

## Does a Final Coverage Check Identify and Reduce Census Coverage Errors?

*Elizabeth Martin<sup>1</sup> and Don A. Dillman<sup>2</sup>*

A national mail out experiment evaluated a final series of questions intended to reduce and identify coverage errors on the census mail form. A random portion of the sample received the experimental series, which included reminders of the reference date and people who might be missed and two coverage questions. Another random portion received a version without reminders or coverage questions. Results from a follow-up interview show that responses to the questions help discriminate between households in which someone was missed or counted in error and those without coverage errors. There is no evidence that the series led to coverage improvements. Item nonresponse for Question 1 (household count) is lower in the experimental panels, suggesting the series stimulated respondents to check their forms and fill in the question when it had been left blank.

*Key words:* Mail questionnaire; decennial census; coverage questions.

### 1. Introduction

Census questionnaires used to count populations in many countries of the world ask householders to report the number and characteristics of people who reside in their households. The short form used in the U.S. Census contains only two general questions (number of residents and whether the residence is rented or owned) plus six questions about each resident (name, gender, age, ethnicity, race, and relationship to the householder). What seems on the surface a simple task is in many ways quite complex. Residency in a particular household is not always clear to respondents (e.g., whether people away at college or in a nursing home should be reported), and they sometimes forget to report residents or leave off people they do not consider “family.” In addition critical instructions are sometimes not read.

The approach traditionally used in the U.S. Decennial Census to obtain an accurate count of household members is to provide rules for inclusion and exclusion and ask householders to follow those rules. Despite the testing of many alternative formats and instructions to obtain more accurate coverage within households, there is strong evidence that errors continue to be made (Linse et al. 2006).

Our purpose in this article is to evaluate whether a different approach, i.e., asking at the end of this short census questionnaire questions that encourage people to review their

<sup>1</sup> 8707 Stockton Parkway, Alexandria, VA 22308, U.S.A. Email: [betsy@folhc.org](mailto:betsy@folhc.org)

<sup>2</sup> Washington State University, Social and Economic Sciences Research Center, Pullman, WA 99164-4014, U.S.A. Email: [dillman@wsu.edu](mailto:dillman@wsu.edu)

**Acknowledgments:** Thanks to Bob Fay for advice on calculation of standard errors and significance tests, to Tammy Adams, Danny Childers, Eleanor Gerber, and Jim Treat for helpful comments on an earlier draft, and to Cynthia Rothhaas and Eli Krejsa for checking and correcting the tabulations and providing useful comments.

answers and provide additional information about situations which gave them pause, will improve the accuracy of reporting. This approach is based upon a general usability principle proposed by Donald Norman (1988) of providing an opportunity for people to correct any errors they might have made.

We report results from an experiment conducted by the U.S. Census Bureau in a 2006 national mail out test in which four brief questions were added to the end of the questionnaire. The questions serve the dual purpose of encouraging respondents to review their answers as well as explain any uncertainty they felt about who should or should not be included as living in their household. The questions were evaluated using follow-up telephone interviews to better understand their ability to reveal coverage errors.

## 2. Background

The U.S. Census Bureau attempts to enumerate people at their “usual residence,” or the place where they live and sleep most of the time. Additional rules determine where people in special circumstances (e.g., in the military, away at college) should be counted in the census. Errors made in applying the rules (along with other errors not explored here, such as errors in address lists) contribute to differential census undercounts of males, young adults, babies, renters, blacks and Hispanics, poor people, and other segments of the population (Robinson et al. 1993; Robinson 2001). Some groups, such as college students and elderly women, are over-counted. The differential effect of coverage errors distorts counts and characteristics of demographic subgroups and geographic places. Coverage errors negatively affect the constitutionally-mandated use of census data for congressional reapportionment, as well as their use for congressional redistricting, distribution of government funds, and demographic analyses of U.S. population counts and characteristics. Errors made in completing rosters of household residents have in past censuses accounted for about a third of all decennial census coverage errors (Hogan 1993).

When people’s lives are complicated, their residence may be ambiguous and respondents may erroneously omit or include them from census rosters. A 1993 survey found that about 9% of persons had “complex living situations” that were vulnerable to misreporting by household respondents (Sweet and Alberti 1994). The types of situations that create enumeration difficulties include the following:

*Multiple residences* may occur for reasons related to family (e.g., children in joint custody), work (e.g., jobs that require temporary residence away from home), or school (e.g., children living away at college or boarding school who return home for holidays and the summer). Determining where a person who moves among different residences should be counted is difficult, and the census rules that apply to such situations often do not accord with respondents’ notions of where the person belongs. Coverage errors are more frequent in *households composed of unrelated individuals*, people with ambiguous household membership, two or more nuclear families, and those formed for the sole purpose of sharing rent or living expenses (de la Puente 1993; Fay 1989; Ellis 1994, 1995). Unrelated individuals may be omitted because they are not regarded as “family” or part of the core household (Rodriguez and Hagan 1991). People who have *tenuous ties to any household* and move from place to place may not have a usual residence at which they can be enumerated (Martin 2007b). People who are frequently absent may be assumed falsely to

have another residence (Gerber 1990) and omitted from rosters by household respondents (Martin 1999). *Residential mobility and life changes* cause errors. People who move from one residence to another around the time of the census are at risk of being included at both locations, or omitted from both, depending on the timing of the move and nonresponse follow-up attempts. It may be difficult to recall accurately when a move occurred, and respondents sometimes ignore the April 1st census reference date (Wellens and Gerber 1996). About 2.1 million in-movers were enumerated in the 1990 Census at the address to which they moved after April 1st, accounting for 20% of estimated erroneous enumerations (Moriarity 1993). Births and deaths that occur after forms are mailed back but before Census Day may lead to omissions or erroneous enumerations. *Inconsistent and counterintuitive rules and confusing terminology* cause enumeration difficulties. The numerous rules that determine where people should be enumerated in the census involve complex and unfamiliar concepts (e.g., “usual residence”) that must be simplified for presentation on the census form. Some respondents do not understand the instructions, while others understand but ignore them because they do not agree with respondents’ own notions of who lives in their households (Gerber 1994, 2004; Gerber, Wellens, and Keeley 1996; Parsons, Mahon-Haft, and Dillman 2005). Respondents often disregard counterintuitive instructions to count college students at their dorms and not at “home,” for example (Gerber, Wellens, and Keeley 1996). Confusing terminology also causes errors. For example, in pretests, a question about “another residence” sometimes elicited “official” or permanent residences where a person spent almost no time, or failed to elicit legitimate second residences if respondents interpreted “residence” too strictly (Wellens and Gerber 1996). Finally, *measurement errors* arise from many sources, including deliberate concealment (Valentine and Valentine 1971; Tourangeau et al. 1997), respondents’ lack of knowledge or failure to recall relevant facts, confusion, and interviewer errors.

### 3. Development of Coverage Questions

In past U. S. censuses, special “coverage questions” have been included on self-administered census forms to identify coverage errors. Such questions usually provide cues to remind respondents of the types of people who might be inadvertently left off the form. In 1970, four coverage questions appeared on the mail form following the 8 lines for listing household members. (The questions were “*If you used all 8 lines – Are there any other persons in this household?*” “*Did you leave anyone out of Question 1 because you were not sure if he should be listed – for example, a new baby still in the hospital, or a lodger who also has another home?*” “*Did you list anyone in Question 1 who is away from home now – for example, on a vacation or in a hospital?*” and “*Did anyone stay here on Tuesday, March 31, who is not already listed?*” Respondents who gave positive answers to the latter three were asked to record the names and reasons on the back of the form). Although their placement at the bottom of the form increased item nonresponse (one third of respondents failed to answer them; Rothwell 1972), the first one nonetheless added about 400,000 people to the census (Davie 1973). (Documentation of the performance of the other coverage probes is not available.) Similar probes were also included in the 1980 and 1990 mail questionnaires and in the 2000 enumerator questionnaire (see National Research Council 2006).

In the 2001 Canadian Census, an undercount question (“Step C” – “Did you leave anyone out of Step B because you were not sure the person should be listed? *For example: a person living at this address who has another home; a person temporarily away.*”) immediately following the household roster in the mail questionnaire successfully identified some census omissions. About 1% of respondents gave positive responses, and about 20% of these mentioned people who were added to household rosters (Roy 2003), for an improvement of about 0.2%. The National Research Council (2006, Appendix B) documents the use of supplementary coverage questions in censuses conducted in New Zealand, Switzerland, the United Kingdom, and other countries. The U.K. Office for National Statistics conducted a test in 2007 to evaluate its supplementary census questions on address and residence, but results are not available at this writing.

Coverage questions are being evaluated for possible inclusion in the mail census questionnaire for the U.S. Census in 2010. An “overcount question” is intended to identify other places each person also might have been enumerated. An “undercount question” is asked immediately after Question 1 to identify possible omissions. Possible errors would be followed up by a Coverage Follow-up interview to correct the error, if any. After the undercount question performed poorly (that is, it did not identify many omissions) in a 2004 census test (Krejca et al. 2005), a revised version (“Were there any *additional* people staying here September 15, 2005 that you *did not include* in Question 1?”) was fielded and performed somewhat better in a 2005 national mail out experiment (Linse et al. 2006). However, it still fails to identify most of the omissions identified in a coverage follow-up interview.

Our approach to the design of coverage questions is inspired by one of seven principles that Norman (1988) proposes for improving the usability of everyday objects: *design for error*. He urges designers to “Assume that any error that can be made will be made. Allow the user to recover from errors, to know what was done and what happened, and to reverse any unwanted outcome” (1988: 200).

Dillman, Gertseva, and Mahon-Haft (2005) apply this principle by building in mechanisms throughout a questionnaire to allow people to correct their errors (pp. 202–203). Similarly, Redline et al. (2003) provide opportunities for respondents to detect branching errors at the point they are made, thereby decreasing the number of such errors by 20–35%.

Here, we give respondents a chance to correct coverage errors they may have inadvertently made, by adding a series of questions with the dual purpose of reducing and identifying coverage errors. Our approach differs from previous approaches in several ways.

*Orientation.* An introduction orients respondents to the task of reviewing the form to be sure it includes everyone in the household by reminding respondents of the reference date and of types of people who might be omitted (see Figure 1, experimental version).

*Structuring the cognitive task.* Final Question 1 (FQ1) presents the task of reviewing the form to identify discrepant counts of the number of people in the household by asking, “Is the number of people for whom you have provided information the same as the number you counted in question 1 on page 1?” If the respondent marks “no”, s/he is asked to “Please briefly explain the reason.” The question is intended to facilitate a more active



Final Question 2 (FQ2) is similar to the second coverage question in the 1970 census and to Step C in the Canadian census, but the “not sure” wording used in those questions is modified to ask, “Did you leave anyone off the form that you thought about including? *For example: a person living at this address who has another home, a person living temporarily away, etc.*” If the respondent marks “yes,” s/he is asked to “Please briefly explain the situation.”

Qualitative tests show that respondents are usually quite confident about who should be reported (or not reported) as members of their households, even when their reports are erroneous according to census rules. Human judgments are biased by overconfidence in many situations, especially those involving difficult judgments (Griffin and Tversky 2002), and survey respondents typically express high levels of confidence in their answers. As Cannell et al. (1989) note, “respondents [do] not appear to doubt their own, often mistaken, interpretations” (p.47). This consideration suggests that omissions may be better identified by a question that does not require respondents to express uncertainty in their answers. A similarly-worded debriefing question successfully identified unreported incidents in the National Crime Survey (Martin et al. 1986).

In pretests, the “thought of” wording was more inclusive than the “not sure” wording (Kerwin and Moses 2006), so it may capture more omissions by avoiding bias due to overconfidence. A probable disadvantage is more mentions of people whom the respondent thought of but decided, confidently and correctly, did not belong on the form.

*Placement.* We believe it makes more sense to place coverage questions at the very end of the questionnaire rather than immediately following Question 1. The latter placement implies to some respondents they have made a mistake or are being asked to second-guess an answer they just provided, and caused confusion or agitation in cognitive tests. It led some respondents to go back and change their answers to Question 1, thereby introducing errors (Gerber 2004; Cantor, Heller, and Kerwin 2003). Following Norman (1988), placement at the end may facilitate correcting errors associated with the action of filling out the questionnaire. Cognitive tests indicate that respondents find this a logical place to review and check their answers, and actually want such items on the census form (Kerwin and Moses 2006).

Placement at the end also provides a clear stopping point in the questionnaire and gives respondents a sense of completion, which is lacking in current designs of the questionnaire (Dillman, Parsons, and Mahon-Haft 2004). A possible disadvantage is that respondents may never find the questions. In previous censuses, coverage questions placed at the bottom of the mail form or on the back of the enumerator form had high rates of missing data (Rothwell 1972; Nguyen and Zelenak 2003). In this study, navigational instructions were revised to route respondents to the back page, and the final series was headed conspicuously by a title indicating the questions should be answered by everyone.

#### 4. Method

To evaluate the Final Questions, Westat conducted a national mail out/mail back test for the U.S. Census Bureau. The test evaluated three experimental factors in four panels, with the sample of 28,380 households equally allocated among each panel.

All panels include a new section on the back page of the questionnaire entitled “Final Questions for Everyone” that requests the respondent’s name and phone number and asks whether the respondent lives in the household or is filling the questionnaire out for the people who do. (Adding the new section required dropping Persons 11 and 12 from the continuation roster on the back page for all panels.)

In Panels 3 and 4, “Final Questions for Everyone” also includes an instruction to respondents to check over their answers, reminders of types of people who might be missed, and the two final questions described above. Panels 1 and 2 do not include the instructions or questions to check on coverage. (See Figure 1 for facsimiles of control and experimental versions of the questions.)

The other panels in the test evaluate the effects of a deadline and compressed mailing schedule on response rates, and the effects of a revised instruction on coverage errors. Those results are not reported in this article (see Martin 2007a; Martin forthcoming for those results). Thus, the four panels are:

- Panel 1: Control panel
- Panel 2: Includes a revised instruction about whom to list as Person 1, otherwise identical to Panel 1. Both Panels 1 and 2 include the control version of the Final Question series (shown in Figure 1).
- Panel 3: Includes the revised instruction about whom to list as Person 1 and the experimental version of the Final Question series (Figure 1).
- Panel 4: Like Panel 3, includes the revised instruction about whom to list as Person 1 and the experimental version of the Final Question series. In addition, all mailings were delayed by a week and respondents were given a deadline for return of the form.

The mail out test was conducted in March and April 2006 in households with city-type addresses that receive mail from the U.S. Postal Service (USPS) that would be eligible for a mail out-mail back short form in the census. Households in Austin, TX, were excluded to avoid interfering with a census test conducted in Austin at the same time. Census Day was April 13, 2006, for all panels.

A sample of 28,380 households was drawn from the USPS Delivery Sequence File (DSF) that contains all delivery point addresses serviced by the USPS. Entries listed with a P.O. Box rather than street address were excluded from the sample, because a P.O. Box is not clearly tied to a single residential housing unit. Although the DSF undercovers new housing and misses units due to resident requests for removal from the list, the imperfections of the list should not affect the results of the experiment. However, they should be kept in mind when comparing results from this test to those from other census tests.

The sample was allocated proportionately across the 50 states and the District of Columbia (except Austin). The frame was implicitly stratified (using as sort variables State, Household size, % Black or Hispanic in the zip code, % High School or less, % earning less than \$20,000 income, and zip code) and a systematic sample selected. After the sample was selected, assignment was made to the four experimental panels, sorting the file using the same sort variables first to ensure that each panel was representative.

Sampled households received an advance letter, an initial questionnaire package, and a thank you/reminder postcard, all delivered by first class mail. A postage-paid return



envelope was enclosed for respondents to mail back their completed questionnaires to Census Bureau headquarters, where they were checked in and keyed. Mailing pieces, including letters, included the U.S. Census Bureau logo in the masthead and were signed by the Director. There was no replacement questionnaire and no follow-up in nonresponding households.

A sample of approximately 600 cases was sent for a Coverage Follow-up interview in order to assess coverage gains in households where responses to Final Questions 1 and 2 indicated a possible error. All cases with a negative response to Final Question 1 or positive response to Final Question 2 or a write-in response to either question were sent for follow-up. In addition, a random subsample of about 300 cases that provided no indication of a coverage issue in their answers to the final questions was sent to Coverage Follow-up. Interviews were conducted by telephone between June 30 and July 21, 2006.

## 5. Results

13,703 completed questionnaires were returned by the cutoff date of May 19, and an additional 436 were returned after that date. Excluding the 1,804 mailing packages (6.4% of the sample) that were returned marked “vacant” or Undeliverable as Addressed, response rates were between 50.3% and 53.1% for the four panels. Response rates do not vary significantly among Panels 1–3, while Panel 4 (which included a deadline for returning the form combined with a later mailing) had a significantly higher response rate than the other panels combined. Analyses below are based on the 13,703 completed questionnaires received by May 19.

Four analyses are conducted to evaluate the performance of the Final Questions:

1. Do respondents find and answer the Final Questions about coverage?
2. Do the write-in responses describe possible coverage errors, as intended?
3. Based on the results of a coverage follow-up (CFU) interview, do the Final Questions identify omissions or other coverage errors?
4. Do the questions and reminders reduce coverage errors?

Standard errors and *t*-statistics are computed using a stratified jackknife replication procedure with random groups using VPLX (Fay 1998). Cases are sorted by the implicit variables used to sort the frame for sample selection, and strata are composed of pairs of adjoining cases on the sorted list. Standard errors are shown in parentheses in tables.

### 5.1. Do Respondents Answer the Final Questions?

Table 1 shows that about 0.5% of respondents answered “no” to FQ1 and 2.4% responded “yes” to FQ2, indicating a potential coverage problem. Item nonresponse rates were 5% to 6% for each question. (Panels 3 and 4 are combined to increase the number of cases; there is no panel difference in response distributions).

Thus, most respondents found and answered the Final Questions, despite their placement at the end of the questionnaire. These rates compare reasonably well with household items that appear on the front page of the questionnaire, including Question 1 and tenure, with item nonresponse rates of 3.7% and 2.4% respectively across all panels. By comparison, item nonresponse was 7.7% for the undercount question placed on the



Table 1. Response Distributions to Final Questions 1 and 2 (Panels 3 and 4)

	1. Is the number of people for whom you have provided information the same as the number you counted in Question 1 . . . ?	2. Did you leave anyone off the form that you thought about including? . . .
Yes	94.5% (0.3)	2.4% (0.2)
No	0.5 (0.1)	91.7 (0.3)
No answer	5.0 (0.3)	5.9 (0.3)
Total	100.0%	100.0%
N	6,974	6,974

front of the form in the 2005 National Content Test. Moving the telephone number from the front to the back of the form also did not harm response to that item. Between 91% and 92% of households in this test provided a phone number, compared to 89% in the 2005 test when the item was on the front of the form. Thus, the placement of the items on the back page apparently did not result in more missing data, and they performed fairly well in this first field test.

To check whether respondents understood that FQ1 was asking about count discrepancies, the rate of actual count discrepancies was compared with respondents' reports. Of those who answered "no" to FQ1, 51% actually had a count discrepancy, compared to fewer than 1% of those who answered "yes." Most respondents who answered "no" apparently understood the question, although some who marked "no" even though the counts were consistent may not have, or may have interpreted the question a different way, or may have marked "no" but then corrected the discrepancy. Additional qualitative testing might suggest wording refinements to clarify the intent of the item.

### 5.2. What Types of Living Situations Are Problematic?

54% of those who checked "no" to FQ1 wrote an explanation in the space provided, as did 91% of those who checked "yes" to FQ2. Some respondents provided a write-in response when none was necessary. Over half of the write-in responses to FQ1 were provided by people who marked "yes" and did not need to write anything. Many respondents explained unnecessarily that they lived alone.

FQ1 and FQ2 target different coverage errors. Since count discrepancies may result either from erroneous inclusions or from omissions, the FQ1 write-in responses may describe either type, or may describe cases in which an incorrect number was given in Question 1 but the number of people reported on was correct. FQ2 is intended to identify omissions, although its inclusive wording also invites reports of nonresidents whom respondents thought of including but correctly left off the form. Some responses clearly describe errors on original census rosters. For instance, one respondent wrote in FQ1, "NO MORE SPACE FOR NAME." She counted 11 people in Question 1, but the form contained space for only 10, so one person was omitted. In a second example, FQ1 identifies an erroneous enumeration: the respondent wrote, "IT SAID TO EXCLUDE

COLLEGE STUDENTS SO I DIDN'T COUNT HIM BUT HE LIVES [HERE].” She did not count her college student son in Question 1, but erroneously included him on the form, creating a count discrepancy. In a third example, FQ2 identifies two omissions: the respondent marked “yes” and wrote, “TWIN BOYS IN THE NICU (BORN APRIL 8, 2006).” Born 5 days before Census Day, these babies still in the Neonatal Intensive Care Unit should have been included on the form but were not.

Most responses are less clear-cut, and information beyond the brief write-in entry would be required to determine whether or not the write-in response described a household resident who should have been included. Although Census Day residence status usually cannot be determined from the write-in responses, the latter can be categorized according to the types of living situation they describe. Results for each question are shown in Table 2.

As discussed above, most of the FQ1 write-in responses were provided by respondents who marked “yes” and did not need to write anything. When these are excluded from the calculation, only 25% of the FQ1 write-ins are unresponsive, rather than 71%, as shown in Table 2. 9% of the FQ1 write-ins explain that the counts are discrepant because there was not enough space on the form, or information was lacking for someone. 9% each describe a mobile or part-time resident, or a person in college, the military, or other group quarters, and 2% describe someone in an adjoining apartment.

About a third (34%) of FQ2 write-ins describe part-time residents (Type 1), 6% describe new or unborn babies (Type 2), 1% describe people left off due to lack of space on the form (Type 3), and 2% describe caregivers or live-in employees (Type 4). These types of situations account for 43% of FQ2 write-ins and should be productive for follow-up to identify missed residents. Responses coded as Type 1 include many complex and ambiguous situations known from past research to contribute to omissions and other coverage errors, such as children in custody arrangements, people in the process of moving

Table 2. Types of Living Situations Described by Write-In Responses to Final Questions

Type	% of all FQ1 write-in responses	% of all FQ2 write-in responses
1. Mobile or part-time resident	9%	34%
2. Unborn or newborn babies	—	6
3. No space on the form; lacked information about person or didn't want to provide it	9	1
4. Caregiver or nanny	—	2
5. Person in college, military, jail, prison, nursing home, or other group quarters	9	40
6. Pets	—	2
7. Missionary abroad	—	1
8. Someone in nearby apartment	2	1
9. Person who died	—	1
10. Name only	—	2
11. Unresponsive (e.g., “I live alone”) or uncodable write-in	71	10
Total	100%	100%
Unweighted <i>N</i>	56	174

(“SON WAITING TO MOVE INTO NEW HOME”), part-time residents (“MY SON LIVES WITH ME ABOUT 50% OF TIME”), frequent or regular visitors (“THE BABIES DAD STAYS BUT LIVES ELSEWHERE”), people with transient lives or lifestyles (“MY SON IS 34 YEAR OLD. HE STAY HERE, THERE AND EVERYWHERE”), people with jobs involving frequent travel (“TRAVELS ALL OVER THE COUNTRY – PLAY IN A BAND”), extended stays or absences (“MY DAUGHTER, US CITIZEN NOW IN INDIA”) and so on.

Just over half of FQ2 write-ins are unlikely to yield coverage improvements because they are unresponsive or uncodable (10%), describe people in group quarters (40%), missionaries abroad (1%), people who died (1%), or pets (2%), none of which are considered Census Day residents. Apparently, the instructions on the form to exclude college students who live away, people in jail or prison, etc., are read and followed by some respondents, even though they think about doing otherwise.

Thus, the Final Questions elicited reports of the types of situations that give rise to coverage errors and should be productive for follow-up; FQ2 also elicited many reports of nonresidents, few of which would be productive for follow-up.

### 5.3. *Do the Final Questions Identify Omissions or Other Coverage Errors?*

In order to determine the productivity of the Final Questions, Coverage Follow-up interviews were attempted by telephone in all households responding “no” to FQ1 or “yes” to FQ2 or providing a write-in response to either question. (The households so selected are labeled as “flagged by FQ.”) Using this criterion, 3.71% of Panels 3 and 4 households were flagged for follow-up in CFU. A random sample of households that were not flagged by their responses to the final questions also was followed up. Interviews were completed in 81.8% of the households sent for follow-up. We compare 201 completed CFU cases in households flagged by FQ (of 242 households sent for follow-up), with 145 completed CFU cases in households not flagged by FQ (of 176 households in the random nonflagged sample of Panels 3 and 4 cases).

In the CFU, interviewers requested an interview with the respondent whose name was provided on the back of the mail questionnaire. Interviewers did not have responses to the Final Questions available to them when they conducted the interviews. The CFU was designed to follow up households in the 2006 Census Test in Austin TX, which did not include the Final Questions. (In that test, follow-up interviews were attempted in households that responded positively to the undercount or overcount questions, in large households, and in households with a count discrepancy.)

Thus, interviewers were “blind” to the experimental treatment; they had no way of knowing whether they were interviewing flagged or not-flagged cases, or what situation led the household to be followed up.

CFU procedures called for the interviewer to review with the respondent the list of persons who had been recorded on the census form for that household. Probes were administered to identify potential adds: people not listed on the original census roster who should be added if further questioning determines they were Census Day residents. (The probes were, “Any newborns or babies; any foster children; any nonrelated children; any other nonrelatives who lived or stayed here; any nonrelatives, roommates, or boarders;

anyone else who stayed here often; anyone else who had no other place to live.”) CFU also included extensive questions to identify other residences and group quarters stays, in order to identify people who had been enumerated in error and delete them from the household roster.

In order to determine whether the Final Questions identify households with missed or erroneously enumerated people, we compared the fractions of households in which people were added or deleted as a result of CFU, in flagged and nonflagged households. We adopted a slightly more restrictive rule for flagging households by dropping households ( $N = 40$ ) that unnecessarily provided a write-in response to FQ1 without checking “no” to FQ1 or “yes” to FQ2. As noted above, most of these write-in responses were extraneous and did not indicate a coverage problem.

A narrower rule that targets households with a “no” to FQ1 or a “yes” to FQ2 selects 2.9% of Panels 3 and 4 households for follow-up. Results in Table 3 show that these households are much more likely to have a coverage error identified in the CFU than the random sample of nonflagged households. Standard errors are shown in parentheses.

The odds of CFU adding someone are 8.4 times greater, and of deleting someone 3.1 times greater, in flagged households ( $t = 2.789$ ,  $p < .01$  and  $t = 2.088$ ,  $p < .05$ , respectively).

As expected, FQ1 more effectively targets erroneous enumerations than FQ2: CFU deleted someone in 21.4% of households responding “no” to FQ1 compared to 5.2% of households responding “yes” to FQ2 ( $t = 2.018$ ,  $p < .05$ ). The latter rate does not differ statistically from the 2.8% rate of deletes in the nonflagged random sample. Rates of adds do not differ for households responding “no” to FQ1 or “yes” to FQ2.

Table 3 implies that, of the 2.9% of households flagged for follow-up using the narrow rule, only 5.6% added someone in CFU – in other words, omissions were corrected in only 0.2% of households. In order to better understand the reasons for this disappointing result, write-in responses to the final questions, and the CFU outcomes, were examined in detail (reported in Martin 2007a).

Most write-in responses ( $N = 15$ ) to FQ1 appear to reflect a good understanding of the question intent and to describe situations that are potential coverage errors and should be productive for follow-up. The CFU appears to have identified and resolved most erroneous enumerations appropriately, but failed to identify and correct omissions in three households due to lack of space on the form or an uncooperative person. These omissions are troubling, since such count discrepancies should be easy to identify and resolve. Two additional ambiguous Type 1 cases should have been identified as possible adds but

Table 3. Percentage of flagged and nonflagged households with persons added or deleted in CFU (Panels 3 and 4)

Household	Percent with one or more people added	Percent with one or more people deleted	$N$
FQ1 = “no” or FQ2 = “yes”	5.6% (1.61)	8.1% (2.20)	161
Random sample of nonflagged households	0.7% (0.7)	2.8% (1.4)	145

were not. CFU appears to have done a better job of correcting erroneous enumerations than omissions in households flagged by FQ1.

Table 4 summarizes the fraction of households in which CFU added someone in households flagged by FQ2, within two groupings of write-in responses. Write-ins coded as Types 1–4 in Table 2 are grouped together as potential “residence situations,” and write-ins coded as Types 5–9 are grouped as “nonresidence situations.”

The CFU outcomes help us evaluate alternative explanations for the low rate of adds in CFU. If CFU is found to identify many potential adds but add few, that suggests that most of the people described turned out not to be residents. If CFU identifies few possible adds, that suggests that CFU failed to identify possibly missed residents.

The second row of Table 4 shows that FQ2 write-ins coded as “residence situations” have a significantly elevated rate (22.4%) of possible adds in CFU. Most of these were weeded out by the CFU residence questions, however, so the final rate of adds was about 9%. This is significantly higher than the add rates for the other rows in Table 4.

Write-ins coded as “nonresidence situations” also are associated with a significantly elevated rate of possible adds in CFU. A larger fraction (about 90%) were weeded out as nonresidents, so the final rate of adds was less than 2%. This rate does not differ significantly from that from nonflagged households.

These results make sense. Write-ins describing someone the respondent thought of including led to an elevated rate of possible CFU adds, whether the person was in a “residence situation” or a “nonresidence situation.” But fewer of the people in “nonresidence situations” were confirmed by CFU questioning to be residents who should be added to census rosters.

The categorization by type of living situation predicts the CFU outcome, with write-ins coded as “residence situations” producing significantly more missed residents, as one would expect. Even so, many possible adds were ultimately determined to be nonresidents, reflecting the effects of the inclusive wording of FQ2.

In addition, CFU identified only 22% of the people in “residence situations” as possible adds. This result suggests that CFU failed to identify a substantial number of potential

Table 4. CFU Possible and Actual Adds in Households Flagged by FQ2 (Panels 3 and 4)

Response to FQ	% with possible adds	% with adds	N
FQ2 = “yes” or write-in provided, and	16.7%	4.7%	150
–write-in describes “residence situation” (Types 1–4)	(3.1) 22.4%	(1.7) 8.6%	58
–write-in describes “nonresidence situation” (Types 5–9)	(5.5) 15.6%	(3.7) 1.6%	64
–write-in is missing or uncodable, or gives name only	(4.7) 7.1%	(1.6) 3.6%	28
Not flagged by FQ	(5.0) 2.8%	(3.5) 0.7	145
	(1.4)	(0.7)	

missed residents. (The 90% confidence interval includes 68.6% to 86.6% not identified in CFU.) An effective coverage follow-up should identify as possible adds many, if not most, of the people described in these write-in responses, even if subsequent CFU questioning determines they were not Census Day residents. A rate of 22% seems too low for a coverage follow-up interview that is intended to correct omissions.

Possibly, some respondents added the person to the form after answering FQ2. If the person was already on the form, he or she would not be identified as a possible add in CFU. Household rosters and responses to FQ2 were reviewed to assess this possibility. In three instances and possibly a fourth, the person described in the FQ2 write-in was included on the form. Two describe people in college and a nursing home, so this explanation accounts for very few of the cases in which CFU did not identify people in “residence situations.”

A sizable fraction (41% according to one expert coder) of the FQ2 write-in entries are ambiguous, and may or may not describe a person who was left off the original census roster in error. To determine if the people described should be added to household rosters, they first must be identified as possible adds in CFU so that questions can be asked to determine their status as Census Day residents or nonresidents. It is problematic that so many were not.

#### 5.4. Does the Final Coverage Check Reduce Omissions and Other Coverage Errors?

The Final Question series asked respondents to review their answers to be sure they provided information about everyone living in the household on April 13, reminding them to include “yourself . . . , new babies, temporary guests with no other place to live.”

*Omitted respondents.* Respondents left themselves off the form in error at the same rate across panels (on average, 0.27% of respondents did so) (see Martin 2007a for complete results). Panels 3 and 4 (with reminders) do not differ significantly from Panels 1 and 2 (without reminders).

*New babies.* Panels 3 and 4 both remind respondents to include new babies, but only Panel 4 elicits more new babies than the control ( $t = 1.67, p < .10$ ), as shown in Table 5. (“New” babies are arbitrarily defined as those born in 2006.)

The difference for Panel 4 may be due to the later questionnaire mailing. Babies born just before Census Day, or just home from the hospital, are more likely to be included when the questionnaire is filled out closer to Census Day.

*Count discrepancies.* Final Question 1 was intended to stimulate respondents to review the form and check whether they had answered questions for each person they counted in Question 1. Reviewing their answers might lead respondents to make corrections to eliminate the discrepancies, although they were not instructed to do so. If so, the rate of count discrepancy should be lower for Panels 3 and 4 than for Panels 1 and 2.

Table 6 shows no significant differences among panels in the fraction of forms with a count discrepancy. However, the item nonresponse rate for Question 1 is significantly lower in Panels 3 and 4. (Question 1 was left blank in 3.1% of forms in Panels 3 and 4, compared to 4.3% in Panels 1 and 2;  $t = 3.8, p < .001$ ). This suggests that the Final Questions did stimulate respondents to look back at Question 1 and in some cases to fill it in when it had been left blank.

Table 5. Percentage of Data-Defined People who are New Babies, by Panel

	Panel 1 (no FQ)	Panel 2 (no FQ)	Panel 3 (FQ)	Panel 4 (FQ)
New baby born in 2006	0.23% (0.053)	0.26% (0.056)	0.31% (0.063)	0.37% (0.066)
All other people	99.77 (0.053)	99.74 (0.056)	99.69 (0.063)	99.63 (0.066)
Total	100.00	100.00	100.00	100.00
Unweighted <i>N</i>	8,125	8,085	8,350	8,543

Note: In 14 households, the flag that identified data-defined people was not applied. 17 people are excluded from this table.

Table 6. Discrepancies between the Count provided in Question 1 and the Number of Data-defined People, by Panel

	Panel 1 (no FQ)	Panel 2 (no FQ)	Panel 3 (FQ)	Panel 4 (FQ)
No count discrepancy	94.6% (0.4)	94.2% (0.4)	95.4% (0.4)	96.0% (0.3)
Q1 not equal to number of people on the form	1.4 (0.2)	1.1 (0.2)	1.3 (0.2)	1.1 (0.2)
Q1 left blank	4.0 (0.3)	4.7 (0.4)	3.3 (0.3)	2.9 (0.3)
Total	100.0%	100.0%	100.0%	100.0%
<i>N</i>	3,391	3,330	3,427	3,538

Note: 14 households in which the flag that identified data-defined people was not applied are not included.

On the other hand, most respondents (62%) whose forms actually contained a count discrepancy erroneously answered “yes” to FQ1, suggesting they did not understand the question, or did not bother to check the consistency of the information they had provided.

## 6. Conclusions

Most respondents (94%–95%) found and answered two Final Questions pertaining to coverage. About 0.5% marked “no” to FQ1 (“Is the number of people for whom you have provided information the same as the number you counted in Question 1. . .?”) and 2.4% marked “yes” to FQ2 (“Did you leave anyone off the form that you thought about including?”), indicating a possible coverage error.

In the 3.7% of households in which respondents marked “no” to FQ1 or “yes” to FQ2 or provided a write-in response to either question, a coverage follow-up interview was attempted to identify census omissions or erroneous enumerations. Interviews were also attempted in a random sample of nonflagged households. Follow-up interviews identify significantly more errors in original census rosters in flagged than in nonflagged households. Thus, the questions help discriminate between households in which a follow-up interview is productive and those in which it is much less so. CFU added people in 5.6% of the households flagged by this rule, compared to 0.7% of the random sample, and deleted people in 6.1% of the flagged households, compared to 2.8% of the random



sample. Overall, CFU added people in 0.2% of households. This rate, while disappointingly low, is comparable to results for other census tests testing similar questions and for Statistics Canada's "Step C."

This disappointing result occurred partly because CFU failed to identify as potential adds most of the people described in the write-in responses. Even among people whose living situations were coded as a "residence situations," only 22 percent were identified as potential adds and 9 percent were added. This suggests that the follow-up interview may have failed to identify a substantial fraction of true census omissions. The magnitude of the problem is unknown, because without asking the detailed CFU residence questions, we do not know how many would turn out to be Census Day residents. This result may help explain why the performance of undercount questions in some census tests has been disappointing, since coverage follow-up interviews have provided the standard for evaluating their performance.

There is no evidence that the coverage check and reminders provided in the Final Question series reduced coverage errors. The frequency with which respondents or new babies were left off the form or counts were discrepant did not vary significantly between panels with the Final Questions and those without. It does appear that FQ 1 may have stimulated respondents to look back at Question 1 and in some cases fill it in when it had been left blank.

In conclusion, we have not found convincing evidence from this test of Norman's (1988) usability principle that providing an opportunity for respondents to correct erroneous reports of the number of people living in their household improves respondent accuracy. Although the questions show promise (they were answered by nearly all respondents, and the open-ended answers describe potential coverage problems that could provide a basis for guiding follow-up interview efforts), these data seem not to be strong enough to recommend widespread use at this time.

We do believe that the evidence supports further development and testing of coverage questions using our approach, and we urge additional research. The results show that the CFU failed to identify most people in residence situations who were described in the write-in responses and who should have been identified as potential adds. Without better measurements of census omissions in the follow-up interview, it is difficult to assess how well or poorly these (or other) coverage questions identify them. Improvements in coverage questions depend on better measurement of omissions in coverage follow-up interviews.

Indeed, coverage questions asked in the mail questionnaire in order to target follow-up need to work effectively in conjunction with the coverage follow-up interview in order to yield coverage improvements. It is apparent that the Final Questions provide the potential for guiding follow-up interview efforts to focus specifically on respondents' comments, rather than simply using independent reinterview techniques to recount members of households. Use of coverage questions in this way should be a priority for future research.

Research and development are needed to improve respondents' ability and motivation to recall and report missed residents in a follow-up interview. The follow-up interview may target omissions more effectively by asking *dependent questions* that incorporate information provided by respondents in the Final Questions. Interviewers can do a better job of probing for omissions if they have available information about people who may

have been missed and situations that gave rise to the follow-up interview. Descriptions of actual situations in respondents' own words would provide a more effective stimulus to recall than a standard set of probes.

In the U. S. Census, it is critically important and terribly challenging to get the counts right. Whereas most surveys can accept reasonably small amounts of error – in fact all sample surveys by definition have sampling error – the United States Constitution mandates reapportionment of Congressional representation every decade based on the decennial census. This requires highly accurate census counts. To illustrate, in the congressional reapportionment after the 2000 Census, Utah was only 857 residents shy of gaining a fourth congressional seat at the expense of North Carolina (New York Times 2001). Knowledge that people may be unsure of their reports and say so provides information that can be used to guide the targeting of households for follow-up interviews aimed at getting the final numbers for the nation, States, and congressional districts correct. Although the current study's design and the findings fall short of providing definitive evidence that their use will improve the accuracy of the census, it does suggest avenues for seeking such improvement.

## 7. References

- Cannell, C., Oksenberg, L., Fowler, F.J., Kalton, G., and Bischooping, K. (1989). *New Techniques for Pretesting Survey Questions*. Final Report of a project funded by the National Center for Health Services Research and Health Care Technology Assessment.
- Cantor, D., Heller, T.H., and Kerwin, J. (2003). *Cognitive Testing of Alternative Rostering Methods, Coverage Items, and Residence Rules Instructions in Five Experimental Versions of the Census Short Form for the 2004 Census Site Test*. Final Report prepared by Westat under contract with the U.S. Census Bureau.
- Davie, W. (1973). (E8): Effectiveness of Questionnaire Item 9 in Mail Areas. 1970 Census: Preliminary Evaluation Results Memorandum No. 40. U.S. Census Bureau, January 12.
- de la Puente, M. (1993). *Why Are People Missed or Erroneously Included by the Census: A Summary of Findings From Ethnographic Coverage Reports*. 1993 Proceedings of the Conference on Undercounted Ethnic Populations. Washington, D.C.: U.S. Department of Commerce.
- Dillman, D.A., Gertseva, A., and Mahon-Haft, T. (2005). *Achieving Usability in Establishment Surveys through the Application of Visual Design Principles*. *Journal of Official Statistics*, 21, 183–214.
- Dillman, D.A., Parsons, N.L., and Mahon-Haft, T. (2004). *Cognitive Interview Comparisons of the Census 2000 Form and New Alternatives*. Prepared under contract for the U.S. Census Bureau.
- Ellis, Y. (1994). *Categorical Data Analysis of Census Omissions*. DSSD 1990 REX Memorandum Series No. PP-10. Washington DC: U.S. Census Bureau.
- Ellis, Y. (1995). *Examination of Census Omission and Erroneous Enumeration Based on 1990 Ethnographic Studies of Census Coverage*. *Proceedings of the American Statistical Association, Survey Research Methods Section*, 515–520.

- Fay, R.E. (1989). An Analysis of Within-Household Undercoverage in the Current Population Survey. Proceedings of the Annual Research Conference. Washington DC: U.S. Census Bureau.
- Fay, R.E. (1998). VPLX Program Documentation, Vol. 1. U.S. Census Bureau. Available at <http://www.census.gov/sdms/www/index.html>
- Gerber, E.R. (1990). Calculating Residence: A Cognitive Approach to Household Membership among Low Income Blacks. Report prepared under contract for the U.S. Census Bureau.
- Gerber, E. (1994). The Language of Residence: Respondent Understandings and Census Rules. Center for Survey Methods Research, U.S. Census Bureau.
- Gerber, E. (2004). Summary Report of Cognitive Testing on Residence Rules Instructions. Statistical Research Division, U.S. Census Bureau.
- Gerber, E.R., Wellens, T., and Keeley, C. (1996). Who Lives Here? The Use of Vignettes in Household Roster Research. Paper presented at the annual meeting of the American Association for Public Opinion Research.
- Griffin, D. and Tversky, A. (2002). The Weighing of Evidence and the Determinants of Confidence. *Heuristics and Biases*, T. Gilovich, D. Griffin, and D. Kahneman (eds). Cambridge: Cambridge University Press, 230–249.
- Hogan, H. (1993). The 1990 Post-Enumeration Survey: Operations and Results. *Journal of the American Statistical Association*, 88, 1047–1060.
- Kerwin, J. and Moses, L. (2006). Cognitive Testing of Proposed Revisions to the Census 2006 Test Short Form. Report prepared by Westat for the U.S. Census Bureau.
- Krejsa, E.A., Linse, K.M., Karl, L., and Van Vleck, C. (2005). Evaluation of Residence Rules and Coverage Questions. Decennial Statistical Studies Division, U.S. Census Bureau.
- Linse, K., Karl, L., Kostanich, M., Van Vleck, C., Heimel, S., Woltman, H., and Kind, R. (2006). Coverage Analysis for the 2005 NCT. Washington DC: U.S. Census Bureau.
- Martin, E. (1999). Who Knows Who Lives Here? Within-household Disagreements as a Source of Survey Coverage Error. *Public Opinion Quarterly*, 63, 220–236.
- Martin, E. (2007a). Final Report of an Experiment: Effects of a Revised Instruction, Deadline, and Final Question Series in the Decennial Mail Short Form. 2010 Census Memorandum. Washington DC: U.S. Census Bureau. Available at <http://www.census.gov/srd/www/byyear.html>
- Martin, E. (2007b). Strength of Attachment: Survey Coverage of People with Tenuous Ties to Residences. *Demography*, 44, 427–440.
- Martin, E. (forthcoming). Can a Deadline and Compressed Mailing Schedule Improve Mail Response in the Decennial Census? *Public Opinion Quarterly*.
- Martin, E., Groves, R., Matlin, J., and Miller, C. (1986). Report of the Development of Alternative Screening Procedures for the National Crime Survey. Washington DC: Bureau of Social Science Research.
- Moriarity, C. (1993). Characteristics of Census Error – Additional Results. 1990 Decennial Census Preliminary Research and Evaluation Memorandum No. 240. Washington DC: U.S. Census Bureau.
- National Research Council (2006). Once, Only Once, and in the Right Place: Residence Rules in the Decennial Census. Committee on National Statistics, D.L. Cork and P.R. Voss (eds). Washington DC: The National Academies Press.

- New York Times. (2001). Judge Calls Counting Missionaries Unfair, March 29.
- Nguyen, N.T. and Zelenak, M.F.E. (2003). Coverage Gains from Coverage Questions C1 and C2 on Enumerator-Completed Questionnaires for Census 2000. Census 2000 Evaluation I.5. U.S. Census Bureau.
- Norman, D.A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.
- Parsons, N.L., Mahon-Haft, T., and Dillman, D.A. (2005). Cognitive Evaluations of Three Census Form Design Features: The Internet Option Message, Roster Instructions, and Identifying Person 1. Social and Economic Science Research Center Technical Report 05-022. Washington State University: Pullman.
- Redline, C., Dillman, D.A., Dajani, A., and Scaggs, M.A. (2003). Navigational Performance in Census 2000: An Experiment on the Alteration of Visually Administered Languages. *Journal of Official Statistics*, 19, 403–419.
- Robinson, J.G., Ahmed, B., Das Gupta, P., and Woodrow, K.A. (1993). Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis. *Journal of the American Statistical Association*, 88, 1061–1079.
- Robinson, J.G. (2001). ESCAP II: Demographic Analysis Results. Executive Steering Committee for A.C.E. Policy II, Report No. 1, Oct. 13. Available at <http://www.census.gov/dmd/www/pdf/Report1.pdf>
- Rodriguez, N. and Hagan, J. (1991). Investigating Census Coverage and Content among the Undocumented: An Ethnographic Study of Latino Immigrant Tenants in Houston. Ethnographic Evaluation of the 1990 Decennial Census Report #3. Available at <http://www.census.gov/srd/www/byyear.html>
- Rothwell, N.D. (1972). Three Classroom Experiments and One Field Test with Census of Population Housing Questionnaires. Questionnaire Research Report – Series A, Report 3. Statistical Research Division Response Research Staff Report No. 72–14. U.S. Census Bureau. Available at <http://www.census.gov/srd/www/byyear.html>
- Roy, L. (2003). 2001 Census, Quality Analysis of Certain Steps of Coverage: Verification of the Persons Entered in Step C (Uncertainty about Including a Person in the Household List). Unpublished report. Statistics Canada: Ottawa.
- Sweet, E.M. and Alberti, N.S. (1994). Coverage Implications of Using the Term “Usual Residence.” DSSD 2000 Census Memorandum Series #H-17. U.S. Bureau of the Census.
- Tourangeau, R., Shapiro, G., Kearney, A., and Ernst, L. (1997). Who Lives Here? Survey Undercoverage and Household Roster Questions. *Journal of Official Statistics*, 13, 1–18.
- Valentine, C. and Valentine, B.L. (1971). Missing Men: A Comparative Methodological Study of Underenumeration and Related Problems. Report prepared under contract with the U.S. Census Bureau, May 3. Available at <http://www.census.gov/sdms/www/index.html>
- Wellens, T. and Gerber, E. (1996). ICM Cognitive Evaluation. Center for Survey Methods Research, Statistical Research Division, U.S. Census Bureau, Feb. 26.

Received March 2007

Revised February 2008