A Feasibility Evaluation of a Web-Based Demographic Survey

Elizabeth K. Griffin and Heather C. Holbert, U.S. Census Bureau
Demographic Surveys Division, Room 3351-3, Washington, DC 20233

Keywords: Internet, household survey, methods panel

Introduction

The Internet has already supplemented and even replaced many traditional forms of communication, and is quickly working its way into the data collection arena. According to figures compiled by *Inside Research*, a market research newsletter, the value of online research has experienced triple-digit percentage growth in each of the past four years (Lipke, 2000). The U.S. Census Bureau's Demographic Surveys Division (DSD) is committed to testing alternative methods of data collection, and thus conducted the Methods Panel Web (MPWEB) survey. The MPWEB tested whether the Internet is a viable reporting method for long, complex demographic survey instruments. The objective was to measure the response rate for members of eligible households (those with at least one college graduate) that indicated willingness to complete a survey on the Internet during a prior MP interview. The results will aid in determining what proportion of the population under the most optimistic scenario is willing to and actually does respond to long, complex household surveys using the Internet. This paper will discuss related research, followed by the MPWEB's survey methodology, results, and recommendations.

Past Research

Studies have been conducted to evaluate the use of electronic surveys, both via e-mail and the Internet. Although such studies reveal improvements in cost, time, and quality, they also reveal concerns about coverage and response rates.

The literature reveals that two of the main factors driving the tremendous growth in online research are lower costs and response time. Compared to traditional research methodologies, online studies are conducted with an average savings of more than 40 percent in cost and cycle time (Rosenblum and Grecco, 1998). Kwak and Radler (2000) reported that the cost of delivering their Web data set was less than one-fourth of the cost for their mail data set, and their Web survey was more than four times faster than their mail survey.

Online surveys combine some of the best features of computer-assisted systems and paper surveys. Like computer-assisted systems, online questionnaires utilize real time skips and edits within the instrument, improving data quality, reducing the time needed to clean and edit the data, and decreasing respondent burden (Sedivi, Nichols, and Kanarek, 2000; Kwak and Radler, 2000;

Abraham, Steiger, and Sullivan, 1998). Like paper surveys, online questionnaires allow respondents to complete the survey at their convenience. They may respond whenever and from wherever they choose, and at their own pace, "putting it down" and returning to complete it at a later time (Slevin, 1997; Abraham, Steiger, and Sullivan, 1998; Meeks, Lanier, Fecso, and Collins, 1998).

While all of these advantages are appealing, response rates remain a major concern. Mail surveys, like e-mail and Internet surveys, are self-administered, and therefore serve as a standard against which electronic surveys are measured. Literature shows that to date, electronic methodologies such as e-mail and Internet surveys have generally failed to meet the standard set by comparable mail techniques (Schaefer and Dillman, 1998). The Census Bureau has also compared online response rates to those of more traditional methods, and has consistently experienced lower than expected response using the Web. Low response rates were also experienced when respondents were pre-screened. The sample for the 1998 Company Organization Survey (COS) reported that they were interested and had the ability to report online, but only 27 percent actually did so (Sedivi, Nichols, and Kanarek, 2000).

Researchers have offered many reasons for the low response rates observed in online research. The most popular reason is security or confidentiality. Nichols and Sedivi (1998) found that security concerns were the primary reason (39 percent) that respondents were unwilling to report via the Web. Other reasons discussed in the literature include technical difficulties (Couper, Blair, and Triplett, 1999), the respondent's inability to determine the questionnaire length (Ramos, Sedivi, and Sweet, 1997; Abraham, Steiger, and Sullivan, 1998), perceived lack of Internet expertise, and cost of Internet service for limited usage plans.

Another possible cause of low response rates, and a concern of online researchers, is coverage. How many people are actually online, and do they differ from those who are not? There are varying estimates of the numbers and types of people who currently have Internet access. The most recent, and probably the most accurate measures are from the U.S. Census Bureau's Current Population Survey (CPS) Internet and computer use supplement, fielded in August 2000. Data from the supplement show that 44.4 percent of individuals in the U.S. have Internet access, up from 32.7 percent in December 1998. If growth continues at this rate, more than half of all Americans will be using the Internet by mid-2001 (U.S.

Department of Commerce, 2000). Almost all groups that have traditionally lagged behind in Internet access are now making dramatic gains. Nonetheless, a digital divide still exists between those with different levels of income and education, different racial and ethnic groups, old and young, single and dual-parent families, and those with and without disabilities. Essentially, the "Internet population" is quite different from the general population of the United States. The fact that half of the country does not have a chance to be selected for online surveys may be a serious threat to attempts to develop probability-based samples of the general population (Couper, 2000).

Such issues may limit the U.S. Census Bureau's ability to take advantage of online data collection, particularly among household surveys. Internet surveys have proven useful for specific types of samples, however the Bureau conducts many demographic surveys which represent the country as a whole. The U.S. Census Bureau's Methods Panel Web (MPWEB) survey tested whether the Internet is a viable reporting method for such surveys, especially those that utilize large, complex instruments. The remainder of this paper will discuss the MPWEB's survey methodology, results, and recommendations.

Methodology

The MPWEB sample was drawn from that of an existing MP demographic survey, the Methods Panel Survey of Income and Program Participation (MPSIPP). The MPSIPP instrument was very long and complex, and collected information about the structure of households, economic status, sources of income, and labor force participation. A series of household level screener questions was added to the end of the MPSIPP instrument to ascertain the ability and willingness of the household to complete a survey online. Respondents eligible for the MPWEB survey were those in the prescreened households who graduated from college with at least a Bachelor's degree and were between the ages of 15 and 76.

Research has continuously shown that people with higher levels of education are more apt to be online. The 2000 CPS data found that people whose highest level of education is a Bachelor's degree or higher had the highest Internet use (74.5 percent). Age has also been cited as a determining factor for Internet use with over 50 percent of people between the ages of 18 and 49 online (U.S. Department of Commerce, 2000). Such statistics led us to believe that the MPWEB sample would be more likely than others to not only have access to, but use the Internet. Thus, the MPWEB would test an Internet survey under the most optimistic scenario.

Census Bureau Field Representatives conducted personal interviews with 3,264 MPSIPP respondents. Of those interviews, only 11% were with eligible respondents whose household had access to the Internet and was willing to complete an online survey (see Table 1).

Table 1 - Selection of Cases

Description	Number of Respondents	%
Total Eligible MPSIPP Respondents (Age 15+)	3372	100%
Complete and Partial Interviews	3264	97%
Eligible MPWEB Respondents	1913	57%
Access to Internet*	579	17%
Willing to Complete Online Survey*	355	11%

* Household level screener questions

The MPWEB instrument was based on a modified version of the National Science Foundation's Survey of Doctorate Recipients (SDR), and asked about work experience, principal employer, career related experiences and recent education. The questionnaire contained less than 90 items and relatively straight forward skip patterns. We selected the SDR questionnaire because its small size and simplicity enabled us to program an online instrument in a relatively short amount of time.

The MPWEB instrument was programmed in Macromedia Flash, a first for the Census Bureau. Other Census Bureau online surveys have used Delphi downloadable executables, HTML, HTML with Java Script, and Java. The decision to use Flash was based on dissatisfaction with the other software's download time, screen appearance, compatibility with and consistency between browsers, and ability to include pop-up features and perform real-time skips and edits (Sedivi, Nichols, and Kanarek, 2000; Nichols and Sedivi, 1998; Kanarek and Sedivi, 2000). The Macromedia Flash player addressed all of those concerns, and provided extra features such as fly-over text which provided the ability to display topic-specific instructions when the respondent pointed to highlighted text. Screen-based design was chosen because of the many skip patterns in the MPWEB instrument. Soft edits were built into the instrument to allow improved data quality without forcing a response.

The first two screens of the survey contained general instructions on how to navigate through the Flash instrument, and could be accessed at anytime from within the survey through the menu bar. A menu bar containing the following buttons: Instructions, FAQ (Frequently Asked Questions), Comments, Print Preview, Save, Exit, and Contact Us appeared at the top of each screen. The save and exit features allowed the respondent to save and exit the survey, and resume at a later time at the same point at which he or she exited. The Flash instrument also

performed an automatic save every five minutes, so a partial interview would be saved even if there were technical difficulties such as a lost Internet connection or a system that locked up.

The MPWEB addressed security at three levels: (1) communications between the respondent and the Census Bureau were protected via 128 bit encryption; (2) respondent data at the Census Bureau was protected by providing all respondents a unique username and password which they used to log in to the server used for data collection; and (3) the Census Bureau network is protected by a firewall. Collectively, encryption, authentication, and a firewall provide a secure method for the Census Bureau to collect data on the Internet (Kanarek and Sedivi, 2000; Sedivi, Nichols, and Kanarek, 2000).

A short debriefing questionnaire was added to the end of the online questionnaire. Respondents who completed the entire survey would also complete the online debriefing questionnaire. Respondents who only completed part of the survey received a paper version of the debriefing questionnaire after the survey closed out. Non-respondents also received a paper debriefing questionnaire.

We sent MPWEB advance letters via priority mail to 355 college graduates in January 2001. The letter asked them to complete the survey within 30 days. We did not give respondents the option of completing the survey via a different mode because the purpose of our test was to identify how many people, in a best case scenario, would respond via the Web. This was feasible because these respondents were selected from households that reported both the ability and willingness to complete the survey online. After 30 days, we sent a reminder letter via first class mail to the non-respondents and partial respondents which asked them to complete the survey within the next 15 days. After 15 days, we mailed a cover letter, paper debriefing questionnaire, and postage paid envelope via priority mail to non-respondents and partial respondents. We asked them to complete the debriefing questionnaire within ten days. At the end of March, we sent each person who either responded to the online questionnaire or the paper debriefing questionnaire a thank you postcard.

Respondents who required assistance were instructed to e-mail a help desk located at Census Bureau headquarters for assistance, or call a toll-free number which would connect them to a help desk.

Results

Like previous online surveys, the MPWEB suffered from low response rates and reports of security concerns and technical difficulties.

Response Rate: The 355 person sample yielded 79 completed online questionnaires (including respondent debriefing) and no partial online questionnaires for a response rate of 22 percent. If the screener question that ascertained willingness to complete an online survey is eliminated, the overall response rate was 14 percent of those who had Internet access (see Table 2).

Table 2 - Response Rate

Description	# of Respon-dents	% per willing-ness	% per access
Access to Internet*	579	—	100%
Willing to Complete Online Survey*	355	100%	61%
Completed Online Int.	**79**	**22%**	**14%**
Completed Debrief. Quest.	144	41%	25%

* Household level screener questions

There is a wide range of research available regarding response rates, some of which was discussed earlier in this paper. Of those discussed, the MPWEB most closely resembles the 1998 Company Organization Survey (COS), which also pre-screened respondents. When beginning with a pre-screened sample, the COS achieved a response rate of 27 percent, which is slightly higher than our response of 22 percent. The COS collects information from businesses, however, which typically have greater Internet access and ability than individuals, and may therefore account for much of the difference (Sedivi, Nichols, and Kanarek, 2000; Sweet and Russell, 1996). Another relevant study was conducted by Westat in 1999. They tested an Internet survey using the National Survey of Recent College Graduates (NSRCG). They prescreened college graduates for Internet access, and they, too, achieved a 27 percent response rate via the Web (Collins and Tsapogas, 2000).

Rate of Return: We received daily progress reports from the Census Bureau's Computer Assisted Survey Research Office (CASRO). These reports showed that we received 64% of the completed interviews within the first 18 days of the survey. After that, we received nothing until we sent out the reminder letter ten days later, which resulted in the remaining 36% of the completed interviews within two weeks.

Respondent Assistance: We were only contacted by eight respondents throughout the data collection period. Five of the eight respondents reported difficulties accessing the survey. Two of those five had problems with their password, one tried to access the survey while the Census

Bureau servers were down, and the other two problems could not be identified. We believe that one of the unidentified problems was a keying error by the respondent. He was able to access the survey the next day, only to discover that he needed the Flash software, and was not willing to download it. Two of the remaining three respondents also needed to download upgraded versions of Flash. One was not willing to do so because of the small memory on her computer. The other downloaded the software and successfully accessed the survey. The eighth respondent requested a telephone interview because he felt more comfortable reporting over the phone than via the Internet. When told that was not an option, he said he would have his wife reply for him since she is more familiar with the Web. This raises an interesting issue for person-based surveys: how can we control that the person selected to be in sample is actually the person that responds?

Debriefing Questionnaire: We received a total of 79 online respondent debriefing questionnaires and 65 paper non-respondent questionnaires for a debriefing response rate of 41%.

The results of the online debriefing questionnaire show that the survey was rated favorably by respondents with an overall score of 4.4 out of 5. The majority of respondents were willing to participate in future Census Bureau surveys and prefer completing a survey online as opposed to other modes. The greatest concern of respondents was the security of data transmission, followed by concerns about installing Macromedia Flash. The greatest difficulty respondents encountered was with logging in to the survey.

The paper non-respondent debriefing questionnaire shows that 45 percent of those who completed the debriefing attempted to access the online survey. The main reasons for not completing the survey were lack of time and difficulty logging in to the survey. These were followed by having an incompatible Internet browser and security concerns. Over half of the people who attempted access also wrote in other reasons for not completing the online survey. A common reason was that they did not want to install Macromedia Flash. It was interesting to note that a few of the write-in responses also said that they or their search engine could not locate the Web site. We inferred from those responses that these people typed the Uniform Resource Locator (URL) into a search engine, rather than in the address toolbar. We realized that some Internet users may not know how to access Web sites using the type of instructions we provided. Some users are used to searching for Web sites using a search engine. We also realized that America OnLine (AOL) users may be used to keywords and may have found our instructions difficult because we did not provide a keyword for the MPWEB.

The paper debriefing questionnaires from those who did not attempt to access the survey online show that the vast majority of people did not attempt to access the survey because they didn't have time. This reason was followed by no access to the Internet (although we screened on that initially) and concerns about security. A common other reason for not attempting to access the survey is that people forgot, although we sent a reminder letter.

The questions about the nature of technical difficulties had a very low response rate, possibly because people who had technical difficulties may not have known what caused them, and those who did not attempt to access the survey didn't know if they would have experienced difficulties.

Fifty-four percent of non-respondent debriefing questionnaire respondents indicated that they would be willing to participate in future Census Bureau surveys, in spite of their non-response to this particular survey. Over half also said they prefer mail surveys, followed by a third who prefer online surveys. The remaining 6% of non-respondents who completed a debriefing questionnaire were evenly split between preferring telephone and personal visits as the mode of completing a Census Bureau survey.

Overall, the issues of technical difficulties and security mirror those faced by other researchers. The ability to better identify specific types of technical problems will help researchers overcome some of the issues related to online surveys. Security issues, however, may be more difficult to overcome. Ultimately, public perceptions about the security of the Internet may be a more significant barrier to participation than actual threats to security (Sweet and Russell, 1996).

Discussion and Recommendations

The MPWEB study has both answered existing questions and raised new ones. There were advantages and disadvantages to each of the choices we made, including our choice of software, instructions for logging into the instrument, mailings, and security.

Choice of Software: Like all software, the use of Macromedia Flash had advantages and disadvantages. Our in-house testing found the survey to be very visually appealing, quick to download, and easy to navigate. However, some respondents, and almost a quarter of non-respondents, were concerned about and/or had difficulty with the installation of the Flash software. Future researchers should keep any respondent downloads to an absolute minimum whenever possible. Although Flash provided an appealing instrument for the end user, it provided many challenges for the authors and programmers. Since Flash is a movie software package, every item was literally a movie scene made up of a series

of frames, each of which had to be individually programmed. For example, a question with all blank entries was one frame, the same question with the first radio button selected was another frame, and so on. Since every variable had to be coded at the frame level, there was not a common area to place code, and any modifications had to be made in every frame where the text or variable appeared. This was very tedious, error prone, and difficult to debug and document. Additional challenges included difficulty programming and controlling skip patterns, navigation, and radio buttons.

We found that it was difficult to make some of our requested changes. This was most likely due to the challenges described above, but would nonetheless present problems for on-going, and continually changing, surveys. Although the MPWEB tested a relatively small, simple survey, it appears that Flash cannot handle some basic survey functionality, such as rostering. Overall, programming, debugging, and documenting the survey was very tedious and error prone. Therefore, we would not recommend the use of Flash for future complex survey instruments.

Logging into the Survey Instrument: Both respondents and non-respondents reported difficulty logging into the survey instrument, due to mistyping URLs, usernames and passwords, and inability to locate the URL using a search engine. These difficulties may have been prevented with more detailed log-in instructions, including examples of where and how to type the URL and passwords. It may also be advantageous for future online surveys to provide a keyword for AOL.

Survey Mailings: Since we received the majority of responses within two weeks after each mailing, extending the length of time the survey is available to respondents does not appear to increase response rates unless respondents are sent frequent reminders. Some literature shows that email may be a better means of communication for electronic surveys, however, we did not have access to respondents' email addresses and were thus unable to attempt such communication (Rosen, Manning, and Harrell, Jr, 1998). We received 64% of responses after the advance letter which was sent via priority mail and the remaining 36% of responses after the reminder letter which was sent via first class mail. The cost of using priority mail must be considered. Does the added cost result in increased response rates? Research conducted by Abreu and Winters (1999) shows that it does, so we would recommend the use of priority mail in the future.

Security: Security remains a top concern among both respondents and non-respondents. We briefly addressed security issues in our Frequently Asked Questions, which

appeared on the back of each letter sent to the sample. Security concerns should be addressed up front, however, and possibly in more detail. Before the Internet can become a widely successful research tool, the public's concerns must be addressed, and negative perceptions must be changed.

Cost: Unfortunately, we do not have specific monetary costs for the MPWEB project because costs were absorbed by other projects. We utilized a great deal of time and resources throughout the project, however. We experienced many challenges programming, testing, and debugging the survey, which ultimately forced us to delay fielding the survey by one month.

Conclusion

Based on the MPWEB results, we do not recommend the Internet as a reporting method for large, complex demographic survey instruments, especially those with nationally representative samples. Considering that only 44.4 percent of the U.S. population is currently online, and their demographics differ substantially from those who are not, the Internet cannot provide estimates of the general population. The MPWEB survey tested a short, simple questionnaire in a sample of people who not only had Internet access, but indicated that they were willing to complete a survey online. Even in such an optimistic scenario, we utilized a great deal of resources and experienced many challenges programming, testing, debugging, and fielding the survey, and ultimately obtained a response rate of only 22 percent. The typical survey conducted by the Census Bureau's Demographic Surveys Division is magnitudes larger and more complex than the MPWEB survey we have tested. Technology is not currently sophisticated enough to handle the complexity of such a large demographic survey instrument, and the complicated skip patterns and rostering that it will entail.

Although the MPWEB tested the Internet as a sole method of data collection, our low response rate combined with technological challenges and limitations indicate that the costs of converting a complex questionnaire to an online survey far outweigh the benefits we may see, even in a multi-mode environment. Therefore, we do not recommend devoting the time and resources necessary to test the Internet on a larger, more complex survey within the Census Bureau, neither as a sole nor multi-mode method.

Future Research

We plan to perform additional analyses on the MPWEB data in the near future. We want to look at item non-response and demographic characteristics of respondents vs. non-respondents. We would also like to compare our response rates and rate of return to a

comparable sample of the Survey of Income and Program Participation (SIPP). Such analyses may help us to better understand who, exactly, does respond via the Internet, and thus the potential for future online research.

References

Abraham, Sameer Y., Darby Miller Steiger, and Colleen Sullivan (1998). "Electronic and Mail Self-Administered Questionnaires: A Comparative Assessment of Use Among Elite Populations." Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 833-841.

Abreu, Denise A. and Franklin Winters (1999). "Using Monetary Incentives to Reduce Attrition in the Survey of Income and Program Participation." Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 533-538.

Collins, Mary and Tsapogas, John (2000). "An Experiment in Web-based Data Collection". Paper presented at the annual meeting of the American Association for Public Opinion Research, 2000.

Couper, Mick P. (2000). "Web Surveys: A Review of Issues and Approaches." Public Opinion Quarterly Volume 64:464-494 by the American Association for Public Opinion Research.

Couper, Mick P., Johnny Blair, and Timothy Triplett (1999). "A Comparison of Mail and E-mail for a Survey of Employees in U.S. Statistical Agencies." Journal of Official Statistics Volume 15, No. 1, pp. 39-56.

Kanarek, Howard and Barbara Sedivi (2000). "Internet Data Collection at the U.S. Census Bureau." 1999 Federal Committee on Statistical Methodology Research Conference: Complete Proceedings, pp. 708-717.

Kwak, Nojin and Barry T. Radler (2000). "A Comparison between Mail and Web Surveys: Response Pattern, Respondent Profile, Data Quality, and Construct Association." Paper presented at the annual meeting of the American Association for Public Opinion Research, 2000.

Lipke, David J. (2000). "You've Got Surveys." American Demographics, Volume 22, No. 11, November 2000, pp. 42-45.

Meeks, Ronald L., Ann T. Lanier, Ronald S. Fecso, and Mary A. Collins (1998). "Web-based Data Collection in National Science Foundation Surveys." Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 349-353.

Nichols, Elizabeth and Barbara Sedivi (1998). "Economic Data Collection via the Web: A Census Bureau Case Study." Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 366-371.

Ramos, Magdalena, Barbara M. Sedivi, and Elizabeth M. Sweet (1998). "Computerized Self-Administered Questionnaires (CSAQs)." In: M. P. Couper, R. P. Baker, J. Bethlehem, C. S. F. Clark, J.Mmartin, W. L. Nicholls II, & J. M. O'Reilly (eds). Computer assisted survey information collection (pp. 389-408). New York: Wiley.

Rosen, Richard J., Christopher D. Manning, and Louis J. Harrell, Jr (1998). "Web-based Data Collection in the Current Employment Statistics Survey." Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 354-359.

Rosenblum, Jeff and Chris Grecco (1998). "The Future of On-line Research." Quirk's Marketing Research Review. July 1998. Article Number 0357.

Schaefer, David R. and Don A. Dillman (1998). "Development of a Standard E-mail Methodology: Results of an Experiment." Public Opinion Quarterly Volume 62:378-379 by the American Association of Public Opinion Research.

Sedivi, Barbara, Elizabeth Nichols, and Howard Kanarek (2000). "Web-based Collection of Economic Data at the U.S. Census Bureau." Proceedings of the Second International Conference on Establishment Survyes, American Statistical Association, pp. 459-468.

Slevin, Jim (1997). "An Integrated Approach." Quirk's Marketing Research Review. October 1997. Article Number 0269.

Sweet, Elizabeth and Chad Russell (1996). "A Discussion of Data Collection Via the Internet." Proceedings of the Section on Survey Research Methods, Vol. II, American Statistical Association, Alexandria, VA, 774-779.

U.S. Department of Commerce (2000). "Falling Through the Net: Toward Digital Inclusion. A Report on Americans' Access to Technology Tools."