

# WHAT DO WE REALLY GAIN FROM USING REPLACEMENT QUESTIONNAIRES TO INCREASE RESPONSE TO MAIL SURVEYS?

Theresa F. Leslie, U. S. Bureau of the Census  
DSSD, Room 2503/2, Washington, DC 20233-7600

**Key Words:** Mail Surveys, Response

## Introduction

The present study seeks to address a gap in the survey literature about the quality of data provided by survey respondents who respond to mail surveys only after receiving four mailings including initial and replacement questionnaires. At least four mailings are recommended by Dillman's Total Design Method (Dillman, 1978). Research conducted by the U.S. Bureau of the Census since 1992 has shown that mailing a replacement questionnaire to initial nonrespondents increases mail response rates by about 10 percentage points (Treat, 1992, 1993a, 1993b, and 1993c). It is important to evaluate the quality of the data collected from these late respondents to learn what effect they have on total error. This study looks at three aspects of data quality--the coverage of persons in the late response universe, the completeness of data provided on the questionnaire, and the reliability of answers provided to the survey questions.

Survey researchers have found many different ways to increase response rates to mail surveys. Examples include respondent-friendly questionnaire design--a questionnaire judged by researchers to be easy for respondents to complete (Dillman, 1991); mailing advance letters to legitimize the survey request, convey the importance of the survey, and invoke a norm of reciprocity (Dillman, Clark, and Sinclair, 1995; Dillman, 1991; Yammarino, Skinner, and Childers, 1991; and Heberlein and Baumgartner, 1978); mailing a reminder postcard (Dillman, Clark, and Sinclair, 1995; Dillman, 1978; Dillman et al., 1984); and mailing a replacement questionnaire to nonrespondents (Dillman, 1991; Goyder, 1982).

While prior evidence is lacking, some have hypothesized that response error (that is, the incorrect responses to items on completed questionnaires) may be greater for late respondents than early respondents to mail surveys (Childers and Moriarity, 1993; Griffin and Moriarity, 1992). This study attempts to compare the quality of data collected for early and late respondents to learn if late respondents (that is, persons who complete a replacement questionnaire)<sup>1</sup> introduce higher levels of error into the survey estimates compared to respondents who complete an initial mail survey questionnaire (early respondents). In particular, this study looks at three aspects of data quality. First, it examines coverage errors (that is, the count of persons missed or counted erroneously during the survey data collection). Second, the study examines the completeness of questionnaires received early or late in the

data collection cycle. Finally, this study uses reinterview data to assess the reliability of the data provided by early and late respondents.

## Methodology

**SURVEY DESIGN**--This study uses data collected from four of seven U.S. 2000 Census Test short-form questionnaire panels. The U.S. 2000 Census Test (Census Day was March 2, 1996) was a national probability sample of 94,500 households that were mailed census questionnaires in the 1990 census. The sample was stratified based on race, Hispanic origin, and tenure variables at the 1990 Census Tract/Block Numbering Area level. Two strata, or partitions, were defined--High Coverage and Low Coverage Areas. The Low Coverage Area stratum contained 17,359,020 households and had many minority persons and renters. The High Coverage Area stratum contained 71,812,378 households and included the residual population. This mail survey was the major vehicle for testing subject content and specific question wording, format, and sequencing of items for Census 2000. The seven short-form panels tested different questionnaires and procedures. The national response rate<sup>2</sup> was 68.91 percent.

The four short-form questionnaire panels used for this study represent an experimental design that provide opportunities to evaluate questionnaire design and rostering methodologies. All four questionnaires were designed using the same principles; that is, they were printed in green, had the same cover page and were mailed in the same plain white envelope, which had the following message printed on the front "U. S. Census Form Enclosed. Your Response is Required by Law". Analysis was done on all four panels to make sure that findings related to the difference in the quality of data provided by early and late respondents were not really functions of the panel design. A description of the experimental design of these four panels appears in Table 1.

**Table 1. Description of the four panels used in this study**

---

Panel 1: 12-page questionnaire with a roster (to list all persons living in the household) and a complete list of residence rules. The roster was used during the 1990 Census to ensure adequate within-household coverage.

Panel 2: 12-page questionnaire with a box to enter the number of persons living in the household (as opposed to a roster) and a complete list of residence rules.

Panel 3: 12-page questionnaire with a box to enter the number of persons living in the household. There was no roster or list of residence rules provided.

Panel 4: 4-page questionnaire with a shortened list of residence rules and a box to enter the number of persons living in the household.

---

For each short-form panel in the U.S. 2000 Census Test, a systematic sample of 2,400 households was selected from the High Coverage Area stratum and a systematic sample of 3,600 households was selected from the Low Coverage Area stratum. The U.S. Postal Service attempted to deliver four mailings to households in the national probability sample—an advance letter (mailed February 23, 1996), an initial questionnaire (mailed February 28, 1996), a reminder postcard (mailed March 4, 1996) and a replacement questionnaire with a letter for nonresponding units (mailed March 25, 1996).

Following the mail survey, a computer-assisted telephone reinterview was conducted for one-half of the mail returns from the High Coverage Area stratum and all of the mail returns from the Low Coverage Area stratum to evaluate the quality of data collected during the U.S. 2000 Census Test. The reinterview sample was designed to measure gross error differences in panel-to-panel multiple comparisons with 90 percent confidence levels. MEASURES—This analysis consisted of three parts. The coverage analysis used data from person records in all four study panels that were classified as census residents or nonresidents during residence status coding by computer or expert coders (N=21,695). Person records that remained unresolved (there was not enough information to classify them as residents or nonresidents) were excluded from this analysis because we do not have enough information to say whether these persons were omissions or erroneous inclusions. For the analysis of completeness, a household-level record was created for each mail questionnaire returned for study panels 1 and 2 that had complete data for the householder (N = 5,710 households). Data from panels 3 and 4 were excluded from this analysis because no information was collected about household tenure or marital status of household members in these panels and preliminary analysis revealed that these

characteristics differed for early and late respondents. Since we could not control for these differences statistically in panels 3 and 4, we excluded them from the completeness analysis. Data for persons who responded to both the mail survey and computer-assisted telephone reinterview were used to evaluate the reliability of the data collected from early and late respondents (N=21,510 persons).

*Coverage.* Within household coverage of persons can be studied by reinterviewing samples of listed households. Households that returned mail questionnaires were reinterviewed to enable staff to identify the census day residents for each household enumerated during the U.S. 2000 Census Test. To evaluate coverage, each person record was evaluated to determine whether it were included in the household erroneously or omitted from the mail questionnaire in error. Person records classified as unresolved were excluded from the analysis since the expert coders did not have sufficient information to determine the person's census day residence status (N=233 persons). The definitions of erroneous inclusions and omissions follow.

1. Erroneous Inclusions = Persons listed on the census questionnaire but determined not to be census day residents through the reinterview. These include mail survey duplicate persons and nonresidents.

2. Omissions = Persons who were not listed on the census questionnaire but determined to be residents on census day (March 2, 1996) through reinterview.

*Completeness.* To evaluate completeness of the questionnaires returned by early and late respondents, a completeness index was created for each mail survey questionnaire returned. This index, a continuous variable ranging in value from 0 to 1, represents the proportion of questions answered for up to seven persons in each household in panels 1 and 2. Since the data analysis file did not identify the mail survey respondent, characteristics of the householder (the person listed in person column 1 on the questionnaire) were used for this analysis since the householder was either the respondent or the respondent's spouse or partner on more than 90 percent of the 1990 census returns (Sweet, 1994). Several records were lost because no data were included for person 1 (N=12).

*Reliability.* To evaluate the reliability of data collected for early and late respondents, the overall gross difference rate or rate of gross disagreement between the census and reinterview was estimated for the early and late respondents for four demographic characteristics collected during the mail survey and then again during reinterview across all four study panels. These included age, sex, race, and Hispanic origin. The differences in the overall gross difference rates for the early and late respondents were then compared to see if they differed.

All results reported here are based on analyses of weighted data, reflecting unequal probabilities of selection.

The reinterview weights were adjusted to account for households responding to the mail survey for which we were not able to complete a reinterview. These include households that we were unable to contact (no phone number provided), did not answer the phone, or refused to be interviewed when contacted, and households not selected for reinterview (one-half of the mail returns in the High Coverage Area stratum). Tests of statistical significance and standard error estimates are based on jackknife replication, reflecting the complexity of the study design using WesVarPC (Brick et al, 1996)

### Results

**ANALYSIS OF COVERAGE**--Logistic regression was used to model the probability of being an erroneous inclusion or omission by panel, stratum, and early and late respondent (separate models were run for erroneous inclusions and omissions.) The variables examined as potential explanatory variables were: stratum (High Coverage Area, Low Coverage Area); panel (1, 2, 3, 4); type of respondent (early, late); and the interaction of stratum and type of respondent.

The models used are shown in Table 2. The first model predicts census erroneous inclusions; the second, census omissions. As can be seen from the first column of Table 2, stratum is the only significant predictor of erroneous inclusions. Persons living in the Low Coverage Area stratum were 1.72 times more likely to be erroneously included on the mail survey questionnaire than persons living in the High Coverage Area stratum ( $p < .0001$ ).

The model of omissions tells a different story. As column two of Table 2 shows, the type of respondent, stratum, and whether the return was in panel two significantly predict whether persons will be omitted from the mail survey questionnaire in error. Persons were 1.42 times more likely to be omitted when the response was a late response than if the response was received earlier in the data collection period ( $p < .05$ ). The second finding is that persons living in the Low Coverage Area stratum were more likely to be omitted on completed mail survey questionnaires than persons living in the High Coverage Area stratum ( $p < .0001$ ). As column two shows, there was no interaction between the type of respondent and the strata they lived in. The third finding from this analysis is that persons enumerated on panel 2 questionnaires were less likely to be omitted than persons enumerated on study panel 1 questionnaires ( $p < .05$ ).

Table 2. Logistic Regression Coefficients on Census Day Erroneous Inclusions and Omissions

Predictors	Erroneous Inclusions (1) Coefficient (standard error)	Omissions (2) Coefficient (standard error)
Intercept	-4.2106 ** (0.1999)	-3.9967 ** (0.2401)
Late Respondent	0.3449 (0.2580)	0.7915 * (0.3096)
Low Coverage Area stratum	0.5434 ** (0.1331)	0.8918 ** (0.1483)
Late x Low	-0.0749 (0.3154)	-0.6469 (0.3402)
Panel 2	-0.1717 (0.2499)	-0.7099 * (0.2867)
Panel 3	0.2924 (0.2258)	-0.2399 (0.2560)
Panel 4	0.2507 (0.2523)	-0.2550 (0.3027)
N	21,695	21,695

\*  $p < .05$  \*\*  $p < .0001$ .

**ANALYSIS OF COMPLETENESS**--Linear regression models were used to determine if questionnaire completeness differed by whether the household responded early or late. The analysis attempted to control for panel effects and differences in demographic characteristics found between early and late respondents during preliminary analysis. The variables examined as potential explanatory variables were: panel (1, 2); stratum (High Coverage Area, Low Coverage Area); type of respondent (early, late); gender; marital status (married, other); race (white, nonwhite); Hispanic origin (Hispanic, not Hispanic); householder's age; total persons in the household (0 to 7); type of housing unit (single versus multi unit); ownership of the housing unit (owned versus rented) and the interaction of stratum and type of respondent (late\*low coverage area). As mentioned under Measures, if characteristic data for the householder were missing, the record was not included in the analysis. This resulted in deleting 12 records.

As the results of the regression analysis show (see Table 3), late respondents return less complete questionnaires than do early respondents ( $p < .05$ ), controlling for differences in demographic characteristics. As can be seen

from Table 3, those householders who provide less complete data are older, Hispanic, part of a large household, and live in rented housing. Householders who respond early, are younger, non-Hispanic, part of smaller households and live in owned housing provide more complete data.

Table 3. Linear Regression Coefficients (and standard errors) for Modeling Questionnaire Completeness

Predictors	Coefficient (standard error)
Intercept	1.0136 (0.0030)
Householder's Age	-0.0002 (0.0000) **
Householder's Sex (1 for female)	-0.0011 (0.0010)
Householder's Hispanicity (1 for Hispanic)	-0.0071 (0.0030) *
Type of Respondent (1 for late)	-0.0037 (0.0017) *
Stratum (1 for Low Coverage Area)	-0.0020 (0.0012)
Type of Unit (1 for multi)	-0.0020 (0.0011)
Race of householder (1 for not white)	-0.0013 (0.0016)
Panel (1 for panel 2)	0.0002 (0.0008)
Tenure of Unit (1 for rented)	-0.0030 (0.0013) *
Householder Marital Status (1 for not married)	-0.0023 (0.0013)
Number of persons in household (0 to 7)	-0.0038 (0.0005) **
Late * low Coverage Area Stratum	0.0034 (0.0026)
N	5,710

\* p < .05; \*\* p < .0001.

ANALYSIS OF RELIABILITY--To evaluate the reliability of data collected for early and late respondents,

the rate of gross disagreement between the census and reinterview (g) was estimated for the early and late respondents for four demographic characteristics collected during the mail survey and then again during reinterview across all four study panels. These included age, sex, race, and Hispanic origin. The differences in the overall gross difference rates for the early and late respondents were then compared to see whether they differed. The analysis, conducted at the National and stratum levels, showed that late respondents provide just as reliable information as early respondents; that is, none of the differences were statistically significant.

#### Discussion

This research sought to address a gap in the survey literature about the quality of data provided by survey respondents who respond to mail surveys only after receiving four mailings about the survey, including two questionnaires. At least four mailings are recommended by Dillman's Total Design Method. The Census Bureau began mailing replacement questionnaires to mail survey respondents in the last five years because of the substantial decline in response to the 1990 census and has consistently found that this methodology increases response to mail surveys by about 10 percentage points. Through the 1990 census, a field data collection operation was conducted to collect data from every household that did not respond after receiving the initial questionnaire and the reminder postcard. The purpose of this study was to look at the data provided by those responding late (completing a replacement questionnaire) to determine how they compared to the data of early respondents (those completing the initial questionnaire). Using data from the U.S. 2000 Census Test, three aspects of data quality were evaluated--coverage, completeness, and reliability.

This study showed that of the three aspects examined, late respondents perform significantly different than early respondents on two aspects of data quality--coverage and completeness. Late respondents (i.e., those who return a replacement questionnaire) are more likely to omit persons from the completed mail survey questionnaire and return less complete questionnaires than respondents who return an initial questionnaire. These findings suggest that it may be worthwhile to use different data edit rules for early and late mail survey returns; that is, late mail returns may require a more thorough edit to evaluate completeness and coverage than early mail returns.

Using 1990 census and Post Enumeration Survey data, Griffin and Moriarity (1992) conclude that the quality of mail and interviewer data deteriorates over time. The rate of erroneous inclusions on mail returns rose from 2.4 percent to 5.3 percent over a four week time period beginning in late March, 1990. For returns completed by interviewers, the rate rose from 3.1 percent to 18.4 percent over a four week period. The present study finds no difference in the rate of erroneous inclusions between

early and late mail respondents. The late respondents act no differently than the early respondents in terms of erroneously including persons on the mail survey questionnaire. Childers and Moriarity (1993) found a significant difference in the omission rate when comparing 1990 census mail and enumerator return data. They report that 1.3 percent of the persons in the mail return universe were omitted from questionnaires in error while 5.0 percent of the persons in the interviewer universe were omitted in error. In this study, nationally, about 1.7 percent of the persons in the early mail response universe were omitted in error while 3.3 percent of the persons in the late mail response universe were omitted in error. This study seems to support the Childers and Moriarity finding and may indicate that late mail survey respondents omit fewer persons on average than census interviewers omitted during the 1990 census.

One outstanding issue that needs further research is whether there is a bias/variance tradeoff for doing multiple questionnaire mailings versus subsampling the initial nonrespondents and collecting the outstanding data through a personal visit interview. As Groves (1989) points out, the proportion of eligible sample persons interviewed is a term in the nonresponse bias expression and is therefore often compared across surveys. Higher response rates can lead to higher, not lower nonresponse bias when those initial nonrespondents who are converted to respondents are very atypical of the full set of initial nonrespondents (Groves, 1989). This study, a first step, shows that the quality of data provided by late respondents differs from the quality of data of early respondents. Additional work must be conducted to evaluate whether there is a bias/variance tradeoff of mailing the replacement questionnaire versus subsampling the nonresponse universe and collecting data through personal visit interviews. One recommendation for looking at this would be to use the data on omissions to look at the effect of subsampling the late respondents on the mean square error.

#### References

- Biemer, P. and S. L. Stokes. (1991). "Simple Response variance: Estimation and Determinants." Chapter 27 in Biemer, Groves, Lyberg, Mathiowetz, and Sudman (eds.), *Measurement Errors in Surveys*. New York: Wiley.
- Brick, J.M, P. Broene, P. James, and J. Severynse. (1996). *A User's Guide to WesVar PC, Version 2.0*. Maryland: Westat.
- Childers, D and C. Moriarity (1993). "Analysis of Census Omissions: Preliminary Results", in *Proceedings of the American Statistical Association Meetings*. Virginia: Alexandria.
- Dillman, D.A., C. Jenkins, B. Martin, T. DeMaio (1996). "Cognitive and Motivational Properties of Three Proposed Decennial Census Forms." Internal Census Bureau Memorandum.
- Dillman, D. A., E. Singer, J. R. Clark, J. B. Treat. (in press). "Effects of Benefits Appeals, Mandatory Appeals, and Variations of Statements of Confidentiality on Completion Rates for Census Questionnaires." Paper presented at the 1994 Annual Meetings of the American Association of Public Opinion Research. In press, *Public Opinion Quarterly*.
- Dillman, D. A., J. R. Clark, and M. A. Sinclair (1995). "How Prenotice Letters, Stamped Return Envelopes and Reminder Postcards Affect Mailback Response Rates for Census Questionnaires," *Survey Methodology*.21(2)1-7.
- Dillman, D.A. (1991). "The Design and Administration of Mail Surveys. *Annual Review of Sociology*. pp.225-249.
- Dillman, D.A. (1978). *Mail and Telephone Surveys: The Total Design Method*. New York:Wiley.
- Dillman, D. A., J. J. Dillman, and C. J. Makela. (1984). "The Importance of Adhering to Details of the Total Design Method (TDM) for Mail Surveys." in *New Directions for Program Evaluation*, ed. Daniel C. Lockhart, pp. 49-64. San Francisco,: Jossey-Bass.
- Fay, R. (1989). "An Analysis of Within-household Undercoverage in the Current Population Survey." *Proceedings of the Bureau of the Census Fifth Annual Research Conference*, U.S. Bureau of the Census, pp. 156-175.
- Griffin, D. and C. Moriarity (1992). "Characteristics of Census Error." in *Proceedings of the American Statistical Association Meetings*. Virginia: Alexandria.
- Goyder, J.C. (1982). "Further Evidence on Factors Affecting Response Rates to Mailed Questionnaires." *American Sociological Review*, 47:550-553.
- Heberlein, T. A. And Baumgartner, R. (1978). "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature." *American Sociological Review*, 43:447-62.
- Sweet, E. (1994). "Who Fills out the Census Questionnaire Study." Unpublished memorandum. Bureau of the Census, Washington, DC.
- Thomas, K. and T. Dingbaum. (1992). "How Good are these Data Anyway? The 1990 Content Reinterview Survey." IN *Proceedings of the American Statistical Association Meetings*. Virginia: Alexandria, pp. 369-374.
- Treat, J.B. (1992). "Final Results of the Mail Response Evaluation for the Simplified Questionnaire Test

(SQT).” DSSD 2000 Census Memorandum Series #E-18. U.S. Bureau of the Census, August 5.

Treat, J.B. (1993a). “Final Results of the Mail and Telephone Mode Test.” DSSD 2000 Census Memorandum Series #E-47. U.S. Bureau of the Census, July 21.

Treat, J.B. (1993b). “1993 National Census Test Appeal and Long Form Experiment (ALFE). Appeals or Short Form Component: Final Report.” DSSD 2000 Census Memorandum Series #E-62. U.S. Bureau of the Census, October 29.

Treat, J.B. (1993c). “1993 National Census Test Appeals and Long Form Experiment (ALFE). Long Form Component: Final Report.” DSSD 2000 Census Memorandum Series #E-63. U.S. Bureau of the Census, November 17.

Yammarino, F., S. Skinner, and T. Childers (1991). “Understanding Mail Survey Response Behavior.” *Public Opinion Quarterly*, 55:613-639.

#### End Notes

1. Late respondents were identified using the questionnaire identification number (ID). The ID included one character to identify whether the questionnaire checked in was mailed during the initial or replacement mailing.
2. The response rate was defined as the number of mail survey questionnaires returned divided by the number of questionnaires mailed out less those questionnaires returned by the U.S. Postal Service after both mailings because they were undeliverable as addressed.