DATA QUALITY ISSUES IN A MULTI-MODE CENSUS: RESULTS FROM THE MAIL AND TELEPHONE MODE TEST (MTMT)

Nancy Bates, U.S. Bureau of the Census
433 Washington Plaza, Washington D.C. 20233

KEY WORDS: Decennial census, telephone response, item nonresponse

Levels of public cooperation with the self-administered mailout/mailback method have declined over the past two censuses. In 1980 the mail return rate[1] to the census was approximately 83 percent compared to only 74 percent in 1990 (Barret 1992). In both censuses, households were first given the opportunity to complete and return a mailed census form. If no response was received by a predetermined cutoff date, a census enumerator was sent to the household to follow-up and conduct a face-to-face interview. The decreased mail response in 1990 greatly increased the number of personal enumerations required and, consequently, resulted in a higher than expected census cost.

In examining ways to curtail these rising census costs, the Census Bureau has initiated a program of research dedicated to exploring new methods of increasing public participation. One strategy calls for increasing the number of response modes offered. Many new methods such as touch-tone data entry (TDE), interactive cable TV (ICTV), home personal computers, facsimile machines (FAX) and telephone voice entry (TVE) have been discussed as additional data collection techniques for the future (U.S. Census Bureau 1993).

Because telephone service is widely available and commonly used by a large segment of the population, telephone response was an attractive candidate for immediate testing. Adding a telephone response option was hypothesized to better facilitate participation by persons with low literacy levels or those with a low "forms literacy" level, those with English language difficulties and those with eyesight problems. It would also serve as a response alternative for those who misplace or lose their forms. On the negative side, however, the telephone method presents potentially serious mode effects and operational complexities with high implementation costs.

This report examines results from an experiment conducted by the Census Bureau in the Spring of 1993 known as the Mail and Telephone Mode Test (MTMT). The experiment was designed to test the impact of providing both mail and telephone as response options in a national census test. Specifically, respondent-initiated telephone response via computer assisted telephone interviewing (CATI) was selected as the test response mode in addition to the traditional mailout/mailback method.

The test had two primary objectives: first, to determine the public's preference for responding by mail versus telephone and second, to determine whether overall response rates could be improved by offering the telephone option as a response mode. A tertiary objective was to determine the data quality effects of answering by telephone. This report presents an evaluation of this third objective by comparing rates of item nonresponse between mail and telephone responses.

MODE DIFFERENCES AND DATA QUALITY

While response levels are an important criteria by which to judge the quality of a census method, other factors should be weighed as well. For example, the data collection technique should ensure completeness and accuracy of data as indicated by low item nonresponse and consistency and reliability of data across items. The current census method already utilizes a mixed mode of mail questionnaires and personal interview. While data quality evaluations such as content reinterview surveys are routinely conducted after each census, these typically do not concentrate on an examination of data differences by mode. The possibility of yet a third method of census data collection further magnifies the need for research into mode influences on data quality and response effects.

While there are potentially several mode effects common to telephone interviews, many of them were minimized given the question content of the MTMT and the reverse-CATI design. Only the short form census questions were asked and these consist only of basic demographic questions and questions inquiring about objective aspects of the house or apartment. Additionally, only exhaustive and mutually exclusive response categories are used, avoiding the use of descriptive quantifiers which can exacerbate mode differences. Effects were also minimized because the questionnaire was mailed to all sampled households and therefore available to respondents during the interview. This provided the advantage of visually reviewing answer categories simultaneous to hearing the response options being read.

The uncertainty of whether phone respondents would complete their forms prior to placing the call was another mode effect consideration. Because information is asked about each individual within a household, telephone respondents who had not pre-completed their forms would be required to answer questions about themselves, their homes and other household members "on the spot." While in the majority of cases, this was not foreseen to present difficulties (since most U.S. households are comprised of related individuals or persons living alone), there still remained households where members were not related, such as group homes and households with boarders and live-in employees.

While "on the spot" proxy reporting was thought to be the largest threat to data quality in the reverse-CATI method, the live interview environment at the same time was believed to have many positive effects. While the self-administered method has no way of probing respondents to answer difficult or sensitive questions or questions for which respondents are unsure of an answer, interviews collected by telephone could encourage respondents to make a best guess. This method also provided immediate questionnaire assistance in helping respondents understand instructions, terms and definitions.

The computer assisted telephone interview also rigidly enforced the form's intended question order, automatically branching respondent's skip patterns. This structured process through the form meant that questions were not inadvertently skipped or left blank and while break-offs were possible during the telephone interview, respondents were more likely to complete the interview without interruptions. This control over the response process, coupled with the ability to probe, were arguably the two biggest advantages the telephone method held in achieving a superior quality of data.

Assuming that English language comprehension is correlated with levels of data quality, the telephone method was hypothesized to have an advantage in this respect as well. For example, persons whose first language is Spanish and have only limited English comprehension may struggle with an English version of the questionnaire, leaving some or all portions incomplete. However, provided that a Spanish-speaking interviewer is present to interpret and provide clarification, the same person may provide better data during a telephone interview. The MTMT provided such a service by including a 1-800 number specifically for Spanish language CATI interviews. Of course for this to be successful, Spanish-speaking respondents had to recognize that such an option was available and notice the Spanish message printed at the bottom of otherwise English questionnaires, letters and postcards.

MTMT METHODOLOGY

From a cost standpoint, the most desirable methodology of the MTMT was to elicit a high mail response early on and then introduce the telephone invitation later. This delay tactic would discourage persons who would normally respond by mail from switching to the more costly telephone alternative. Consequently, the MTMT treatments were designed to compare improvements in response resulting from a telephone invitation introduced at different stages of the implementation process.

The experiment consisted of five different treatments. The questionnaire used in all five treatments contained the content of the 1990 decennial short form but was fashioned after a "user-friendly" modified questionnaire developed as a result of an earlier census test (see the 1992 Simplified Questionnaire Test, Dillman, Sinclair and Clark 1992). The first treatment served as the control for the experiment and did not offer a telephone response option. Instead, a prenotice, questionnaire, reminder/thank-you card and targeted replacement form were used, none of which mentioned the reverse-CATI option.

Treatment 2 consisted of a prenotice, initial form and a reminder/thank-you postcard. The option of telephone response was introduced only on the reminder/thank-you postcard. Operators were available to take CATI interviews seven days a week between 8:00 a.m. and 9:00 p.m. The second treatment group did not receive the replacement form.

Treatment 3 increased the number of telephone invitations by sending an additional follow-up letter after the reminder/thank-you postcard. Both follow-up pieces carried the telephone invitation. Treatment 3 did not include the replacement questionnaire. The fourth treatment introduced the telephone option three different times -- in the reminder/thank you postcard, in a follow-up letter and on an insert accompanying a targeted replacement form. Treatments 2, 3 and 4 all introduced the telephone invitation at varying points in time between reminder/thank you and replacement form, being careful to postpone introduction of the telephone response with the intention of preventing those persons who would normally respond by mail from substituting the telephone.

The fifth and final treatment measured pure "choice" between modes by introducing the telephone option at every step of implementation (on the prenotice letter, on an insert accompanying the first form, on the follow-up postcard and on an insert accompanying the replacement form). By allowing respondents to select the telephone from the beginning, the MTMT had a pure measure of what percent of the population simply

preferred calling versus filling out and mailing back a paper questionnaire.

The sample consisted of 22,500 housing units selected nationwide. The sample was divided into two strata, one consisting of households from low mail response areas to the 1990 census (LRAs) and the second comprised of households from all other geographic areas (referred to as high response areas or HRAs). Households from LRAs were characterized by a 64 percent combined black and/or Hispanic minority population. Contrastingly, areas from HRAs had about a 15 percent combined black/Hispanic population. This stratification allowed for an examination of mode differences based on prior census behavior and thus was particularly useful in making inferences about households from areas with historically low mail response. Each of the two strata were divided equally into the five treatment groups for a total of 2,250 housing units per panel per strata.

In this paper, data quality between the mail and telephone method is evaluated by comparing rates of item nonresponse. Item nonresponse is defined as cases where 1) the question was left blank 2) an answer of "don't know" was given or 3) a refusal was given. Data for these analyses come from the 12,423 responses received by the close-out date. These consist of forms mailed back and reverse-CATI interviews conducted. All returned mail forms were included in the analyses with the exception of 69 cases determined to be returned completely blank. Data from the mailed forms were keyed and merged with data captured by the CATI instrument.

Estimates of item nonresponse were generated using Variance Estimates for Complex Samples (VPLX), a software estimation package which produces standard errors adjusted for the MTMT's clustered sample design. Estimates referred to as "overall" have been weighted to reflect the LRA and HRA sample stratification. Weighted estimates reflect the approximate number of housing units nationwide in the census mailback universe (approximately 88 million households). Significant differences were determined using t-tests with a confidence interval of 90 percent.

Before presenting a discussion of the results, an important qualification is necessary. It is recognized that the types of people using the telephone as a response mode may differ from those using the mail method. Therefore, differences observed in the rates of item nonresponse between modes could be a result of differences in the characteristics of the respondents and not the mode itself. For example, because of poor eyesight or a greater desire for social interaction, elderly respondents may be inclined to use the telephone option more often than younger respondents. However,

because of memory loss or other health-related difficulties, elderly respondents may also have higher item nonresponse rates regardless of the data collection mode.

Because of the MTMT's voluntary nature of mode selection, it is of interest to examine the characteristics of persons responding by mail versus telephone before comparing the rates of item nonresponse between them. This report provides only a preview by examining the age and ethnicity characteristics of the respondent for the household.

RESULTS

The overwhelming majority of MTMT responses were received by the self-administered mail method. Of those who responded in panel 5 (which offered the choice of responding by mail or telephone in every contact), only 8 percent chose the telephone (West 1993). Of the entire 12,423 responses received from all 5 panels, approximately 95 percent were mail responses, and the remaining 5 percent were reverse-CATI call-ins. Table 1 presents the age and ethnicity distributions, with standard errors, for respondents who used the mail method versus those who participated by phone. These characteristics reflect the person who placed the CATI call or who indicated they had filled out the mail questionnaire for their household. In cases where the household respondent was not known, a best guess was made by assigning the first person on the form as the respondent. All telephone responses are combined in the same category regardless of the treatment they were in.

Household Respondent Age and Ethnicity by Method of Response

While the reverse-CATI service was in operation, there was anecdotal information from headquarters observers to suggest that the telephone method was being used disproportionately by elderly respondents. However, as Table 1 indicates, the overall respondent age distribution was very similar across modes. No significant differences were found overall, but two differences were discovered within the Low Response Areas (LRAs). Contrasted to LRA mail respondents, LRA telephone respondents were significantly more likely to fall into one of the more extreme end age categories (19-29 or 70+). No significant differences in the age distributions between modes were found in HRAs.

Five percent of the respondents overall reported a Spanish origin of Mexican, Cuban, Puerto Rican, or some other Spanish ancestry[2]. The distribution did not significantly differ by mode of response, overall or within strata. As would be expected given the strata

738

design and selection, the LRAs reflected a much higher proportion of Spanish respondents compared to the HRAs but the difference between modes within LRAs was not significant.

Item Nonresponse for Population Questions

Table 2 contains estimates of item nonresponse and standard errors for the sex, marital status, date of birth, relationship, race and ethnicity questions by type of data collection method. For the most part, the telephone method had consistently lower rates of missing information compared to data gathered from mailed questionnaires. Item nonresponse was generally 1 percent or less for the population questions asked via CATI but ranged as high as 9 percent for certain items from mailed questionnaires. Overall, item nonresponse was significantly less over the telephone for date of birth[3], relationship, Spanish origin and race. The lower telephone nonresponse was consistent for these 4 items within both strata and in addition, item nonresponse for marital status was found to be lower for CATI in LRAs. The higher completion rates for CATI are probably due to the encouragement of "best guesses" in place of missing information, for example, the acceptance of age as a proxy for the exact day, month and year of birth.

Item Nonresponse for Housing Questions

For the first five housing questions (type of structure, acreage, commercial property, number of rooms and tenure), the level of missing information was significantly lower for telephone cases compared to the same information gathered from mail responses. This was true overall, and within both strata (see Table 3). However, neither the rent amount nor property value question showed a significant difference in item nonresponse between method of response. For the rent amount item, the majority of CATI item nonresponse consisted of refusals whereas the majority of nonresponse for property value came from "don't know" answers. (Nonresponse in mail cases are not so easily categorized into refusals, "don't knows", or inadvertent skips, since the majority are simply left blank without explanation).

It is interesting that the only two questions failing to exhibit a higher rate of response for CATI were the questions about monthly rent amount and property value. Both are monetary questions that conceivably reflect some measure of income and, as such, are arguably two of the more sensitive short-form questions. This supports the notion of respondents being more hesitant to provide information perceived as personal or sensitive in a live interview. It may also be the case

that property value is a difficult question to answer "on the spot," particularly if the respondent for the household is not the householder. Finally, the way in which these two questions were presented differed slightly between modes. Because both questions contained 26 answer categories, both were asked as open-ended questions during CATI. The interviewer then used the answer given to select one of the categories. For persons who did not have the questionnaire in front of them, this may have made responding somewhat more difficult. Despite this, we should be mindful that while the telephone method failed to obtain _more_ data for these two items, it did not obtain significantly _less_ than the self-administered method.

CONCLUSIONS

Despite the experimental offer to participate by phone, the overwhelming majority of responses to the MTMT came from the traditional self-administered method. The mail response evaluation failed to show that the addition of a telephone response option improved participation rates above and beyond what they would have been without it (Clark, Dillman, and West; 1993). While this certainly makes the telephone method appear less desirable as an option for Census 2000, the initial research questions concerning characteristics of the telephone respondent population and potential mode effects are still of interest.

A glance at two selected demographics, age and ethnicity, suggests that respondents were similar across modes. Overall, the age distribution for the mail and telephone respondents did not significantly differ. Likewise, the distribution of persons reporting Spanish origin did not significantly vary depending upon whether the response was received by mail or reverse-CATI -- this fact remained when examining ethnicity within strata. It should be noted that these illustrations of age and ethnicity do not necessarily reflect the characteristics of respondents who _preferred_ the telephone method because all five treatments were combined here and broken out only by method of response. Consequently, this examination does not control for the timing, method, or frequency of the telephone invitation.

The analysis next examined the question of whether the two response methods differed in terms of data quality. The measure used in this report, item nonresponse, provides only a partial answer because it does not address accuracy or content differences. It does, however, provide an indication of whether one mode is superior in obtaining more complete data than another. We found that for 4 of the 6 population questions, phone responses had significantly lower

levels of missing information; the same was true for 6 of the 8 housing questions. For at least one question (Spanish origin), the reduction of missing information was quite substantial (from over 6 percent item nonresponse overall to less than 1 percent). However, the questions of monthly rent amount and property value failed to obtain better item response rates when asked over the phone suggesting that the strength of telephone data quality weakens somewhat when the item in question is sensitive or requires an above standard level of knowledge. These findings are especially relevant if the census long form is ever considered in a multi-mode approach. The trend of lower item nonresponse levels remained constant when compared by strata, suggesting that the data quality gains for the phone method are realized in both LRAs and HRAs.

Based on this examination of data quality, it is evident that the reverse-CATI cases yielded more complete responses than those received by the mail. However, due to the self-selection nature of the telephone respondents, we cannot be certain that this is entirely attributable to mode differences. It is plausible that the telephone respondents were more motivated than mail respondents and, consequently, represented a better than average respondent. It should also be considered that only a small minority of households used the CATI option, that there was no additive boost in mail response as a result of offering the additional mode, and that the costs associated with starting-up, staffing and maintaining such a telephone operation are extremely high. Consequently, the evidence of better data quality reported here must be viewed in the context of many other factors.

REFERENCES

Barret, D.F. (1992) "1990 Census Mail Return Rates." U.S. Bureau of the Census, Decennial Statistical Studies Division REX Memorandum Series #Q-13.

Bates, Nancy (1993) "The 1993 Mail and Telephone Mode Test: Respondent Characteristics by Method of Response," U.S. Census Bureau report to S. Miskura, Year 2000 Research and Development Staff, August 30.

Clark, J.R., Dillman, D.A. and West, K. K. (1993) "Influence of an Invitation to Answer by Telephone on Response to Census Questionnaires." A paper presented at the American Statistical Association Annual Meetings, San Francisco, CA.

Dillman, D.A., Sinclair, M.P., and Clark, J.R. (1992) "Mail-Back Response Rates for Simplified Decennial Census Questionnaire Designs." A paper presented at the American Statistical Association Annual Meetings, Boston, MA.

Tarnai, J. and Dillman, D.A. (1992) "Questionnaire Context as a Source of Response Differences in Mail versus Telephone Surveys," in Context Effects in Social and Psychological Research, Schwarz, N. and Sudman, S. (eds), Springer-Verlag: New York.

U.S. Census Bureau (1993) "U.S. Bureau of the Census Technology Assessment of Data Collection Technologies for the Year 2000: Final Report." A report prepared by Ogden Government Services and IDC Government, Task No: N1S132010.

West, Kirsten (1993) "1993 National Census Test Mail and Telephone Mode Response Evaluation: Final Report." U.S. Bureau of the Census, DSSD 2000 Census Memorandum Series #E-47.

NOTES

[1.] The mail return rate is the ratio of the number of households returning a census questionnaire by mail to the total number of occupied housing units that should have received a questionnaire by mail or by a census enumerator.

[2.] Determination of Spanish origin was based only upon the check-box portion of the ethnicity question.

[3.] Nonresponse to date of birth was defined as cases missing all three components (day of birth, month of birth, and year of birth).

Table 1. Household Respondent Characteristics by Mode

RESPONDENT CHARACTERISTIC	OVERALL (%)		1990 LRAs		1990 HRAs	
	Mail	Telephone	Mail	Telephone	Mail	Telephone
Age: < 18	0.36% (.07)	.76% (.44)	.50% (.10)	0.0% (0.0)	.35% (.07)	0.9% (.49)
19-29	10.0 (.37)	11.4 (1.5)	10.8 (.50)	14.3 (2.0)	9.9 (.40)	11.1 (1.7)
30-44	31.3 (.57)	34.1 (2.3)	30.3 (.71)	27.1 (2.5)	31.4 (.62)	34.9 (2.6)
45-59	24.6 (.50)	23.5 (2.0)	24.3 (.65)	20.7 (2.3)	24.6 (.55)	23.9 (2.3)
60-69	14.5 (.41)	12.0 (1.6)	15.7 (.54)	15.0 (2.0)	14.4 (.44)	11.7 (1.7)
70+	19.2 (.53)	18.2 (1.9)	18.4 (.63)	22.9 (2.5)	19.3 (.57)	17.6 (2.1)
Ethnicity: Non-Spanish	95.0 (.25)	94.0 (1.0)	78.2 (.83)	79.9 (2.3)	96.5 (.27)	95.7 (1.1)
Spanish	5.0 (.25)	6.0 (1.0)	21.8 (.83)	20.1 (2.3)	3.5 (.27)	4.3 (1.1)

Table 2. Item Nonresponse Rates for Population Questions by Mode

CENSUS QUESTION	OVERALL (%)		1990 LRAs		1990 HRAs	
	Mail	Telephone	Mail	Telephone	Mail	Telephone
Sex	.51% (.05)	.49% (.21)	.52% (.07)	.60% (.36)	.51% (.05)	.47% (.24)
Marital Status	1.2 (.10)	1.0 (.33)	2.5 (.21)	1.1 (.43)	1.1 (.10)	.94 (.37)
Date of Birth	.88 (.07)	.26 (.15)	1.0 (.11)	.48 (.29)	.86 (.08)	.24 (.17)
Relationship	1.7 (.13)	.29 (.19)	2.0 (.18)	1.0 (.51)	1.7 (.14)	.20 (.20)
Spanish Origin	6.7 (.27)	.71 (.26)	9.1 (.45)	.72 (.38)	6.4 (.30)	.71 (.29)
Race	2.6 (.15)	1.0 (.40)	4.9 (.34)	.72 (.38)	2.3 (.16)	1.1 (.45)

Table 3. Item Nonresponse Rates for Housing Questions by Mode

CENSUS QUESTION	OVERALL(%)		1990 LRAs		1990 HRAs	
	Mail	Telephone	Mail	Telephone	Mail	Telephone
Type of Structure	4.3% (.22)	.25% (.25)	7.4% (.39)	0.0% (0.0)	4.0% (.24)	.28% (.28)
10+ Acres	1.9 (.18)	.21 (.11)	5.6 (.47)	2.3 (1.2)	1.6 (.19)	0.0 (0.0)
Commercial Property	2.2 (.20)	0.0 (0.0)	3.0 (.35)	0.0 (0.0)	2.1 (.21)	0.0 (0.0)
Number of Rooms	3.6 (.21)	.25 (.25)	5.9 (.35)	0.0 (0.0)	3.4 (.22)	.28 (.28)
Tenure	7.8 (.30)	.79 (.44)	11.2 (.47)	.32 (.32)	7.5 (.32)	.85 (.49)
Rent Amount	2.2 (.33)	4.3 (1.7)	1.8 (.32)	.70 (.71)	2.2 (.37)	4.8 (2.0)
Board	4.8 (.47)	0.0 (0.0)	5.2 (.53)	0.0 (0.0)	4.7 (.53)	0.0 (0.0)
Property Value	5.4 (.31)	5.2 (1.4)	7.7 (.53)	7.3 (2.1)	5.3 (.33)	5.0 (1.5)

741