

Consequences of Reducing Nonresponse in a National Telephone Survey

Scott Keeter, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser
January 15, 2000

Scheduled for publication in *Public Opinion Quarterly* in summer or fall 2000.

Contact author: Scott Keeter, Dept. of Public and International Affairs, George Mason University, Fairfax, VA 22030. Phone (703) 993-1412. Fax (703) 832-0209. E-mail skeeter@gmu.edu.

Scott Keeter is professor of government and politics and chairperson of the Department of Public and International Affairs at George Mason University. Carolyn Miller is an independent research consultant. Andrew Kohut is director of the Pew Research Center for the People and the Press. Robert M. Groves is professor of sociology and research scientist at the Institute for Social Research at the University of Michigan, and director of the Joint Program in Survey Methodology. Stanley Presser is professor of sociology and director of the Survey Research Center at the University of Maryland. The authors wish to thank Darlene Montemarano for excellent research assistance, Mark Schulman and the survey firm of Schulman, Ronca, and Bucuvalas, Survey Sampling, Incorporated, John Brehm, and Greg Flemming and Kim Parker of the Pew Research Center for the People and the Press. The questionnaire and data for this study are available from the Pew Research Center web site at <http://www.people-press.org/resprpt.htm>.

Abstract

Critics of public opinion and media polls often claim that methodological shortcuts taken by pollsters to collect timely data produce biased results. This study compares two RDD national telephone surveys that used identical questionnaires and the same interviewing facility but with very different levels of effort: a “Standard” survey conducted over a five day period that used a sample of adults who were home when the interviewer called, and a “Rigorous” survey conducted over an eight week period that used random selection from among all the adult household members. Response rates for the two studies, computed according to AAPOR guidelines, were 60.6% for the Rigorous and 36.0% for the Standard study. The Rigorous survey achieved both a higher contact rate (92.0% vs 68.5%) and a higher cooperation rate (73.7% vs 58.1%) than the Standard survey. Despite these differences, the two surveys produced similar results. Across 91 comparisons, no difference exceeded 9 percentage points and the average difference was about 2 percentage points. Most of the statistically significant differences were among demographic items. Very few significant differences were found on attention to media and engagement in politics, social trust and connectedness, and most social and political attitudes, including even those towards surveys. The most notable finding from an analysis of the effects of the number of calls it took to contact a household and the household’s initial willingness to be interviewed was that respondents in reluctant households were less favorable toward racial minorities than were those in amenable households. Additional analyses suggested that this difference was partly the result of differences in the interviewer groups conducting the two surveys.

Introduction

In extolling the virtues of the sample survey, Sidney Verba has written that “Surveys produce just what democracy is supposed to produce - - equal representation of all citizens. The sample survey is rigorously egalitarian; it is designed so that each citizen has an equal chance to participate and an equal voice when participating” (Verba 1996, 3). Verba acknowledges that the people interviewed in surveys are not truly random, but sees surveys as much closer to the egalitarian ideal than those in any other venue from which citizens can be heard.

Nevertheless, there is considerable skepticism about the representativeness of contemporary opinion polls. Such concerns range from the dismay expressed by scholars and practitioners regarding declining response rates, to fears by liberals that the poor and dispossessed are often “out of frame” and thus out of mind, to the pointed charge that “conservatives are more likely than others to refuse to respond to polls, particularly those polls taken by media outlets that conservatives consider biased” (Barone 1997).

From a methodological vantage point, it would not be surprising if respondents to media and political polls were unrepresentative, because the need for timely results demands that the polls have very short data collection periods. Short time spans not only make high response rates difficult to attain, but are also partly responsible for the use of nonrandom within-household selection methods, such as choosing respondents from among those who are home when the interviewer calls. But although there are reasons to expect both low response rates and nonrandom selection methods to produce unrepresentative respondent pools, there is little scientific evidence on these issues.

It is well known, of course, that nonresponse error is a function of both the nonresponse rate and the difference between respondents and nonrespondents on the statistic of interest. High nonresponse rates can still yield low nonresponse errors (if the difference between respondents and nonrespondents is small), and low nonresponse rates can yield high nonresponse errors (if respondents and nonrespondents differ dramatically). Although 100% response rates obviously eliminate nonresponse error, it is not clear what reduction of error results from moving between other response rates, e.g., moving from 50% to 70% or from 30% to 50%.

While there is evidence that those not responding to some election surveys are less interested in politics (Couper 1997; Brehm 1993), and that even high response rate surveys can systematically miss certain kinds of households (Groves and Couper 1998), methodologists have not yet developed theories that are capable of predicting when nonresponse rates imply nonresponse error and when they do not. One viewpoint assumes that efforts to increase response rates will gradually reduce nonresponse bias by bringing into the respondent pool more diverse groups of respondents, obtaining, as a function of the response rate, a progressively more balanced representation of the entire survey population. This implies that distributions of survey variables change as the response rate increases.

Another viewpoint assumes there is a set of characteristics that affect the survey participation decision (e.g., topic, burden, sponsorship, interviewer behavior), and that persons for whom those attributes are not attractive will be nonrespondents. Thus as the response rate increases, the respondent pool is being increased by more of those persons located and contacted for whom the survey conditions are minimally acceptable. The rest remain nonrespondents. This implies that survey statistics stay relatively fixed as the response rate increases.

Compared to changes in response rate, the effect of nonrandom selection methods, in

particular choosing from among those at home when the interviewer calls, might seem straightforward. Yet the consequence for a survey statistic will depend not only on the relationship between the statistic and time spent at home, but also on the nature of the survey's calling pattern, as well as the association between time spent at home and willingness to be interviewed. Thus, as with response rates, the absence of empirical studies makes it difficult to gauge the effect of nonrandom selection within households.

To assess the impact of low response rates and nonrandom within-household selection methods, we undertook an experiment in which identical questionnaires were administered by the same survey firm in two different studies: one using "Rigorous" procedures; the other "Standard" procedures. Both studies used a list-assisted random digit dial sample of the continental United States, though the Rigorous frame included numbers in banks with one or more published listings, whereas the Standard frame included numbers in banks with three or more listings. As less than one percent of the Rigorous interviews were from banks with one or two listings, this difference does not figure importantly in any of our results.

The "Standard" study was designed to complete 1,000 interviews in five days. Using a protocol typical of Pew Center surveys every number was called a minimum of five times during the period consisting of Wednesday and Thursday evenings, June 18-19, 1997, and the mornings, afternoons, and evenings of Friday through Sunday, June 20-22, 1997. Interviewers asked to speak to the youngest adult male who was home (and, if there was no male present, the oldest adult female who was home). One follow-up call was made to households that refused. The response rate at the end of the 5 days was 36.0%¹.

The "Rigorous" study began at the same time but was conducted over eight weeks, June 18-August 12, 1997, employing a much more exhaustive effort to locate and interview individuals difficult to reach during a short field period and those initially reluctant to participate. Respondents in the Rigorous study were selected randomly from among all adults who lived in the household. In addition, in order to increase cooperation, households with listed telephone numbers were sent an advance letter that included a \$2 bill. The response rate was 60.6%.²

¹ We use the rate referred to as RR3 in the AAPOR guidelines for the computation of response rates. $RR3 = \text{Completed interviews} / ((\text{completed interviews} + \text{partial interviews}) + (\text{refusals} + \text{noncontacts} + \text{other}) + e(\text{unknown if household or occupied} + \text{unknown other}))$, where "e" is an estimate of the proportion of unknown outcomes that are eligible. We set e on the basis of a comparison of outcomes after five days with outcomes after the entire field period of nearly two months. After five days, 319 cases had a disposition of no answer on all attempts. Of these, 319 (19.7%) were identified as possibly households on subsequent attempts. Thus we chose e to be .20.

Response rates for most media polls are generally not made public, but there is some evidence that the rate for the Standard survey is not unusual (Brady and Orren 1992).

² Interviewing was conducted by the survey research firm of Schulman, Ronca, and Bucuvalas, Inc., from their facility in New York. Interviewers who worked on this project were blind to the purpose of the experiment, but were aware that two surveys with different respondent selection methods were being conducted simultaneously with the same questionnaire. A single pool of interviewers was trained on the questionnaire and the different respondent selection procedures. These interviewers were used interchangeably on the two studies. At the beginning of each shift interviewers were randomly assigned to

Many of our analyses compare the Standard five day study and the Rigorous study. But we also take advantage of the fact that data collection on the Standard study continued until July 23 in an effort to maximize the response rate in this study as well. After the first five days of both studies, households that refused (and for which an address was available) were sent a letter appealing for cooperation, and two more efforts were made to persuade most refusals (both listed and unlisted) to consent to an interview. Table 1 shows the overall contact, cooperation, and response rates for the Standard Five Day, Standard Complete, and Rigorous studies, along with complete call disposition information.

*** Table 1 about here ***

Findings

Overview

The questionnaire included a wide range of topics often found on opinion polls of the U.S. public. A special effort was made to include items on which we might expect to find differences between amenable and reluctant respondents, and between the easy-to-reach and those who are more difficult to contact (DeMaio 1980; Goudy 1976; Goyder 1987; Stinchcombe, Jones, and Sheatsley 1981). The survey asked 96 items taken from studies conducted by the Pew Center or other major national polls. These included:

- 34 Political and social opinion items;
- 5 Electoral behavior measures;
- 8 Media use items;
- 3 Knowledge items;
- 11 Social integration measures (7 objective and 4 subjective);
- 4 Crime-related items;
- 4 Items about polling;
- 23 Demographic characteristics; and
- 4 Interviewer ratings of the respondent.

Five of the political items were time-sensitive (e.g., presidential approval) whose distributions changed in other surveys during the extended field period for the Rigorous survey (between June and August 1997). After removing these five, 91 items were left for which we could compare Rigorous and Standard estimates.

Table 2 shows that 14 of the 91 distributions were significantly different ($p < .05$) between the Standard 5 Day and Rigorous studies: seven demographics, five opinion items, one interviewer rating, and one social integration measure (a behavior). But the differences were

either the Standard or the Rigorous survey and would stay on that survey for the entire shift. Due, in part, to the comparatively short field period for the standard survey, and the long field period for the rigorous survey, only 60 of 160 interviewers actually worked on both. As we note later in the paper, the interviewing teams for the two studies differed somewhat on race and gender. For the Standard Five Day survey, 18% of interviews were conducted by white males, 30% by non-white males, 3% by white females and 49% by non-white females. For the Rigorous survey, 34% of interviews were conducted by white males, 28% by non-white males, 5% by white females and 33% by non-white females.

generally small, none exceeding 9 percentage points.³

As one way to test whether the small differences between the studies were systematic, we combined individual items to produce a three-item media use scale, a three-item knowledge scale, a six-item conservative/liberal scale, a five-item social activity scale (using the objective social integration measures) and a three-item trust scale (using the subjective social integration measures). None of the scale distributions differed between the Rigorous and Standard 5 Day studies (p value for the trust scale was .11; for the rest the p value was greater than .48).

*** Table 2 about here ***

Demographic differences

Demographic variables provide an especially useful basis of comparison between the Standard and Rigorous surveys. Demographic characteristics are intrinsically important since they are related to many attitudes and behaviors studied in public opinion surveys. Moreover, for many demographics, we have independent, more accurate information about the population parameters and thus can judge not only how the two surveys differ but which is closer to reality (as defined and documented by official government statistics).

Seven of the questionnaire's 23 demographic items differed significantly ($p < .05$) between the Rigorous and Standard 5 Day studies: income, home ownership, education, race, having been short of money for clothes in the past year, having been short of money for food in the past year, and listed vs. unlisted phone. Estimates for the first four items are also available for telephone households from the March 1996 Current Population Survey (CPS), and the distributions from all three studies are shown in Table 3a.⁴ Compared to the Rigorous study, a greater proportion of Standard 5 Day respondents were nonwhite, less educated, lower income, and renters. In general, the Standard distributions on these items resemble the CPS more closely than do the Rigorous distributions, but the differences are modest. The average absolute percentage point difference between the CPS and the Standard sample for these four comparisons is 3.1, compared with an average difference of 3.9 between the CPS and Rigorous sample.

*** Table 3 about here ***

Table 3b presents the same comparisons for the six nonsignificant demographic differences between Rigorous and Standard studies for which estimates are also available in the CPS. There is little evidence of differential bias in the Standard and Rigorous surveys on these

³ These are unweighted results. The chi-square tests were conducted without collapsing any of the categories of the variables; "don't know" and "no answer" responses were excluded and are discussed in Table 4. We also carried out the analysis using weighted data. For the standard sample, we followed normal Pew procedure, which is to post-stratify to Current Population Study totals using age by sex, education by sex, education by age, and the marginals of region and race. For the Rigorous sample, we adjusted to those totals and also corrected for differential numbers of phone lines and adults (which are not used in Pew's standard procedure). The differences between the weighted estimates are very similar to those reported in the text based on the unweighted data. (Of course, for demographic variables used in the weighting, differences between the samples were eliminated when the weighting was applied.)

⁴ Estimates from the Current Population Survey are weighted and adjusted, using independent estimates of the civilian noninstitutional population, by the Bureau of the Census.

variables.

Differences in political and social attitudes and Electoral Behavior

We found only the barest support for the charge that Standard procedures - - relative to the Rigorous protocol - - underrepresent conservatives and Republicans. On the one hand, the only five significant differences on political and social issues were a result of respondents in the Standard 5 Day survey giving more favorable ratings (summing “very” and “mostly”) to Labor Unions, Militia Groups, and Blacks, less favorable ratings to the Republican Party, and more endorsement (summing “strong” and “not strong”) to the position that the poor “have hard lives because government benefits don’t go far enough to help them.” On the other hand, none of these differences exceeded 4 percentage points, and respondents in the Standard and Rigorous surveys were nearly identical on party affiliation, self-described ideology, candidates voted for in recent elections, and a host of other political attitudes (Table 2). (Although the difference on party identification is just shy of $p < .05$, the effect stems from somewhat more Standard study respondents saying they were independents; the ratio of Republicans to Democrats does not differ between the studies.)

Media use, Knowledge, and Engagement in Politics

Citizens who are uninterested in politics, who pay little attention to the news, or have little information about current events may decline to participate in media-related surveys, either out of concern that they will appear uninformed or because they do not want to devote the time necessary to participate (Brehm 1993). By this logic, a survey that reaches more reluctant respondents will provide a more accurate view of public opinion. And, presumably, such a survey will reflect lower overall levels of voter participation, attention to the news, and knowledge of politics. Yet, the Rigorous and Standard 5 Day studies generally did not differ on these dimensions. There were essentially no differences between the studies in the percentages who reported reading a newspaper, listening to radio and tv news or talk shows, knowing Bob Dole loaned Newt Gingrich some of the money he needed to pay his ethics fine, knowing Republicans have a majority in the U.S. House, being able to identify Bill Gates, or voting in the 1996 elections. Finally, although the Rigorous survey contained somewhat more computer and Internet users, the differences just missed being statistically significant.

Social Integration

Willingness to participate in surveys is plausibly related to social integration. On the one hand, people low on social integration tend to be less trusting of others, which might discourage participation in surveys. On the other hand, people high on social integration may engage in many more activities and thus either have less time to participate in surveys or spend less time at home and thus be harder to contact. Based on a comparison of the Standard 5 Day and Rigorous surveys, neither theory - - at least as stated in these forms - - receives much support.

There were no differences between surveys on three standard trust items: whether “most people can be trusted” or “you can’t be too careful in dealing with people;” whether “most people would try to take advantage of you if they got the chance” or “would they try to be fair;” and whether “most of the time people try to be helpful” or they are “mostly just looking out for themselves.” Respondents in the two surveys also reported a nearly identical level of interaction

with other people (visited with family or friends on the day before the survey, and called a friend or relative just to talk the prior day); fairly similar numbers of other people the respondent could count on for support; and a virtually identical level of attendance at religious services. Finally, although there was little difference between studies in the percentage reporting having done volunteer work in the past year, significantly more respondents in the Standard 5 Day than in the Rigorous reported doing regular, as opposed to occasional, volunteer work ($p < .05$).

Crime-related Items

Many survey researchers believe that concerns about safety affect some potential respondents' willingness to participate in surveys. Telemarketing fraud is a serious problem, with unscrupulous persons and organizations engaging in unethical or illegal activities under the guise of conducting a survey. In addition, some people may fear that a potential burglar or mugger might employ a survey as a ruse to ascertain whether or not an individual is at home, who else might be in the household, and what property might be available. Accordingly, we might expect that, compared with the Standard survey, the Rigorous survey would have picked up more respondents who were concerned about crime or who had experienced it. Among other differences, the pre-survey letters sent to many of the respondents in the Rigorous survey should have helped ease fears about the legitimacy of the survey (though, of course, such letters went only to people whose phone numbers were listed with an address in the telephone directory).

But the extent of fear of neighborhood crime, and the familiarity or experience with criminal victimization in the neighborhood, did not differ significantly between the Standard 5 Day and Rigorous surveys. Nor was there a significant difference between the surveys in the percentage of respondents who reported that they keep a firearm in their house.

Attitudes toward surveys and Interviewer Ratings

We might expect more reluctant respondents to have more skeptical attitudes about surveys in general, and perhaps to be less likely to have ever participated in a public opinion poll before. Yet despite these expectations, there were no statistically significant differences between the surveys in responding that polls work for the best interests of the general public, and that a random survey of 1,500 or 2,000 people can accurately reflect the views of the nation's population. Nor were there significant differences in reporting having been interviewed in a poll in the past, or expressing a willingness at the end of the interview to do the interview if they had the choice again.

While respondents in the two surveys reported having similar attitudes toward surveys, the interviewers themselves differed in their views of the respondents. More respondents in the Standard survey received ratings of "very high" or "above average" interest than in the Rigorous survey. Indeed, this difference, 9 points, was the largest we found.

Effects of Accessibility versus Amenability

Survey nonresponse is largely the result of two broad problems: some sample persons are relatively inaccessible to the surveyor, unamenable to cooperating, or both. Our results demonstrate the effect of efforts to deal with these problems on rates of nonresponse. In the five day field period of the Standard survey, contact was made with 68.5% of telephone numbers assumed to be working residential phones. By the end of the Standard study, the contact rate rose

to 89.3% (and was 92.0% for the Rigorous study). Similarly, at the end of the five day field period, the cooperation rate for the Standard study was 58.1%. Through the use of such techniques as a letter to households that refused and persistence in attempting to convert refusals, the cooperation rates at the conclusion of the project were 68.9% for the Standard study and 73.7% for the Rigorous study.

While the comparisons between the five day Standard and the full Rigorous studies provide a “bottom line” measure of the impact of survey nonresponse, it is possible to gauge the consequences of accessibility and amenability more directly by examining each phenomenon separately. Therefore we shall compare the accessible with the inaccessible and the amenable with the reluctant.

Amenability

Before proceeding to the analysis of amenability, a brief discussion of method is warranted. We will be combining across the Standard and Rigorous samples and comparing data from households that never refused with households where one or two refusals occurred. For all cases except the one-person households, we do not know whether the same person refused and later granted a completed interview; in many instances different people will have been involved. Accordingly, we refer to “amenable households” and “reluctant households.”

It comes as no surprise that compared to those from reluctant households, respondents from amenable households were rated by interviewers as more cooperative and more interested in the interview and more likely to say, having completed the survey, that they would agree to do such a survey again. Aside from these almost tautological effects, the remaining 88 items show only about the number of statistically significant differences ($p < .05$) between amenable and reluctant households that would be expected by chance (5 of 88): Two demographics (listed/unlisted status and employment status) and three opinion items. But two of the three opinion items are about blacks--the favorability rating of blacks (that also differs between Standard 5 Day and Rigorous studies) and the judgment about whether racial discrimination is responsible for the condition of blacks (Q4d, which was of borderline significance in the Standard 5 Day versus Rigorous comparison). Moreover, the difference between amenable and reluctant households in ratings of Asians just misses being significant. In each of these cases, fewer reluctant households express sympathetic attitudes toward the minority groups.

Accessibility

In the combined Rigorous and Standard samples, number of calls to first contact with the household is significantly related ($p < .05$) to 20 of our items. Two of the largest associations are with age and education. The young are more difficult to reach, as are the better educated. Those households containing one adult and those with employed respondents also required more calls to first contact. These patterns are well established in the literature (Groves and Couper 1998). Most of the other 16 differences involve items that are strongly related to age and education: 5 media use items, 1 knowledge item (with a second just missing significance), 2 social integration measures, 5 other demographics, 1 interviewer rating (of the respondent’s hearing), and only 2 opinion items.

Analysis of Item Nonresponse

Aside from substantive differences between amenable and reluctant respondents, we might also expect to find differences in ability or willingness to provide responses to questions. Reluctance to participate might translate into reluctance to answer certain questions, either because of an unwillingness to reveal one's opinions or disclose details about one's personal situation, or because of a genuine inability to answer questions. Evidence to support this expectation has been found by Triplett et al., (1996) and by Blair and Chun (1992).

Two approaches were used to measure differences in item nonresponse. First, we computed an index of the number of questions on which the respondent declined to provide a substantive response, and then examined the relation of this index to our measures of amenability and accessibility. As Table 4 shows, the mean item nonresponse for all respondents in both surveys was 2.9 items (out of 89). Mean nonresponse for amenable households was 2.8; for reluctant households it was 3.2, a difference that just missed statistical significance ($p < .08$). A greater difference was seen when reluctant households were divided according to whether there had been one or two refusals. Single refuser households had a mean item nonresponse of 3.0, while the mean for double refusers was 3.5. The difference between double refusers and amenable households met the conventional test for statistical significance ($p < .05$). When accessibility is also considered, we find that item nonresponse was actually lower among respondents in households where more than 10 calls were required to obtain an interview.⁵ The lowest level of item nonresponse was found in households where no refusal occurred but more than 10 calls were made.

*** Table 4 about here ***

A second indicator entailed a comparison of item nonresponse percentages for each item in the survey.⁶ Of 85 comparisons on which differences could be computed, the mean absolute difference in nonresponse percentage between amenable and reluctant respondents was 0.9 percentage points, with a range from -4.2% (a minus indicating that reluctant respondents had more nonresponse than amenable respondents) to +3.9%; the median difference was 0.5 percentage points. Of the 85 comparisons, 16 were significant at the .05 level, though it should be noted that 7 of these 16 involved percentage differences smaller than 0.5 percent.⁷ More of the differences favor the hypothesis that nonresponse will be greater in reluctant respondents than in amenable ones, but the difference is quite minimal.

Multivariate Ransacking

Following the analyses above, which were to some extent guided by the past literature on patterns of nonresponse error in surveys, we constructed two approaches at statistical ransacking of the data. The purpose of these exercises was to leave no stone unturned in an effort to

⁵This is almost certainly a consequence of the fact that inaccessible respondents tended to have higher levels of education.

⁶The three political knowledge items were not included in this analysis, since nonresponse has a direct substantive interpretation on these questions.

⁷In some of these comparisons, 4 or 5 of the reluctant respondents refused to answer compared with none of the amenable respondents.

discover patterns of relationships that might reveal differences between the two surveys.

Both ransacking approaches set up a dependent variable that was coded “1” if the respondent case was from the Rigorous survey and “0” if it was from the Standard 5 Day survey. We first used stepwise OLS regression procedures to locate single variables that were highly correlated with that dependent variable (comforted by the fact that the near 50-50 split of the dependent variable would not threaten homoskedasticity assumptions too violently and using $p < .10$ cutoffs for inclusion of a predictor). In essence, this asked the question of what groups were most differentially distributed between the two samples. The second ransacking procedure used a stepwise interaction detection procedure (CHAID in SPSS, with $p < .10$ cutoffs and cell size minima of 100) that sequentially broke the respondent pool into groups that were maximally different in the percentages in the two surveys. One can think of the first procedure as identifying main effects, and the second as identifying interaction effects, on the likelihood of being a respondent under the two protocols.

The forward stepwise regression identified interviewer experience (number of months) as the single most powerful predictor, with more experienced interviewers more likely to conduct interviews in the rigorous study. The second predictor was interviewer race. In short, more experienced, white interviewers tended to conduct the rigorous interviews and less experienced, black interviewers conducted the standard.

The CHAID analysis produced results quite compatible with the OLS regression analysis. Interviewer experience, race, and sex were the dominant predictors. After these were entered into the model, the only substantive variable was opinion of the Republican party, but it was a significant discriminator between the standard and rigorous samples for only a small subset of interviewers.

The surprise from this exercise was that it revealed interviewer staffing differences in the two surveys. Further investigation revealed that only about 60 of the approximately 160 interviewers worked on both the standard and rigorous surveys. White male interviewers did a much higher proportion of the rigorous survey interviews than they did of the standard.

The reader will recall that less racially liberal answers were provided by respondents that came into the data set after refusal conversions. Race of interviewer effects on racial attitude questions have been well documented (e.g., Schuman and Converse 1971; Hatchett and Schuman 1975), and were observed on racial attitude questions here.⁸ For that reason, we fit a set of logistic models that predicted the two racial attitude items that showed a significant amenability effect (opinion toward blacks and agreement with the statement that racial discrimination is the main cause of blacks not getting ahead). Bivariate logit models again showed significantly fewer racially “liberal” responses for those interviewed after a refusal conversion than for those who provided the interview with less reluctance. Multivariate logit models revealed that when statistical controls for interviewer race and gender were introduced, the original relationships

⁸In the Rigorous sample, white respondents interviewed by black interviewers were 9 percentage points more likely than those interviewed by white interviewers to say they had a “very favorable” opinion of blacks; in the Standard sample the difference was 6 points. In the Rigorous sample, white respondents interviewed by black interviewers were 16 percentage points more likely than those interviewed by white interviewers to agree that racial discrimination is the main reason why blacks can’t get ahead; in the Standard sample the difference was 7 points.

diminished, though in both cases they were still near the conventional .05 significance cutoff (.08 and .04). Thus the finding that the reluctant respondents were less racially tolerant appears to be at least partly due to interviewer staffing differences.

Cost Comparisons

A final consideration in assessing the results is the relative cost of the Standard and Rigorous surveys. The limited time and resources available to conduct polls make compromises and trade-offs necessary in their design and implementation. Researchers are faced with difficult decisions about the best use of resources to obtain the most accurate and representative snapshot of the public at a given time. Resources must be divided among survey tasks such as questionnaire design, sample design, length of field period, call design, respondent selection and precontact, among others. Given these practical constraints of conducting public opinion polls it is important to understand the cost implications of project design decisions in order to arrive at a reasoned approach to the juggling of competing priorities.

Two direct costs associated with conducting these two surveys can be compared for the purposes of making a general cost comparison across the surveys. First, the direct costs of matching the RDD sample to obtain addresses where possible, the cost of advance letters, incentives, and refusal conversion letters can be estimated. All of these costs were incurred for the Rigorous design only. Second, the interviewing productivity of the two surveys can be compared and translated into another direct cost, interviewing hours.

Comparing the Standard five day survey and the Rigorous survey on these direct costs illustrates the impact of the Rigorous design on the cost of the survey. Table 5 shows the productivity rates (calculated as number of completes divided by number of interviewing hours) for the two surveys during different time periods throughout the data collection. As the field period progressed, the productivity rates dropped off for both surveys, as would be expected. The difference between the two surveys at comparable points in time is not large - - suggesting that the respondent selection and sample differences do not have a big impact on productivity, and that it is mostly the extended field time and call design that lowers productivity.⁹

*** Table 5 about here ***

The overall productivity rate for the Standard five day survey is .99 compared with the overall Rigorous survey productivity of .79. On a survey of 1000 cases, this difference in productivity would translate into a difference of 256 interviewing hours (approximately a 25% increase). In addition, these interviewing hours would also require additional supervisory time, data processing time, CATI time etc. In conjunction with the additional costs of the advance letters and incentives, and refusal conversion letters (approximately \$6,000 for these costs in a 1000 case survey), this represents a significant cost increase of the Rigorous design over the Standard five day design.

Conclusion

⁹The productivity of the two studies during Phase I is similar, but it should be remembered that part of the Rigorous sample received an advance letter and monetary incentive. This may have boosted productivity, offsetting the advantages provided by the at-home respondent selection scheme in the Standard survey.

Rates of participation in sample surveys are one of the most pressing issues in the field. This is true both because of concerns about nonresponse bias in survey estimates, but also because efforts to maintain response rates in the face of growing reluctance to participate greatly increases the costs of surveys. This study addressed a very practical question: “What differences arise in point estimates subject to different response rates?”

Like most experimental designs, this study offers a demonstration conditional on a set of features of the measurement situation. The two surveys achieved different contact, cooperation, and overall response rates. The rigorous achieved a higher response rate with a specific set of features (i.e., advance letters and incentives for listed numbers, repeated callbacks and refusal conversions over an extended time period). The 36.0% vs. 60.6% response rates are substantially different, but do not permit us to make inferences about the effect of other field efforts on nonresponse error. This is partly because the two studies used different within household selection methods. More importantly, the results offer only two observations on the response rate continuum from 0 to 100%, and it would be inappropriate to use the evidence to note that the same results would apply if the five day protocol had been reduced to one or two days, or the rigorous protocol had achieved a response rate of 75%.

The value of the experimental evidence, we believe, will be in stimulating other work attempting to discover under what circumstances and for what measures nonresponse rate differences imply disparate nonresponse errors. We expect that such work requires theory development that links the decision to participate with the purposes of the survey. Past research demonstrating nonresponse error is of that ilk (e.g., membership surveys showing disproportionate response from active vs. inactive members and election surveys that found higher cooperation from the politically active (Brehm 1993). How those mechanisms are manifested in successive waves of effort to measure sample persons, however, has not been well-developed in the literature. This is the work that needs to be done in the future.

References

Barone, Michael. 1997. The Weekly Standard, March 10.

Blair, Johnny and Young I. Chun. 1992. “Quality of Data from Converted Refusers in Telephone Surveys.” Presented at the American Association for Public Opinion Research Meeting, St. Petersburg.

Brady, Henry E. and Gary R. Orren. 1992. “Polling Pitfalls: Sources of Error in Public Opinion Surveys.” In Media Polls in American Politics, ed. Thomas E. Mann and Gary R. Orren. Washington, D.C.: Brookings.

Brehm, John . 1993. The Phantom Respondents: Opinion Surveys and Political Representation. Ann Arbor: University of Michigan Press.

Couper, M.P. 1997. “Survey Introductions and Data Quality,” Public Opinion Quarterly 61: 317-338.

DeMaio, Theresa. 1980. "Refusals: Who, Where and Why." Public Opinion Quarterly 44: 223-233.

Goudy, Willis J. 1976. "Nonresponse Effects on Relationships Between Variables." Public Opinion Quarterly 40: 360-373.

Goyder, John. 1987. The Silent Minority: Nonrespondents on Sample Surveys. Boulder: Westview.

Groves, Robert M., and Couper, M.P. 1998. Nonresponse in Household Interview Surveys, New York: John Wiley and Sons.

Hatchett, Shirley, and Howard Schuman. 1975. "Race of Interviewer Effects Upon White Respondents." Public Opinion Quarterly 39: 523-528.

Schuman, Howard, and Jean Converse. 1971. "The Effects of Black and White Interviewers on Black Responses in 1968." Public Opinion Quarterly 35: 44-68.

Stinchcombe, Arthur L., Calvin Jones, and Paul Sheatsley. 1981. "Nonresponse Bias for Attitude Questions." Public Opinion Quarterly 45: 359-375.

Triplett, Timothy, Johnny Blair, Teresa Hamilton, Yun Chiao Kang. 1996. "Initial Cooperators vs. Converted Refusers: Are There Response Behavior Differences?" Proceedings of the American Statistical Association: Section on Papers presented at the 51st Annual Conference of the American Association for Public Opinion Research: 1038-1041.

Verba, Sidney. 1996. "The Citizen as Respondent: Sample Surveys and American Democracy." American Political Science Review 90: 1-7.

Table 1. Response Rate Calculations and Call Dispositions by Survey Protocol

Calculation of Response Rates Using AAPOR Standards

Final Disposition Category	Standard Five Day	Standard Full	Rigorous
Total numbers dialed	4177	4193	3062
I=Complete	1000	1527	1201
P= Partial	0	0	0
R=Refusals/Breakoffs	720	689	429
NC=Non-contacts	801	239	125
Resp away/unavailable			
Answering machine			
Call back			
O=Other	181	243	193
Health			
Language			
UH=Unknown HH eligibility	371	284	166
Busy all attempts			
No answer all attempts			
UO=Unknown other	0	0	0
NE=Not eligible	1104	1211	948
Not working/disconnect			
Business/govt			
Computer/fax			
No one 18 or older			
Response Rate 3	36.0%	55.4%	60.6%
Cooperation Rate 3	58.1%	68.9%	73.7%
Contact Rate 2	68.5%	89.3%	92.0%

The American Association for Public Opinion Research. 1998. Standard Definitions: Final Disposition of Case Codes and Outcome Rates for RDD Telephone Surveys and In-Person Household Surveys. Ann Arbor, Michigan: AAPOR.

**Table 2. Tests of Homogeneity for Survey Variables Between Unweighted 5-Day Standard and Rigorous Studies
(sorted by p-values within groups)**

Item Number	Item Description	Modal Category	Modal % Standard 5-day	Modal % Rigorous	χ^2 all categories	df	p value
Political and Social Opinion							
Q21d	Opinion of Militia Groups	Very Unfavorable	62%	61%	9.74	3	0.02
Q21j	Opinion of Republican Party	Mostly favorable	46%	52%	9.60	3	0.02
Q4b	Poor People Have it Easy/Poor People Have Hard Lives	Poor people have it easy	37%	38%	8.42	3	0.04
Q21b	Opinion of Labor Unions	Mostly favorable	46%	47%	8.05	3	0.04
Q21f	Opinion of Blacks	Mostly favorable	66%	69%	7.91	3	0.048
D13	Political Affiliation	Democrat	35%	36%	5.97	2	0.050
Q4d	Racial Discrimination/Responsible for Own Condition	Feel strongly Blacks are responsible for own condition	46%	48%	6.95	3	0.07
Q21g	Opinion of Hispanics	Mostly favorable	66%	63%	6.95	3	0.07
Q5	Abortion Should Be Generally Available/Restricted/Not Permitted	Restricted	35%	31%	6.92	3	0.08
Q21c	Opinion of Business Corporations	Favorable	65%	68	5.47	3	0.14
Q35c	Answer for Sins on Judgement Day/Don't Have to Answer for Sins	Feel strongly will be called to answer for our sins	80%	77%	5.20	3	0.16
Q4f	Business Corporations Make Too Much Profit/Fair and Reasonable Profit	Feel strongly Business Corp. Make too much profit	46%	44%	5.03	3	0.17
Q35a	Homosexuality Should be Accepted/Rejected	Feel strongly should be discouraged	43%	40%	4.82	3	0.19
Q4g	Most Elected Officials Care What People Think/Don't Care	Feel strongly elected officials don't care	58%	53%	4.80	3	0.19
Q21h	Opinion of Asians	Mostly favorable	64%	68%	4.02	3	0.26
D14	Lean More Towards Democrat/Republican	Democrat	53%	58%	1.23	1	0.27

Item Number	Item Description	Modal Category	Modal % Standard 5-day	Modal % Rigorous	X^2 (all categories)	df	p value
Political and Social Opinion (continued)							
Q35b	Books Should be Banned from Public Libraries/Libraries Should Carry What they Want	Feel strongly books that contain dangerous ideas should be banned	41%	44%	3.61	3	0.31
Q21a	Opinion of Congress	Mostly favorable	51%	55%	3.50	3	0.32
Q21e	Opinion of Jews	Mostly favorable	62%	65%	3.47	3	0.32
	Index of Conservative Opinions (Q4a, Q4b, Q5, Q22, Q35a, Q35b. Alpha=.49)	Mean*	2.57	2.52	1.67	6	0.95
Q6	Expectations of Taxes for the Average American in 2000	The average American will pay more taxes	93%	94%	0.75	1	0.39
Q7	Best Ideas on Middle Class Tax Cuts	Republican congressional leaders	37%	37%	2.75	3	0.43
Q4a	Govt. Always Wasteful and Inefficient/Govt. Does a Better Job	Feel strongly Govt. Is always wasteful and inefficient	51%	49%	2.68	3	0.44
Q22	Political Ideology	Moderate	42%	42%	2.98	4	0.56
Q8a	Can Govt. Reduce Budget Deficit and Cut Taxes	Can't do it	51%	50%	0.29	1	0.59
Q4e	Immigrants Today Strengthen Country/Burden Country	Feel strongly immigrants burden country	40%	41%	1.56	3	0.67
Q35d	No Excuse for Oklahoma City Bombing/Understand Anger and Frustration	Feel strongly No excuse	86%	86%	1.40	3	0.70
Q8	Highest Priority Cutting Taxes for Middle Class/Reduce Budget Deficit	Reduce budget deficit	57%	57%	0.45	2	0.80
Q4c	Position of Blacks has Improved/Hasn't Been Much Real Progress	Feel strongly Position of Blacks has improved	58%	59%	0.84	3	0.84
Q21I	Opinion of the Democratic Party	Favorable	55%	54%	0.51	3	0.92

Item Number	Item Description	Modal Category	Modal % Standard 5-day	Modal % Rigorous	χ^2 all categories	df	p value
Electoral Behavior							
D15	Vote in 1996	Voted in 1996	74%	76%	1.05	1	0.30
D16	Presidential Vote 1996	Voted for Clinton in 1996	55%	58%	3.33	3	0.34
D17	House Vote 1996	Voted for Republican	44%	46%	1.83	3	0.61
D12	Voting Frequency	Always Vote	44%	43%	1.66	4	0.80
D11	Registered Voter	Registered to vote	79%	79%	0.00	1	0.98
Media Use							
Q13a	Use a Computer to Connect to Online Services	Connects to online services	49%	54%	3.53	1	0.06
Q13	Use a Computer	Use a computer	63%	66%	2.69	1	0.10
Q12c	Watch Daytime Talk Shows (Lake, Springer, Jones)	Never watch	59%	61%	5.78	3	0.12
Q12a	Listen to Rush Limbaugh's Radio Show	Never listen	73%	72%	3.02	3	0.39
Q11	Listen to Radio Yesterday	Did not listen to radio news	50%	52%	0.69	1	0.41
Q12b	Listen to Religious Radio Shows	Never listen	57%	55%	1.95	3	0.58
	Index of Media Usage (Q9, Q10, Q11. Alpha=.35)	Mean*	1.64	1.61	1.15	3	0.77
Q10	Watch News/News Program Yesterday	Watched Yesterday	66%	65%	0.04	1	0.83
Q9	Read Newspaper Yesterday	Did not read a newspaper	52%	52%	0.03	1	0.86
Knowledge							
	Index of Knowledge (Q18, Q19, Q20. Alpha=.60)	Mean*	1.36	1.40	2.42	3	0.49
Q18	Political Party Having Majority in House of Representatives	Republicans	89%	90%	0.23	1	0.63
Q20	Know Who Bill Gates is	CEO/head of Microsoft	88%	88%	0.09	1	0.77
Q19	Individual Who Lent Newt Gingrich Money to Pay Ethics Fine	Bob Dole	88%	88%	0.01	1	0.93

Item Number	Item Description	Modal Category	Modal % Standard 5-day	Modal % Rigorous	χ^2 all categories	df	p value
Social Integration							
Q28a	Frequency of Volunteer Work	Volunteer on occasional basis	44%	53%	10.46	2	0.005
Q26	Amount of Social Support	Can turn to many people	43%	47%	4.74	2	0.09
Q14	Most People Can be Trusted/Can't be Too Careful	Can't be too careful	56%	52%	2.72	1	0.10
Q15	Most People Take Advantage/Try to be Fair	Most people try to be fair	60%	63%	2.20	1	0.14
Q16	Most People Try to be Helpful/Looking Out for Themselves	People try to be helpful	60%	63%	2.05	1	0.15
Q17b	Exercise Yesterday	Did not exercise yesterday	63%	60%	1.54	1	0.22
Q17c	Call a Friend or Relative Yesterday	Called a friend or relative	66%	68%	1.39	1	0.24
	Index of Trust (Q14, Q15, Q16. Alpha=.72)	Mean*	1.54	1.66	6.07	3	0.11
	Index of Social Activity (Q17a, Q17b, Q17c, Q17d, Q28. Alpha=.31)	Mean*	2.67	2.71	3.99	5	0.55
Q17d	Go Shopping for Something Other Than Food or Medicine Yesterday	Did not go shopping	70%	71%	0.32	1	0.57
Q17a	Visit With Family or Friends Yesterday	Visited w/family and friends	74%	73%	0.16	1	0.69
Q28	Have Done Volunteer Work in the Past Year	Yes	60%	60%	0.01	1	0.94
Q27	Frequency of Religious Service Attendance	Once a week	28%	28%	0.48	5	0.99

Item Number	Item Description	Modal Category	Modal % Standard 5-day	Modal % Rigorous	χ^2 all categories	df	p value
Crime and Guns							
Q33b	Assaulted This Past Year	Has not been physically assaulted or mugged	89%	90%	1.33	2	0.52
Q32	How Safe is Neighborhood When Walking After Dark	Neighborhood is very safe	46%	47%	1.58	3	0.66
Q33a	Property Stolen This Past Year	Did not have money or property stolen	67%	68%	0.71	2	0.70
Q34	Guns or Revolvers at Home	Do not have guns or revolvers in home	61%	61%	0.14	1	0.71
Polling							
D27	Willing to do Interview Again	Would participate again	79%	81%	1.37	1	0.24
Q37	Random Sample of 1,500 or 2,000 People Sufficient	Random sample insufficient	70%	69%	0.27	1	0.60
Q36	Opinion Polls Work For/Against Public Interest	Opinion polls work for the best interests	77%	78%	0.22	1	0.64
Q38	Ever Done a Public Opinion Poll	Have never been questioned in an opinion poll	53%	53%	0.06	1	0.81
Demographics							
Q30	No Money for Clothes	Not without money to buy clothes	80%	86%	11.37	1	0.001
D25	Listed/Unlisted	Listed in telephone directory	73%	79%	13.00	2	0.002
D18	Own/Rent	Homeowner	67%	72%	10.00	2	0.007
D10	Household Income	\$50,000-\$74,999	16%	19%	19.12	7	0.008
D5	Respondent's Race	White	79%	83%	10.53	3	0.015
D3	Respondent's Education	High school graduate	38%	34%	8.73	3	0.04
Q29	No Money for Food	Not without money to buy food	82%	85%	4.27	1	0.04

D23a	Number of Adults Living in Household	Two	52%	55%	8.95	4	0.06
Q31	No Money for Medical or Health Care	Not without money to pay for medical or health care	78%	81%	2.87	1	0.09
D9a	Born Again or Evangelical Christian	Would not consider self a born-again or evangelical Christian	59%	62%	2.31	1	0.13
D2	Age	Age 35-44	24%	23%	8.24	5	0.14
D4	Hispanic Origin or Descent	Not of Hispanic origin or descent	93%	94%	1.73	1	0.19
D1	Respondent's Gender	Female	56%	58%	1.62	1	0.20
D22	Years lived in community				6.52	5	0.26
D7	Marital Status	Married	54%	56%	4.96	4	0.29
D21	Lifetime in Community	Has not lived in community entire life	69%	71%	0.86	1	0.35
D23	Number of People Living in Household	Two	32%	33%	4.20	5	0.52
D6	Parents or Grandparents From a Country Other Than the U.S.	Parents/grandparents born in U.S.A.	63%	61%	1.14	3	0.77
D20	Urbanicity	Small city or town	37%	36%	1.12	3	0.77
D9	Religious Preference	Protestant	56%	57%	0.85	3	0.84
D19	Employment Status	Employed full-time	55%	56%	0.34	2	0.84
D8	Parent or Guardian of Children in Household	No children in household	64%	64%	0.04	1	0.84
D24	Number of Telephone Numbers	One	79%	78%	0.22	2	0.90

Interviewer Ratings of Respondent

D29	Respondent's Interest	Average	32%	35%	25.18	4	0.000
D31	Respondent's Hearing	No degree of hearing loss	93%	95%	4.21	2	0.12
D28	Respondent's Cooperation	Very good	60%	55%	6.16	4	0.19
D30	Respondent's Facility with English	Fully conversant	97%	96%	1.20	2	0.52

Table 3a. Unweighted Respondent Distributions on Demographic Variables by Survey Protocol, Compared to 1996 Current Population Survey Distributions (significant differences between Standard and Rigorous)

Subgroup	1996 CPS (Telephone households only)	Standard 5 Day	Rigorous
(n)		1000	1201
Household Income			
Less than \$20k	27%	26	21
\$20-\$29.9k	15	17	17
\$30-\$49.9k	24	29	27
\$50-\$74.9K	18	16	19
\$75K+	16	12	16
Total	100%	100%	100%
Home Ownership			
Own	68%	66%	72%
Rent	31	30	24
Other	1	4	4
Total	100%	100%	100%
Education			
Less than HS	18%	11%	9%
HS Graduate	37	38	34
Some College	23	23	24
College plus	22	28	33
Total	100%	100%	100%
Race			
White	85%	79%	83%
Black	11	13	9
Other	4	8	8
Total	100%	100%	100%

Table 3b (nonsignificant differences)

Subgroup	1996 CPS (Telephone households only)	Standard 5 Day	Rigorous
(n)		1000	1201
Ethnicity			
Hispanic	9%	7%	6%
Not Hispanic	91	93	94
Total	100%	100%	100%
Marital Status			
Married	59%	54%	56%
Not Married	41	46	44
Total	100%	100%	100%
Employment			
Full time	54%	55%	55%
Part time	12	11	12
Not employed	34	34	33
Total	100%	100%	100%
Number of Adults			
One	30%	30%	27%
Two	55	52	55
Three or more	14	18	18
Total	100%	100%	100%
Sex			
Male	48%	44%	42%
Female	52	56	58
Total	100%	100%	100%
Age			
18-24	12%	12%	9%
25-34	21	20	19
35-44	22	24	23
45-54	17	15	18
55-64	11	12	12
65+	17	17	19
Total	100%	100%	100%
Gender by Age			
Female, 18-29	11%	10%	9%
Female, 30-49	22	24	27
Female, 50+	19	21	22
Male, 18-29	11	11	7
Male, 30-49	21	21	19
Male, 50+	16	13	16
Total	100%	100%	100%

Table 4. Mean Number of Items with Missing Data by Survey Protocol, Number of Prior Refusals before Interview, and Refusal count and Calls to First Contact

	<u>Mean</u>	<u>SE</u>
Entire sample (N=2728)	2.9	.09
Standard (N=1000)	3.0	.15
Rigorous (N=1201)	2.8	.13
No refusals (N=2186)	2.8	.10
One refusal (N=342)	3.0	.27
Two refusals (N=200)	3.5	.37
No refusals, ≤ 10 calls (N=1841)	2.9	.11
No refusals, > 10 calls (N=345)	2.6	.21
Refusal, ≤ 10 calls (N=372)	3.4	.28
Refusal, > 10 calls (N=170)	2.9	.32

**Table 5. Mean Number of Completed Interviews Per Hour
by Time Periods During Data Collection, by Survey Protocol**

	Standard	Rigorous
Short Field Period	0.99	0.97
First Week After	1.01	1
Second Week	0.79	0.59
Third Week	0.58	0.62
Fourth Week	0.39	0.4
Fifth Week	0.26	0.23
Sixth Week	NA	0.13
Seventh Week	NA	0.24
Eighth Week	NA	0.75
TOTAL	0.85	0.79
Total for Five Day	0.99	