

Methodology

Hoosier 2008 Poll

Prepared by Princeton Survey Research Associates International
for Ball State University

November 2008

SUMMARY

The Hoosier 2008 Poll, sponsored by Ball State University, obtained telephone interviews with a representative sample of 601 adults living in Indiana state telephone households. The survey was conducted by Princeton Survey Research International. The interviews were conducted in English by Princeton Data Source, LLC from November 12 to November 16, 2008. Statistical results are weighted to correct known demographic discrepancies. The margin of sampling error for the complete set of weighted data is $\pm 4.6\%$.

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

DESIGN AND DATA COLLECTION PROCEDURES

Sample Design

The sample was designed to represent all continental U.S. telephone households. The telephone sample was provided by Survey Sampling International, LLC (SSI) according to PSRAI specifications. The sample was drawn using standard *list-assisted random digit dialing* (RDD) methodology. *Active blocks* of telephone numbers (area code + exchange + two-digit block number) that contained three or more residential directory listings were selected with probabilities in proportion to their share of listed telephone households; after selection two more digits were added randomly to complete the number. This method guarantees coverage of every assigned phone number regardless of whether that number is directory listed, purposely unlisted, or too new to be listed. After selection, the numbers were compared against business directories and matching numbers purged.

Questionnaire Development and Testing

The questionnaire was developed by PSRAI in collaboration with staff of Ball State University. In order to improve the quality of the data, the questionnaire was pretested with a small number of respondents using RDD telephone number sample. The pretest interviews were monitored by PSRAI staff and 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 November 12 to November 16, 2008. 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 household received at least one daytime call in an attempt to find someone at home. In each contacted household, interviewers asked to speak with the youngest adult male currently at home. If no male was available, interviewers asked to speak with the youngest female at home. This systematic respondent selection technique has been shown to produce samples that closely mirror the population in terms of age and gender.

WEIGHTING AND ANALYSIS

Weighting is generally used in survey analysis to compensate for patterns of nonresponse that might bias results. The weight variable balances the interviewed sample of all adults to match Indiana state parameters for sex, age, education, race, Hispanic origin, region¹, and population density. The region distribution came from a special analysis of the Census 2000 Redistricting Data (Public Law 94-171) Summary File. The other parameters came from the Census Bureau's 2007 Annual Social and Economic Supplement (ASEC) that included all households in the state of Indiana that had a telephone.

¹ Region was broken into three groups using the FIPS county codes. The groups were "Northern", "Central", and "Southern" Indiana.

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 of Indiana. Table 1 compares weighted and unweighted sample distributions to population parameters.

Table 1: Sample Demographics

	Parameter	Unweighted	Weight
<u>Gender</u>			
	Male	49.7	49.8
	Female	50.3	50.2
<u>Age</u>			
	18-24	10.6	8.4
	25-34	16.8	17.7
	35-44	20.1	20.0
	45-54	21.7	22.4
	55-64	15.7	16.0
	65+	15.1	15.5
<u>Education</u>			
	Less than HS Grad.	13.2	11.4
	HS Grad.	45.5	45.9
	Some College	20.3	20.8
	College Grad.	21.0	21.9
<u>Region</u>			
	Northeast	33.9	34.2
	Central	42.5	42.4
	Southern	23.6	23.5
<u>Race/Ethnicity</u>			
	White/not Hispanic	87.3	88.0
	Black/not Hispanic	7.3	7.3
	Other/Hispanic	5.4	4.6
<u>Population Density</u>			
	1 - Lowest	19.5	19.8
	2	32.2	32.4
	3	26.1	26.5
	4	8.0	7.7
	5 - Highest	14.2	13.6

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 1.31.

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 $\pm 4.6\%$. 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 five 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 sample. 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:²

- Contact rate – the proportion of working numbers where a request for interview was made³
- 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 this survey was 27 percent.

² PSRAI's disposition codes and reporting are consistent with the American Association for Public Opinion Research standards.

³ 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

5,698	Total Numbers Dialed
251	Business/Government/Non-Residential
219	Fax/Modem
2	Cell phone
2,487	Other Not-Working
412	Additional projected NW
<hr/>	
2,327	Working numbers
40.8%	Working Rate
<hr/>	
127	No Answer
10	Busy
268	Answering Machine
17	Other Non-Contacts
<hr/>	
1,905	Contacted numbers
81.9%	Contact Rate
<hr/>	
75	Callbacks
1,143	Refusals
<hr/>	
687	Cooperating numbers
36.1%	Cooperation Rate
<hr/>	
20	Language Barrier
<hr/>	
667	Eligible numbers
97.1%	Eligibility Rate
<hr/>	
66	Break offs
<hr/>	
601	Completes
90.1%	Completion Rate
<hr/>	
26.6%	Response Rate