research Associates International (PSRAI). Interviews were done in English by Issues and Answers from October 2-15, 2017. Statistical results are weighted to correct known demographic discrepancies. The margin of sampling error for the complete set of weighted data is ± 5.3 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 Indiana 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 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 2-15, 2017. As many as seven attempts were made to contact every sampled landline telephone number. For cellular telephone numbers, as many as five attempts were made. 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. 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. The interviewers also confirmed the respondent was a current resident of Indiana.

For the cellular sample, interviews were conducted with the person who answered the phone. Interviewers verified that the person was an adult currently residing in Indiana and in a safe place before administering the survey.

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 sample was weighted to match adult general population parameters for Indiana. A two-stage weighting procedure was used to weight this dual-frame sample.

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.¹ This weighting 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, only a cell phone, or both kinds of telephone.

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

$$WT_i = \left[\left(\frac{S_{LL}}{F_{LL}} \times \frac{1}{AD_i} \times LL_i \right) + \left(\frac{S_{CP}}{F_{CP}} \times CP_i \right) - \left(\frac{S_{LL}}{F_{LL}} \times \frac{1}{AD_i} \times LL_i \times \frac{S_{CP}}{F_{CP}} \times CP_i \right) \right]^{-1}$$

Where S_{LL} = the size of the landline sample

F_{LL} = the size of the landline sample frame

S_{CP} = the size of the cell sample

F_{CP} = the size of the cell sample frame

AD_i = Number of adults in household i

$LL_i=1$ if respondent has a landline phone, otherwise $LL=0$.

$CP_i=1$ if respondent has a cell phone, otherwise $CP=0$.

The second stage of weighting balanced sample demographics to population parameters. The sample is balanced to match parameters for sex, age, education, race, Hispanic origin, region², population density, number of adults in the household, and telephone usage. The basic weighting parameters came from the US Census Bureau's 2015 American Community Survey data³ that included all households in Indiana. The population density parameter we derived from 2010 Census data at the county level and the region parameters were derived from 2011-2015 American Community Survey data. The telephone usage parameter came from an analysis of state-level estimates released by NHIS.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 population. Table 1 compares weighted and unweighted sample distributions to population parameters.

² Region was broken into three groups using the FIPS county codes. The groups were "Northern", "Central", and "Southern" Indiana.

³ ACS analysis was based on all adults excluding those living in institutional group quarters.

⁴ NCHS, National Health Interview Survey, 2011–2015; U.S. Census Bureau, American Community Survey, 2010-2014; and InfoUSA.com consumer database, 2011-2015.

Table 1: Sample Demographics

	<u>Parameter</u>	<u>Unweighted</u>	<u>Weighted</u>	
<u>Gender</u>				
	Male	48.3	44.0	49.2
	Female	51.7	56.0	50.8
<u>Age</u>				
	18-24	13.5	4.5	10.7
	25-34	16.7	7.7	16.7
	35-44	16.4	9.3	16.0
	45-54	17.5	15.7	18.2
	55-64	17.1	19.7	18.2
	65+	18.8	43.2	20.1
<u>Education</u>				
	HS Grad or less	45.5	35.8	43.0
	Some College/Assoc degree	31.3	30.3	32.1
	College Graduate	23.3	33.8	24.9
<u>Race/Ethnicity</u>				
	White/not Hispanic	82.6	89.2	84.1
	Black/not Hispanic	8.4	7.2	8.3
	Hispanic/Other	9.0	3.7	7.6
<u>Region</u>				
	Northern	32.4	32.0	33.1
	Central	43.8	46.3	44.3
	Southern	23.8	21.7	22.6
<u>County Pop. Density</u>				
	1 - Lowest	22.5	24.3	21.7
	2	23.5	26.2	24.4
	3	32.5	28.3	32.1
	4	7.6	6.5	7.6
	5 - Highest	13.9	14.7	14.1
<u>Household Phone Use</u>				
	LLO	8.5	5.2	7.5
	Dual	38.5	68.8	40.9
	CPO	53.1	26.0	51.6
<u># of adults in HH</u>				
	One	19.6	27.3	19.8
	Two	54.8	54.7	55.6
	Three +	25.6	18.0	24.6

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 a disproportionate sample design and systematic non-response. The total sample design effect for this survey is 1.75.

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 ± 5.3 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 5.3 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, 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 sample that was ultimately interviewed. Response rates are computed according to American Association for Public Opinion Research standards.⁵ Thus the response rate for the landline samples was 9 percent. The response rate for the cellular samples was 10 percent.

⁵ American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR.

Table 2. Sample Disposition

<u>Landline</u>	<u>Cell</u>	
403	106	OF = Out of Frame
403	106	Non-residential/Business
1,353	2,557	NWC = Not working/computer
1,069	2,483	Not working
284	74	Computer/fax/modem
1,186	436	UHUO _{NC} = Non-contact, unknown if household/unknown other (NA/busy all attempts)
2,060	2,281	UO _{NC} = Non-contact, unknown eligibility
2,060	2,281	Voice mail
0	0	Other non-contact (deaf/disabled/deceased)
1,478	1,321	UO _R = Refusal, unknown if eligible
1,299	1,046	Refusals
179	275	Callbacks
74	62	O = Other (language)
68	245	SO = Screen out
	72	Child's cell phone
68	173	Not an Indiana Resident
20	26	R = Refusal, known eligible (breakoffs and qualified CBs)
360	240	I = Completed interviews
7,002	7,274	T = Total numbers sampled
69.8%	61.1%	$e1 = (I+R+SO+O+UO_R+UO_{NC}) / (I+R+SO+O+UO_R+UO_{NC}+OF+NWC)$ - Est. frame eligibility of non-contacts
84.8%	52.1%	$e2 = (I+R) / (I+R+SO)$ - Est. screening eligibility of unscreened contacts
40.9%	42.6%	$CON = [I + R + (e2*[O + UO_R])] / [I + R + (e2*[O + UO_R + UO_{NC}]) + (e1*e2*UHUO_{NC})]$
21.2%	24.3%	$COOP = I / [I + R + (e2*[O + UO_R])]$
8.7%	10.4%	AAPOR RR3 = $I / [I + R + [e2*(UO_R+UO_{NC}+O)] + [e1*e2*UHUO_{NC}]] = CON*COOP$