

## **An Experiment Comparing Computer-Assisted and Paper Modes of Data Collection for the Short Form in Census 2000**

Sid J. Schneider\*  
David Cantor\*  
Lawrence Malakhoff\*\*  
Carlos Arieira\*  
Paul Segel\*  
Luu Nguyen\*  
Jennifer Guarino\*\*

\* **Westat**, Rockville, Maryland

\*\* **Bureau of the Census**, Suitland, Maryland

Presented at the 57<sup>th</sup> Annual Conference of the American Association for Public Opinion Research, St. Pete Beach, Florida, Friday, May 17, 2002, 2:00 p.m.

Direct inquiries to Sid J. Schneider, Westat, 1650 Research Boulevard, Rockville, Maryland 20850. Email [schneis1@westat.com](mailto:schneis1@westat.com)

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

## BACKGROUND

**Research questions.** New computer technologies are transforming the field of survey data collection. The expansion of the Internet has made web-based surveys more feasible. Improvements in development software have facilitated computer-assisted telephone interviewing (CATI). New speech processing systems have made Interactive Voice Response (IVR) surveys possible.

These technologies have not been widely offered as alternate response modes in large scale surveys which rely on paper forms for data collection, such as the decennial census. However, computer-assisted data collection methods may have the potential to improve response rates in such surveys, if some of the respondents who are unwilling or unable to respond by paper-and-pencil questionnaires use these computer-mediated methods instead.

The goal of the present study, called the Response Mode and Incentive Experiment (RMIE), was to assess the public's willingness to provide Census 2000 data using these computer-mediated data collection methods.

This paper presents the results of the RMIE concerning these research questions:

1. Does offering an alternative Internet, CATI, or IVR response mode increase response rates to the decennial census? If so, do any of these modes have a larger effect than the others?
2. What is the effect of offering an incentive, in the form of a calling card, for using an alternative mode?

**Computer-assisted response modes.** In web-based surveys, respondents access a survey web site, enter codes to authenticate their identities, and fill out the survey. They respond to multiple choice questions by clicking on their responses. They respond to text-entry questions by typing in their responses.

With inbound CATI, the respondents call a toll-free number to reach a CATI interviewer who administers the survey. A software system guides the interviewer through the interview by displaying the wording for each question. The interviewers read the questions on the computer screen and enter the responses. The interviewers can note any unusual responses or situations, such as broken-off interviews. The software also allows the interviewers to keep track of each respondent's status. For example, the interviewers can identify the respondents who have called or who are scheduled for follow-up calls.

An IVR survey enables respondents to take a survey over the telephone by interacting with a "talking computer." The respondents hear digitized voice files that present the survey instructions and questions. The respondents answer by speaking. Speech recognition software transforms the respondents' spoken replies into responses recorded in a database. Contemporary

speech recognition software is speaker-independent; it attempts to recognize spoken words regardless of the accent, gender or identity of the speaker. To some extent, IVR questionnaires emulate face-to-face interviewing. For example, the respondent might hear the IVR system ask, "What is your child's sex?" The computer would accept the respondent's spoken replies, such as "a daughter" or "she's a girl" or "my child is female." Of course, an IVR system can also accept responses entered with the familiar "touch tone" buttons. For example, the respondent might hear, "Please enter your telephone number, using the touch-tone keys." The respondent would reply by pressing the buttons.

**Rationale.** CATI, IVR, and Internet-based surveys have some advantages and some drawbacks relative to paper-and-pencil questionnaires. One of their advantages is their automatic ability to follow the survey branching logic, no matter how complex. The computer-based methods prevent respondents from making common errors, such as selecting more than one response for "choose one" multiple choice questions, filling out multiple surveys, or submitting surveys with contradictory or illegible information. Also, the three computer-assisted survey methods store responses in a database, eliminating transcription costs.

An intriguing but untested advantage of computer-assisted data collection methods is their potential ability to increase response rates. For example, respondents who are unable or unwilling to use a paper questionnaire form might use computer-assisted methods of responding by telephone. For example, CATI and IVR systems can accommodate respondents with limited literacy skills. Results of the National Adult Literacy Survey has suggested that the reading proficiency of about 30 million English-speaking Americans was at the lowest defined level. About one-fifth of this group have trouble reading because of visual difficulties (Kirsch, Jungeblut, Jenkins, and Kolstad, 1993). Many people who have limited literacy skills have particular problems filling out forms. However, telephone-based data collection methods do not require respondents to read or fill out any forms. With the IVR survey and inbound CATI available, respondents need only be able to provide their data verbally in response to a series of questions over the telephone.

Some respondents may perceive IVR and Internet applications as interesting or "high tech" data collection methods and be eager to provide their census data this way when given the option. IVR applications are quickly becoming more numerous and popular. Callers are likely to encounter IVR automated-attendant systems when they telephone the customer service departments of airline or mutual fund companies, or other firms that handle a large number of customer inquiries. Similarly, some people may prefer web surveys over paper-and-pencil forms. The Internet is an essential part of daily life for a large and increasing number of people.

Some respondents might find computer-assisted data collection systems to be more private and less burdensome than paper-and-pencil surveys. The results of the Survey of Census Participation, which studied the 1990 census, suggest that some people failed to provide their census data because they believed that the forms required a great deal of time or because they feared that the Census Bureau might not keep the information confidential (Couper, Singer, and Kulka, 1998; Singer, Mathiowetz, and Couper, 1993). However, in one study, respondents rated

an IVR application as the “easiest” way to provide personal data (Turner, Miller, Smith, Cooley, and Rogers, 1996). Several studies suggest that respondents are more willing to reveal personal data on computer-administered surveys than in person-to-person surveys or on paper forms (e.g., Turner, Ku, Rogers, Lindberg, Plaeck, and Sonenstein, 1998).

In summary, computer-assisted data collection methods may appeal to many potential survey respondents, including ones who would not respond using paper forms. Nonetheless, computer-mediated data collection systems also have drawbacks. For example, web surveys can be inconvenient to many potential respondents. Most American households still do not have Internet access, even though the proportion of households with Internet access rose from 26.2 percent in December 1998 to 41.5 percent in August 2000 (National Telecommunications and Information Administration, 2000). Those who do not have access to the Internet at home or work may have access at a library, community center, or “cybercafé,” but these locations may not be convenient. Some elderly, infirm, and physically challenged persons who do not have access to the Internet at home may have particular difficulty gaining access to the Internet outside of their homes. In addition, some persons may not feel confident that they have the skills needed to fill out a web-based survey.

IVR systems also present some problems. Speech recognition technology is new and has not yet been perfected. It can be slow or error-prone, particularly in surveys that contain questions with many possible responses. Also, some respondents may equate IVR surveys with annoying call routing systems.

CATI systems, too, may not appeal to some respondents. Many respondents might prefer not to reveal personal information to another person over the telephone.

Since computer-assisted methods have both advantages and drawbacks, their impact upon response rates is uncertain in a large data collection effort. The present study, RMIE, was intended to assess the impact of CATI, IVR, and Internet-based data collection methods upon the response rates in Census 2000.

## METHOD

**Households.** A total of 35,376 households were randomly selected for this study from the Decennial Master Address File (DMAF) developed for Census 2000. All of these households were from the 94.3 million households in mailout/mailback areas, where the Census Bureau collects data by mail with paper forms. All of the selected households were scheduled to receive the short version of the form.

**Experimental Design.** Of the households selected for this study, 15,737 were randomly dispersed among six panels in a 3 by 2, fully factorial design. The first factor, response mode, had three levels: CATI, IVR, and Internet. The second factor, incentive, had two levels: incentive and no incentive.

All of the households had the option of mailing in their census data on the usual paper forms. In addition, one-third were given the option of using a CATI survey, one-third were given the option of using an IVR survey, and one-third were given the option of using a web survey. Half of the households were offered an incentive, in the form of a telephone calling card, for using the alternative, computer-assisted response mode, and half were not offered an incentive.

The number of households assigned to each of the six panels were as follows:

CATI with no incentive	2,621
IVR with no incentive	2,621
Internet with no incentive	2,626
CATI with incentive	2,621
IVR with incentive	2,625
Internet with incentive	2,623

**Census Control Group.** The remaining 19,639 households that were selected for this study comprised the Census Control Group (CCG). The CCG received mailings that contained a cover letter and a census short form, like households that were not selected for this study. The mailings offered the CCG households neither a calling card incentive nor a computer-assisted response mode option. The CCG served as a group against which the six panels in this study could be compared.

**Mailings.** The Census Bureau mailed a short form for Census 2000 and a cover letter to each household in the six experimental panels and the CCG around March 13, 2000, the same time that census forms were mailed to all households in the nation. The cover letter explained that the household could provide census data in either of two ways. The household could mail in the data in the usual way, using the paper form enclosed in the mailing. Alternatively, the household could use a computer-assisted method. The cover letters to the two CATI panels explained that the household could provide data over the telephone by dialing a toll-free number. The letters did not explicitly mention that the data would be collected by a CATI operator. Similarly, the cover letters to the two IVR panels explained that the household could provide data over the telephone by calling a toll-free number (different from the CATI number). The letter did not explicitly mention that the data would be collected by an IVR system. The cover letters to the two Internet panels explained that the household could provide data via a web-based questionnaire, and provided its address (URL). The questionnaire appeared on a single web page. The respondent scrolled down the page, which presented the same questions that appeared on the paper census form.

The mailings to the three incentive panels contained the paper Census form, a cover letter, and an insert, printed in color on heavy stock paper. The calling card could be peeled off of the insert. The cover letter and insert explained that if the household provided its census data using the computer-assisted method, the calling card would be activated, giving it value worth 30 minutes of domestic calls. The mailings to the three no-incentive panels contained only the paper Census form and a cover letter, with no calling card or insert. The mailings to the

incentive panels were assembled and posted by the Census Bureau; the mailings to the no-incentive panels were the responsibility of a contractor.

The paper census forms sent to the households in all six panels provided a toll-free “Operator Assistance” (OA) number to call with questions about Census 2000 generally and about the incentive and alternative response modes. This number was different from the toll-free OA number that appeared on the forms received by households that were not in this study.

**Leakage across response mode conditions.** Some leakage could occur across the response mode conditions. If respondents called OA and asked to provide their census data, the OA operator transferred the call to a CATI operator who collected the respondents’ data. Even respondents who were in the CCG or the IVR or Internet conditions could provide their data via CATI in this way.

Also, several news organizations and public officials disclosed the URL of the Internet-based census form. Consequently, all households in the country, including those that were assigned to the IVR or CATI conditions in this study, could provide their data via the Internet survey.

If the speech recognition software was unable to understand the utterance of an IVR respondent at a criterion level of certainty, the system repeated the question. If the response was still unintelligible to the software, the system transferred the call to a CATI operator who completed the interview. Thus, a respondent in the IVR condition could ultimately provide census data via CATI.

## RESULTS

**Last year’s presentation.** At last year’s AAPOR annual conference in Montreal, our research team reported that we found very low response rates for the IVR-no incentive panel. We were not yet able to explain this result. We have since determined that the mailing for this panel was mishandled. Almost no data were received within the Census 2000 data collection time period from any household in this panel located in nine Midwestern and Western states and in large parts of three other states. Apparently the mailings were somehow delayed or lost for households in those areas.

This problem spoiled the data for the IVR-no incentive panel and forced us to drop the panel from the response rate analyses. In order to maintain a balanced, factorial design, we also dropped the IVR-incentive panel. The remainder of this paper reports the findings for the other four panels.

**Response rates and weighting.** Response rates were calculated three ways. “Overall response rates” considered all modes of responding: paper forms, Internet, IVR, and CATI. “Assigned mode response rates” considered only the computer-assisted alternative response mode of the

respondent's panel. "Mail-only response rates" considered only the mailed-in paper forms and ignored the computer-assisted response modes. In all three response rate calculations, households were considered nonrespondents if they failed to respond at all, or if they provided data that were too incomplete to meet the criteria for a "sufficiently complete response" for Census 2000 according to the criteria of the Census Bureau. Households whose mailings were returned by the U.S. Postal Service as undeliverable as addressed were dropped from the analyses. A chi square test revealed no significant differences among the panels as to the proportion of households whose mailings were undeliverable as addressed.

The data analysis took into account the leakage across the response mode conditions. When the assigned mode response rates were calculated, households that were assigned to a CATI panel and that provided data to a CATI operator were always counted as respondents, even if they reported their data after having called OA and been transferred from the OA operator.

The following households were not counted as respondents when the assigned mode response rates were calculated:

- Those who were assigned to an Internet panel and who provided data to a CATI operator after having called OA and been transferred from the OA operator.
- Those who were assigned to a CATI panel who provided their data on the web survey.

All estimates were weighted by the inverse of the probability of selection. Significance tests were computed with a replicate variance estimation method, using a Westat-authored program, WesVar Version 4, which used a jackknife balanced replication procedure.

**Effect of the alternative response mode upon the overall response rate.** Chi square analyses were performed to assess the impact of offering a computer-assisted alternative response mode option, without the calling card incentive. The comparisons involved three groups: the CCG, which was offered no alternative response mode, the CATI-no incentive panel, and the Internet-no incentive panel.

The overall response rates of both the CATI-no incentive panel (73.50 percent) and the Internet-no incentive panel (73.90 percent) were greater than the overall response rate of the CCG (71.44 percent; chi square = 5.38,  $p < .02$  and 6.33,  $p < .02$  respectively). Thus, offering a CATI survey or a Internet survey option increased the proportion of households that provided their census data; some of the households that would not have provided their census data on the paper forms did provide their data when the Internet or CATI response mode was offered.

The overall response rate of the Internet-no incentive panel (73.90 percent) did not differ from that of the CATI-no incentive panel (73.50 percent; chi square = 0.10, n.s.). Thus, increase in the overall response rate associated with the CATI survey option was not significantly different from the increase in the overall response rate associated with the Internet survey option.

The mail-only response rate of the Internet-no incentive panel (69.62 percent) was lower than



that of the CCG (71.44 percent; chi square = 3.83,  $p < .05$ ). Thus, when households were offered the option of using an Internet survey, some of the households that would have provided their census data on the traditional paper forms switched to the Internet survey instead.

The mail-only response rates of the CATI-no incentive panel (72.00 percent) and the CCG (71.44 percent) did not differ (chi square = 0.37, n.s.). Thus, when households were offered the option of using a CATI survey, households that were willing to provide their census data on the traditional paper forms did not tend to switch to the CATI survey.

**Effect of the incentive upon the assigned mode response rate.** Additional chi square analyses were performed to assess the impact of the incentive.

The incentive that was offered for using the alternative response mode greatly increased the proportion of households responding via the alternative response mode. When households were offered no calling card incentive for using the CATI survey, only 1.45 percent used it. When households were offered the incentive, 17.68 percent used it (chi square = 417.94,  $p < .0001$ ). When households were offered no incentive for using the Internet survey, only 3.98 percent used it. When households were offered the incentive, 15.37 percent used it (chi square = 205.26,  $p < .0001$ ).

Thus, considering both modes together, the incentive brought about a 13.81 percentage point increase in the assigned mode response rate, from 2.71 percent with no incentive to 16.52 percent with an incentive (chi square = 575.07,  $p < .001$ ).

When no incentive was offered, households were less likely to use the CATI survey option than the Internet survey option (1.45 percent versus 3.98 percent, chi square = 23.94,  $p < .0001$ ). When the incentive was offered, households were more likely to use the CATI survey option than the Internet survey option (17.68 percent versus 15.37 percent, chi square = 6.13,  $p < .02$ ).

**Effect of an incentive upon the overall and mail-only response rates.** The overall response rate of the two panels that were offered an incentive (71.65 percent) was lower than the overall response rate of the two panels that were not offered an incentive (73.70 percent, chi square = 4.32,  $p < .05$ ). Even though the incentive was associated with a 13.81 percentage point increase in the assigned mode response rate, that increase was offset by a larger, 15.85 percentage point reduction in the mail-only response rate (70.81 percent for the no-incentive panels, 54.96 percent for the incentive panels, chi square = 275.56,  $p < .001$ ). Consequently, the incentive brought about the 2.05 percentage point reduction in the overall response rate.

However, this effect did not reach statistical significance when the data for each response mode were analyzed separately. The overall response rate of the CATI-incentive panel (71.81 percent) was somewhat lower than the overall response rate of the CATI-no incentive panel (73.50 percent), but this effect was not statistically significant (chi square = 2.05, n.s.). Similarly, the overall response rate of the Internet-incentive panel (71.50 percent) was somewhat lower than the overall response rate of the Internet-no incentive panel (73.90 percent), but again this effect was not statistically significant (chi square = 2.23, n.s.).

When the incentive was offered, the overall response rate of the Internet panel (71.50 percent) did not differ from that of the CATI panel (71.81 percent; chi square = 0.05, n.s.). Thus, the effect of the incentive on the proportion of households who provided their data was not significantly greater for either of these two response modes.

## Discussion

**Alternative response modes.** The results suggest that offering an Internet or CATI survey option in the decennial census may increase the proportion of households who provide their data. The increase was small in this study, only about 2.5 percent over the proportion achieved when no alternative response modes were offered. However, even a small increase in the initial response rate could result in substantial cost savings by reducing the workload of enumerators who must visit the non-responding households.

The CATI and the Internet alternative survey options brought about roughly equal increases in the overall response rate. That is, they had similar appeal to households that would not report their data on the paper forms.

Among respondents who would be willing to use the paper forms, however, only the web survey had significant appeal. Some of these respondents switched from the paper forms to Internet when it was offered, apparently because they preferred a web survey over a paper-and-pencil survey. Respondents who were willing to use the paper forms did not detectably switch to the CATI survey, however, when it was offered.

**Incentive.** The results show that offering a calling card incentive can greatly increase usage of an alternative response mode. This increase was greater for the CATI survey than for the web survey. The reason for this difference may be that a larger proportion of households have convenient access to a telephone than to the Internet. As a result, a larger proportion of households in the CATI panels than in the Internet panels were in a position to take advantage of the calling card offer.

Even though the incentive brought about this large increase in the response rate via the alternate mode, it did not bring about an increase in the overall response rate. The reason was that the 13.81 percentage point increase in the assigned mode response rate (for the CATI and Internet panels combined) was offset by a larger, 15.85 percentage point decrease in the mail-only response rate. This finding may suggest that when some households were offered the calling card, they resolved to use the CATI or Internet response mode rather than the paper forms. In some cases, however, the respondents found that they were unable to use the CATI or Internet survey as they intended. Perhaps some potential respondents called the CATI system after hours, when no interviewers were on duty. Perhaps some potential respondents could not locate a computer with Internet access, or could not get through to their Internet Service Providers. These people may have put their census mailings aside, with the intent of trying again later to complete the Internet or CATI survey. Some of these people, however, may have neglected to return to the mailing; they eventually responded neither on the paper form nor via the alternative response mode. As a result, the incentive brought about a small decrease in the overall response

rate, for both the CATI and Internet panels.

The effect of the incentive in increasing the assigned mode response rates might be attributable to the calling card itself, or to the insert upon which it was attached. These inserts described the computer-mediated response alternative in large, colorful text. The mailings to the households in the no-incentive condition included only a cover letter and a census form, without any insert. People in the no-incentive households who ignored the cover letter would not have learned that they had the option of responding using a computer-mediated method. The inserts in the mailings to the households in the incentive condition, however, were not easily ignored. Because of these inserts, people in the incentive-condition households may have been more likely than people in no-incentive households to be aware that they could respond by an alternative, computer-mediated mode. The inserts may have also conveyed a sense that the computer-mediated response alternative was important and desirable.

### Conclusions

**Alternative response modes.** Should either CATI or Internet alternative response modes be offered for the decennial census? Both options appear to boost the response rate by a few percentage points, converting some households that would ignore the paper forms into respondents. Therefore, both may reduce the expense of sending enumerators into the field after the initial data collection period. Both eliminate the costs of handling and scanning paper forms. However, the CATI survey can be expensive itself. The telephony system for the CATI survey must be developed and maintained. Interviewers must remain on duty continually, even during periods when few respondents are calling. By contrast, an Internet survey requires a web-based data collection system, but not interviewers. For that reason, the Internet alternative response mode may be more cost-effective than the CATI mode.

In addition, some respondents who were willing to use the paper forms switched to the Internet survey, but not the CATI survey, when these alternatives were available. Collecting census data via the Internet survey may be somewhat less expensive than obtaining the data on the paper forms by mail. These savings could further justify a web survey in the decennial census.

**Incentive.** Should households be offered an incentive for using an alternative response mode in the decennial census? The results of this study do not suggest that the incentive would increase the proportion of households who provide their census data. To the contrary, the incentive simultaneously boosted use of the alternative computer-assisted response mode and decreased use of the traditional paper forms. The cost of the incentives therefore could not be offset by savings engendered by an increased response rate.

### REFERENCES

Couper, M.P., Singer, E. & Kulka, R.A. (1998) Participation in the 1990 decennial census: Politics, privacy, pressures. *American Politics Quarterly*, 26, 59-80.

Kirsch, I.S., Jungeblut, A., Jenkins, L. & Kolstad, A. (1993) *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*. Washington, DC: National Center for Education Statistics.

National Telecommunications and Information Administration (2000) *Falling through the net: Toward digital inclusion. A report on Americans' access to technology tools*. Washington, DC: Author.

Singer, E., Mathiowetz, N.A., & Couper, M.P. (1993) The impact of privacy and confidentiality concerns on survey participation: The case of the 1990 U.S. Census. *Public Opinion Quarterly*, 57, 465-482.

Singer, E., Hoewyk, J. V., Gebler, N., Raghunathan, T. and K. McGonagle (1999) "The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys." *Journal of Official Statistics* 15, 217 - 230.

Turner, C.F., Miller, H.G., Smith, T.K., Cooley, P.C., and Rogers, S.M. (1996) *Telephone audio computer assisted self-interviewing and survey measurement of sensitive behaviors: Preliminary results*. Technical Paper 21, Research Triangle Institute.

Turner, C.F., Ku, L., Rogers, S.M., Lindberg, L.D., Pleck, J.H., & Sonenstein, F.L. (1998) Adolescent sexual behavior, drug use and violence: Increased reporting with computer survey technology. *Science*, 280, 867-873.