WEB SURVEY RESEARCH: CHALLENGES AND OPPORTUNITIES

Mick P. Couper
Survey Research Center, P.O. Box 1248, Ann Arbor, MI 48103

Key words: Web surveys, Internet, Total survey error

1. Introduction

Web surveys have already had a substantial influence on the survey industry. Much research is focusing on the potential benefits and possible drawbacks of Web surveys for a wide variety of applications in market, academic and government research alike. This paper presents a brief review of some of the key challenges of Web surveys for the industry, and some of the opportunities offered by the new method. This is done within the context of the total survey error framework (see, e.g., Groves, 1989), which apply equally to Web as to other modes of survey data collection.

Many claims and counter-claims are being made with respect to Web surveys. To determine the veracity of these claims and to identify the pros and cons of various approaches, it is important to be explicit about the context of the arguments. Web survey types are many (see Couper, 2001) and, as with other modes of data collection, vary in terms of quality.

2. Web Surveys and Data Quality

Any comparison of Web surveys to other methods of survey data collection should take the following into account:

1) The standards of comparison should be made explicit. For example, if the comparison is to a mall intercept survey, Web surveys do indeed offer the opportunity to collect much more (and richer) data at far less cost. On the other hand, if the comparison is to a high response rate probability sample of the general population, the failures of the Web method become rapidly evident. The question here is, To what are we comparing the Web survey?

2) The criteria for evaluation should be made explicit. Web surveys, like all other methods of survey data collection, involve trade-offs. Survey methodology has a well-developed model of total survey error that can serve as a useful departure point for such comparisons. In terms of costs, Web surveys may be unequaled. But such benefits often come at the expense of low response rates, poor coverage of the general population, and the difficulties of selecting probability samples. So, here the question is, What source of error is the focus of attention?

3) The type of Web survey should be made explicit. There are many different ways to implement Web surveys, and some are better (by some of the evaluation criteria) than others. For instance, probability samples *can* be developed for Web surveys, but these may suffer from other problems such as high nonresponse. Here we should ask, What type of Web survey are we talking about?

Any discussion of the benefits and drawbacks of Web surveys thus requires explicit details of both the source and target of the comparison, and of the criteria being used to compare the two. The way one asks the questions about Web surveys may determine the likely answer. Survey quality is not an absolute, but should be evaluated relative to other features of the design (such as accuracy, cost, timeliness, etc.) and relative to the stated goals of the survey. The relative quality of a particular approach must be judged in light of alternative designs aimed at similar goals and with comparable resources.

3. Coverage and Sampling Error

These represent key challenges for probability-based Web surveys of broad populations (e.g., the full U.S. population, or even the population of Internet users). However there are a number of groups for which coverage is of little concern and for which sampling frames can be readily developed or obtained.

It is with respect to sampling that one sees a broad range of approaches. Some fully acknowledge their limitations and make no attempt to generalize to the full population. But others make claims that would be difficult to justify statistically. Two approaches that have received much press, and are discussed by others in this session, are those of Knowledge Networks and Harris Interactive. The former aims at a representative sample of the U.S. population by using RDD sampling methods and providing respondent Web access in exchange for participation. Harris Interactive's approach begins with a large group of self-selected persons, and uses propensity modeling to calibrate results to parallel telephone surveys. But many Web survey organizations do not even go this far, being content with large numbers of respondents obtained cheaply.

One method of sampling ideally suited to the Web (if one's interest is users of or visitors to a particular site) is transaction or intercept sampling. The selection probabilities are known and selection can be well controlled. The key drawback is nonresponse.

Coverage error is a function both of the proportion of the target population that is not covered by the frame and of the difference on the survey statistic between those covered and those not covered:

$$Y_c = Y_t + \frac{N_{nc}}{N_t}(Y_c - Y_{nc})$$

Where Y_c is a statistic estimated on the covered portion of the population, Y_t is the statistic we wish to estimate for the full population, N_{nc}/N_t is the noncoverage rate, and Y_c -Y_{nc} the difference between the covered and the not covered on the statistic of interest.

Many arguments are being made to downplay the coverage problem, but it remains clear that a substantial proportion of the general population does not have Web access *and* that those without access differ from those with access on a number of important dimensions, including education, income and race (NTIA, 2000), not to mention probably countless other attitudes, behaviors, attributes and other variables we wish to measure (see Table 1).

While issues of sampling and coverage remain of key concern for Web studies aimed at broad populations, we should be careful not to focus exclusively on this issue. Probability sampling is a necessary but not sufficient condition for representation. Other sources of error (e.g., nonresponse) may still produce a sample that is patently dissimilar from the population it purports to represent.

4. Nonresponse Error

Nonresponse error arises through the fact that not all people included in the sample are willing or able to complete the survey. As with coverage error, nonresponse error is a function both of the rate of nonresponse and of the differences between respondents and nonrespondents on the variables of interest (Groves and Couper, 1998): Nonresponse error can be represented as follows:

$$Y_r = Y_t + \frac{N_{nr}}{N_t}(Y_r - Y_{nr})$$

Where Y_r is a statistic estimated on the respondents, Y_t is the statistic we wish to estimate for the total population, N_{nr}/N_t is the nonresponse rate, and Y_r - Y_{nr} the difference between respondents and nonrespondents on the statistic of interest.

A high response rate does not guarantee an absence of nonresponse error. But the lower the rate, the greater the potential for differences between respondents and nonrespondents on the statistic of interest to affect estimates derived from the sample.

The early evidence from split-sample mode comparisons suggests that response rates to Web surveys do not reach the levels of equivalent paper-based surveys (see Couper, 2001b; Table 2). However, this may well change over time as we learn to adapt response-enhancing strategies to the Web, and as the populations of interest become more comfortable with responding using this medium.

Even among members of opt-in panels, response rates vary widely, but do not appear to reach levels of other modes of data collection. For example, Schmidt (2000) reported response rates ranging from 20%-60% for the Greenfield Online panel, while Terhanian (2000) reported rates ranging from 20% to 25% for Harris Interactive's online panel. Schonlau (2000) found that 14% of those invited started the survey and 12% completed it among California members of the Harris Interactive panel. Using Survey Sampling Inc.'s panel, we obtained a 20.3% response rate with an e-mail invitation and single reminder; a second survey of similar content with no reminder obtained an 11.9% response rate (Couper, Tourangeau, & Steiger, 2001). Similar results are reported for Japan (Yoshimura & Ohsumi, 1999).

Pop-up or intercept surveys typically fare no better. Comley (2000) reports response rates ranging from 9% to 48%, with an average of 24% for pop-up surveys in the U.K. MacElroy (2000) and McLaughlin (2000) report similar rates for intercept surveys in the U.S.

Response rates for banner-advertized surveys are even lower. In an experiment of different types of appeals, Tuten, Bosnjak and Bandilla (2000) obtained click-through rates ranging from 0.13% to 0.44%. MacElroy (2000) supports this finding with estimates of response rates around 0.5% for banner-advertized surveys.

Measured against the yardstick of many market research efforts, these response rates may not be cause for great concern. But for most academic and government surveys these rates are unlikely to inspire confidence in the new methods.

To date, very little work has focused on nonresponse error in Web surveys. As the sampling and coverage problems become less salient, the nonresponse problem is likely to become relatively more important. Furthermore, the focus of research should turn toward understanding why some respond to Web surveys while others do not, and whether and in what ways these two groups may differ on key variables of interest.

5. Measurement Error

Measurement error is the deviation of the answers of respondents from their true values on the measure. The measurement process is one of the least explored areas of Web surveys, but one that offers great potential for improving data quality. In some ways Web surveys are like other methods of survey data collection, but in other ways they are unique.

First, Web surveys are self-administered. In interviewer-administered surveys, well-trained interviewers can often explain unclear terms to respondents, keep them motivated, reassure them of the confidentiality of their answers, probe incomplete or inadequate responses, and so on. In self-administered surveys there is no such intermediary, and the survey instrument itself serves to convey the researcher's questions and expectations to the respondent. Similarly, the respondents' only means of communication with the researcher is through the medium of the instrument, and their answers and intentions must be taken at face value. This places greater focus on the self-administered survey instrument. The instrument must be easy to understand and to complete, must be designed to keep respondents motivated to provide optimal answers, and must serve to reassure respondents regarding the confidentiality of their responses. On the other hand, there are documented advantages of self-administered methods, particularly for the collection of sensitive information and the reduction of social desirability effects, that are likely to apply for Web surveys too.

Second, Web surveys are computerized. In contrast to mail surveys which are static or passive instruments, Web surveys can make use of the full power of computer-assisted interviewing (CAI) methods. These include automated branching or skipping, randomization of questions or response options, tailored fills or question wording, range and edit checks, feedback to respondents, and so on. While not all Web survey designs make use of these features, they are nonetheless potentially powerful tools for interacting with respondents, assisting them in the completion of the task, and motivating them to continue with the survey.

In these two respects, Web surveys are much like computer assisted self-interviewing (CASI) methods. However, Web surveys differ from CASI in that there is no trained interviewer present, and the survey organization does not provide the equipment. This latter point means that respondents are completing the survey on a variety of different hardware platforms and using a variety of software systems, with the potential that the instrument does not look and act in an identical manner for all respondents. This again puts more burden on design.

Finally, Web surveys have the power to extend the visual elements of presentation beyond what is usually feasible in paper surveys. Web survey instruments no longer only (or primarily) consist of verbal features (words and numbers) but can also make use of rich visual features. These visual enhancements include still and moving images, animation, line drawings, pictures, color, shapes, etc., not to mention true multimedia which includes both sound (aural) and pictures (visual) features. The graphical nature of the Web frees the survey designer from the traditional constraints of paper-based questionnaires (order, font, color, etc.). It's not that these design features could not be used before, but they were expensive and time-consuming to develop and reproduce in large quantities, and were thus used sparingly. In contrast, embedding a color photograph or image in, or changing the background color, design or layout of, a Web survey is a relatively trivial task.

All these elements combine to make the Web a unique medium for the presentation of survey questions and for the elicitation of responses. In fact, the Web permits the extension beyond the traditional survey "question" to include a wide variety of stimulus material. While the Web provides a wonderful opportunity to "think out of the box" and expand the variety of ways information can be presented to respondents, this freedom may come at a price. For example, while images are increasingly being used in Web surveys to enhance the user experience and motivate respondents to continue with the survey, the addition of images may have unintended consequences for the survey questions and the responses being elicited. Even when the image is explicitly designed to supplement the question text, the effect may be different than that desired. Thus, with the increased range of tools comes the heightened possibility of inadvertently introducing measurement error.

Visual enhancements are widely used in Web surveys, both because they are easy to implement and because Web survey designers believe that aesthetically pleasing sites are needed to motivate respondents to complete the survey. But the possible effects of these design choices on measurement error are largely unexplored, and as yet no evidence supports the response rate arguments for including them (see Dillman et al., 1998).

In addition, while the task of designing or developing a Web survey rich in visual features may be relatively simple, the effort of ensuring that all respondents see the survey in the same way and are able to navigate, view and complete the survey in a consistent fashion should not be underestimated.

One of the basic tenets of survey measurement is the notion of standardization. Survey questionnaires or instruments are designed to present a standard stimulus to all respondents, and interviewers are trained, monitored and supervised in order to ensure standardized oral delivery. The Web offers much less control for the survey designer and, in fact, places such control in the hands of the respondent (explicitly or otherwise). Thus, variations in modem speed, computer operating system, Internet service provider (ISP), browser type and version, availability of plug-ins, and a host of hardware and software settings under the control of the individual user may all affect the way the survey instrument is received and viewed by the respondent. Three brief examples may serve to illustrate this point:

1) Users have a variety of choices of security settings in browsers. One of these is to alert the user whenever data are transmitted to an insecure server. Many Web surveys do not make use of secure servers, and this may mean that respondents need to respond to several dialogue boxes in order to complete a single question.

2) The size of the browser window (full screen versus reduced-size window) is under user control, as are font size settings (e.g., both small vs. large font selections in the Windows display settings, and font size adjustments within the browser). In addition, monitor resolution (e.g., 800×600 or 1024×768) is not consistent for all respondents. Together these variations may results in some parts of the question not being visible to some respondents while fully visible to others, in uneven columns in tables, or in other variations in the presentation of questions across respondents.

3) The user has control over the colors used for the browser (both background and foreground). If the designer is not careful, and does not explicitly override these settings, the survey may be unreadable or the color enhancements may not have their intended effects.

There are many other user-defined options and variations in systems that may change the survey experience across respondents. Table 3 lists some of characteristics of browsers and operating systems, and shows the diversity of systems and settings that users bring to Web surveys. Several surveys are designed to require plug-ins (e.g., video or sound players, viewers for proprietary images, Java applets, etc.) that rarely work as intended for all respondents. In addition, there are a host of emerging wireless or portable Internet access technologies that will only serve to make the design of Web surveys even more challenging.

In terms of color, all browsers are *not* created equal. What may work on one browser may be unreadable on another. There are ways to avoid these problems (e.g., using browser-safe colors), but again this requires careful attention to design details that are often overlooked in the frenetic world of Web surveys where speed and cost appear to be key driving factors.

Thus, while the Web offers great promise for creative and innovative survey design and measurement, great care must be taken to ensure that respondents receive, view, and interact with the instrument in a similar way.

6. Summary and Conclusions

For many market research applications, the sampling, coverage, nonresponse and measurement problems may be of little consequence. The researcher or client may not be interested in broad generalizations to a larger population, and may simply want to gain a sense of general reaction to a product or advertisement from as large and diverse a group as possible in a short time and with relatively little cost. This may well be sufficient to inform decisions about packaging, product placement, pricing, and so on. For such uses, Web surveys may well be an ideal method of data collection. One area where this may be especially true is in advertising research. The capacity of the Web to deliver full color images, sound, video, and other rich audio-visual stimuli to large numbers of respondents in a controlled (i.e., randomized, timed, ordered, standardized) fashion is unparalleled. Similarly, the Web offers a great tool for experimentation where the focus is on randomization rather than representation.

On the other hand, these same methods may not (yet) be suitable for informing major public policy decisions where accuracy and representation are critical. Estimates of unemployment, welfare recipiency and program participation, crime victimization, consumer expenditures, and a host of other government survey programs would be ill-served by data based on such surveys. For such policy research, most Web surveys cannot compete with traditional interviewer-administered approaches in terms of accuracy and the richness of the information that can be obtained. Yet, for such sectors of the industry, Web surveys may well serve a supplemental or supporting role, particularly in mixed-mode applications or surveys of specialized populations (e.g., business surveys).

When evaluating Web surveys, it is important to make the standards of comparison explicit. Several industry bodies, including ESOMAR, the Council of American Survey Research Organizations (CASRO) and the National Council on Public Polls (NCPP) have issued guidelines or standards for methodological disclosure of Web surveys. If adhered to, these may serve as a basis for evaluating the claims made by the purveyors of Web surveys.

Given the relatively low statistical literacy on the part of the lay public and policymakers alike, all data, regardless of their provenance or quality, may be valued equally. Few can distinguish good surveys from bad—indeed many have trouble distinguishing survey data from those gathered by other less-rigorous methods. Under such circumstances, size (i.e., the number of cases) becomes the only accessible measure of quality, and price becomes the only measure of value, a dangerous state of affairs indeed.

Undoubtedly, the Web allows us to collect information from large numbers of respondents cheaply and quickly, and therein lies its key benefits for survey research. In addition, Web surveys permit us to extend the traditional survey instrument in ways we have yet to fully explore. Yet, these considerable benefits of the new technology mush be weighed against the threats to representation, whether because of noncoverage or nonresponse, and the potential for measurement error arising from the use of new and relatively untested measurement tools. Clearly much work remains to explore the full potential, and avoid the pitfalls, of this new method.

References

Bason, J.J. (2000), "Comparison of Telephone, Mail, Web, and IVR Surveys." Paper presented at the annual conference of AAPOR, Portland, OR, May.

Comley, P. (2000), "Pop-Up Surveys. What Works, What Doesn't Work and What Will Work in the Future." Paper Presented at the ESOMAR Net Effects Internet Conference, Dublin, April.

Couper, M.P. (2001a), "Web Surveys: A Review of Issues and Approaches." *Public Opinion Quarterly*, 64 (4), 464-494.

Couper, M.P. (2001b), "The Promises and Perils of Web Surveys." In A.Westlake et al. (eds.), *The Challenge of the Internet*. London: ASC, pp. 35-56.

Couper, M.P., Tourangeau, R., & Steiger, D.M. (2001), "Social Presence in Web Surveys." *Proceedings of CHI '01, Conference on Human Factors in Computing Systems*, Seattle, April. New York: ACM, pp. 412-417.

Dillman, D.A., et al. (1998), "Influence of Plain vs. Fancy Design on Response Rates for Web Surveys." *Proceedings of the Survey Research Methods Section, American Statistical Association*.

Groves, R.M. (1989), *Survey Errors and Survey Costs*. New York: Wiley.

Groves, R.M. & Couper, M.P. (1998), *Nonresponse in Household Interview Surveys*. New York: Wiley.

Guterbock, T.M., et al. (2000), "Web Versus Paper: a Mode Experiment in a Survey of University Computing." Paper presented at the annual conference of AAPOR, Portland, OR, May.

Jones, R. & Pitt, N. (1999), "Health Surveys in the Workplace: Comparison of Post, Email and World Wide Web Methods." *Occupational Medicine*, 49 (8):556-558.

Kwak, N. & Radler, B.T. (2000), "Using the Web for Public Opinion Research: A Comparative Analysis Between Data Collected via Mail and the Web." Paper presented at the annual conference of AAPOR, Portland, OR, May.

MacElroy, B. (2000), "Variables Influencing Dropout Rates in Web-Based Surveys." *Quirk's Marketing Research Review*, July/August.

McLaughlin, T. (2000), "Customer Database Research: Guidelines for Complete and Ethical Data Collection." Paper presented at the annual conference of AAPOR, Portland, OR, May.

National Telecommunications & Information Administration (NTIA) (2000), *Falling Through the Net: Toward Digital Inclusion*. Washington, DC: U.S. Department of Commerce.

Schonlau, M., et al. (2000), "Comparing Random Digit Dial with Web Surveys: The Case of Health Care Consumers in California." Santa Monica, CA: RAND, unpublished paper.

Schmidt, J. (2000), " 'Bill of Rights of the Digital Consumer': The Importance of Protecting the Consumer's Right to Online Privacy." Paper presented at the annual conference of AAPOR, Portland, OR, May.

Terhanian, G. (2000), "How To Produce Credible, Trustworthy Information Through Internet-Based Survey Research." Paper presented at the annual conference of AAPOR, Portland, OR, May.

Tuten, T.L., Bosnjak, M., & Bandilla, W. (2000), "Banner-Advertised Web Surveys." *Marketing Research*, 11 (4): 17-21.

Weible, R. & Wallace, J. (1998), "Cyber Research: The Impact of the Internet on Data Collection." *Marketing Research*, 10 (3): 19-24, 31.

Yoshimura, O. & Ohsumi, N. (1999), "Some Experimental Surveys on the WWW Environments in Japan." Paper presented at the International Symposium on New Techniques of Statistical Data Acquisition.

Table 1. Percent Online by Selected Demographic Characteristics, November/December 2000

Characteristic	Percent online
Age:	
18-29	75%
30-49	65%
50-64	51%
65+	15%
Income	
Under $30,000	38%
$30,000-$50,000	64%
$50,000-$70,000	72%
$70,000 +	82%
Education	
High school or less	37%
Some college	71%
College degree or more	82%

Source: Pew Internet Project: Internet Tracking Report (February, 2001)

Table 3. Key Browser Statistics, January 2001

Characteristic	Percent of browsers
Screen resolution:	
800x600	54%
1024x768	30%
640x480	7%
Other, unknown	9%
Javascript and Java:	
Javascript enabled	81%
Java enabled	78%
Browsers:	
MSIE 5.x	72%
MSIE 4.x	12%
Netscape 4.x	10%
Other	6%
Operating system:	
Win 98	68.92%
Win 95	13.08%
Win NT	6.67%
Win 2000	4.54%
Unknown	3.55%
Mac	2.10%
WebTV, Linux, Unix	0.98%

Source: http://www.thecounter.com

Table 2. Example Web Survey Response Rates

Source	**Target Population, Topic and Design**	**Response Rate by Mode**
Kwak and Radler (2000)	Univ. of Wisconsin students; 1999 survey on campus computing; n=1000 per mode	Web: 27.4% Mail: 41.9%
Radler (2000)	Univ. of Wisconsin students; 2000 survey on campus computing; n=1000 per mode	Web: 28.0% Mail: 52.6%
Guterbock et al. (2000)	University of Virginia students; survey on university computing; mode comparison	Web: 36.8% Mail: 47.6%
Bason (2000)	Univ. of Georgia students; survey on drug and alcohol use; mode comparison, n=200 per mode	Web: 15.5% Mail: 27.7% IVR: 17.4% Phone: 23.9%
Jones and Pitt (1999)	Staff at 10 English universities; mode comparison	Web: 19% E-mail: 24% Mail: 72%
Weible and Wallace (1998)	MIS faculty; mode comparison, n=200 per group	Web: 26% E-mail: 24% Mail: 35% Fax: 25%