

GAUGING THE IMPACT OF GROWING NONRESPONSE ON ESTIMATES FROM A NATIONAL RDD TELEPHONE SURVEY

SCOTT KEETER
COURTNEY KENNEDY
MICHAEL DIMOCK
JONATHAN BEST
PEYTON CRAIGHILL

Abstract Declining contact and cooperation rates in random digit dial (RDD) national telephone surveys raise serious concerns about the validity of estimates drawn from such research. While research in the 1990s indicated that nonresponse bias was relatively small, response rates have continued to fall since then. The current study replicates a 1997 methodological experiment that compared results from a “Standard” 5-day survey employing the Pew Research Center’s usual methodology with results from a “Rigorous” survey conducted over a much longer field period and achieving a significantly higher response rate. As with the 1997 study, there is little to suggest that unit nonresponse within the range of response rates obtained seriously threatens the quality of survey estimates. In 77 out of 84 comparable items, the two surveys yielded results that were statistically indistinguishable. While the “Rigorous” study respondents tended to be somewhat less politically engaged, they did not report consistently different behaviors or attitudes on other kinds of questions. With respect to sample composition, the Standard survey was closely aligned with estimates from the U.S. Census and other large government surveys on most variables. We extend our analysis of nonresponse to include comparisons with the hardest-to-reach respondents and with respondents who terminated the interview prior to completion.

SCOTT KEETER is with the Pew Research Center. COURTNEY KENNEDY is with the Pew Research Center and the Joint Program in Survey Methodology. MICHAEL DIMOCK is with the Pew Research Center. JONATHAN BEST is with Princeton Survey Research Associates International. PEYTON CRAIGHILL is with ABC News. Andrew Kohut, president of the Pew Research Center, originated this research project in 1997 and directed the replication study in 2003. The project has benefited from the contributions of numerous Pew staff members, including Carroll Doherty, Elizabeth Mueller Gross, Nilanthi Samaranyake, Nicole Speulda, Jason Owens, and Kate DeLuca. Several other people provided valuable advice in the design of the study. An advisory committee chaired by

doi:10.1093/poq/nfl035

© The Author 2006. Published by Oxford University Press on behalf of the American Association for Public Opinion Research. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org.

Introduction

Confronted with a growing number of unsolicited telephone calls and armed with increasingly sophisticated technology for screening their calls, more Americans are refusing to participate in random digit dial (RDD) telephone surveys than was the case just a few years ago. Survey organizations using this methodology are experiencing declining response rates, with notable drop-offs taking place within the last 10 to 15 years (Curtin, Presser, and Singer 2005; U.S. Department of Education 1997).

Of primary concern for survey researchers are the consequences of lower response rates for estimates of population characteristics. Unfortunately, measuring the relationship between nonresponse and the accuracy of a survey statistic is generally both complex and expensive. To date, there have been few rigorously designed studies with actual empirical evidence to inform theoretical expectations about the consequences of lower response rates.

Recent research, however, has yielded encouraging results. Two research teams using very different experimental designs found little evidence for a relationship between response rate and nonresponse bias in RDD surveys (Curtin, Presser, and Singer 2000; Keeter et al. 2000). Curtin, Presser, and Singer (2000) used detailed call records from the Index of Consumer Sentiment to compare estimates based on a low-effort restricted-call design with estimates based on a relatively high-effort "all-call" design. They found almost no change in the estimates when the additional interviews obtained in the all-call design were omitted. Keeter et al. (2000) conducted a randomized design using identical questionnaires in two parallel surveys: one featuring a "Standard" 5-day design and the second featuring a "Rigorous" design that allowed for more callback and refusal conversion attempts. In comparing the two surveys, they found an average difference of 2 percentage points for the 91 statistics measured.

Meta-analyses have yielded similar findings. Groves (2006) notes that the collective body of empirical work suggests no consistent relationship between response rates and nonresponse bias. Holbrook, Pfent, and Krosnick (2003) concluded that RDD telephone surveys with low response rates generally still have excellent demographic representativeness. Judged against the actual vote totals, most preelection telephone surveys in 2000 and 2004 yielded very accurate forecasts of both national and state-level presidential voting.

Diane Colasanto included Richard Kulka, Warren Mitofsky, Richard Morin, Linda Piekarski, Mark Schulman, Evans Witt, and Cliff Zukin. Jon Rochkind of Princeton Data Source, LLC and Larry Hugick and Mary McIntosh of Princeton Survey Research Associates International provided methodological guidance for the project. The contribution of Survey Sampling International, which donated the telephone sample and demographic data for the project, is gratefully acknowledged. The data on which this study is based are available at <http://people-press.org>. Address correspondence to Scott Keeter; e-mail: skeeter@pewresearch.org.

These results provide compelling evidence that response rate is not necessarily an indicator of survey quality. It is unclear, however, under what circumstances—or for how long—this finding will hold. The survey research environment has been evolving since its inception and has continued to change since these studies were conducted. Curtin, Presser, and Singer (2005) and Holbrook, Pfent, and Krosnick (2003) document a steady increase in nonresponse over the past several years, and the findings presented herein are consistent with these trends. Beyond nonresponse, RDD telephone surveys face a growing noncoverage problem as the percentage of people who can be reached only by cell phone has risen steadily since 2001 (Blumberg and Luke 2006). Consequently, what once was true about the potential for bias may not be the case today. The study reported here was an effort to understand the current state of affairs and to identify trends regarding nonresponse error in telephone surveys that focus on social and political topics.

Research Design

This article reports on a 2003 replication and extension of the Pew Research Center's 1997 study of the relationship between nonresponse and survey accuracy described in Keeter et al. (2000). Both the original experiment and the replication compare samples of respondents obtained through Pew's usual methodology with samples obtained with a more rigorous survey effort over a much longer field period. In both years, an identical questionnaire was used in two separate surveys. In the current study, a "Standard" survey was conducted among 1,000 adults from June 4 to June 8, 2003, using the same call design and effort level as in most Pew survey projects. The sample was list-assisted, and respondents within households were selected using a procedure that asked first to interview the youngest male currently at home; if no male was at home, the interviewer asked for the oldest female. A minimum of 10 call attempts was made for each sampled telephone number, and one refusal conversion attempt was normally made. Interviews were conducted in English only. The "Rigorous" survey began at the same time and ran for more than 21 weeks through October 30, 2003, collecting interviews with adults in 1,089 sampled households. In order to maximize response rates in the Rigorous survey, a number of procedures were implemented, including advance letters (some of which included a \$2 incentive), refusal conversion letters, answering machine messages, and a greatly extended calling period. Since the two survey questionnaires were identical, comparisons of the personal attitudes, behaviors, and characteristics can be made, though questions on which opinions were subject to change over the long field period of the Rigorous survey, such as President George W. Bush's job approval or interest in news stories, are not used in the analysis.

Though the survey is not explicitly described as political in nature, informants and respondents for both the Standard and Rigorous surveys were told in the introduction, “We are conducting a telephone opinion survey for leading newspapers and TV stations around the country.” The first substantive question in the study asked if the respondent approves or disapproves of the way George W. Bush is handling his job as president. Thus, the social and political nature of the study is clearly made salient to the potential respondent.¹

In addition to the main experimental comparison, this study design facilitates several other detailed nonresponse analyses. To go beyond the two discrete samples, we isolate as a point of analysis a subset of respondents who were especially difficult to interview. The 494 people designated as the “hardest-to-reach” respondents refused the interview at least twice before complying and/or required 21 or more calls to complete. While many of these hardest-to-reach respondents were interviewed as part of the Rigorous survey, the research design included continued efforts to reach households in the Standard survey sample beyond the 5-day field period. Though not included in the Standard 5-day estimates, many of the interviews completed in this extended field period are included here in the analysis of the hardest-to-reach.

Smith (1983) demonstrates that extrapolation based on difficult cases has important limitations. He finds that respondents who are difficult to contact or refuse temporarily are not necessarily representative of final nonrespondents. Comparisons between the Standard sample and the hardest-to-reach, therefore, only reflect the effect of the extra recruitment effort described above; they are not pure measures of nonresponse bias.

The final component of this analysis addresses the cost and benefit of a rigorous research design for RDD telephone surveys. Increasing the response rate by collecting data over a span of several months and using incentives and additional nontelephone contacts was very costly, especially in comparison with the 1997 study. We raise—but cannot fully answer—the question of whether such effort is ultimately worthwhile.

Findings: Response Rates and Demographic Representativeness

CONTACT AND COOPERATION

The response rates in this study reflect the increasing difficulty that telephone survey researchers face contacting Americans in their homes and the growing reluctance among the public to cooperate when they are contacted. As shown in table 1, the Standard 5-day survey, employing techniques used by most opinion polling organizations, obtained interviews with people in 25

1. The full introduction is available from the authors.

Table 1. Comparison of Rates across Studies

	Standard 5-Day Survey		Rigorous Survey	
	1997	2003	1997	2003
Response Rate	36%	25%	61%	50%
Cooperation Rate	43%	34%	72%	58%
Contact Rate	90%	79%	94%	91%
Sample Size	(1,000)	(1,000)	(1,201)	(1,089)

NOTE.—Figures computed according to American Association for Public Opinion Research (AAPOR 2005) standard definitions of response rate 3, cooperation rate 3, and contact rate 2.

percent of the sampled households, a result comparable to that of other Pew studies during this time frame.² That represents a decrease of about 11 percentage points (on average) from the late 1990s. This decline results from a combination of both lower contact and cooperation rates when compared with the 1997 benchmark. The new study was able to make contact with 79 percent of residential households during the 5-day survey period, down from 90 percent in the original 1997 study. And among contacted households, just a third (34 percent) resulted in completed interviews, down from 43 percent in 1997.

The Rigorous survey achieved twice the overall response rate of the Standard 5-day study (50 percent versus 25 percent), but this was significantly lower than the 61 percent response rate achieved in the 1997 benchmark survey. Virtually all of this decline can be attributed to people's growing unwillingness to participate in telephone surveys. Because of its extended design, the Rigorous study was able to achieve a contact rate comparable to 1997 by making contact with virtually all households in the sample (91 percent). But even with a 4-month field period, refusal conversions, letters, incentives, and other efforts to boost respondent participation, the final cooperation rate was 58 percent, down from 72 percent in 1997.

2. Response rates for other standard Pew surveys conducted in 2003 ranged from approximately 22 percent to 32 percent. This analysis describes several different measures of success in the conduct of surveys; all rates are computed using American Association for Public Opinion Research (AAPOR 2005) standards. The "contact rate" is the percentage of known or assumed residential households in which contact was made with a person. The "cooperation rate" is the percentage of contacted households in which an interview was completed. The "response rate" is the overall percentage of known or assumed households in which an interview was completed. A telephone number was presumed to be servicing a residential household if contact was established with an eligible respondent, a nonbusiness answering machine was reached, or the final disposition was "call back." In addition, 47 percent of the numbers of unknown residential eligibility are also counted as residential (per AAPOR response rate guidelines for RR3). For details on this calculation, please contact the authors.

SAMPLE REPRESENTATIVENESS

To assess the representativeness of our sample relative to the U.S. adult population, we compared demographic and other characteristics of the unweighted survey samples with national parameters established in surveys that obtain response rates of 90 percent or more, including the Current Population Survey (CPS) and the National Health Interview Survey (NHIS). The CPS parameters are based on telephone households, while the NHIS parameters are based on all households.³ Compared with the benchmarks, the demographic composition of the samples of survey respondents were generally very accurate (see table 2). On the important characteristic of race, African-Americans comprised 10 percent of the Standard 5-day sample, which almost exactly matches the national parameter (11 percent of all adults age 18 and older). People identifying as Hispanic, however, were substantially underrepresented in the Standard sample (7 percent compared with the U.S. Census estimate of 12 percent of adults), a function, in part, of the fact that interviews were conducted only in English.

Younger adults (ages 18–34) were also underrepresented in the Standard sample, but the difference was relatively small (28 percent in the sample, 31 percent in the population). The biggest mismatch between the sample and the population pertained to educational levels: 16 percent of U.S. adults have not completed high school, compared with only 8 percent of survey respondents. Similarly, one-quarter of the public has at least a 4-year college degree, but fully 34 percent in the sample reported this level of education.

Despite its higher response rate, the Rigorous sample was not closer to population parameters than the Standard sample in every comparison. It did reach more African-Americans (12 percent) and Hispanics⁴ (11 percent) than the Standard survey, and the Rigorous sample was somewhat more reflective of the general public's education, as well as current levels of cigarette smoking. But it was no closer to the population in terms of the income and age distribution and was slightly less accurate than the Standard survey in its estimate of food stamp utilization, health insurance status, and home ownership.

3. In 2003 the CPS telephone question did not make a distinction between cell phones and landlines, and thus "telephone households" in the CPS may include at least some cell-only households, which are unreachable in a landline survey such as Pew's. The magnitude of this error is likely to be small, however, since cell-only households constituted just 2.8 percent of all households in the first half of 2003 (Blumberg and Luke 2006). Similarly, the NHIS estimates for health insurance coverage and smoking include adults in nonphone households (1.4 percent) and cell-only households (2.8 percent), as well as those reachable by landline.

4. The Rigorous survey offered respondents the opportunity to be interviewed in Spanish, while the Standard survey did not; 34 Spanish language interviews were conducted, accounting for 3 percent of all cases in the Rigorous survey. The majority of these individuals said that they would not have been able to do the interview in English, and an additional 26 percent (9 cases) said it would have been difficult for them.

Table 2. Demographic and Social Characteristics

	CPS ^a %	Standard %	Rigorous %	Hardest-to-Reach %
Race ^b				
White	83	82	81	74
Black	11	10	12	15
Other	6	8	7	9
Ethnicity ^{b,c}				
Hispanic	12	7	11	9
Not Hispanic	88	93	89	91
Age ^b				
18–24	13	12	9	10
25–34	18	16	17	19
35–44	21	20	18	20
45–54	19	18	20	20
55–64	13	14	15	15
65+	16	18	19	14
Education ^b				
College graduate	25	34	31	31
Some college	23	24	25	26
High school graduate	36	34	33	34
Not a high school graduate	16	8	11	8
Own or Rent				
Own home	73	70	66	61
Rent/other	26	30	34	39
Food Stamps				
Yes	5	6	7	5
No/Don't know/Refused	95	94	93	95
Health Insurance from Employer				
Yes	66	67	69	69
No/Don't know/Refused	34	33	31	31
Current Smoker				
Yes	22	26	23	22
No/Don't know/Refused	78	74	77	78
County Population Density				
Highest quintile	20	17	17	18
2nd highest	20	17	18	22
Middle quintile	20	21	19	18
2nd lowest	20	23	25	22
Lowest quintile	20	22	21	20
		(1,000)	(1,089)	(494)

NOTE.—“Hardest-to-reach” cases had refused the interview at least twice and/or required 21 or more calls to complete.

^aAll CPS (Current Population Survey) figures are from March 2003 Annual Social and Economic Supplement, except smoking (National Health Interview Survey).

^bBased on unweighted data.

^cStandard survey conducted only in English; Rigorous survey offered respondents the option to be interviewed in Spanish.

Similarly, the Rigorous protocol did not overcome a common geographic disparity in telephone surveys caused by lower response rates in urban areas. As documented by Dimock, Samaranayake, and Best (2005), counties with higher population density are systematically underrepresented in many RDD surveys. Whereas 40 percent of interviews should come from the two highest-density quintiles of counties, just 34 percent and 35 percent of completed interviews in the Standard and Rigorous samples, respectively, came from these counties.

Harder-to-reach respondents were somewhat more likely to live in places with higher population density. The geographic distribution of the 494 hardest-to-reach respondents was much closer to the national parameters, but these respondents are not numerous enough even in this study to offset the bias toward lower-density areas.

In what follows, analyses of attitudinal and behavioral measures in the study are based on data weighted to census distributions of age, gender, education, race/ethnicity, and population density, thus showing the magnitude of differences that persist after known demographic discrepancies have been corrected (as they are in most national RDD surveys). Because the demographic differences between the studies were so small, however, the resulting weights have very little impact on the substantive comparisons. Estimates for the hardest-to-reach respondents are based on unweighted data because there are no parameters to which, a priori, the hardest-to-reach respondents could be compared.

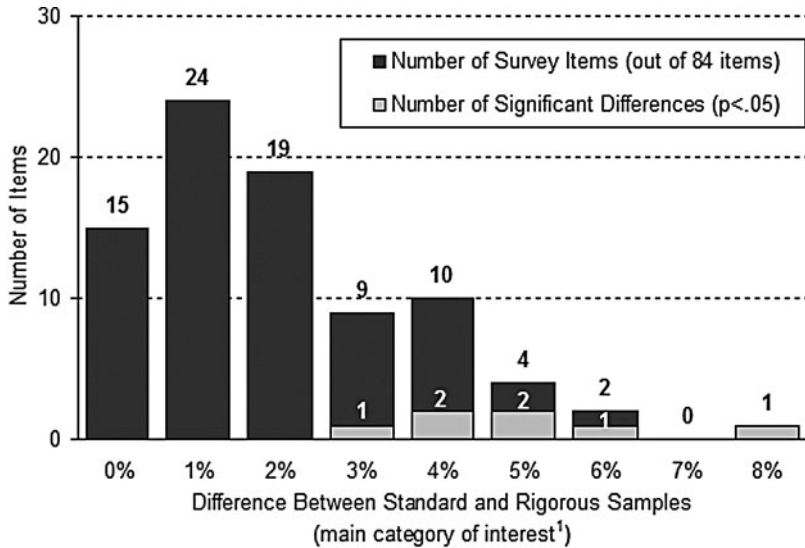
Findings: Attitudes and Behaviors

OVERVIEW

As in the 1997 experiment, comparisons of estimates from the Standard and Rigorous surveys show that achieving a higher response rate does not yield significantly different estimates for the vast majority of questions. The distribution of observed differences for 84 comparable survey items is presented in figure 1.⁵ The median difference in results between the two surveys was 2 percentage points, and nearly half of the survey items (46 percent) resulted in differences of 0 or 1 percentage points.

To determine which cross-survey differences were statistically significant, we used a chi-square test to account for variance across all response categories

5. In most cases, figure 1 shows the response option within each survey item with the *widest* margin between the Standard and Rigorous estimates (excluding “don’t know/refused” or other volunteered categories). The exception is on list items (such as favorability ratings) where we chose the same category of comparison across all items within the same list. For continuous variables such as respondent age and household size, we tested the analytic grouping with the widest margin (e.g., proportion age 18–29 or proportion of households with three or more people).



¹ See footnote 6 for a discussion of the criteria used to select the particular response categories that were compared.

Figure 1. Histogram of the differences between Standard and Rigorous survey estimates.

simultaneously. Overall, 7 of the 84 comparable survey items resulted in a difference that exceeded a 95 percent confidence interval for the population mean. Comparisons that yielded a *p*-value less than or equal to .05 in a chi-square test of homogeneity are represented by the lighter shading in figure 1.

Each of these significant differences is described individually in table 3. While we observed approximately twice as many significant differences as might be expected from chance alone, the magnitude of even the significant differences tended to be modest. The most important and consistent difference was seen in a pair of measures of political engagement: respondents in the Rigorous survey were less likely to say they voted in the 2002 midterm elections and were less likely to report being registered to vote at all. Consistent with the results of the 1997 study, respondents in the Rigorous survey are somewhat less trusting of others and express less favorable views of minorities, though the only difference that exceeds the .05 threshold is on views of Jews, whereas in 1997 the significant gap was on favorability ratings of African-Americans.

Although the comparison of the Standard with the Rigorous surveys provides the best estimate of the net effect of increased survey nonresponse, it

Table 3. Items with Statistically Significant Differences across Samples

Relative to the Standard Sample, the Rigorous Sample Is . . .	χ^2 Test (<i>p</i> -value)	Weighted Survey Estimates		
		Standard %	Rigorous %	Difference
Less likely to have voted in 2002 (certain)	0.000	56	48	-8
Less likely to be Republican	0.005	32	26	-6
Less likely to have 3+ adults in the household	0.015	28	23	-5
Less likely to be registered to vote (certain)	0.018	72	67	-5
Less likely to trust most people	0.019	35	31	-4
Less favorable opinion of Jews	0.03	25	22	-3
More likely to be ideologically moderate	0.05	36	40	+4
		(<i>N</i> = 1,000)	(<i>N</i> = 1,089)	

NOTE.—This is a comprehensive list of the items that had $p \leq .05$ in χ^2 tests of homogeneity. Prior to the χ^2 testing, the weights were adjusted so that the weighted sample size equaled the actual number of cases in each χ^2 table. To eliminate small cells, nonresponse categories (such as “other” or “don’t know”) were removed before conducting the χ^2 test. All survey estimates are weighted to adjust for known variation from census demographic parameters, and based on the total sample.

may understate ways in which respondents who are particularly difficult to reach differ because the Rigorous study includes a number of people who were contacted relatively easily. For this reason, the following analysis of survey differences across the full range of topics includes a comparison with the “hardest-to-reach” respondents—the 494 households that refused the interview at least twice before complying and/or required 21 or more calls to complete.

TRUST AND SAFETY CONCERNS

Just as the politically engaged may be more likely to respond to surveys that have at least a partial focus on politics and public affairs, people with low levels of interpersonal trust may be less likely to respond to almost any type of survey because of privacy concerns. Respondents in the Rigorous sample were somewhat less trusting of others than those in the Standard sample, and the hardest-to-reach respondents reported even less interpersonal trust. As shown in table 4, more than a third of respondents in the Standard sample (35 percent) said that generally speaking “most people can be trusted.” Significantly fewer of those in the Rigorous sample and people who were most difficult to reach expressed that sentiment (31 percent and 28 percent,

Table 4. Measures of Personal Trust and Everyday Concerns

	Standard %	Rigorous %	Hardest-to-Reach %
Trust most people	35	31	28
Can turn to many people	37	37	37
Neighborhood not safe	12	12	11
Victim of property crime	14	12	12
Been mugged or assaulted	2	2	2
Worry about computers invading privacy	39	37	40
	(<i>N</i> = 1,000)	(<i>N</i> = 1,089)	(<i>N</i> = 494)

NOTE.—Items in boldface indicate figures that are significantly different from the Standard survey at $p \leq .05$.

respectively). The General Social Survey, with a response rate of 70 percent, found 34 percent in 2002 saying that most people could be trusted.

There were no differences between the samples in the extent to which respondents said they had many people they could turn to if they needed help, or in their reported incidence of criminal victimization. The Standard sample respondents and the hardest-to-reach reported similar levels of neighborhood crime and expressed similar levels of concern about computers and technology invading their privacy.

PARTY AND IDEOLOGY

The partisan makeup of the two surveys differed significantly, with the Rigorous survey reaching more nonpartisans and fewer Republicans than the Standard survey (table 5). Overall, 44 percent of Rigorous respondents identified themselves as either independents or expressed no partisan preference, compared with 37 percent in the Standard survey. There is a corresponding gap in the percentage identifying themselves as Republican—26 percent in the Rigorous survey and 32 percent in the Standard survey. The percent Democratic was roughly even in both surveys (30 percent of Rigorous, 31 percent of Standard).

A follow-up question asking nonpartisan respondents whether they currently “lean” toward either party further confirmed that Rigorous respondents were less partisan. While equal proportions in both surveys replied that they leaned Democratic or Republican, 17 percent of Rigorous respondents volunteered that they had no partisan leaning at all, compared with 11 percent of Standard survey respondents. The partisan balance differs even more noticeably among the hardest-to-reach, where just 23 percent identified as Republican, 33 percent as Democrats, and 19 percent refused any party identification or leaning altogether.

Table 5. Political Ideology and Party Affiliation

	Standard %	Rigorous %	Hardest-to-Reach %
Party Identification			
Republican	32	26	23
Lean Republican	13	14	14
Independent, Refused to lean	11	17	19
Lean Democrat	13	13	11
Democrat	31	30	33
	100	100	100
2002 House Vote ^a			
Republican candidate	47	44	39
Democratic candidate	40	42	46
Other/Independent	2	2	3
Don't know/Refused	11	12	12
	100	100	100
Ideology			
Conservative	39	35	35
Moderate	36	40	37
Liberal	21	18	20
Don't know/Refused	4	7	8
	100	100	100
Opinion of Democratic Party			
Favorable	54	58	55
Unfavorable	38	32	33
Don't know/Refused	8	10	12
	100	100	100
Opinion of Republican Party			
Favorable	57	56	54
Unfavorable	34	35	33
Don't know/Refused	9	9	13
	100	100	100
Difference between Parties?			
Great deal	29	27	26
Fair amount	49	49	50
Hardly any	20	19	18
Don't know/Refused	2	5	6
	100	100	100
	(N = 1,000)	(N = 1,089)	(N = 494)

NOTE.—Items in boldface indicate figures that are significantly different from the Standard survey at $p \leq .05$.

^aBased on those who report having voted.

While this cross-survey difference in party composition was statistically significant in the 2003 study, there are three important caveats to keep in mind. First, the original 1997 study found no comparable gaps in partisan balance across surveys. In fact, in 1997 there were slightly more nonpartisans in the Standard than the Rigorous survey (40 percent versus 37 percent), and there was no difference in the partisan balance between Republicans and Democrats. Second, as reported earlier, there are geographic disparities in the sample that may be related to these patterns. Both the Standard and the Rigorous samples overrepresented rural areas, where Republicans are more prevalent, while the hardest-to-reach were more evenly distributed geographically. And third, other measures of partisanship in the 2003 study elicited much smaller cross-survey differences. The parties received similar favorability ratings in the two surveys, and the difference between the Standard and Rigorous estimates of the 2002 congressional vote was not statistically significant. In fact, when the survey estimates of the 2002 House vote are compared with the actual outcome on Election Day, all three are well within two standard errors of the national parameter—a 4.6 percentage point Republican victory.⁶

POLITICAL ATTITUDES

Across a range of other social and political topics, participants in the Standard and Rigorous samples were similar in their attitudes and values. There were virtually no differences in opinion toward government, the poor, business, and other issues (table 6). Moreover, respondents in the Standard and Rigorous samples, and the hardest-to-reach respondents, differed very little in attitudes about immigrants or about whether the Islamic religion is more encouraging of violence. The hardest-to-reach were, however, less likely to believe that African-Americans who “can’t get ahead in this country are mostly responsible for their own condition.”

Hardest-to-reach respondents were slightly less likely to hold favorable opinions about Jews, African-Americans, Muslims, and Asians, compared with the Standard sample. But they did not hold more *unfavorable* opinions. Instead, they were slightly less likely to offer an opinion. Compared with the Standard sample, the ratio of positive-to-negative views among those expressing an opinion was not significantly different for the hardest-to-reach.⁷

6. Election results available at <http://www.fec.gov/pubrec/fe2002/hseparty.htm> (accessed June 3, 2006).

7. When this experiment was first conducted in 1997, the hardest-to-reach respondents reported less favorable views of minority groups than did more willing respondents, a result, in part, of the fact that the hard-to-reach were disproportionately interviewed by white interviewers. In 2003 roughly equal proportions of the Standard survey respondents and the hardest-to-reach were interviewed by white interviewers (73 percent and 69 percent, respectively). Even when the analysis of the 2003 data is restricted to white respondents only, there is no evidence that the least amenable and available respondents were more hostile toward minority groups.

Table 6. Political Attitudes

	Standard %	Rigorous %	Hardest-to-Reach %
Political and Social Issues			
Government wasteful	48	48	50
Poor have it easy	34	35	35
Business profits unfair	51	53	53
Elected leaders care	33	33	34
Accept homosexuality	47	51	48
Ban dangerous books	45	44	43
Protect gun owners	42	42	37
U.S. single world leader	12	11	10
Race, Immigration, and Islam			
African-Americans mostly responsible for own condition	64	62	58
Immigrants burden on United States	44	44	43
Islam encourages violence	35	37	36
Most Muslims are anti-American	24	24	24
	(<i>N</i> = 1,000)	(<i>N</i> = 1,089)	(<i>N</i> = 494)

NOTE.—Items in boldface indicate figures that are significantly different from the Standard survey at $p \leq .05$.

LIFESTYLES

There were almost no significant differences between the Standard and Rigorous samples in responses to several questions about personal behaviors, family finances, health status, and church attendance (table 7). Roughly equal numbers of the easy and hardest-to-interview use the Internet, attend church every week, watch reality shows on television, or smoke cigarettes. Similar numbers describe their family as “struggling” and report not having had enough money for food or health care at some point in the past year. Difficult to reach respondents, however, were significantly more likely to report going out three or more days a week in the evening (49 percent), which is when survey organizations conduct most of their interviewing. That compares with 45 percent of the Rigorous sample and 42 percent of the Standard sample.

POLITICAL ENGAGEMENT AND VOTING

One consistent pattern that the survey experiment revealed is that the hardest-to-reach respondents are somewhat less interested and engaged in politics than those who readily consent to an interview (table 8). A majority in the Standard sample (56 percent) say they voted in 2002, compared with just 48 percent in the Rigorous sample and 46 percent among the hardest to interview; as with

Table 7. Lifestyle Measures

	Standard %	Rigorous %	Hardest-to-Reach %
Go out 3+ nights/week	42	45	49
Exercised yesterday	40	40	44
Attend church weekly	37	36	38
Goes online	67	65	67
Never go out	15	14	13
Not enough money for food	16	17	18
Not enough money for health care	25	21	25
Watch reality TV	33	35	34
Watch late night TV	33	32	32
Listen to religious radio	31	32	35
Excellent/good health	80	78	80
Currently smokes	26	23	22
Ever smoked marijuana	38	37	39
	(<i>N</i> = 1,000)	(<i>N</i> = 1,089)	(<i>N</i> = 494)

NOTE.—Items in boldface indicate figures that are significantly different from the Standard survey at $p \leq .05$.

Table 8. Measures of Political Engagement

	Standard %	Rigorous %	Hardest-to-Reach %
Voting Behavior	%	%	%
Voted in 2002 election	56	48	46
Registered to vote	72	67	64
Always/nearly always vote	65	63	61
Political Knowledge			
Knows Republican Party controls Senate	63	62	57
Media Attentiveness			
Read newspaper yesterday	39	39	41
Watched TV news yesterday	64	63	63
Listed to radio news yesterday	44	42	42
Regularly watches “O’Reilly Factor”	23	27	25
Attitude Item Nonresponse			
Mean number of “don’t know” answers^a	1.22	1.33	1.56
	(<i>N</i> = 1,000)	(<i>N</i> = 1,089)	(<i>N</i> = 494)

NOTE.—Items in boldface indicate figures that are significantly different from the Standard survey at $p \leq .05$.

^a Based on 19 attitudinal questions.

nearly all surveys of this type—including the federal government’s postelection voter survey—all of these estimates exceed the proportion of adults who actually voted (Bernstein, Chadha, and Montjoy 2001; Traugott 1989). Those most difficult to interview were also less likely than those in the Standard sample to say they are registered to vote (64 percent versus 72 percent in the Standard survey), and to know that the Republican Party has a majority in the U.S. Senate (57 percent versus 63 percent).

Respondents in the Rigorous survey, as well as those who were most difficult to reach, had higher levels of item nonresponse, which may also indicate political disengagement. On 19 questions measuring political attitudes, the mean number of “no opinion” responses was 1.22 for respondents in the Standard survey. The average among the hardest-to-reach was 1.56. While statistically significant, the difference is modest in real terms—if asked 100 attitudinal questions, the Standard survey respondents will decline to give a response an average of 6.4 times; average nonresponse for the Rigorous respondents would be 8.2 items.

Further Evidence about Political Engagement and Nonresponse from an Analysis of Terminated Interviews

Survey nonresponse consists not only of people never reached and those who refuse to participate but also those who start the survey and then quit before the interview is complete. Terminated interviews are not an insignificant part of overall nonresponse. Out of 2,626 respondents who began interviews in both the Standard and Rigorous samples, 282 (or 11 percent) broke off the interview at some point prior to the last question. Of these, our interviewers were subsequently able to persuade 115 to complete the survey, while 167 were never completed. While the standard Pew survey typically makes at least one attempt to recontact and complete survey terminations—and succeeds in converting at least some of them—the analysis below compares people who ever broke off the interview with those who completed the interview uninterrupted in order to gauge whether respondents with a tendency to break off are different in any way from those who demonstrate full cooperation. A number of key demographic and attitudinal questions were placed near the beginning of the survey so that important characteristics of the break-off cases might be captured before they terminated. A subset of questions asked of a minimum of 191 people who broke off was retained for this analysis.

The most notable characteristic of people who initially consent to an interview but then quit is their greater tendency to answer “don’t know” to the attitude questions and in general to be more disengaged from public affairs. On 10 of the 15 comparable attitude measures, people who terminated the interview were at least twice as likely to say “don’t know” as those who finished

the survey. In terms of political attentiveness, nearly two-thirds (65 percent) of those who completed the interview without interruption were able to identify the Republican Party as the party holding a majority in the Senate, compared with 55 percent of those who terminated their interviews. In separate analysis of data from a large ($N = 3,000$) media consumption survey, Kennedy, Funk, and Keeter (2006) found that break-off cases were significantly less likely than completed cases to report reading, watching, and listening to the news even after controlling for differential item nonresponse.

While break-offs were less interested and engaged, there were relatively few differences in the balance of opinion between people who did and did not terminate. Table 9 compares demographic and attitudinal survey responses from those who terminated the interview at least once with those from respondents who never broke off; “don’t know” responses are excluded from these tabulations. Demographically, older and less educated respondents were

Table 9. Characteristics of Respondents Who Broke Off and Those Who Did Not

	Never Broke Off %	Broke Off ^a %
Demographics		
Age 65+	18	25
High school education or less	41	61
Male	44	48
Attitudes		
Most people can be trusted	37	27
Islam doesn’t encourage violence	59	46
Government is wasteful	49	43
Racism keeping African-Americans from getting ahead	26	28
Don’t censor books in school libraries	56	58
Immigrants strengthen country	49	48
Accept homosexuality	50	52
Elected officials don’t care	64	63
Businesses make fair profits	46	41
U.S. should be most active nation	40	36
Poor people have hard lives	64	64
Control gun ownership	57	62
	($N = 1,741-2,328$)	($N = 125-282$)

NOTE.—Figures based on unweighted data and recalculated with “don’t know/refused” responses excluded. Items in boldface indicate that those who broke off are significantly different from those who never broke off at $p \leq .05$.

^aFigures for “Broke Off” include two groups of respondents: (1) those who terminated and never completed the survey ($n = 167$) and (2) those who terminated but were later recontacted and completed the full interview ($n = 115$).

significantly more likely to terminate the interview. Respondents who broke off the interview were significantly more likely to believe that “you can’t be too careful in dealing with people” and that “the Islamic religion is more likely than others to encourage violence among its believers.” No significant differences were observed with respect to a range of other social and political issues.

These results are consistent with the notion that the willingness of potential respondents to participate in a survey is related, at least in part, to the perceived content of the survey (Groves, Presser, and Dipko 2004; Groves, Singer, and Corning 2000). People who are uninterested in a survey topic, or who have not thought about the topic enough to have views to express, are significantly more likely to terminate the interview prematurely when the subject matter becomes apparent to them. With extensive effort, roughly 4 in 10 of these terminations can be convinced to complete the interview at a later time. Without this effort, a small but potentially important bias in measures of citizen engagement may be introduced.

The Cost and Benefit of a Rigorous Design

In evaluating the relative validity of the Standard and Rigorous surveys, it is worth considering the practical limitations of the Rigorous survey. Not only does such a design require an extended data collection period, which may undermine the validity of measures of attitudes that are changing during the field period, but the overall costs in terms of fieldhouse effort raise questions about the value of the additional effort required, relative to the benefits in possible reduction of error or of alternative uses of the resources expended. As noted in table 1, the 2003 Rigorous study achieved a response rate of 50 percent, down from 61 percent in 1997. This decline occurred despite the application of more than twice the effort in terms of fieldhouse resources in the 2003 study. Where the 1997 Rigorous study completed 1,201 interviews after making 31,385 calls, the 2003 Rigorous study completed 1,089 interviews after 72,485 calls. This decline in productivity is consistent with previous research showing that the number of calls needed to achieve a given response rate has increased dramatically in recent years (Brick et al. 2003; Curtin, Presser, and Singer 2000).

The need for greater effort in data collection is partially a consequence of a lower eligibility rate for RDD sample numbers, though most ineligible numbers are detected on the first call and thus account for only a small portion of the extra effort required in the 2003 study.

The combination of growing respondent inaccessibility and resistance to surveys is requiring a greater level of fieldhouse effort to achieve response rates comparable to those of just a few years ago. Even among households that cooperated with the survey request, interviews were more difficult to achieve

in 2003. The mean number of calls placed to telephone numbers where an interview was completed increased significantly from 2.99 (SE = 0.08) calls in 1997 to 3.40 (SE = 0.09) in 2003.

The costly effort of the Rigorous survey design to deal with these problems is not without benefits. The Rigorous design allows the fieldhouse to resolve a number of ineligible telephone numbers that would otherwise count against the overall response rate had this effort not been made. More important, fully 17 percent of interviews in the 2003 Rigorous study were completed only after 20 or more attempts had been made. Just 7 percent of completed interviews in the 1997 study required this much fieldhouse effort.

The question is whether these “hardest-to-reach” respondents provide a significant enough improvement in data quality to merit the costs necessary to gain their cooperation. There is substantial evidence beyond the current study that pursuing hard-to-reach respondents provides only limited, if any, benefit in terms of the demographic or political representativeness of the overall survey sample (Craighill and Dimock 2005).

Summary and Discussion

As in the 1997 study, the 2003 Pew experiment found few significant differences in estimates produced by its Standard survey and by one employing more rigorous techniques aimed at obtaining a high rate of response. In terms of objective measures such as demographic characteristics and certain personal behaviors, the findings are generally reassuring. When compared with government benchmarks, the demographic and social composition of the sample in the Standard survey was quite representative on most measures.

There is also reassurance in most of the findings regarding the measurement of attitudes. The relatively small magnitude of the differences between the Standard sample, the Rigorous sample, and the subset of hardest-to-reach cases provides evidence that, within the limits of the experimental conditions, nonresponse did not introduce substantial biases into the estimates. And, of course, judged by their accuracy in forecasting voter behavior on Election Day 2004, political surveys conducted using methods similar to Pew’s Standard survey just before an election continue to be highly valid.⁸

But important limitations and caveats must also be noted. The absence of large differences between estimates from the Standard and Rigorous surveys does not, in itself, demonstrate that the Standard survey is unbiased. With a Rigorous survey response rate of just 50 percent, much of the population remained beyond our view in this experiment. Comparisons with national

8. For a review of the accuracy of preelection polling in recent election cycles, see analyses by the National Council on Public Polls: <http://www.ncpp.org>. Results from the final 2004 preelection surveys of major polling organizations are available at http://www.realclearpolitics.com/bush_vs_kerry.html (accessed June 4, 2006).

parameters are reassuring, but they do not cover all of the topics addressed in this survey or in similar polls. With respect to measures of political engagement, the results are less encouraging—though even here, the apparent nonresponse bias is not especially large. For this study, which prospective respondents are told is an “opinion survey for leading newspapers and TV stations around the country,” the hardest-to-reach and those who begin but break off the interview are less engaged politically than those who cooperate and finish the interview. But we cannot conclude from this experience that all surveys, regardless of topic, will overestimate political engagement. Empirical tests of leverage-salience theory (Groves, Singer, and Corning 2000), particularly with respect to the prominence of survey topic (Groves, Presser, and Dipko 2004), suggest that the nature of nonresponse bias is apt to be dependent on what topics are made salient when the survey is introduced to potential respondents.

Accordingly, even the encouraging findings of the present study cannot necessarily be generalized to surveys introduced as focusing on other, nonpolitical topics. Just as a survey introduced as focused on politics may overstate engagement in politics, so too might surveys introduced as focusing on travel, cooking, or religion overstate engagement or facility with those subjects. One practical recommendation from our research might be that the additional effort undertaken in a Rigorous study be reserved for situations in which there is a strong theoretical expectation that the level of interest in the survey topic is likely to lead to nonresponse bias on key measures in the study.

References

- American Association for Public Opinion Research. 2005. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Lenexa, KS: AAPOR.
- Bernstein, Robert, Anita Chadha, and Robert Montjoy. 2001. “Overreporting Voting: Why It Happens and Why It Matters.” *Public Opinion Quarterly* 65:22–44.
- Blumberg, Stephen J., and Julian V. Luke. 2006. “Wireless Substitution: Preliminary Data from the 2005 National Health Interview Survey.” National Center for Health Statistics. Available online at <http://www.cdc.gov/nchs/products/pubs/pubd/hestats/wireless/wireless2005.htm> (accessed June 3, 2006).
- Brick, J. Michael, David Martin, Patricia Warren, and John Wivagg. 2003. “Increased Efforts in RDD Surveys.” Paper presented at the annual meeting of the American Association for Public Opinion Research, Nashville, TN.
- Craighill, Peyton, and Michael Dimock. 2005. “Tough Calls: Potential Non-Response Bias from Hard-to-Reach Respondents.” *Public Opinion Pros*, Aug.
- Curtin, Richard, Stanley Presser, and Eleanor Singer. 2000. “The Effects of Response Rate Changes on the Index of Consumer Sentiment.” *Public Opinion Quarterly* 64:413–28.
- . 2005. “Changes in Telephone Survey Nonresponse over the Past Quarter Century.” *Public Opinion Quarterly* 69:87–98.
- Dimock, Michael, Nilanthi Samaranayake, and Jonathan Best. 2005. “Are We Under-Counting Blue Counties? Correcting for Disproportional Response Rates at the County Level.” Paper presented at the annual meeting of the American Association for Public Opinion Research, Miami Beach, FL.
- Groves, Robert M. 2006. “Nonresponse Rates and Nonresponse Bias in Household Surveys.” *Public Opinion Quarterly* 70:646–75.

- Groves, Robert M., Stanley Presser, and Sarah Dipko. 2004. "The Role of Topic Interest in Survey Participation Decisions." *Public Opinion Quarterly* 68:2–31.
- Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration." *Public Opinion Quarterly* 64:299–308.
- Holbrook, Allyson, Alison Pfent, and Jon Krosnick. 2003. "Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms." Paper presented at the annual meeting of the American Association for Public Opinion Research, Nashville, TN.
- Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser. 2000. "Consequences of Reducing Nonresponse in a Large National Telephone Survey." *Public Opinion Quarterly* 64:125–48.
- Kennedy, Courtney, Carolyn Funk, and Scott Keeter. 2006. "Survey Terminators: They Won't Be Back But Does It Matter?" Paper presented at the annual meeting of the American Association for Public Opinion Research, Montreal, Canada.
- Smith, Tom W. 1983. "The Hidden 25 Percent: An Analysis of Nonresponse on the 1980 General Social Survey." *Public Opinion Quarterly* 47:386–404.
- Traugott, Santa. 1989. *Validating Self-Reported Vote: 1964–1988*. ANES Technical Report Series, no. nes010152. Available online at http://www.electionstudies.org/resources/papers/technical_reports.htm (accessed October 29, 2006).
- U.S. Department of Education, National Center for Education Statistics. 1997. "An Overview of Response Rates in the National Household Education Survey: 1991, 1993, 1995, and 1996." NCES 97–948, by J. Michael Brick, Mary Collins, and Kathryn Chandler, Washington, DC: U.S. Department of Education.