

Subject Acquisition for Web-Based Surveys

R. Michael Alvarez, Robert P. Sherman, and Carla VanBeselaere

California Institute of Technology,

Pasadena, CA 91125

e-mail: rma@hss.caltech.edu

e-mail: sherman@hss.caltech.edu

e-mail: carla@hss.caltech.edu

This article provides a basic report about subject recruitment processes for Web-based surveys. Using data from our ongoing Internet Survey of American Opinion project, two different recruitment techniques (banner advertisement and subscription campaigns) are compared. This comparison, together with a typology of Web-based surveys, provides insight into the validity and generalizability of Internet survey data. The results from this analysis show that, although Internet survey respondents differ demographically from the American population, the relationships among variables are similar across recruitment methods and match those implied by substantive theory. Thus, our research documents the basic methodology of subject acquisition for Web-based surveys, which, as we argue in our conclusion, may soon become the survey interview mode of choice for social scientists.

1 Introduction

Are Web-based surveys the future of survey research? On one hand, Web-based surveys show great promise. They can be very cost-effective ways to collect public opinion data, provide rapid turn-around time, allow for easy presentation of complex textual or multimedia material to respondents, provide flexibility for respondents, and may alleviate much of the time and social pressures present in conventional interview formats (Dillman 2000; Krosnick and Chang 2001; Berrens et al. 2003). Also, some evidence indicates that Web surveys might produce highly reliable survey responses (Krosnick and Chang 2001). Finally, Web surveys promise to democratize the process of survey data collection: “Not only can researchers get access to undreamed numbers of respondents at dramatically lower costs than traditional

Authors' note: We thank the IBM Corporation's University Matching Grants Program, the USC–Caltech Center for the Study of Law and Politics, and Caltech for their support of this research. Garrett Glasgow and audience members at the 2001 Annual Meetings of the American Political Science Association provided excellent commentary on this research, as did participants at the Fall 2001 meetings of the Southern California Political Methodology Program at the University of California, Riverside. Dorothy Kaslow, Chris Malek, and Catherine Wilson all provided assistance with various components of this research. We also thank Chey Hahn and ValueClick, Inc. for their assistance with the research reported in this paper, and for the use of ValueClick, Inc.'s UltraLeads e-mail model; finally, we thank Michael Halberstam, Alice Strauss, and Michael Meeks of Interviewing Services of America, Inc., for their assistance with our telephone interviewing research.

Copyright 2003 by the Society for Political Methodology

methods, but members of the general population too can put survey questions on dedicated sites offering free services and collect data from potentially thousands of people” (Couper 2000, pp. 464–465).

Furthermore, there is increasing evidence of problems with traditional survey methodologies. Brehm (1993, p. 17) reported

nonresponse rates are getting worse.... Total nonresponse rates for the DAS (Detroit Area Studies) have risen from 12.5 percent in 1952 to 31.6 percent in 1988.... Refusal rates for the NES (National Election Studies) have climbed from well under 8 percent at inception (in the early 1950s) to current refusal rates near 25 percent. The refusal rates for the Surveys of Consumer Affairs climb at a virtually identical rate to the NES. Similar increases have been observed for other organizations.

Data from the Current Population Survey (CPS)—one of the highest quality periodic surveys in existence—show noninterview rates of between 4% and 5% from 1964 through 1994, but an increasing noninterview rate thereafter, reaching almost 8% in 2001; the refusal rate in 1964 was barely more than 1% in the CPS monthly surveys, but reached more than 4% by 2001.¹ Americans are increasingly difficult to contact at home and are less willing to participate in traditional surveys. One hope is that Web-based surveys might alleviate, or possibly counteract, these trends.

On the other hand, Web-based surveys are potentially problematic. There are technological hurdles that can plague Web-based surveys. For example, the fact that there are many different Web browsers used by potential respondents raises compatibility problems. Interface design issues can arise (Couper et al. 2001) and the lack of easy give-and-take between interviewer and respondent may make resolution of ambiguities in the survey instrument impossible. Also, of course, there are significant problems regarding sampling and representativeness of Web survey samples (Couper 2000).

What is clear at this stage of research on Web survey methodology, however, is that these issues are still awaiting scientific examination. Questions about whether Web-survey methods are superior to existing face-to-face, telephone, or mail surveys are difficult to answer because Web-based survey methodologies are not well developed. What is an appropriate sampling frame for a Web-based survey? How can subjects be contacted for participation in a Web-based survey? What standards are appropriate for Web-survey design, layout, and length? Again, these proximate questions are still awaiting research, and without answers to these questions it is difficult to compare Web-surveys to other established survey methodologies.

In this article, we tackle only one of these questions. Using data collected from our continuing Web-based survey project, we discuss and examine two different methodologies for recruiting subjects to participate in Internet surveys. In the next section, we discuss subject acquisition for Web-based surveys and present the details of the two different procedures we used. In the following sections, we examine the profiles of the subjects we recruited using each methodology, and also look at how these subjects respond to Web-based surveys. We conclude with an evaluation of these methodologies for recruiting Web survey subjects.

¹The CPS is conducted by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics. Data for 1964 through 1996 are from the U.S. Department of Labor, Bureau of Labor Statistics 2000; response data for 2001 are from www.bls.census.gov/cps/basic/perfmeas/folder.htm.

2 Subject Acquisition for Web Surveys

2.1 Survey Errors and Traditional Methods for Surveying

When discussing subject acquisition methods, it is appropriate to start with some simple definitions and how they apply to the traditional surveys that are familiar to social scientists. Three main concepts are of concern regarding subject acquisition for surveys:

- *the survey population*: the entire set of units—whether individuals, households, organizations, institutions, geographic entities, or others—for which the researcher wants to make generalizations or inferences;
- *the sampling frame*: a list of population members;
- *the sample*: the units from the population that are drawn from the sampling frame to be included in the survey.

By way of example, consider a typical telephone survey of the adult U.S. population. The objective of this type of survey is to understand American political preferences and behaviors. The survey population in this case is the adult American population. If a list of all Americans existed, this could be used as a sampling frame. Unfortunately, no such list exists. Instead, random digit dialing (RDD) is often used to obtain random samples of households.² RDD enables researchers to approximate random samples of the American population. Assuming that most households have a telephone and given that RDD techniques ensure that each residential phone number has an equal probability of being drawn, RDD generates a random sample of potential respondents.³

Surveys are subject to various types of error—survey results may incorrectly reflect the population of interest. Groves (1989) identifies four major sources of error in surveys: coverage error, sampling error, nonresponse error, and measurement error. The previously mentioned example of RDD surveys of the adult American population is useful for understanding these four sources of error.

Coverage error is simply the deviation between the sampling frame and the survey population. If there were a list of all American adults (the sample population), there would be virtually no mismatch between the sampling frame and the population of interest. However, because respondents in this example are contacted by telephone, there will be some coverage error because there is a small percentage of Americans who do not own telephones.⁴

Sampling error is error that is introduced because not all of the units in the sampling frame are included in the sample. There will always be variability or sampling error in estimates if the whole population is not examined because there will always be units in the sampling frame not in the sample. If another sample were drawn, different units would be observed and this might induce slightly different results in a subsequent empirical study.

²RDD involves the random generation of any possible telephone numbers including unlisted numbers and new numbers. RDD often uses information about which exchanges will likely contain residential numbers as a basis for selecting a sample of telephone numbers.

³However, RDD is greatly complicated by many aspects of contemporary telephony in the United States. One problem is the explosion in the use of multiple telephone lines for residences; in 1988, only 2.7% of households with telephone service had multiple lines, but by 1998, 22.3% had multiple lines (assuming that the use of multiple residential lines has continued to grow at roughly that same rate, at least 25% of households now have multiple residential phone lines) (Federal Communications Commission 2000, Table 20.4). Furthermore, the use of multiple residential lines is likely to be concentrated in certain types of households, like ones with higher incomes and larger numbers of residents. Unless attempts are made to deal with the explosion of multiple residential lines, simple RDD is problematic, because not every residence will have the same selection probability.

⁴According to the Federal Communications Commission (2002), the nationwide telephone penetration rate was 94.5% in March 2001.

For example, if two random samples of 500 individuals were drawn from a population of 1500, the two samples would likely contain different individuals, and slightly different results could be obtained in the empirical analysis.⁵

Nonresponse error arises because not all of the units in the sample actually complete the survey. Whether this is because sample members were unable or unwilling to complete the survey, their absence from the survey dataset produces a third form of error. The extent of nonresponse error in a survey is based on both the incidence of nonresponse and on how nonrespondents differ from respondents on variables of potential interest. To return to the telephone survey example, once contacted by telephone, potential respondents might refuse to participate in the survey.⁶ Nonresponse can confound analytical results by confusing the behavioral parameters of interest with the factors affecting participation. Brehm (1993) argued that nonresponse in telephone and face-to-face surveys affects the predictions of several political behavior models.

Measurement errors are a result of inaccuracies in responses recorded. Groves (1989) categorizes measurement errors by their different sources—the interviewer, the respondent, the questionnaire, and the mode of data collection. In a telephone survey, the interviewer might induce error by failing to read the question correctly, leading the respondent's choice of answer, and failing to record the answer correctly. Phone survey respondents might introduce error because they do not exert enough effort to obtain an accurate response. The wording of the survey can also affect the quality of respondents' answers.⁷ Finally, the mode of data collection appears to affect responses. For example, telephone respondents appear to exert less effort in responding to questions and have a tendency to provide shorter answers relative to face-to-face interviews.⁸

A survey researcher's task is easiest when the population is known, the sampling frame is known, and there is total compliance with the survey—the only errors that exist in such a survey are sampling error and measurement error. In some cases, the population and sampling frame are known with certainty: examples of these types of situations usually involve smaller populations, like interest groups, students, employees, and state or local governments. Unfortunately, these conditions rarely, if ever, exist in many important survey applications of interest, especially when they are about large populations like the universe of American adults. Thus, survey researchers must make important decisions about the relative costs and benefits of minimizing each source of error when planning their survey analysis.

For many large populations of interest, a simple list—for the sampling frame—does not exist. As discussed previously, there is no general list of addresses and telephone numbers for the adult American population. This led to the development of technologies like RDD for telephone surveys. When American adults were home in the evening, and were not bombarded with telemarketing calls, RDD was effective at minimizing all four types of error: coverage error, sampling error, response error, and measurement error. However, today, telephone interviewing faces enormous obstacles, ranging from an increasing reliance on technologies that screen out unwanted telephone calls to a simple unwillingness to answer

⁵Sampling error is accounted for by reporting standard errors or confidence intervals.

⁶The response rate is calculated as the number of complete interviews divided by the number of eligible respondents. The American Association for Public Opinion Research (2000) developed standard definitions for final dispositions of case codes and outcome rates for surveys.

⁷Research has found that small, apparently innocuous, changes in wording of a question can lead to large changes in responses (Groves 1989).

⁸For a discussion of this, see Green et al. (2001).

Table 1 2001 CSLP telephone survey response information

<i>Disposition</i>	<i>Number</i>	<i>AAPOR codes</i>
Contacts:		
Completed interviews	1500	(1.10)
Refusals	1469	(2.10)
No one over 18	151	(4.70)
Other	429	(2.35)
Total contacts:	3549	
Still active:		
No answer	3018	(3.13)
Call back	1378	(2.21)
Answering machine	615	(2.22)
Busy	223	(3.12)
Suspended interview	199	(2.12)
Other	46	(3.90)
Total still active:	5479	
Bad numbers:		
Disconnected	1902	(4.30)
Business	807	(4.51)
Fax	529	(4.20)
Language barrier	236	(2.33)
Change number	28	(4.41)
Wrong number	20	(3.90)
Other	270	(3.90)
Total bad numbers:	3792	
Never dialed:	275	(3.21)
Total numbers:	13,095	

the phone at home. Thus, although all households may have an equal chance of being called and included in a contemporary telephone sample, RDD is increasingly problematic because nonresponse rates have been steadily increasing since the early 1980s.

For a sense of the difficulties involved in contemporary RDD, we present data in Table 1 on interview attempts from a recent telephone survey conducted March 26 through April 5, 2001. The data we present in Table 1 are rarely, if ever, available for secondary users of survey data from either commercial or academic sources. The telephone survey was a collaborative project of the USC–Caltech Center for the Study of Law and Politics, and the telephone interviewing was conducted by Interviewing Service of America, Inc. (ISA), a market research firm that routinely conducts computer assisted telephone interviewing (CATI) surveys. ISA obtained 1500 complete interviews during the interviewing period, having begun with a randomly drawn set of 13,095 residential telephone numbers.⁹

To get the 1500 complete telephone interviews, 3549 contacts were made: 1469 ended by the respondent's refusal to participate, 151 ended because there was no one over the age of 18 in the household, and 429 ended for a variety of other reasons. Notice in Table 1,

⁹ISA used list-assisted RDD techniques in attempt to restrict the sample to residential phone numbers. In a simple RDD techniques, computer software would generate a random telephone number from an algorithm designed to determine random sequences of residential telephone numbers. List-assisted RDD techniques begin with a large list of residential phone numbers (here, 13,095 numbers), and interviewers draw numbers randomly from this list.

however, the high number of “still active” telephone numbers (5479) that received up to six call-backs. Of the “still active” numbers, 3018 were never answered, 1378 resulted in a request for a call-back, 615 reached only an answering machine, 223 were busy, and 199 had interviews that were begun but were suspended before completion (39 of the suspended interviews were eventually concluded, however).

Even more troubling in Table 1 are the 3792 “bad numbers” that arose in the telephone interviewing. Of these, 1902 were phone lines that were disconnected; 807 reached a business; 529 were fax numbers; 236 reached households with a language other than English; and 318 were bad for a variety of other reasons. It is clear from the data in Table 1 that RDD is a difficult enterprise in contemporary America and, although in principal it can minimize sampling and coverage errors, the tradeoff with nonresponse error may be worse today than is acknowledged by survey methodologists.

Traditional survey methods—face-to-face interviewing, mail or other self-completion paper surveys, and RDD telephone surveys—all aim to minimize sampling, coverage, and nonresponse errors. However, the extent to which these traditional methods successfully minimize these different sources of error can vary dramatically, and traditional survey methods can generate substantial errors. Unfortunately, often little is known by secondary users of survey data in social science about the extent of sampling, coverage and nonresponse errors in the data they use. These errors can proliferate in traditional surveys and, as we document in Table 1, even researchers using RDD techniques are today finding it difficult to recruit participants into their surveys. The difficulties of recruiting subjects for traditional telephone surveys implies that many RDD telephone surveys might be more error laden than is assumed by practitioners.

2.2 *Web-Survey Typology*

Couper (2000) provides an excellent overview of the various types of Web surveys that have been developed to date. He breaks Web surveys into two basic types: probability and nonprobability Web surveying. Probability approaches are much like those discussed previously under the rubric of “traditional” survey methodologies: the researcher identifies the population, develops a sampling frame, and uses the sampling frame to generate a random research sample. Using this approach, the probability that any unit of the population will be sampled is known; probability-based Web surveys can thus be used to make generalizations about the population on which they are based.

For surveys of the general American population, probability-based sampling over the Internet is more complicated than RDD for telephone surveys. Obviously, virtually every American household has a telephone; but, as a result of the “digital divide,” many American households do not have computers or Internet access. Recent U.S. Department of Commerce estimates indicate that at least 51% of American households have computers, and at least 41.5% of households had Internet access in August 2000 (U.S. Department of Commerce 2000). Thus, because Internet access is limited to only some households and neither universal e-mail directories nor easy techniques to contact Internet users randomly exist, there is no simple way to develop a complete sampling frame of Internet users.

Thus, two approaches have arisen to conduct probability-based surveying on the Web. One is to restrict the population to only Internet users and devise methods of randomly selecting Internet users into a sampling frame. The second is to use other approaches to contact a broader population and, from that, develop a sampling frame of survey respondents.

There are at least five different types of probability-based Web surveys (Couper 2000):

1. Intercept-based surveys of visitors to particular Web sites
2. Known e-mail lists
3. Prerecruited panels
4. Mixed-mode survey designs
5. Probability population samples

The first, the intercept-based approach, is based on interview techniques used in exit poll surveys or many types of market research. With a sampling frame being all visitors or users of a particular Web site, the sample is a randomly selected set of visitors who are asked to participate in some form of survey. Known e-mail lists are a second form of probability-based Web survey, and these are extremely useful when the population is one that has universal Internet access. Student university faculty, and employee surveys are examples of known e-mail list surveys; surveys like these can minimize sampling and coverage errors.

The remaining approaches for probability-based Web surveys are all based on having a probability sample and then using the probability sample to obtain Web-survey subjects. In the prerecruited panel approach, researchers use other techniques of probability sampling, like RDD telephone surveys, to recruit Web-survey samples; such an approach works well for studies of the Internet-using population if respondents with Internet access agree to participate in subsequent Web surveys by providing their e-mail address and then participate when contacted by e-mail. Mixed-mode approaches simply offer Web surveys as one of a multitude of modes for their participants to use (in addition to telephone or other modes). Probability population samples, of the sort developed by Knowledge Networks (www.intersurvey.com), begin with RDD telephone surveys and recruit participants into their Web-based surveys by providing them with Internet access.

Nonprobability Web surveys, however, are the most ubiquitous surveys on the Internet. These are surveys in which the population is unknown, and hence it is impossible to derive a sampling frame or a probability sample. Thus, traditional forms of response rate measures are impossible to estimate. In addition, as the population is unknown and there is no traditional sampling frame, any inferences made about population parameters from nonprobability surveys are potentially problematic.

There are three types of nonprobability Web surveys identified by Couper (2000):

1. Entertainment surveys
2. Self-selected surveys
3. Volunteer surveys panels

The first, entertainment surveys, are found all over the Internet. Generally, they are not intended for scientific surveying, but for the entertainment of visitors to Web sites. Self-selected surveys are those on the Internet that give visitors to a Web site the opportunity to participate in a survey; thus, only visitors to the site are possible subjects and only if they actively initiate the interview. The third type of nonprobability Web survey is volunteer survey panels, in which respondents are recruited on the Internet through advertisements of various types. Individuals go to a particular Web site and provide some information about themselves, including their e-mail addresses, which goes into a database. In this last type of nonprobability Web survey, respondents can be sampled from the database, contacted, and asked to visit a Web site in order to participate in some form of survey.

The issue, of course, with the nonprobability Web surveys is that, as they are not starting with a known population and fixed sampling frame, it is difficult to know whether they can be used to make generalizations about larger populations. In particular, the question is whether volunteer survey panels can be used when appropriate weighting or selection corrections have been made (see VanBeselaere 2001, for an example). Given that volunteer survey panels are widely used by a number of commercial survey research firms, more research is needed regarding how they can be used to make inferences about broad populations (Couper 2000).¹⁰

Although nonprobability Internet surveys are not based on rigorous sampling procedures, they can and are being used in situations in which researchers desire to exploit within-sample variance in a situation in which statistical power can be maximized. For example, Internet survey samples can be used to examine priming or framing, especially studies that might involve graphical or multimedia materials. In such designs, thousands or tens of thousands of subjects might be included in a potential study and, as long as these subjects are assigned to control and experimental groups using some type of probability assignment protocol, this can produce powerful experimental results.

2.3 Recruiting Subjects for the Internet Surveys of American Opinion (ISAO)

For the remainder of this article, we discuss data collected as part of our Internet Surveys of American Opinion (ISAO) project on Web surveys (survey.caltech.edu). Our survey project “went live” on January 25, 2000, at which time we began enrolling subjects and soliciting survey interviews from them. Working with ValueClick, Inc., we embarked on two different procedures for recruiting subjects for our Web surveys, both of which were designed to construct what Couper’s (2000) typology would call a *volunteer panel nonprobability Web survey*.

2.3.1 Banner Advertisement Experiments

Our first experiment with subject recruitment involved the use of banner advertisements placed throughout the Web.¹¹ We developed an animated banner advertisement. Using a yellow background and black print, the first part of the banner simply read “Internet Surveys of American Opinion”; after 2 sec, the banner switched to read “[http:// survey.caltech.edu](http://survey.caltech.edu)” on top, with “Click Here!” “Say Your Piece” “Win Free Stuff!” along the bottom. Any individual who clicked on the banner advertisement would be directed to our survey Web site. The banner advertisement can be seen at survey.caltech.edu/banner6a.gif.

Using this banner advertisement, we ran four banner campaigns in 2000: April 25 through May 11, May 16 through June 1, July 3 through July 10, and October 4 through October 18. The first three campaigns aimed at obtaining about 10,500 “click-throughs”; the fourth

¹⁰Harris Interactive (vr.harrispollonline.com/register/main.asp) and Greenfield Online (greenfieldonline.com) are perhaps the best known volunteer panels, but the technique is used by many other commercial survey research firms.

¹¹In all of our experiments conducted by ValueClick, Inc., we used what is called *run of network advertising*; that is, our advertisements were run across all of the sites that had contracted with ValueClick, Inc., to conduct advertisements. To ensure the privacy of our subjects, we make no attempt to track where visitors to our Web site come from, nor do we use any type of tracking device (for example, “cookies”) to gather any information about our subjects. We require a valid e-mail address for an individual to be included in our database, and we request that they provide their e-mail to us when they return to complete a subsequent survey. The e-mail address allows us to identify each survey respondent uniquely and filter out multiple responses. If someone does not provide an e-mail address in a subsequent survey, that response is not included in our data set for that study.

campaign was twice as large, aiming to obtain roughly 21,000 click-throughs. In order to obtain the 10,500 clicks in a typical banner advertising campaign (the number of times individuals actively clicked on the banner advertisement and were directed to our Web site), more than 3 million “impressions” were necessary (the number of times the banner advertisement was displayed).¹² Considering the final campaign as two separate smaller campaigns, an average of 686 subjects “registered” (the number of individuals who completed our respondent information form and entered into our subject database during the banner advertising period) for each 10,500 click campaign.¹³

The first banner advertisement period, from April 25 through May 11, 2000, provides a good example of the efficacy of our experiments with banner advertisement as a method for subject recruitment. During this period, there were 10,652 clicks on our banner advertisement—the result of 3,521,561 impressions of our banner advertisement on the Internet. These figures imply a yield of 0.3% of clicks per impression. In the first period, we obtained 599 new subjects in our panel, or about 5.6% of the individuals who clicked through our banner and came to our Web site.

Roughly similar numbers hold for the other banner advertisement experiments. Across the banner advertisement experiments, we ran 17,908,189 banner advertisement impressions, producing 53,285 clicks, and yielding a total of 3431 new subjects. The total percentage yield of new subjects per click was 6.4%, whereas the percentage yield per impression was 0.2%.

For comparison, in our RDD telephone survey, we included a short series of questions designed to test the possible response rates for our banner advertisement questions. These results are presented in Table A1, available on the *Political Analysis* Web site.

The first question asked all of the 1500 respondents whether they had Internet access at home or in any other location; 978 (about 65%) of the respondents in this sample reported having Internet access. Of those who had Internet access, we asked how frequently they used the Internet. Almost 58% reported using the Internet every day, 32% said they used it at least once a week, and 10% said they used it at least once per month.

The third, fourth, and fifth questions (the exact wording of each question is in Table A1) go to the heart of our banner advertisement experiment. Our goal was to re-create the Internet experience over the phone and then, among other things, compare the telephone results to the Internet results. We first gave respondents a brief description of our banner advertisement. Question 3 asks about the likelihood (0%, 25%, 50%, 75%, 100%) of clicking on the banner. Given a positive likelihood in Question 3, Question 4 asks about the likelihood (again, 0%, 25%, 50%, 75%, 100%) of participating in the survey. Given a positive likelihood in both Questions 3 and 4, Question 5 asks for an e-mail address to enable participation in the survey.¹⁴

¹²The fourth campaign actually generated almost 8 million impressions. The figures for this campaign are divided in two to compare them with the previous campaigns.

¹³We must point out, however, that we do not know with certainty that subjects who registered during each banner advertisement period came into our subject sample as the result of viewing a banner advertisement; it is possible that some of these subjects found our Web site by chance, by a search engine, or by word-of-mouth. Likewise, it is possible that some subjects might have viewed a banner, in one period, wrote down our url or remembered it, and then came to our registration page after the particular banner advertisement campaign ended. In fact, during the January 25 through October 18, 2000, period, we obtained a total of 3942 unique registered subjects; because 513 of these subjects entered our sample outside the banner advertisement periods, we assume that they did not enter our sample because of banner advertisements. Of the 513 subjects, 233 (45.4%) entered the sample before we began the first banner campaign; 280 (54.6%) entered between campaigns.

¹⁴Note that the results in Question 5 can be used to check the consistency of responses to Questions 3 and 4.

We are interested in three different analyses with these survey questions: first, comparing telephone and banner advertisement response rates; second, comparing demographics of various groups from the telephone and Internet samples (e.g., Internet versus non-Internet users, banner clickers versus non-banner clickers, and participants versus nonparticipants); and last, checking the consistency of responses to Questions 3 and 4 by using the results in Question 5.

We start with the first question. From the results in Questions 3, we see that the expected proportion of banner clickers from the telephone survey is about 20% (198/978). Compare this with the 0.3% click rate for the actual Internet banner ads. Only 127 telephone survey respondents provided their e-mail address, indicating a participation rate among Internet users of about 13% (127/978). Compare this with the 0.02% participation rate for the Internet banner ads. These substantial differences in response rates suggest that our attempt to recreate the “look and feel” of the banner ad over the phone may have been too successful. These differences also suggest that the Internet banner ad participants may be different in important ways from the telephone survey participants.

Columns 4 and 5 of Table 2 contain tabulations of various demographics for phone survey respondents who “would register” (those who provided e-mail addresses to enable participation in the survey) and actual Internet banner recruits. There appear to be substantial differences with respect to age and education. In fact, χ^2 tests indicate that the distribution of age and education levels is different across these two samples. However, the distribution of race and gender does not differ significantly.

Table 2 Banner advertisement telephone experiment results

<i>Attribute</i>	<i>Telephone survey</i>				<i>Actual Internet banner subjects</i>
	<i>All adults</i>	<i>Internet user</i>	<i>Banner clicker</i>	<i>Would register</i>	
Gender					
Male	47.0	51.6	45.3	48.0	48.1
Female	53.0	48.4	54.7	52.0	51.9
Age					
18–29	20.9	25.2	31.1	28.4	27.2
30–39	19.2	22.6	23.1	33.1	19.4
40–49	21.0	24.9	24.0	18.9	22.8
50–59	17.4	16.7	12.9	13.4	17.5
60–69	10.7	7.8	6.7	4.7	8.4
Over 69	10.9	3.2	2.2	1.6	4.8
Race					
Non-white	21.1	20.4	28.9	21.3	22.5
White	78.9	79.7	71.1	78.7	77.5
Education					
No high school	7.4	2.2	2.3	4.0	3.9
High school	32.4	26.1	35.5	29.4	19.0
Some coll. voc.	23.0	26.2	28.1	23.8	36.9
Coll. graduate	23.4	28.5	22.6	27.0	23.4
Post-graduate	13.9	16.9	11.5	15.9	16.8

Note. These data are from unique registrants received before 10 A.M. Nov. 1, 2001. Only respondents over 18 are included in this analysis.

Columns 2, 3, and 4 of Table 2 compare the different types of phone survey respondents. χ^2 tests suggest the Internet users in the telephone sample are different from non-Internet users with respect to age, education, and sex. These results also show that banner clickers in the telephone sample are different from non-banner clickers with respect to age, race, and education. However, we also see in Table 2 that phone respondents who “would register” and nonparticipants do not differ significantly with respect to these four demographic variables.

Finally, we address our third question. Using the results in survey Questions 3 and 4, we see that the expected number of participants in the survey is 165. Compare this with the 127 respondents who gave their e-mail addresses in response to Question 5. This suggests that respondents somewhat overestimate their likelihood of participating.¹⁵

2.3.2 Subscription Campaign Experiments

In 2001, we began using a different type of subject acquisition procedure. We call this process the *subscription campaign*. Working with ValueClick, Inc., we used a multistage process whereby Internet users who were registering for some service on a multitude of Web sites were given the opportunity to check a box on their subscription form indicating their interest in participating in Internet surveys.¹⁶ Once this registration form was processed, a new pop-up window would appear on the individual’s computer desktop. This new window thanked the individual for their interest in Internet surveys and informed them that if they provided their e-mail address, they would be sent materials about how to participate in our Internet surveys. Once an individual typed their e-mail address in the provided box and clicked “submit,” two different processes were initiated. First, they were immediately e-mailed a welcome letter that contained our Web site address (URL). Second, their e-mail address was sent to us.

We conducted two recruitment drives using this process in 2001, the first from March 23 through April 8, and the second on June 15 through July 22. Both were set up to deliver more than 10,000 e-mail addresses. The first drive (March 23–April 8) netted 10,021 e-mail addresses, or “subscriptions.” Of these, 3064 unique individuals came to our Web site and entered our subject database by providing the requested information about themselves. That is a yield of 31% for the first subscription drive. The second drive produced 11,357 subscriptions, and 3725 of those individuals entered our subject database. The second drive netted a 33% yield. Across the two subscription campaigns, we received 21,378 subscriptions and a total of 6789 new subjects in our database—a 32% yield.

Of course, there were 14,228 subscribers who provided us with their e-mail addresses but did not follow our initial e-mail’s suggestion that they come to our Web site to sign up as a survey participant. Accordingly, during a 6-week period from September 18 through October 31, 2001, we e-mailed these initial nonsubscribers again and reminded them of our project. During this period, 1078 of these initial nonsubscribers (8% of the initial nonsubscribers) came to our site and registered as a subject. This increased the total yield of the subscription campaign to 7867 subjects, for a total yield of 37%.

As was the case for banner advertisement subjects, individuals who enter our subject pool through this subscription process must take a series of active steps. First, they must indicate their interest in participating in Internet surveys as they provide information on

¹⁵Formal statistical tests of this conjecture will appear in future research.

¹⁶Again we used ValueClick’s *run of network* distribution system; this placed our subscription line on a randomly selected set of Web sites that were participating with ValueClick in this program during our subscription campaign periods.

some other Web page. Second, they must give their e-mail addresses to be contacted to participate in an Internet survey. Third, they must respond to our e-mail solicitation to sign up as a subject. Each of these stages might introduce selection biases into our subscription subject pool. Unfortunately, at this stage of our research we have not conducted a telephone survey test of the potential representativeness of the subscription campaign so that we can compare it to the banner subscription process.

However, there are four different ways that we can compare these two different methodologies for recruitment of Web survey subjects. First, their efficiency: How quickly do they obtain subjects? Second, their cost: How much does it cost per subject? Third, their representativeness: How closely do the subject samples resemble the overall Internet-using population? This third issue, of course, is what we discussed earlier as coverage error. Finally, fourth, their reliability: How reliable are the survey responses from each type of subject recruitment process?¹⁷ We address the first two questions here; the second two are discussed in subsequent sections of this paper.

On the efficiency and cost comparisons, the subscription process clearly dominates the banner advertising process. Each banner campaign took between 2 and 3 weeks to fulfill our objective, but the yield rate for the banner advertisement campaign (in terms of subjects who actually entered our subjects database per click) was relatively low. The overall yield rate was 6.4%, and across three campaigns of around 10,500 clicks and one campaign of 21,000 clicks, we obtained 3429 subjects. If we consider each campaign to be 10,500 clicks—and thereby split the last effort into two separate campaigns—we have five different banner campaigns. This means that we averaged 686 subjects per campaign.

The efficiency of the subscription process is clear. We ran only two campaigns, running between 2 and 3 weeks. The yield was much higher across these two campaigns, with 32% of those who provided their e-mail addresses entering our subject pool immediately and 36% entering after a reminder e-mail. The two campaigns yielded 7867 subjects, an average of 3934 per campaign. Thus, the subscription process yielded almost twice as many subjects as we obtained from the various banner advertisement campaigns, and the per campaign yield for the subscription process is almost five times greater than for the banner advertisement process.

Regarding cost, we were able to obtain a reduced advertising rate from ValueClick, Inc.—essentially an academic advertising rate for our project. Each of the 10,500 click banner advertising campaigns cost \$5,000; also, each of the two subscription campaigns cost \$5,000. Thus, we spent a total of \$25,000 on banner advertising, and \$10,000 on subscription-type advertising. This produces per-subject costs that heavily favor the subscription process: We obtained 3429 subjects in the banner advertisement campaigns, at a cost of about \$7.29 per subject; we obtained 7867 subjects in the subscription campaign, at a cost of about \$1.27 per subject.

Thus, the subscription campaign generated subjects for our Web-survey project more efficiently, and almost six times less costly, than our banner advertisement strategy. So on these two criteria, we conclude that the subscription campaign dominates the banner advertising strategy for producing Internet survey subjects.

¹⁷There is a second type of reliability that we recognize: How effective is each type of recruitment process at obtaining subjects who will be most likely to remain as subjects in our data set for a long period of time. In other words, does one approach have a lower rate of panel attrition? Although this is another important way to judge these two recruitment strategies, we leave this question for future work on Web survey panels. After we have had the chance to conduct additional surveys with members of each group, we can better assess these effects.

Table 3 Chi-squared comparison across recruitment methods—demographic variables

<i>ISAO category</i>	<i>Banner v. Subs.</i>	<i>Banner v. Other</i>	<i>Banner v. Subs. Conv.</i>	<i>Subs. v. Subs. Conv.</i>	<i>Subs. v. Other</i>	<i>Other v. Subs. Conv.</i>
Gender	796.36*	32.52*	351.14*	109.11*	8.05*	98.28*
Age	129.82*	23.87*	52.09*	31.56*	84.00*	75.87*
Race	100.72*	26.29*	64.92*	0.02	8.04*	3.92*
Education	573.40*	44.94*	105.54*	507.77*	24.93*	175.20*
Marriage	18.65*	1.74	0.01	1.14	8.50*	1.44
Region	76.41*	86.09*	37.41*	74.47*	3.20	30.50*
Kids	51.97*	18.99*	19.36*	70.07*	0.44	44.22*

*Indicates a χ^2 test that is statistically significant at the $p \leq .05$ level. These data are from unique registrants received before 10 A.M. Nov. 1, 2001. Only respondents over 18 are included in this analysis.

3 Examining ISAO Subjects

3.1 Subject Profiles

Although it is clear that the subscription campaign generated respondents more quickly and at a lower cost, the question remains as to what type of subjects these two different methodologies provide. As noted previously, the two different recruitment methods were undertaken at different periods. Fortunately, subjects received the same set of demographic questions when they initially registered to participate in the ISAO. Using these data, it is possible to compare the demographic characteristics of the samples generated by the different recruitment methods.¹⁸

Table 3 contains a comparison between the respondents recruited through banner advertisements and those recruited by the subscription method. We compare four samples of subjects: those recruited in the banner advertising campaign (“Banner”), those recruited in the subscription campaign who registered immediately (“Subscription”), the conversion of subscription campaign e-mail addresses that did not initially register (“Subscription conversions”), and subjects who found our project Web site by other means (“Other”).

The χ^2 values in Table 3 test whether the distribution of each demographic characteristic is the same for the different recruitment methods. From this table, it is evident that the different methodologies generate samples that have different characteristics. Importantly, of the 42 tests in Table 3, we see only 7 instances in which the χ^2 values are not statistically significant.¹⁹

There are three demographic characteristics in which we consistently see substantial differences between the different recruitment samples. The first regards gender: In all six comparisons, the χ^2 test for gender differences is strongly significant. The education levels are distributed differently across all six samples. Furthermore, for age, we see six highly significant χ^2 statistics. These results indicate that the different recruitment schemes generate very different demographic samples.

In fact, these results suggest that the recruitment methodologies appeal to different subsets of the population. Although this empirical fact is interesting, it does not represent a real concern unless the respondents differ in some unmeasured dimension. Because the

¹⁸The data used in this analysis are from all unique registrations received before 10 A.M. on Nov. 1, 2001.

¹⁹See Table A2 (available on the *Political Analysis* Web site) for detailed distributions.

objective is to generalize the results from this survey to the Internet population as a whole, we are primarily concerned with the possibility that either of these recruitment methods generates coverage or nonresponse error. If there is something inherent in the two recruitment methods that results in different sample selection, then this must be considered before we can make general statements based on the analysis of these surveys. For example, it is possible that one recruitment methodology generates more reliable survey responses, which is the topic of the next analysis.

3.2 *Predictive Validity and Measurement Reliability*

Using data from the ISAO political opinion survey that ran from March 23 to April 8, 2001, it is possible to examine and model evaluations of political figures and offices for both recruitment methodologies.²⁰ Our analysis of predictive validity and measurement reliability focusses on six different questions from this spring 2001 political opinion survey: three focused on institutional approval, two were feeling thermometers, and one was a vote recall question regarding the 2000 presidential election:

- Do you approve or disapprove of the way George Bush has been handling his job as president?
 - 1 Approve
 - 2 Disapprove
 - 9 Don't Know
- Do you approve or disapprove of the way the U.S. Congress has been handling its job?
 - 1 Approve
 - 2 Disapprove
 - 9 Don't Know
- Do you approve or disapprove of the way the U.S. Supreme Court has been handling its job?
 - 1 Approve
 - 2 Disapprove
 - 9 Don't Know
- Please rate the Democratic and Republican candidates for president in the November 2000 election using the feeling thermometer. You can choose any number between 0 and 100. The higher the number, the warmer or more favorable you feel toward that person. The lower the number, the cooler or less favorable you feel toward that person. You would rate the person at the 50 degree mark if you feel neither warm nor cold toward them.²¹
 - Al Gore (George W. Bush)
 - George W. Bush (Al Gore)

²⁰In March 2001, respondents who had registered during the banner advertisement campaigns were e-mailed and asked to complete a new political opinion survey. At the same time, new subjects recruited through the first subscription campaign were requested to complete the same survey. All subjects were e-mailed twice during the survey period. We obtained a total of 1492 valid responses to the survey, 1069 of which had e-mail addresses that could be integrated with the subject information database. Thus, we can use these data to compare and model responses based on how subjects were recruited.

²¹The question order was rotated so half the subjects were asked to rate Al Gore first and the other half rated George Bush first.

Table 4 Chi-squared comparison across recruitment methods—political opinion questions

<i>ISAO category</i>	<i>Banner v. Subs.</i>	<i>Banner v. Other</i>	<i>Subscription v. Other</i>
Approve of George Bush	19.31*	2.55	17.41*
Approve of Congress	22.83*	0.47	12.63*
Approve of Supreme Court	29.30*	0.95	10.72*
Identify political offices	247.43*	4.96	124.24*

*Indicates a χ^2 test that is statistically significant at the $p \leq .05$ level. This analysis considers only March 2001 survey respondents who are over 18. The “subscription conversion” respondents did not complete this survey, and are therefore excluded from this analysis.

- Regardless of whether or not you voted on November 7, whom did you support for president?²²
 - 1 Al Gore
 - 2 George W. Bush
 - 3 Ralph Nader
 - 4 Pat Buchanan
 - 5 Harry Browne
 - 6 Another Candidate
 - 8 Don't Know
 - 9 Did Not Vote

We begin by examining the responses to the institutional approval questions, contained in Table 4, for banner, subscription, and other subject samples.²³ From the χ^2 tests, it is clear that these subject groups differ in terms of their sentiments concerning how politicians are handling their jobs. In Table 4, we see that the subscription subjects clearly have different evaluations relative to both banner and other subjects; however, banner and other subjects do not have different evaluations. Specifically, banner recruits are more likely to disapprove of Bush's and Congressional performance, but to approve of the Supreme Court's performance, than subscription recruits.²⁴ These differences, although noticeable, are slight. Of particular importance, subjects recruited through the banner advertisements are less likely to answer that they do not know whether they approve or disapprove of political figures and offices. These differences in “Don't know” rates are the primary force driving the statistically significant χ^2 tests reported in Table 4. It is also interesting to note that the banner recruits are more politically informed when political informedness is measured by respondents' ability to identify the political offices of various public figures. This may be attributable to the fact that these respondents were recruited during the election campaign in 2000, and might have participated in an earlier survey.

Table 5 contains summary statistics for feeling thermometers and vote choice. Regarding the two feeling thermometers, we see minor differences. The banner recruits rate Bush higher (47.64) and Gore lower (44.71) than the other recruit types. Bush received the lowest thermometer score (39.97) from other recruits, whereas Gore received the highest score from the subscription recruits (50.97). On the vote recall question, of those claiming to support Gore or Bush, a slight majority supports Gore.

²²Only respondents supporting Gore or Bush are considered in subsequent analyses.

²³This analysis considers only March 2001 survey respondents who are older than 18.

²⁴See Table A3 (available on the *Political Analysis* Web site) for detailed distributions.

Table 5 Comparison across recruitment methods—thermometers scores and vote choice

<i>ISAO variable</i>	<i>Banner recruits</i>	<i>Other recruits</i>	<i>Subscription recruits</i>
Bush thermometer			
Mean	47.64	39.97	45.97
Std. Dev.	33.76	35.07	34.98
Total sample size	393	145	887
Gore thermometer			
Mean	44.71	48.92	50.97
Std. Dev.	30.17	32.22	34.50
Total sample size	387	146	895
Vote B v G (B = 1)			
Mean	0.49	0.43	0.45
Std. Dev.	0.50	0.50	0.50
Total sample size	416	155	1071

Note. Negative thermometer scores were recoded as zero. This analysis considers only March 2001 survey respondents who are over 18. The “subscription conversion” respondents did not complete this survey and are therefore excluded from this analysis.

Table 6 provides another examination of measurement reliability. Given that we do not yet have panel measurements on the subscription recruits, we cannot compare over-time stability in survey responses,²⁵ but we can examine the consistency of responses across different questions. The expectation is that some survey responses (e.g., the vote recall and the two feeling thermometers) should have strong and predictable correlations. If the recruitment samples produce correlations that are similar, we will be more confident in their respective survey responses.

The banner recruits are presented in the top panel, the subscription recruits in the middle, and the other recruits in the bottom panel. We see that most of the correlations are strong and in the expected direction, and that all of the correlation coefficients have identical signs across the recruitment samples. For example, we see a strong positive correlation between the Bush thermometer and the vote recall (probability of supporting Bush) in both samples (.86 for banner recruits, .87 for subscription recruits, and .89 for other recruits) as well as a strongly negative correlation between the Gore thermometer and the vote recall in both samples (−.79 for banner recruits, −.81 for subscription recruits, and −.81 for other recruits). We also see relatively strong negative correlations between Bush approval and the vote in both samples, as well as a strong negative correlation between the evaluation of Bush’s performance and the performance of the Republican Congress. All of these correlations help to underscore the conclusion that these samples of recruits are providing roughly similar levels of survey response consistency by this measurement strategy.

In Table 7 we present results on predictive validity using each of these six survey responses as dependent variables. The three approval variables had discrete responses (Approve, Disapprove, and Don’t know), so we used a multinomial logit model; the two thermometers were roughly continuous, so we used regression; and the vote choice was binary, so we used probit. In each case, we estimated a model including a wide array of variables on the right-hand side.²⁶ We tested whether these predictive models were statistically

²⁵Panel attrition will be discussed in future research.

²⁶These variables were dummy variables for independents, Republicans, men, whites, married respondents, residents of three regions (northeast, midwest, and south), Jews, Catholics, and Protestants; as well as categorical

Table 6 Measurement reliability

	<i>Vote</i>	<i>Bush</i>	<i>Gore</i>	<i>Bush app.</i>	<i>Con. app.</i>	<i>Court app.</i>
<i>Banner recruits</i>						
Vote B v G (B = 1)	1.00					
Bush thermom.	0.86	1.00				
Gore thermom.	-0.79	-0.80	1.00			
Bush app.	-0.86	-0.89	0.75	1.00		
Congress app.	-0.47	-0.43	0.36	0.53	1.00	
Court app.	-0.47	-0.44	0.38	0.50	0.60	1.00
<i>Subscription recruits</i>						
Vote B v G (B = 1)	1.00					
Bush thermom.	0.87	1.00				
Gore thermom.	-0.80	-0.76	1.00			
Bush app.	-0.79	-0.82	0.65	1.00		
Congress. app.	-0.49	-0.53	0.41	0.54	1.00	
Court. app.	-0.44	-0.47	0.33	0.44	0.55	1.00
<i>Other recruits</i>						
Vote B v G (B = 1)	1.00					
Bush thermom.	0.89	1.00				
Gore thermom.	-0.81	-0.76	1.00			
Bush app.	-0.91	-0.86	0.74	1.00		
Congress app.	-0.47	-0.43	0.33	0.52	1.00	
Court app.	-0.61	-0.62	0.47	0.57	0.40	1.00

Note. This analysis considers only March 2001 survey respondents who are over 18. The “subscription conversion” respondents did not complete this survey and are therefore excluded from this analysis.

the same when applied to each sample of recruits, and we report the predictive results in Table 7.

We see that in only 3 cases—the vote choice model for subscription versus banner and subscription versus other, and the Gore thermometer model for banner and other recruits—do our tests indicate that the different samples of recruits have different predictive models of these variables. In the other 15 cases, we reject heterogeneity across subject recruitment samples: strongly in 12 cases, but close to not rejecting in 3 cases.

The results in this section provide some analysis of the measurement and predictive validity of the subject recruitment strategies. The conclusion of this analysis is that although there are some differences of note, especially the decidedly higher rate of political awareness among the banner subjects, the recruitment strategies appear to generate samples that answered our surveys in very similar ways.

3.3 Comparison with the American Internet Population

In addition to the comparison across recruitment methodologies, it is possible to examine how the ISAO respondents’ demographics differ from those of the adult American Internet population. To obtain information about the American Internet Population, responses from the August 2000 Current Population Survey (CPS), which included the Computer and

variables for education, age, past economic perceptions, future personal finance perceptions, and political awareness. The actual model results are available on the *Political Analysis* Web site.

Table 7 Predictive validity—banner v. subscription recruits

<i>Model</i>	<i>Hypothesis tests</i>		
	<i>Test stat.</i>	<i>Degrees of freedom</i>	<i>p level</i>
Banner v. subscription recruits			
Bush approval (MNL)	$\chi^2 = 33.34$	32	0.40
Congressional approval (MNL)	$\chi^2 = 37.23$	32	0.24
Court approval (MNL)	$\chi^2 = 29.52$	32	0.59
Bush thermometer (OLS)	$F = 0.72$	15, 1246	0.77
Gore thermometer (OLS)	$F = 1.44$	15, 1245	0.12
Bush–Gore vote (Probit)	$\chi^2 = 24.49$	7	0.00
Banner v. other recruits			
Bush approval (MNL)	$\chi^2 = 25.71$	32	0.78
Congressional approval (MNL)	$\chi^2 = 39.49$	32	0.17
Court approval (MNL)	$\chi^2 = 33.06$	32	0.42
Bush thermometer (OLS)	$F = 0.95$	15, 505	0.51
Gore thermometer (OLS)	$F = 1.69$	15, 499	0.05
Bush–Gore vote (Probit)	$\chi^2 = 5.98$	7	0.54
Subscription v. other recruits			
Bush approval (MNL)	$\chi^2 = 45.00$	32	0.06
Congressional approval (MNL)	$\chi^2 = 44.82$	32	0.07
Court approval (MNL)	$\chi^2 = 41.48$	32	0.12
Bush thermometer (OLS)	$F = 0.69$	15, 1000	0.80
Gore thermometer (OLS)	$F = 1.14$	15, 1005	0.31
Bush–Gore vote (Probit)	$\chi^2 = 21.06$	7	0.00

Note. This analysis considers only March 2001 survey respondents who are over 18. The “subscription conversion” respondents did not complete this survey and are therefore excluded from this analysis.

Internet Usage Supplement, are used.²⁷ Before undertaking this comparison, however, it is important to consider the differences between the CPS and the ISAO.

The CPS is a monthly panel survey of about 60,000 households conducted by the Bureau of the Census. Unlike the CPS, the ISAO surveys individuals and not households. Therefore, to correct for overrepresentation of family members in the CPS survey, the data included contain only the primary household member. To identify individuals with Internet access, it was necessary to construct an indicator variable from two questions in the CPS supplement: “Does . . . use the Internet outside the home?” and “Does anyone in this household use the Internet from home?” A positive response to either of these questions resulted in a respondent being coded as having Internet access. CPS respondents were weighted by probability of selection for the survey with an adjustment for special sampling situations and failure to obtain interviews from eligible households.²⁸

Comparisons between these two surveys are further complicated by the fact that each survey asks slightly different questions and provides different response options. For example, instead of asking respondents a single question concerning race, the CPS asks several questions about race and ethnic identification. In order to overcome these differences it was

²⁷The August 2000 CPS is the most recent CPS that includes the Computer and Internet Survey Supplement, and is roughly comparable with the time frame of our surveys.

²⁸For more details about the 2000 CPS, see www.bls.census.gov/cps/computer/2000/sdata.htm.

Table 8 Demographic comparison between ISAO and CPS demographic variables

<i>ISAO variable</i>	<i>Banner recruits</i>	<i>Other recruits</i>	<i>Subscription recruits</i>	<i>Subscription recruits converted</i>	<i>Weighted CPS with Internet access</i>
Gender					
Female	48.06	59.53	76.84	80.84	52.3
Male	51.94	40.47	23.16	19.16	47.7
Age					
18–29	27.24	33.94	26.73	17.32	21.6
30–39	19.38	21.67	23.98	20.53	23.9
40–49	22.76	20.23	25.06	29.77	26.5
50–59	17.47	14.49	17.30	19.75	16.8
60–69	8.36	6.92	5.19	9.05	7.2
Over 69	4.79	2.74	1.75	3.60	4.1
Race					
White	77.52	85.90	85.72	89.01	80.8
Non-white	22.48	14.10	14.28	10.99	19.2
Education					
No high school	3.88	2.09	7.63	4.67	5.9
High school grad	19.00	12.14	33.03	30.64	26.8
Some college–voc.	36.88	34.20	38.82	38.81	32.4
College graduate	23.42	27.68	15.06	18.39	22.9
Graduate prof.	16.81	23.89	5.46	7.49	12.0
Marriage					
Married w/spouse	51.35	48.69	46.66	51.56	76.2
Other	48.65	51.31	53.34	48.44	23.8
Kids					
None	38.17	45.43	30.93	30.64	54.4
One	14.03	14.62	16.70	16.44	18.5
Two	21.95	20.37	23.47	24.42	18.6
Three or more	25.86	19.58	28.90	28.50	8.4
Region					
North East	16.44	17.49	16.98	18.00	19.7
Midwest	17.85	21.28	25.08	24.81	23.8
South	44.21	27.42	37.17	34.73	33.3
West	21.51	33.81	20.77	22.47	23.2

Note. This data is from unique registrants received before 10 A.M. Nov. 1, 2001. Only respondents over 18 are included in this analysis.

necessary to convert variables so that their values matched as closely as possible across data sets. For some variables, such as race and marital status, it was deemed appropriate to compress the response categories to ensure that this matching was as accurate as possible. However, it is still possible that some of the differences in distributions across surveys may be attributed to the differences in question wording.

Table 8 presents a comparison across the ISAO recruitment methodologies and the probability weighted estimates of demographic characteristics for the Internet population as measured by the August 2000 CPS. It is clear from this tabular comparison that the ISAO samples differ from CPS respondents with Internet access.²⁹

²⁹Studies suggest that the Internet population is changing rapidly (see U.S. Department of Commerce 1995, 1998, and 2000, for example). It is possible that some of the differences identified across these tests result from the

These differences across samples suggest that sample selection may be an important issue for volunteer Internet surveys, in particular, the ISAO. The observed differences across samples must either be attributable to coverage error or nonresponse error. If the Internet users who viewed the banner advertisement or the subscription window are representative of the Internet population—a fact that may be difficult to verify because no data are available for nonrespondents—then it may be possible to correct statistically for potential selection biases (see VanBeselaere 2001, for an example).

Correction for coverage error may be more difficult. The important issue to consider is how Web surveys compare with traditional telephone survey methodologies. From Section 2.1, it is clear that RDD is not a panacea for survey error. The fact that the ISA was only able to connect with and get interviews from 11.5% (1500/13,095; see Table 1) of the phone numbers in their original frame suggests that contemporary telephone surveys undoubtedly suffer from coverage error as well.

4 Conclusion

This paper provides a basic report about one of the early steps in Web survey techniques—obtaining subjects on the Internet for social science surveys. Specifically, we reported on our research using two different recruitment techniques: banner advertisement and subscription campaigns. We found that the subscription campaigns appear more efficient and cost-effective. We found little evidence to support the hypothesis that one method produced subjects with more reliable survey responses. In particular, the correlation structure among the variables typically found in public opinion or political behavior surveys appears similar across recruitment methods (see Tables 6 and 7). This is important because, as social scientists, we are interested in relationships among variables implied by particular theories. However, we did find that both subject samples seem different from the current Internet population (see Table 8). These differences may be resolved by weighting or using appropriate empirical methodologies that account for selection.

Clearly, more research on subject acquisition processes is needed. Given that these processes are being used by our own project as well as others, we must turn our attention to ways in which we can use these survey subjects in Web-based surveys. In particular, can we correct for the types of differences we see in the subject samples relative to other estimates of the Internet population? There has been some early work on this question (VanBeselaere 2001), and much more is necessary.

As we discussed previously, contemporary telephone survey technology using RDD is becoming increasingly problematic. As it becomes more difficult to get Americans to participate in telephone surveys, that type of survey technique raises serious questions about coverage and sampling errors. Web-based surveys, however, have coverage and sampling problems of their own; thus, Web-based surveys are not a simple substitute for telephone surveys. Yet Web-based surveys can be inexpensive, rapid, and can allow for interview modes that are impossible to implement in telephone surveying.

In the near future, Web surveys may very well overtake the current RDD telephone survey technology as the methodology of choice for social scientists. If that transition is to happen successfully, there are many questions about how Web surveys can be conducted that must be studied carefully and answered before Web surveys can be widely used. Given that many

changing profile of the Internet population. This is a particular concern for the comparison with the subscription recruits who joined the ISAO in 2001.

of the issues regarding Web surveys are so different from RDD methodologies and from other self-completion methodologies, perhaps it is appropriate to call for the development of new survey methodologies for Web-based surveys.

References

- The American Association for Public Opinion Research. 2000. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Ann Arbor, MI: AAPOR. (Available from www.aapor.org/ethics/stddef.html.)
- Berrens, R. P., A. K. Bohara, H. Jenkins-Smith, C. Silva, and D. L. Weimer. 2003. "The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples." *Political Analysis* 11:1–22.
- Brehm, John. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation*. Ann Arbor: The University of Michigan Press.
- Couper, Mick P. 2000. "Web Surveys: A Review of Issues and Approaches." *Public Opinion Quarterly* 64:464–494.
- Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. "Web Survey Design and Administration." *Public Opinion Quarterly* 65:230–253.
- Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. 2nd ed. New York: John Wiley and Sons.
- Federal Communications Commission. 2000. "Trends in Telephone Service." Industry Analysis Division, Common Carrier Bureau, Washington, DC.
- Federal Communications Commission. 2002. "Trends in Telephone Service." Industry Analysis Division, Common Carrier Bureau, Washington, DC.
- Green, Melanie C., Allyson L. Holbrook, and Jon A. Krosnick. 2001. "The Survey Response Process in Telephone and Face-to-Face Surveys: Differences in Respondent Satisficing and Social Desirability Response Bias." Manuscript, Ohio State University.
- Groves, Robert M. 1989. *Survey Errors and Survey Costs*, New York: John Wiley and Sons.
- Krosnick, Jon A., and Lin Chait Chang. 2001. "A Comparison of the Random Digit Dialing Telephone Survey Methodology with Internet Survey Methodology as Implemented by Knowledge Networks and Harris Interactive." Manuscript, Ohio State University.
- U.S. Department of Commerce. 1995. "Falling Through the Net." (Available from www.ntia.doc.gov/ntiahome/fallingthru.html.)
- U.S. Department of Commerce. 1998. "Falling Through the Net II: New Data on the Digital Divide." (Available from www.ntia.doc.gov/ntiahome/net2/falling.html.)
- U.S. Department of Commerce. 2000. "Falling Through the Net: Towards Digital Inclusion." (Available from www.ntia.doc.gov/ntiahome/digitaldivide/.)
- U.S. Department of Labor, Bureau of Labor Statistics. 2000. "Current Population Survey: Design and Methodology," Technical Paper 63. (Available from www.bls.census.gov/cps/tp63.html.)
- VanBeselaere, Carla. 2001. "Sample Selection Issues in the Internet Survey of American Opinion (ISAO)." Manuscript, California Institute of Technology.