15

The Future of Polling

Relational Inference and the Development of Internet Survey Instruments

JAMES WITTE AND PHILIP E. N. HOWARD

New media technologies have generated much excitement about ways of improving communication between leaders and citizens. During the last election cycle, pundits announced the arrival of electronic democracy—a political environment in which public deliberation and the measurement of public opinion would be facilitated by multimedia technologies like the Internet. Along with electronic town halls and e-mail listservs, Internet surveys and polls have been hailed as new ways of conducting and measuring public deliberation. For many practitioners, these innovations are akin to those that replaced the magnifying glass with the microscope in the biologist's lab. Most of the major survey houses have recast themselves as interactive, multimedia polling agents with the ability to employ a widening range of stimuli to help elicit responses from subjects.

Internet-based survey and polling techniques also require reexamination of arguments against polling in general regardless of the mode of data collection. There is a long tradition of doubting polling sciences, and their generalizability, of the utility of defining a "public opinion," and of the reduction of individual opinion to commensurable inputs (Binkley 1928; Crossley 1937; Gosnell 1937; Katz and Cantrill 1937; Lohman 1988). Some criticize polling for reducing public opinion to a distillate of unrelated individual opinions based on snapshot responses to simple close-ended questions. Political ethnography has long recognized that public opinion is a social process frequently led by a few individuals with strong opinions and persuasive abilities (Mansbridge 1983; Herbst 1993). In this view, a

sample design that "democratically" assigns equal weights to individual responses arithmetically ignores the relative social stature of respondents (Converse 1987). Traditional polling techniques placed this kind of social information outside the research frame of empirical study; it was considered auxiliary, if not immeasurable. However, the large and potentially diverse sample frame of the Internet has made it possible to expand the theoretical sample of "opinion leaders" and permits some Internet polling firms to claim to be able to measure opinion leadership by building a peer-review process into the survey design itself.

Other chapters in this collection have discussed polls and other tools of public opinion measurement. Here, we are specifically concerned with the role of contemporary survey methods and technologies in honing the science of opinion assessment and expanding the range of applications for the art of political polling. We begin with a brief discussion of Internet survey instruments and their advantages and disadvantages. A review of historical lessons learned in polling follows, leading to a discussion of the challenges presented by modern techniques. We highlight the methodological issues associated with sampling public opinion, emphasizing that the nonrandomness of most Internet surveys needs to be viewed in light of a realistic assessment of the randomness of other survey sampling methods, particularly telephone surveys.

In the next section of this chapter, we illustrate the significance of sampling techniques for Internet surveys, using data from Survey2000, a large-scale Internet Survey conducted on the National Geographic Society's (NGS) Web site in fall 1998. Over 80,000 individuals participated in the study worldwide, yielding more than 50,000 completed surveys. Respondents were asked a series of questions to determine patterns of individual mobility over the life course, attitudes toward community, patterns of social contact, and mutual assistance and questions about their musical, literary, and culinary tastes. Though the Survey2000 sample is basically a large convenience sample of visitors to the NGS web site, outreach efforts were conducted to diversify the sample. In fact, over 400 respondents indicated that the first time they had gone on-line was to complete the survey. Moreover, a number of the Survey2000 questions were based on existing survey instruments, most notably the General Social Survey (GSS), as a means to provide external benchmarks of the sample's representativeness. The chapter concludes by emphasizing that representativeness is the aim of all sampling techniques—random and nonrandom—and suggesting that a well-reasoned approach to relational inference is the key to generalization, regardless of the sampling technique employed.

AN INTRODUCTION TO INTERNET SURVEY INSTRUMENTS

There are several clear advantages to Internet survey instruments. Even some of the more traditional polling firms are staking their future on these new developments, renaming themselves and rethinking the way they do business. First,

they allow researchers to collect huge samples, some of hundreds of thousands of respondents. This makes it easier to purposively subsample special sections of the population on a range of variables, from demography, medical conditions, and mobility history to musical, literary, or culinary tastes. Second, the instrument itself permits a range of interesting testing stimuli, allowing researchers to control a respondent's exposure to different media and the flow of information to the respondent, such that Internet survey instruments take on the better features of traditional qualitative research methods. Firms and lobbyists test entire commercials, product packaging, and political arguments on population segments. For example, one firm expects to help political consultants estimate when the public will call a candidate "too negative" by testing small populations for their reactions to entire campaign ads. Ideally, campaigns will be able to compose ads with a range of strong messages, and interactive surveys will help them pick the strong and aggressive, but not too negative ads. Third, Internet survey instruments extract richer information from respondents. Traditional question types, new question types, and the quality of information from open-ended questions can exceed that from Computer-Assisted Telephone Interview (CATI) or personal interviews. This point is well illustrated by the literature section of Survey2000. Questions intended to measure respondents' knowledge of local literature were customized to include the names of authors associated with each respondent's state of residence, including current residence and past residence. Even a well-trained interviewer would be hard-pressed to follow the related questionnaire skip patterns. Similarly, when Survey2000 respondents were asked about their preference for particular music genres, they were not only told the name of the genre but were also given the opportunity to listen to a short audio clip representative of that genre.

Fourth, Internet techniques can quickly reconnect with respondents should follow-up be necessary, whereas both the time involved in the initial survey and any follow-up is often less lengthy than that of other methods. Fifth, more complex experimental designs (e.g., testing for instrument or measurement effects) can easily be accommodated with Internet survey instruments. An experiment of this type will be conducted in Survey2001 using questions, about respondents' attitudes regarding endangered and threatened species. Following the model of the Survey2000 literature questions, it will ask respondents to indicate their support for preserving species associated with their state of residence. Moreover, while the standard question prompt also includes a photographic image of the species, randomly selected respondents will be presented the question without the image prompt. These questions concern the extent to which attitudes toward the environment are influenced by perceptions and emotions surrounding particular species rather than a consistent environmental attitude. The experimental component will consider the extent to which visual images increase the likelihood of inconsistent response patterns. Sixth, the interactive potential of Internet surveys may lead respondents to feel more invested in the research. They tend to provide helpful critiques of the surveys, and as the media allows results to be

quickly returned to respondents, they may revisit the project to learn about the results of the research in which they participated. Here, too, Survey2000 provides a useful example. A comment and suggestion e-mail icon was included on every page of the survey. During the ten-week data collection period, nearly 2,000 e-mail comments and suggestions were received, many constructively addressing the survey technology and content.

However, there are also disadvantages to Internet survey instruments. First, the survey environment created at the terminal comes at the expense of a survey environment regulated by an interviewer. Although interviewer effects are known, it is less known if the environment in which an Internet survey instrument is administered affects response patterns. One of the advantages of having questions asked by a human interviewer is that the agent can gauge the degree to which the respondent is answering without prompting. Many of the new techniques for home surveying are designed to be innocuous, increasing the possibility that something in the environment, uncontrollable by the survey managers, may affect the respondent. A challenge for the future will be learning how to monitor and mitigate instrument and environment effects in new survey contexts.

A related concern raised about Internet surveys and polls is the possibility that interest groups might attempt to deliberately distort the outcome, encouraging members to vote "early and often" to ensure an outcome favorable to the group's position. Thinking about this concern in terms of stuffing "the ballot box," however, places it in perspective. Just as a minimum of standard election procedures make it difficult to manipulate a paper ballot, features common to all but the simplest of on-line surveys and polls are sufficient to thwart the electronic equivalent. Few on-line polls and surveys today consist of simple electronic documents that could be readily copied and anonymously submitted. Programming a certain level of interactivity into a survey or poll not only makes the instrument engaging, but also ensures that a live respondent, or at least a semi-intelligent program, is on-line to complete the instrument. To return to the paper ballot analogy, stuffing an electronic ballot box requires getting in line each time to submit a ballot. Clearly, an electronic survey or poll is not monitored by an election judge, who could spot repeat voters. However, identifying information (e.g., IP address, host name, browser labels, and so on, not to mention "cookies" stored on a respondent's computer) regularly exchanged as part of being on-line is available to identify and exclude efforts to overrun the instrument with many responses from a single source.

The most obvious potential limit of Internet survey methods concerns sampling, discussed at length in the following sections. At this point, however, one point is worth emphasizing: telephone samples, face-to-face survey samples, and Internet survey samples are all the products of considerable self-selection. Quotas and stratified sampling assure "representative" findings in terms of key demographic criteria, but high levels of respondent refusal conceal biases that remain embedded within demographic categories. However, in traditional survey methods, these concerns are largely ignored. These methods have acquired a taken-

for-granted legitimacy that may never have been warranted, based primarily on avoiding three important lessons about mistaken sampling.

HISTORICAL LESSONS IN POLLING

In each political campaign, pollsters learn most from the errors of previous campaigns. Over time, pollsters have learned three basic lessons. Early on, the industry learned to improve techniques for sampling the right population, then they learned about the importance of sampling for interesting populations, and more recently they learned about the importance of sampling for a trendline.

The reader might recall the famous *Literary Digest* presidential election poll that predicted Landon's victory over Roosevelt in 1936. This poll had a sample size of more than 2 million, but still came to the wrong conclusion. The *Literary Digest* poll made no effort to assess the representativeness of its sample, and it clearly sampled the wrong people. But an oft-overlooked fact is that in 1920, 1924, 1928, and 1932, the *Literary Digest* used the same techniques to accurately predict the outcome of presidential elections. Beyond the truth that a large sample does not guarantee accurate results, a nonrandom sample does not amount to a recipe for invalid results. The most important lesson is that sampling must always be assessed; it can be random or purposive, *but it cannot be purposeless.*

In contrast to the *Literary Digest* debacle, the National Election Study came under attack for not sampling a population many people considered most interesting after the 1964 presidential election of Lyndon Johnson. One of the main new political forces of the time was the John Birch Society, and scholars interested in studying this radical new right-wing group were unhappy to find that the national sample included one member of the John Birch Society. In their defense, survey organizers argued that a perfect national sample would have at most two members of the John Birch Society, and that no other national survey had caught the rise of this conservative grassroots movement in the previous four years. Nonetheless, academics with good theoretical reasons for being interested in the John Birch Society were dissatisfied with a representative survey that failed to capture the impact of an influential political group (Converse 1987).

The practical restrictions of traditional polling techniques permit researchers to take a snapshot of respondent demographics, behavior, or preferences. Although two data points *can* make a trendline, they do not always. In 1980, pollsters correctly predicted a Reagan victory using snapshot data, but they did not construct a useful trendline, such that the margins of his victory were significantly underestimated. In the last few days before the election, there was a shift in popular support for Reagan that the polls did not catch. If the shift had been in the other direction, this polling error might have achieved the same infamy as the *Literary Digest* error.

MODERN TECHNIQUES, MODERN CHALLENGES

Telephone surveys, particularly with the introduction of CATI systems, transformed public opinion research. Today, however, telephone survey research has become increasingly problematic, most obviously due to low response rates. A concern over declining response rates is not new among survey researchers; however, recently this issue has also captured broader public attention with articles such as the November 1999 *New York Times* piece titled "Polling's 'Dirty Little Secret': No Response." A low response rate will increase the cost of a survey but does not necessarily undermine its validity. A real problem, however, occurs when nonresponse is not randomly distributed but is instead correlated with respondent characteristics. Weighting and statistical adjustment provide possible corrections when differences in response patterns are observed, but heterogeneity based on unobserved criteria is equally problematic. In addition, from a theoretical sampling perspective, the increasing prevalence of multiple phone lines raises a new question. This clearly breaks down the assumption of a one-to-one correspondence between telephone numbers and respondent households that has been the methodological mainstay of adherents of RDD (random digit dialing).

In the election of 2000, polls again proved their worth as a democratic institution in their own right by providing succinct, albeit simplified, measures of public preferences (Crespi 1989; Gallup and Rae 1940). The campaigns of all the major candidates, especially those who could afford good pollsters, devoted significant resources to assessing public policy priorities. Along with internalizing results and reshaping their platforms, political consultants also used polls to "prime" their candidate, to confirm his popularity, and to help them select which campaign ideas to emphasize (Jacobs and Shapiro 1994a). Organizationally, polling for the presidential candidates was conducted by trusted professional advisers with clear ideological affiliations with their teams. Poll results were used to demonstrate popularity, to evaluate and improve the effectiveness of candidate's media strategy, and to review and challenge media coverage. Internally, polling was also used by powers within campaigns to guide strategic decisions about campaigns and policy (Altschuler 1986; Jacobs and Shapiro 1995b).

Often viewed as a means of reading public opinion, polling clearly has also become a tool to shape public opinion. Good internal polling by the Bush presidential campaign in 1988 helped staffers to realize that public knowledge of what Dukakis stood for was minimal, so Republicans might define him to their advantage by creating a negative image of him (Richard 1994). Push polls, developed more as a way of planting ideas with specific publics than as a means of measuring preexisting opinion, have become a common phenomena. Most pollsters are embarrassed by this application of polling techniques, but they are still used. Even in the last ten days of the 2000 campaign season, stories emerged of a Republican push poll asking residents of New York if they would vote for Hillary Clinton, even if they knew she had accepted campaign contributions from an Islamic group

with members who sympathized with the use of violence in the Middle East. Given that terrorist bombing of the *USS Cole* was still in the news, this tactic tried to associate Clinton with terrorist sympathizers under the guise of testing the resilience of Clinton supporters. The immense political implications of interpreting—or misinterpreting—exit polls were lived out during the presidential election of 2000. The politics of who called what for which campaign and for what reason, with little regard for estimable margins of error, will be debated for years to come.

Over the course of the 2000 election, Internet survey instruments allowed pollsters to refine the science of politics in three ways. First, traditional measures of political involvement—listing forms of political involvement, looking for respondent reaction to public issues and personalities—were greatly refined. Second, the strategic formula for legislative success developed by lobbyists increasingly employed these refined polling instruments. Lobbyists now claim to have growing insight into statistical relationships that connect public opinion, voter turnout, membership in political action groups, political volunteerism, donorship patterns, correspondence with legislative representatives, and legislative outcomes. The refined science of polling is now applied to both test and prod the progression of issues: from their conceptualization in the lobbyist's boardroom to the seeding of grassroots movements and the targeting of issue campaigns in specific congressional districts. Finally, political pollsters can now gather such detailed information and match this information with existing databases, so that they no longer offer demographic analysis to issue and candidate campaigns, but "psychographic" analysis that includes detail about voter sophistication and how people learn and process political information.

METHODOLOGICAL ISSUES IN SAMPLING PUBLIC OPINION

This section emphasizes polling issues associated with sampling, coverage, and nonresponse, that is, who is polled. However, the full range of methodological issues in research, including instrument, interview, environment and measurement effects, potentially take on a new cast in an on-line polling environment.

Who Is Polled: Sampling, Coverage, and Nonresponse in Telephone Surveys

Though there are a variety of types of errors associated with survey research, including coverage error, measurement error, instrument effects, interviewer effects, and nonresponse, issues associated with sampling error have clearly received the most attention from survey researchers. Not surprisingly, as Herbst (1993) points out, this is the type of error that survey researchers are best able to measure and quantify, though not necessarily the most serious type of error.

Beyond sampling error, other types of errors related to sample selection are

no less serious in their consequences; however, they are more difficult to detect and remedy. For example, standard practices regarding sampling error rest on an assumption of probability sampling and complete and adequate *coverage* of the population of interest. Clearly Internet-based surveys and polls face their greatest challenge here, but one ought not assume that traditional survey techniques have at last resolved questions of coverage. Coverage refers to the extent to which a sampling design ensures that all individuals in a population have a known probability of selection into a sample. Simply thinking in terms of sample selection—ignoring the issue of nonresponse for the moment—questions of coverage have typically been treated as a tractable problem. Not everyone has a telephone and, to the extent that surveys and polls are interested in samples that represent all segments of the population, alternative data collection methods (e.g., door-to-door surveys) may be used to collect data from those who would otherwise not be contacted.

On the other hand, new telephone technologies and user habits (answering machines, call forwarding, cell phones, and multiline households) have raised a whole new set of concerns regarding coverage. A description of these concerns is presented in considerable detail in guidelines established by the American Association for Public Opinion Research (AAPOR 2000). Though some researchers follow the explicit guidelines, a cavalier approach to the problem is certainly more common. Simply think back to the last few times you were contacted by a telephone survey or poll. How often were you asked about the number of phone lines associated with your household or your use of call forwarding? Rarely, if at all? However, this information is essential to maintain the integrity of RDD sampling procedures. For example, studies conducted in 1997 and 1998 found that about one-third (31.3%) of households nationwide and about one-quarter of households (25.7%) in the Pacific Northwest have two or more telephone lines (Oregon Survey Research Laboratory 1998). Industry sources in another twelve-state region of the United States indicate that multiple line penetration is currently at 17% with another 6.5% of the households subscribing to calling features that associate multiple numbers with a single line. In addition, as one would expect, the prevalence of multiple phone lines was not randomly distributed but was positively correlated with income, education, and home ownership. Though these studies did not explicitly ask whether there was a unique number or a distinctive ring associated with each line, multiple phone lines pose a significant challenge to any RDD study that simply assumes a single selection port for each household. Moreover, in some areas where multiple phone lines are included as part of a local calling package, household members may themselves be unaware of the fact that their household is associated with multiple phone numbers.

Furthermore, once a sample has been selected, surveys and polls still need to confront issues associated with response and nonresponse. Some of these issues are definitional; for example, AAPOR (2000) distinguishes six different methods of calculating responses rates. The rate may vary, on the one hand, with whether one includes partial interviews with complete interviews in the numerator. On

the other hand, it will vary based on the definition of the denominator, which contains partial and complete interviews as well as the number of noninterviews (refusals and break-offs) and cases of unknown sample eligibility. Not only is the line between partial interviews and break-offs necessarily arbitrary but also caller identification and answering machine technology have blurred the line between refusals and instances of unknown sample eligibility. Finally, the calculated response rate will also vary with the assumptions made about survey eligibility among households and individuals for whom there is insufficient information to establish eligibility.

To think about this from a practical perspective, consider the pool of randomly selected numbers needed to complete a telephone survey of a given size. For example, using equal probability sample selection methods and assuming a medium-length interview on a nonsensitive topic, nearly 7,000 randomly selected phone numbers would be needed to obtain 1,000 complete surveys.[3] In other words, on average, seven numbers yield one completed survey, so we can assume that significant nonrandom effects deplete the value of random selection because nonrandom instrument, interviewer, or environment effects must account for the remaining six incomplete surveys.

Focusing particularly on refusal rates as a key aspect of nonresponse, it is widely accepted that nonresponse is not random. Accordingly survey researchers and pollsters invest considerable effort in call-backs and conversions, prepaid incentives, and advance letters (Frey 1989, Dillman 2000, Singer, Van Hoewyk, and Mahar 2000). Interestingly, one recent experiment has found that while rigorous efforts were able to significantly boost response rates (from 36.0% to 60.6%), there were few differences in response patterns found primarily in demographic characteristics not in substantive areas (Keeter et al. 2000). Nevertheless, even with rigorous efforts, approximately 40% remain nonrespondents, and more than one in four of those who were contacted and eligible refused to be interviewed. Moreover, the apparent similarities between ready and reticent respondents does not preclude real and important differences between these groups and those who remain nonrespondents.

In sum, sampling error, coverage problems, and nonresponse bias are all ways in which traditional, telephone survey, and polling methods commonly fail to live up to the assumptions of probability sampling. In evaluating the potential of Internet-based methods, our point of comparison should not be the textbook ideal for those standard methods but rather everyday practices in the field.

Who Is Polled: Sampling, Coverage, and Nonresponse in Internet Surveys

Until now, Internet-based surveys and polls that seek to tap public opinion were necessarily limited because only about half of Americans have Internet access, while nearly one-third of those not on-line say they have no plans to gain access (Howard et al. 2001). Even though many of the Internet holdouts are older, it is

unlikely that Internet market penetration will reach that of the telephone any time soon. Moreover, turning from the general population to the Internet population, there is no mechanism to specify members of the on-line population, for example, through e-mail addresses, that would be exhaustive and mutually exclusive to serve as the basis for a probability sampling procedure. In this regard, any discussion of sampling error is meaningless. Similarly, discussions of coverage and nonresponse are limited when one cannot define the size of the population or specify its members.

Against this backdrop, one ought to consider an important advantage of Internet-based polling methods: Internet-based methods can survey large numbers of respondents at relatively low costs. Internet-based polls are self-administered, so there are no variable interviewer costs, only fixed programming costs. Internet-based delivery, as compared to traditional postal delivery, is essentially free. Moreover, with an Internet-based system, the pollster need not consider the effect of sample size on data entry costs, as there are no data entry costs, or more accurately, these costs are borne by the respondents who enter the data in the course of completing the poll or survey.

It is easy to underestimate the benefits of increased sample size, if one forgets that polls and surveys are often interested in more than the overall outcome. More often than not, analysts are interested in subgroups in a population and not just the population at large. For analyses of this type, the potential advantages of Internet-based polls and surveys are striking. For example, the widely used GSS in 1993 had only 179 African Americans, while the GSS African American oversamples in 1982 and 1987 included 354 and 353 African American respondents, respectively. Even in 1996, when the GSS sample was doubled, African Americans number only 402. Survey2000, on the other hand, resulted in 538 surveys completed by African Americans at a fraction of the cost of the GSS. Although Survey2000 makes no claims to randomness, the diversity of its sample certainly adds richness. Moreover, this tradeoff needs to be considered in light of a realistic assessment of the extent to which traditional sampling methods are increasingly falling short of the assumption of randomness.

A Nonrandom Road to Representativity

The goal of a poll or a survey is to collect data on a sample that represents a population. Randomness does not guarantee representativeness; rather, it provides the means to quantify the level of confidence with which one can say that the sample *does not* represent the population. Moreover, across the social sciences, the random selection of research subjects is by no means the only approach to generalizability. If one were to discount all findings not based on a random sample, then one would be choosing to ignore much of the historical and comparative work that makes up a sizeable segment of the social science literature. Whereas survey researchers, pollsters, and demographers emphasize randomness nearly to the point of fetishism, others are more catholic in their approach. Trochim (2001)

describes the random sample approach to generalizing as one of two primary approaches:

> I'll call the second approach to generalizing the Proximal Similarity Model. . . . The term proximal similarity was suggested by Donald T. Campbell as an appropriate relabeling of the term external validity. . . . Under this model, we begin by thinking about different generalizability contexts and developing a theory about which contexts are more like our study and which are less so. Notice that here, we can never generalize with certainty—it is always a question of more or less similar.

Like the notion of relational inference discussed at the end of this chapter, the proximal similarity model depends on a theoretically informed assessment of the nonrandom nature of a sample to avoid the pitfalls of a pure convenience sample. The *Literary Digest* polls made no effort to assess the representativeness of its sample, whereas Internet surveys such as Survey2000 explicitly incorporates items designed to measure the gradient of similarity. Such surveys collect data on standard demographic characteristics (e.g., gender, age, race, education, etc.). Combinations of these attributes for the sample can be compared to data collected by standard techniques. In addition, a proximal similarity framework may include other factors—such as attitudes and values toward community and cultural preferences—that cut across standard demographic variables. For this reason, a number of items used in Survey2000 come from the 1993 GSS to provide an external benchmark.

No matter how large the sample, size never guarantees representativeness. A Web survey such as Survey2000, which basically relied on a modified convenience sample, will not yield a random sample, and we will not "know" the selection probabilities for sample members. However, this *does not mean that the survey cannot yield representative social science data.* In fact, a sample of any size may be representative. Focus groups often include well under a dozen members; a single key informant may accurately represent an entire group. The advantage of a random sample design (where the intent is to use a random sample and then quantify the probability with which one's sample does or does not represent the population) is that sufficient sample size may be determined more exactly. Moreover, when one moves from sampling in theory to sampling in practice, a random design is no guarantee that the realized sample will be random. In the following section, we highlight results from two recent surveys to illustrate the significance of sampling strategies for Internet-based surveys.

THE INTERNET AND COMMUNITY: WHAT YOU SEE DEPENDS ON HOW YOU SAMPLE

The Internet itself has been one of the first areas explored using Internet survey techniques. At first glance, however, findings from these efforts are difficult to reconcile. On the one hand, a recent paper based on the Survey2000 data argues that "people's interaction online supplements their face-to-face and telephone communication, without increasing or decreasing it. However, heavy Internet use is associated with *increased participation* in voluntary organizations and politics" (Wellman, Haase, Witte, and Hampton 2001). By contrast, several months earlier, Norman Nie, political scientist and SPSS statistical software founder, made national headlines when he announced, "The Internet could be the ultimate isolating technology that further reduces our participation in communities even more than television did before it." Our contention is that these different findings can be traced back to each study's different approaches to sampling.

Although Survey2000 does not claim to be a random sample, Nie and others describe the InterSurvey sample as a national random sample of 4,113 American adults in 2,689 households. Respondents were provided with free Internet access and WebTV connections to facilitate the survey. In addition, "to avoid contamination of the results due to the fact that the study was itself conducted over the Internet, the results on Internet use presented in this study are based ONLY on the responses of participants who had Internet access prior and independent of the WebTV access installed by InterSurvey" (Nie and Ebring 2000). Even though the InterSurvey sample offered WebTV access to a random sample of individuals, there is no guarantee that those who accepted the company's offer to participate in the survey constitute a random sample. As in the case of a telephone survey, unobserved selection processes may dramatically alter the random character of the sample. For example, an individual with established and regular patterns of Internet use and on-line interaction would presumably be far less willing to adopt a new method of access than a less committed user. In short, just as those Internet users who frequent the National Geographic Society's homepage, which was the primary point of access for Survey2000, are not a perfect mirror of American society, so too those individuals who adopt WebTV as a means to access the Internet are not necessarily a random subset of Americans, even if they were randomly selected to be offered access.

To assess the distinctiveness of those who use WebTV to access the Internet, the subset of 359 Survey2000 respondents who were WebTV users will be compared to all other Survey2000 respondents. Table 15.1 indicates clear differences in social attitudes and engagement, as well as in political participation, between WebTV and non-WebTV respondents in the Survey2000 sample. For example, as shown in the top section of table 15.1, all Survey2000 respondents were asked whether they agreed or disagreed with three statements that indicate attachment to one's community. WebTV respondents were significantly less likely to agree with the statements that they felt close to their community or saw their community as a source of comfort. Meanwhile, WebTV respondents were significantly more likely to indicate that their daily activities did not create something of value for their community. When we compare non-WebTV respondents, these differences are not only statistically different but also each is of a magnitude of about 10%, suggesting real, substantive differences between the two segments of the sample.

Table 15.1 Differences in Social and Political Participation between WebTV Users and Other Survey2000 Respondents

Social and Political Participation	*WebTV Respondents (%)*	*Non-WebTV Respondents (%)*
Agreeing with the statement:		
I feel close to other people in my community.	59.1**	68.6
My daily activities do not create anything worthwhile for my community.	41.3**	30.7
My community is a source of comfort.	59.8**	68.3
Index of social participation (mean value)[a]	3.6**	5.1
Political participation in the past 12 months:		
Signed a petition	65.0*	59.4
Attended a public/town or school meeting	29.5**	38.5
Wrote an elected official	32.6	35.1
Attended a political rally or speech	12.4*	17.2
Served on a local organization committee	12.7*	16.9
Served as a club or organization officer	17.7**	29.6
Worked for a political party	6.7	5.8
Wrote a letter to a newspaper or magazine	22.0	20.2
Gave a speech	15.6**	27.9
Was a member of a group for better government	17.2	13.8
Wrote a newspaper or magazine article	9.5*	13.2
Engaged in political discussion on the Internet	27.7**	20.2

Survey2000 data based on 359 WebTV respondents and 33,851 Non-WebTV respondents for value items, 367 WebTV respondents and 34,755 Non-WebTV respondents for the index of social participation, and 349 WebTV respondents and 32,745 Non-WebTV respondents for political participation items.

[a] Index of social participation based on membership and active membership in 20 different categories of organizations and social groups.

*p. < .05.

**p. < .01.

The middle section of table 15.1 contrasts WebTV and non-WebTV respondents with regard to the mean value of an index of social participation for each segment of the sample. All Survey2000 respondents were queried about participation in twenty different types of groups or social organizations.[3] Further, respondents were asked to identify whether they were members or active members—an active member holds a leadership position, contributes money beyond regular dues, or regularly attends meetings. The index of social participation is the sum total of reported memberships, with active membership receiving double the weight of simple membership. As table 15.1 shows, WebTV respondents participated in significantly fewer organizations (1.5 fewer organizations on average) than non-WebTV respondents. Moreover, this finding remained robust with various specifications of the index of social participation (e.g., with log and exponential transformations of the totals, as well as with varied weights for regular and active membership). Regardless of the specification, the *t* statistic for the estimated difference never fell below 4, approximately double that needed to reject the null hypothesis at the conventional .05 level.[4]

The lower section of table 15.1 summarizes reported political participation in the past twelve months, also contrasting WebTV and non-WebTV Survey2000 respondents. In this instance, for four of the twelve reported forms of political participation, there is no significant difference between WebTV and non-WebTV respondents. Moreover, with regard to two forms of political participation (signing a petition and engaging in political discussion on the Internet), WebTV respondents are actually more involved than others in the sample. Nonetheless, for some of the more active and engaged forms of participation (attending public meetings, attending rallies and speeches, serving as local organization officers or committee members, giving a speech, or writing articles for newspapers or magazines), higher levels of participation occur among non-WebTV respondents than among WebTV respondents.

One plausible explanation for the differences between the WebTV and non-WebTV segments of the Survey2000 sample, as summarized in table 15.1, could be that the demographics of the two samples differ systematically. Indeed, we expect this to be the case; table 15.2 summarizes clear demographic differences between the two segments of the samples. WebTV respondents are significantly less likely to be women and are more heavily concentrated in the older age cohorts than non-WebTV respondents. Moreover, under a third of the WebTV respondents have a bachelor's degree or higher, whereas well over half the non-Web TV respondents have reached this level of educational attainment. Consistent with the observed age differences, WebTV respondents are also more likely to be employed and to live in households with children under the age of nineteen than non-WebTV respondents.

However, as table 15.3 indicates, the differences between WebTV and non-WebTV respondents extend beyond the demographic criteria commonly employed in the selection of a stratified sample. Table 15.3 reports the estimated coefficients obtained when the index of social participation is regressed on the set of demographic characteristics summarized in table 15.2, along with a variable indicating whether a particular respondent belonged to the WebTV segment of the sample. In table 15.3, we see that each of the demographic variables has a significant effect on the predicted value for the index of social participation. Participation in social organizations is higher among women than men. Membership is also greater among younger and older individuals than among individuals in the reference age group (25- through 34-year-olds). Compared to individuals with some college but no degree, respondents with lower levels of education participate at lower rates, whereas individuals with a bachelor's degree or greater participate at a higher rate. Furthermore, respondents engaged in full-time employment participate at lower rates than individuals who are employed part-time, are retired, or are in any other employment status. The index of social participation is also significantly higher among individuals living with children age 18 or younger. However, even after controlling for this set of significant de-

Table 15.2 Differences in Demographic Characteristics between WebTV Users and Other Survey2000 Respondents

	WebTV Respondents (%)	*Non-WebTV Respondents (%)*
Female respondents*	39.3	48.9
Age cohort*		
16–19	2.9	7.0
20–24	5.3	9.6
25–34	22.3	25.0
35–44	20.5	23.4
45–54	22.9	20.8
55–64	12.2	9.6
65 and older	14.1	4.7
Educational attainment*		
High school degree or less	23.9	12.9
Some college but no degree	43.6	31.1
Bachelor's degree or higher	32.5	56.1
Employment status*		
Full-time employed	52.5	61.6
Part-time employed	7.8	14.7
Retired	18.9	16.4
Other (laid off, unemployed, military, homemaker)	20.8	7.3
Live with children 18 or younger in household*	15.8	29.4

Percentages for each type of respondent. Survey2000 data based on 359 WebTV respondents and 33,851 Non-WebTV respondents for value items.
*$p < .01$, for difference between types of respondents.

mographic characteristics, being a WebTV respondent has a strongly significant, negative effect on the index of social participation. Moreover, the effect size of the estimated WebTV respondent coefficient (−0.930) is not trivial. Larger than the effect of gender, employment status, and household composition, it is equal to roughly two-thirds the overall difference between the average WebTV and non-WebTV respondents, as reported in table 15.1.

Efforts to reconcile the different views of the Internet suggested by the InterSurvey study and Survey2000 illustrate just how little is known about using the Internet to gather survey data. Differences in sampling strategies are an easy and obvious explanation for these two different views of the impact of the Internet. However, given how little experience we have with Internet survey methods, which are so new that there has been little systematic study of their operation, we should be cautious about simply accepting the easy and obvious. Years of experience and detailed analyses with other modes of survey data collection (e.g., face-to-face, telephone, paper-and-pencil self-administered) point toward a host of instrument and mode effects (cf, Dillman 2000). Similar issues probably affect the results of Internet surveys, such that it would be premature to attribute all differences between studies to sampling issues.

Table 15.3 Estimated Ordinary Least Squares Coefficients for Survey2000 Index of Social Participation Regressed on Demographic Characteristics and WebTV Respondent Status

Respondent Characteristics	*Unstandardized Coefficient*	*Standard Error*
Constant	2.351*	0.090
Female respondents	0.258*	0.045
Age cohort		
16–19	4.031*	0.118
20–24	0.919*	0.088
25–34 (reference category)		
35–44	0.940*	0.065
45–54	1.558*	0.066
55–64	1.889*	0.090
65 and older	2.487*	0.144
Educational attainment		
High school degree or less	−0.954*	0.220
Some college, but no degree (reference category)		
Bachelor's degree or higher	1.444*	0.167
Employment status		
Full-time employed (reference category)		
Part-time employed	0.853*	0.070
Retired	0.260*	0.116
Other (laid off, unemployed, military, homemaker)	0.462*	0.067
Live with children 18 or younger in household	0.880*	0.052
WebTV respondent	−0.930*	0.220

Survey2000 data based on 359 WebTV respondents and 33,851 Non-WebTV respondents for value items.
*$p < .01$.

CONCLUSION: NEW TECHNOLOGIES, NEW METHODS, AND RELATIONAL INFERENCE

Internet-based survey instruments have the great advantage of widening the range of stimulus for respondents. However, survey designers hoping to take advantage

of the new technologies must still face the challenge of building representative samples. From InterSurvey's WebTV sample to the Survey2000 on-line sample and the Harris Poll Online panel, which currently claims 7 million volunteer e-mail respondents, new technologies have made it possible to deploy increasingly creative and complex survey instruments and build enormous samples of respondents. For a growing number of Americans, Internet tools have become a significant conduit of their social life, allowing people to build their social networks by extending and maintaining friend and family relationships (Howard, Rainie, and Jones 2001). Pollsters are also looking at new ways to exploit Internet sociability to improve snowball samples through techniques such as affinity networks.

Further use of on-line surveys is inevitable, but researchers still need to think critically about the process of survey design, aware that representativeness, not randomness, is the important goal in sample selection. In coming years, wider and deeper Internet market penetration will mitigate sampling problems. In the meantime, however, researchers must remain aware of the implications of possible sample bias. For example, widespread use of on-line political polling may focus parties and platforms away from the needs of particular subsets of the population. Furthermore, instrument effects (e.g., question order, definition of response categories), which have been convincingly documented for other survey modalities, are virtually unexplored for on-line surveys and polls.[5] One also must keep in mind that "interviewer effects" (e.g., respondents' inclination to provide a socially desirable response) do not simply disappear because a survey is computer-mediated.

Perhaps the most intriguing issue for the future is how on-line surveying will move beyond the adaptation of traditional paper-and-pencil and telephone survey techniques to a new technology. On-line surveying raises the possibility of completely new approaches to data collection, particularly the use of engaging, interactive techniques to collect information in such a way that respondents may not even realize the extent of the data they are providing. However, as such techniques develop, researchers need to be reminded that their human subject and data confidentiality practices need to be updated.

On the issues of sampling reviewed in this chapter, Internet survey techniques necessarily raise a challenge to "business as usual" within survey research circles. If you read only within the quantitative survey research literature, you find a strong emphasis on the notion that validity is chiefly concerned with *external* validity. As whole, however, social science has a more expansive notion of validity that focuses on the extent to "which relationships between research operations, or the degree to which generalizations about higher order constructs can be made from research operations."[6]

Rather than thinking of a population as improperly sampled, pollsters and survey researchers should emphasize the extent to which valid conclusions may be drawn by *relational inference*—sensible inferences for a larger diverse population drawn from a well-known, controlled subpopulation, which may or may not have been randomly selected. For example, in Survey2000, the sample respondents are clearly better educated and better off than the general population. One may also assume that they are better informed and more likely to be familiar with a wide range of cultural influences. Within this sample, researchers have found important differences in cultural tastes and preferences that vary with geographical region. If one finds clear and significant differences in this sample, skewed as it is, then one can plausibly argue that similar and quite likely stronger regional differences are to be found among the population at large. Purposive sampling begins with the intent to select a specific sample relevant to the interesting research question at hand. Relational inference differs from purposive sampling in that it is guided as much by the relationships in the data as the purpose behind the study. The researcher is mindful of the characteristics of the sample obtained, and how the sample differs from the general population. The aim, as with much of social science outside the area of survey research, is to make tempered generalizations about the attributes of a larger population.

NOTES

1. Further detail on Survey2000, including information on obtaining the data, may be found in Witte et al. (2000).

2. An accessible and practical discussion of estimating sample size for telephone surveys may be found at http://www.surveysampling.com/ssi_home.html, the web site for Survey Sampling Inc., an industry leader in telephone survey sample selection.

3. Respondents were queried regarding the following types of groups: fraternal groups, service clubs, veterans groups, political clubs, labor unions, sports groups, youth groups, school service groups, hobby or garden clubs, school fraternity or sorority, nationality/ethnic groups, farm organizations, environmental groups, community/neighborhood groups, social advocacy groups, literary groups, art groups, discussion/study groups, professional/academic societies, religious groups, computer clubs, or any other groups.

4. In statistical terms, one of the likely consequences of the nonrandom nature of the Survey2000 sample is biased sample estimates of the population parameters. Standard corrections for bias of this type, including robust standard errors, inflate the estimated standard error, which is equivalent to requiring a larger test statistic to be able to reject the null hypothesis.

5. A notable exception is the Gilens chapter in this volume. In addition, Survey2001, a National Science Foundation follow-up study to Survey2000, includes several experimental components in its design, ranging from rather simple studies of the relative functionality of radio buttons and drop down menus to the impact of including photographic images of plants and animals in questions regarding endangered species. Most important, Survey2001 will also field a parallel telephone survey to assess the overall impact of Internet versus telephone-based data collection.

6. Mueller (1997) develops and tests a set of logistic regression models based on a nonrandom, purpose sample of German daily newspapers.