

**Online Appendix for
“Are Survey Experiments Externally Valid?”
Jason Barabas and Jennifer Jerit
Florida State University**

This appendix provides details on the methodology underlying each survey, sample characteristics, randomization, and attrition.¹

Survey Methodology

The cross-sectional survey on the 2007 Medicare announcement was conducted by Knowledge Networks, an internet research and opinion polling firm. To select a sample, Knowledge Networks identified potential respondents from its KnowledgePanel®. The KnowledgePanel® is a nationally representative, probability-based web panel based on dual-frame sampling that combines traditional random-digit-dialing telephone surveying techniques with an address-based technique that allows the sample to be representative of cell-phone-only households as well as those with land-lines. Potential panel members are chosen via a statistical sampling method and using known published sampling frames that cover 99% of the U.S. population. According to Knowledge Networks, the KnowledgePanel® consists of about 50,000 U.S. residents, age 18 and older, including cell phone-only households and those who are of Hispanic origin that were selected probabilistically. When a person agrees to participate, they are provided with free Internet access and are given the necessary hardware for as long as they remain in KnowledgePanel®.

Between March 2 and March 10, 2007, Knowledge Networks completed 805 interviews out of 1,143 eligible respondents who were contacted for an interview. This results in a 70.4% completion rate. While the American Association for Public Opinion Research response rate standards have not been formally established for web panels, the completion statistic corresponds

¹ Please contact the authors for data to replicate the analyses in this online appendix.

to AAPOR Response Rate 3. In the second wave of the survey, a new cross-section of 817 respondents completed interviews between April 26, 2007 and May 3, 2007 out of 1,143 eligible who were contacted ($817/1,143 = 71.5\%$).

For the second survey on Medicare, the authors contracted with Polimetrix (which subsequently became known as YouGov/Polimetrix PollingPoint) to conduct a panel survey during the spring of 2007. The first wave was conducted from March 1 to 21, 2007. YouGov/Polimetrix PollingPoint interviewed thousands of respondents from their panel—a pool of several hundred thousand individuals who volunteered or were recruited to participate in occasional online polls. Although exact techniques are proprietary, 1,500 participants from the initial survey were selected based upon their gender, age, race, education, party identification and ideology to match known statistics for the U.S. general population from the 2005 American Community Survey. For the second wave from April 26 to May 16, 2007, Polimetrix reinterviewed 64% of the respondents who had previously completed Wave 1 of this study. Of the initial 1,500 respondents, only 509 cases were employed in the analyses here, 346 who had been randomly assigned into the survey treatment condition described in the text as well as 163 individuals in the control group who were followed over time. Given the nature of the topic (Medicare), the sample was designed to include an over-sample of adults aged-55 and older, but the over-sample is distributed randomly across the various conditions.

Finally, the third survey on immigration was conducted as a part of the Cooperative Campaign Analysis Project (CCAP). The CCAP includes teams of over 60 political scientists at 25 academic institutions. This dataset is a six-wave panel study with an over-sample in highly contested battleground and early primary states (FL, IA, MN, NV, WI, NH, NM, OH, PA). In

that sense, the sample more closely represents states with competitive elections rather than the nation as a whole.

CCAP conducted a baseline survey in December of 2007 with over thirty thousand respondents. Subsequent panel waves took place in January, March, September, October, and November of 2008. At each wave, additional cases were added to deficient cells to achieve approximately 30,000 interviews, although the number participating in any given institutional module was much smaller. The analyses in the paper employ 1,039 respondents from the September and October 2008 waves. Detailed response rates are not available.

Respondents were chosen from the YouGov/Polimetrix PollingPoint Panel using a cross-classification system. In particular, YouGov/Polimetrix constructed a sampling frame from the 2007 American Community Study using data on age, race, gender, education, marital status, number of children under 18, family income, employment status, citizenship, state, and metropolitan area. The final set of completed interviews were then matched to the target frame via stratification along age, race, gender, education, and state (with battleground states double sampled) by simple random sampling within strata, excluding all non-registered persons. For background details on the sampling matching and weighting techniques, see Rivers (2006).

Comparing the Samples to the U.S. Census Data

Appendix Table 1 presents data on national population statistics from the U.S. Census Bureau in the mid-2000s near the time of the surveys. The first column displays combined 2005 to 2007 American Community Study estimates on selected demographic characteristics for percentage female, African-American, married, education (bachelor's degree or higher), households with income below \$50,000, and median age. The categories were chosen to align

with what was available in the surveys.² The next nine columns contain comparative data for each sample. Within each sample, the first column shows the overall survey experiment, the second depicts the entire natural experiment group before the announcement, and the third shows the natural experiment group after the announcement.

All three samples are diverse. Every survey is split roughly evenly between women and men. The percentage African-American is generally less than the national average while the percentage married is typically higher. The respondents in the surveys are typically more educated, wealthy and older than the nation as a whole. Although the purpose of the table is to compare the samples to the population, it is also possible to see that the composition remains relatively similar across time for the natural experiment cases.

On some demographic characteristics, the samples might deviate from being nationally representative for logical reasons. First, all three surveys were conducted via the internet, and internet users tend to be unrepresentative in ways mirrored in the table. Second, in the case of the immigration study, the sampling design was targeted toward *registered voters* in presidential primary battleground states instead of the nation as a whole (e.g., Iowa and New Hampshire have proportionally fewer black residents than the nation and voters tend to be more educated, wealthy, etc.). Similarly, the 2007 Polimetrix respondents appear to be older, but this makes

² The 2005-07 figures are close to the 2000 U.S. Census (the corresponding 2000 Census numbers are 50.9 percent female, 12.5 percent black, 54.4 percent married, 24.4 percent with a B.A. degree or higher, 57.9 below \$50,000 in household income, and median age of 35.3). Missing demographic responses for the surveys were recovered via multiple imputation using the Amelia II software program (see King et al. 2001).

sense since sample was selected to include an over-sample of adults aged 55 and older. In general, the Knowledge Networks sample appears to be the most nationally representative, which is to be expected given the probabilistic polling design and the lack of sampling restrictions. While the analyses in the paper were conducted without sampling weights, the experimental effects retain their same substantive significance in analyses with sampling weights. Moreover, although nonprobability internet surveys may be unrepresentative or subject to bias (see Malhotra and Krosnick 2007), the samples are more heterogeneous than the convenience samples often used in laboratory experiments.

Randomization and Attrition

Across the three datasets, randomization into treatment and control conditions generally succeeded. In a model predicting treatment assignment status relative to the control condition, the Polimetrix data for the Medicare and immigration cases showed no significant differences across education, income, age, gender, race, marital status, and partisanship ($p > .10$; two-tailed). However, in the Knowledge Networks sample, the treatment group was slightly more likely to contain women than men ($p < .10$; two-tailed). Also, over time panel attrition appeared to be random as was the composition of the KN cross-sections at each time point; the only exception was a lower percentage of blacks in the t_2 control group in KN. Reexamining the survey experiment and natural experimental effects in models with demographic covariates as controlling factors does not change the substantive conclusions reported in the text.

References

- King, Gary James, Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95 (Mar.): 49-69.
- Malhotra, Neil, and Jon A. Krosnick. 2007. "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples." *Political Analysis* 15(3): 286-323.
- Rivers, Douglas. 2006. "Sample Matching: Representative Sampling from Internet Panels." Polimetrix White Paper Series. Available at: http://www.polimetrix.com/documents/Polimetrix_Whitepaper_Sample_Matching.pdf

Appendix Table 1. Descriptive Information Comparing U.S. Census to Samples (% in Each Category)

Demographic Characteristic	U.S. Census 2005-07 ACS	Study 1: 2007 Medicare Announcement						Study 2: 2008 Immigration		
		Polimetrix			Knowledge Networks			Polimetrix/CCAP		
		Survey Exp.	Natural Exp. t ₁	Natural Exp. t ₂	Survey Exp.	Natural Exp. t ₁	Natural Exp. t ₂	Survey Exp.	Natural Exp. t ₁	Natural Exp. t ₂
Female	50.8	50.3 (2.2)	44.8 (3.9)	44.8 (3.9)	51.6 (1.8)	49.7 (2.0)	50.3 (2.0)	47.6 (2.1)	48.4 (2.9)	51.2 (2.3)
African-American	12.4	10.8 (1.4)	8.0 (2.1)	8.0 (2.1)	9.7 (1.0)	10.1 (1.2)	9.6 (1.2)	3.8 (0.8)	3.6 (1.1)	4.2 (1.0)
Married	50.5	61.9 (2.2)	64.4 (3.8)	64.4 (3.8)	56.8 (1.7)	57.0 (2.0)	55.0 (2.0)	59.9 (2.2)	63.5 (3.1)	67.0 (2.5)
College Degree +	28.0	31.4 (2.1)	30.1 (3.6)	30.1 (3.6)	24.8 (1.5)	25.5 (1.8)	24.9 (1.8)	41.3 (2.0)	43.1 (2.8)	39.8 (2.3)
Income <\$50,000	50.0	43.0 (2.2)	38.0 (3.8)	38.0 (3.8)	58.1 (1.7)	56.5 (2.0)	59.6 (2.0)	37.5 (2.3)	35.7 (2.8)	36.2 (2.3)
Age (median)	36.4	53.0 (1.3)	52.0 (1.9)	52.0 (1.9)	47.0 (1.2)	47.0 (1.2)	46.5 (1.0)	54.0 (0.7)	54.0 (1.0)	54.0 (0.9)
Population or N	298,757,310	509	163	163	805	604	606	584	304	455

Note: The Census data are the combined three year American Community Survey (ACS) estimates collected between January 2005 and December 2007. Standard errors appear in parentheses. The standard errors on the age medians were bootstrapped.