The Future of Polling:
Relational Inference and the Development of Internet Survey Instruments
7/13/01

James Witte, Sociology, Clemson University
Philip E. N. Howard, Sociology, Northwestern University

[In: Jeff Manza, Faye Lomax Cook, and Benjamin Page, Eds., Navigating Public Opinion: Polls, Policy and the Future of American Democracy
(New York: Oxford University Press, 2002)]

New media technologies have generated much excitement about ways of improving communication between and among 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 email listservs, Internet survey and polls have been hailed as new ways of conducting and measuring public deliberation. For many practitioners, these new 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.

Other chapters in this collection have dealt with various aspects of 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 the Internet survey instruments including the advantages and disadvantages of these methods. This is followed by a review of historical lessons learned in polling leading to a discussion of the modern challenges presented by modern techniques. Particular attention is paid to 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. Next, the significance of sampling techniques for Internet surveys is illustrated using data from Survey2000, a large-scale Internet Survey conducted on the National Geographic Society’s web site in the fall of 1998. The chapter concludes by emphasizing that representativeness is the aim of all sampling techniques—random and non-random—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 sub-sample special sections of the population based 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 ads that are strong and aggressive, but not too negative. 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 CATI (Computer Assisted Telephone Interviewing) or personal interviews. Fourth, Internet techniques can quickly reconnect with respondents should follow-up be necessary, while both the time involved in the initial survey and any follow-up is often less 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. 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 since 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.

However, there are also potential disadvantages to Internet survey instruments as well. 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 a 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.

New techniques also have important implications for those critical of polls. There is a long tradition of doubting polling sciences, or at least being highly critical of their generalizability, the utility of defining a ‘public opinion’ and the reduction of individual opinion to commensurable inputs (Binkley, 1924; Crossley, 1937; Gosnell, 1937; Katz and Cantrill, 1937; Lohman, 1988). Some criticize polling for reducing public opinion by extracting unrelated individuals from larger communities for their snap-shot responses to simple close-ended questions. Political ethnography has long recognized that public opinion is a social process frequently lead by a few individuals with strong opinions and persuasive abilities (Mansbridge, 1983; Herbst 1993), and a sample design that democratically assigns equal weights to individual responses saves a step of calculation.
by ignoring 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 unmeasurable. However, the large and potentially diverse sample frame of the Internet has made it possible to expand the theoretical sample of ‘opinion leaders’ and some firms claim to be able to measure opinion leadership by building a peer-review process into the survey design itself.

The most obvious potential limit of Internet survey methods have to do with sampling, as discussed at length below. At this point however, one point is worth emphasizing. Telephone samples (and face-to-face survey samples) are the products of considerable self-selection—quotas and stratified sampling assure "representative" findings in terms of key demographic criteria, but high levels of refusal are likely to conceal biases that remain imbedded within demographic categories. However, in the case of the 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 kinds of lessons about mistaken sampling.

**Historical Lessons in Polling**

In each political campaign, pollsters learn most from the errors of previous campaigns. Over time, pollsters 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. 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 1936 truth that a large sample does not guarantee accurate results, it ought to be emphasized that also 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 LBJ. 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 four years prior. Nonetheless, academics with good theoretical reasons for being interested in the John Birch Society were dissatisfied with a representative survey that nonetheless 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 do so. The Literary Digest poll was always conducted several weeks before the election date, though it is primarily faulted for its sample population, not its timing. In 1980, pollsters correctly predicted a Reagan victory using snapshot data, but 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 November 1999 *NY Times* piece titled “Polling’s ‘Dirty Little Secret’: No Response.” It is important to emphasize that a low response rate will increase the cost of a survey, but does not necessarily undermine its validity. A real problem, however, occurs when non-response 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 respondents 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 help them select which campaign ideas to emphasize and which to de-emphasize (Jacobs and Shapiro, 1994). 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, 1995).

Often viewed as a means of reading public opinion, clearly polling has also become a tool to shape public opinion. Good internal polling by the Bush presidential campaign in 1988 helped them realize that public knowledge of what Dukakis stood for was minimal,
so that 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 pre-existing 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 10 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 was to associate Clinton with terrorist sympathizers under the guise of testing the resilience of Clinton supporters.

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, 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 non-response, i.e., who is polled. It should be noted, however, that the full-range of survey methods issues, including instrument, interview, environment and measurement effects, potentially take on a new cast in an online polling environment.

Who is Polled: Sampling, Coverage and Non-response 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 non-response, 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 it is by no means 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, 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 it is in this regard that Internet-based surveys and polls face their greatest challenge, but one ought not assume that traditional survey techniques have once and for all time 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 non-response for the moment--questions of coverage have typically been treated as a tractable problem. Not everyone has a telephone and, to the extent that a surveys and polls are interested in samples that represents 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 multi-line households) have raised a whole new set of concerns regarding coverage. A description of these concerns is laid out in considerable detail in guidelines established by the American Association for Public Opinion Research (AAPOR, 2000). While the guidelines are very explicit and are followed by some researchers, 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, if one is to maintain the integrity of RDD sampling procedures. For example, studies conducted in 1997 and 1998 found that about one-third (31.3 percent) of households nationwide and about one-quarter of households (25.7 percent) in the Pacific Northwest have two or more telephone lines (Oregon Survey Research Laboratory, 1998). Industry sources in another twelve-state region of the US indicate that multiple line penetration is currently at 17 percent, with another 6.5 percent 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. While these studies did not explicitly ask whether or not 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 one 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 be themselves 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 a variety of issues associated with **response and non-response**. 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 non-interviews (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, it is useful to 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 an interview a medium length interview on a non-sensitive topic nearly 7,000 randomly-selected phone numbers would be needed. In other words, on average, seven numbers yield one completed survey, so we can assume that significant non-random effects deplete the value of random selection because non-random instrument, interviewer, or environment effects must account for the remaining six incomplete surveys.

Focusing particularly on refusal rates as a key aspect of non-response, it is widely accepted that non-response is not random and that survey researchers and pollsters invest considerable effort in call-backs and conversions, prepaid incentives and advance letters (Frey 1989, Dillman 2000, Singer et al. 2000). Interestingly one recent experiment has found that while rigorous efforts were able to significantly boost response rates (from 36.0 percent to 60.6 percent) there were few differences in response patterns and these were primarily found in demographic characteristics and not in substantive areas (Keeter et al. 2000). Nevertheless, even with rigorous efforts approximately 40 percent remain non-respondents, 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 non-respondents.

In sum, sampling error, coverage problems and non-response 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 Non-response in Internet Surveys

Up until now, Internet-based surveys and polls that seek to tap public opinion are necessarily limited by the fact that only about half of Americans have Internet access, while nearly one-third of those not online say they have no plans to do so (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 more or less accurately specify members of the online population, for example through email 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

---

1 A 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.
meaningless. Similarly discussions of coverage and non-response 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 respondent 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 General Social Survey (GSS) in 1993 only has 179 African Americans, while the GSS African-American over samples in 1982 and 1987 include 354 and 353 African-American respondents respectively. Even in 1996, when the GSS sample was doubled, African-Americans only number 402. Survey2000, on the other hand, resulted in 538 surveys completed by African Americans at a fraction of the cost of the GSS. While Survey2001 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 that underlies these methods.

**A Non-random 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, looking 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 that were 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. While survey researchers, pollsters and demographers emphasize randomness nearly to the point of fetishism, others are more catholic in their approach. Trochim (2000) describes the random sample approach to generalizing as one of two primary approaches to generalizing:

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.
Similar to 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, while 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, while 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 [not] represent the population) is that sufficient sample size may be determined more exactly.

**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 et al. 2001). By contrast, several months earlier, political scientist and SPSS statistical software founder, Norman Nie, 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 very different findings can be traced back to the different approaches to sampling employed in each study.

While 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). While the InterSurvey sample offered WebTV access to a random sample of individuals there is no guarantee that those who accepted the
company's offer actually participated in the survey constitute a random sample. As in the case of a telephone survey, a variety of 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 online 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.

In an effort 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 1 indicates that there are 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, looking at the top panel of Table 1, all Survey2000 respondents were asked whether they agreed or disagreed with three statements that indicate a sense 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. Compared to Non-WebTV respondents these differences are not only statistically different, but also each is of a magnitude of about ten percent suggesting real substantive differences between the two segments of the sample.

Table 1 about here

The middle panel of Table 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 with regard to the participation in twenty different types of groups or social organizations. Further, respondents were asked to identify whether they were members or active members, where 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 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

---

2 Respondents were queried regarding the following types of groups: fraternal groups, service clubs, veterans groups political clubs, labor unions, sports groups, youth groups, church, civic, recreational groups, and any other groups.
estimated difference never fell below four, approximately double that needed to reject the null hypothesis at the conventional .05 level.3

The lower panel of Table 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, when it comes to 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 are found among Non-WebTV respondents than among WebTV respondents.

One very plausible explanation for the differences between the WebTV and Non-WebTV segments of the Survey2000 sample, as summarized in Table 1, could be that the demographics of the two samples differ systematically. Indeed, we expect this to be the case and Table 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 one-third of the WebTV respondents have a bachelor’s degree or higher while well over one-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 more likely to live in households with children under the age of nineteen than Non-WebTV respondents.

Table 2 about here

However, as Table 3 indicates, the differences between WebTV and Non-WebTV respondents extend beyond the demographic criteria that are commonly employed in the selection of a stratified sample. Table 3 reports the estimated coefficients obtained when the index of social participation is regressed on the set of demographic characteristics summarized in Table 2, along with a variable indicating whether or not a particular respondent belonged to the WebTV segment of the sample. From Table 3 we do 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 of 25 through 34 year-olds. Compared to individuals with some college but no degree, respondents with lower levels of education participate at lower rates, while individuals with a bachelor’s degree or greater participate at a higher rate. Furthermore, respondents engaged in full-time employment participate

3 In statistical terms one of the likely consequences of the nonrandom nature of the Survey2000 sample are 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.
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. Importantly, however, even after controlling for this set of significant demographic 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 1.

Table 3 about here

Efforts to reconcile the very 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 very 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 (c.f., Dillman 2000). There is every reason to believe that similar issues affect the results of Internet surveys, such that it would be premature to attribute all differences between studies to sampling issues.

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, these 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 online sample and the Harris Poll Online panel, which currently claims 7 million volunteers email 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, 2001). Pollsters are also looking at new ways to exploit Internet sociability to improve snowball samples through affinity networks.

Further use of online surveys is inevitable but researchers still need to think critically about the process of survey design, being 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 online 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 online surveys and
polls. In this vein, 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 to consider the extent to which online surveying will move beyond the adaptation of traditional paper and pencil and telephone survey techniques to a new technology. Online 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 to keep up with these new possibilities.

Focusing more narrowly on the issues of sampling reviewed in this paper, Internet survey techniques necessarily raise a challenge to “business as usual” within survey research circles. If you only read 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.”

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 the case of Survey2000 the sample respondents are clearly better educated and better off 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. Now, 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 found 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

---

4 A notable exception is the Gilens chapter in this volume. In addition, Survey2001, a National Science Foundation followup study to Survey2000 include 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 importantly Survey2001 will also field a parallel telephone survey to assess the overall impact of Internet versus telephone based data collection.

science outside the area of survey research, is then to make tempered generalizations about the attributes of a larger population.
Table 1: Social and Political Differences between WebTV Users and Other Survey2000 Respondents

<table>
<thead>
<tr>
<th>Agreeing with the statement:</th>
<th>WebTV Respondents (%)</th>
<th>Non-WebTV Respondents (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel close to other people in my community.</td>
<td>59.1**</td>
<td>68.6</td>
</tr>
<tr>
<td>My daily activities do not create anything worthwhile for my community.</td>
<td>41.3**</td>
<td>30.7</td>
</tr>
<tr>
<td>My community is a source of comfort.</td>
<td>59.8**</td>
<td>68.3</td>
</tr>
</tbody>
</table>

Index of social participation (mean value) ^1^ 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

** p < .01, * p < .05. 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. ^1^ Index of social participation based on membership and active membership in 20 different categories of organizations and social groups.
Table 2: Demographic Differences between WebTV Users and Other Survey 2000 Respondents

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

Percentages for each type of respondent. ** p < .01, for difference between types of respondents. Based on 359 WebTV respondents and 33,851 Non-WebTV respondents for value items.
Table 3: Estimated OLS Coefficients for Index of Social Participation Regressed on Demographic Characteristics and WebTV Respondent Status

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

** p < .01, * p < .05. Based on 359 WebTV respondents and 33,851 Non-WebTV respondents for value items.
References


Witte, James, Lisa Amoroso and Philip Howard, "Method and Representation in Internet-Based Survey Tools: Mobility, Community, and Cultural Identity in Survey2000," Social Science Computer Review, Summer 2000 Special Issue, 18(2).