

COVERAGE IMPROVEMENT FROM EXPERIMENTAL RESIDENCE QUESTIONS

Anne T. Kearney, Roger Tourangeau, Gary M. Shapiro, Lawrence R. Ernst*

A. Kearney, G. Shapiro, L. Ernst, Bureau of the Census, Washington, DC 20233

R. Tourangeau, National Opinion Research Center (NORC)

KEYWORDS: roster, undercoverage, usual residence

1. Introduction

Undercoverage occurs in two main forms -- entire households are missed and individuals are missed in households that are partially covered. Whole-household undercoverage occurs when a household eligible for interview is excluded from the survey. Within-household undercoverage occurs when one or more members of a sample household are omitted from the household roster.

This paper focuses on within-household undercoverage, the major topic of a study conducted on behalf of the Bureau of the Census by the NORC. In brief, the study used three questionnaire versions. One was a control version and the other two used an expanded set of experimental roster questions. One of these experimental versions required full names and the other allowed "anonymous" responses. The most important findings are that the anonymous questionnaire resulted in a higher average number of Black males listed as usual household residents (on the order of 33% higher). There was no evidence, however, that the expanded roster questions by themselves had any effect.

Survey undercoverage arises through several processes, but two seem to be central. As Hainer and his coworkers (1988) put it: "One reason people are missed is motivational. Black (and Hispanic) males are deliberately omitted from household rosters because of the potential loss of household income if the men were known to authorities ... A second cause of undercoverage is the lack of fit between the census definitions of household and residency, and people's actual living situations."

Consistent with this account, coverage of young Black males is worse than coverage of older Black males; similarly, young Black males are more likely to be missed than their non-Black counterparts. Coverage of Black females is better than coverage of Black males, but is not as good as coverage of non-Black females (Shapiro, Diffendal, and Cantor, 1993).

Confusion about who to count can arise because under current Census Bureau definitions, a person's usual residence is the dwelling where he or she lives and sleeps most of the time. Not everyone associated with the household may fit these criteria; for example, it may be unclear whether to count a person staying in

the respondent's home temporarily. In other cases, people who float between several residences may be included or left off the roster at the respondent's discretion. Furthermore Census Bureau procedures call for individuals with no usual place of residence who are currently staying in the respondent's residence to be included in the final roster in addition to "usual residents," but the household respondents themselves may often be unaware of this rule.

This study examines situations like this and is a follow-up to two earlier studies by Shapiro, et al. (1993). In the next section, we describe the method for this study and provide an overview of the analysis presented in Sections 3 and 4. A longer version of this paper containing additional analysis is available (Kearney, et al. 1993)

2. Methodology

2.1 Sample Design

The sample for the experiment consisted of 644 occupied dwelling units spread across a total of 49 blocks in three sites--the Chicago and Washington, D.C., metropolitan areas and the City of Baltimore. The Field Division of the Census Bureau selected and listed blocks that were predominantly Black and low income. Based on listings of each of these 49 blocks, NORC staff selected 15 housing units in each block using a systematic selection procedure; of the 735 dwellings selected in this way, 644 turned out to be occupied and thus eligible for interview.

Interviews were completed at a total of 509 of these dwellings, for a response rate of 79.0%. NORC interviewers collected information on 1949 individuals in the 509 interviews.

The experiment compared three versions of the questionnaire, which were systematically assigned to the housing units.

2.2 Questionnaires

The first, or standard, version of the questionnaire begins with items drawn directly from the SIPP control card. In Version 1, an initial roster of residents is compiled based on responses to the *standard roster questions* ("What are the names of all persons living or staying here? Have I missed any babies or small children? ... Does (NAME) usually live here?") and demographic information is then collected for each person on the roster. Next, Version 1 included a series

of *experimental roster questions* intended to identify residents omitted from the initial roster; these experimental roster questions ask:

- How many people besides those you've already listed stayed here last night?
- Is there anyone else who stayed at least one night during the past month?
- Is there anyone else who usually stays here but was not here during the last month?
- Is there anyone else who ate here at least once during the last week?
- Is there anyone else who usually eats here but who did not eat here during the last week?
- Is there anyone else who you consider to be a member of this household?
- Is there anyone else who considers himself or herself to be a member of this household?

New people identified by the experimental roster questions were then added to the roster. Once the roster was complete, Version 1 continued with a series of questions intended to clarify the relation of each person listed to the household.

- How many nights did NAME stay here during the last month?
- On how many days did NAME eat here during the last week?
- Does NAME contribute money, food, or other help to this household?
- Do you consider NAME to be a member of this household?
- Was NAME here at all during the last day?
- Does NAME consider himself/herself a member of this household?
- Does NAME usually live here?
- Does NAME have other places where he/she frequently stays?

After these *follow-up items* are a series of items on labor force participation.

The other two versions of the questionnaire contain essentially the same items as Version 1 but administer them in a different order. In Versions 2 and 3, the experimental roster questions are administered first, then the follow-up items, and finally the standard roster questions. (Both the experimental roster questions and the standard items have been reworded slightly in Versions 2 and 3 to fit their new context.) Versions 2 and 3 differ from each other in only one respect: Version 3 requires respondents to identify household members by their full names, whereas Version 2 allows respondents to use initials, nicknames, or other means of identifying individuals instead.

2.3 Generalizability of Results and Standard Errors

We took a purposive sample of blocks from three

arbitrarily chosen sites (though we did select equal probability samples of units within each sampled block). We chose to treat this as a universe equally distributed among the 49 blocks. Thus, the results cannot be properly generalized to any universe beyond that specified. We therefore computed standard errors assuming a simple random sample of housing units. If a reader wants to generalize these results to a larger universe, then under most reasonable models the standard errors we used underestimate the variability in the estimates; as a result it's possible that some results we claim as being significant would no longer be so.

2.4 Overview of the Analysis

We analyzed the data from this study to determine which method, on the average, produced more residents listed on the roster. In particular, the analysis compares standard roster questions to experimental roster questions and full name interviewing to anonymous interviewing. These issues are discussed with respect to the total number of people initially rostered (Section 3.1) the total number of respondents classified as usual residents (Section 3.2), and the total number of people on the final roster under an experimental roster definition (Section 3.3).

Finally, we sought to determine which specific questions account for any gains produced by the experimental versions. This issue is discussed in Section 4.

3. Analysis of Data on Rostered Persons

3.1 Number of Total Persons Rostered

Overall, the experimental roster questions (Versions 2 and 3) added about a person compared to the standard items (Version 1). This is a highly significant difference; a one-way analysis of variance yields an $F(2,506)$ of 8.04 ($p < .001$). (In using the F test here, we recognize that we are ignoring the lack of normality of the underlying distribution.)

The rosters compiled under Versions 2 and 3 included more Black males than those compiled using the standard questions in Version 1. (See Table 1.) Version 2 increased the number of Blacks males rostered per household by more than 60% relative to Version 1; the increase for Version 3 as compared to Version 1 was about 38%. The differences across versions are highly significant-- $F(2,506)$ of 7.88 ($p < .001$). A contrast comparing Version 2 with Version 1 yields a highly significant t of 3.92 ($p < .001$). A contrast between Version 3 and Version 1 is also significant with a t of 2.46 ($p < .01$). (In performing tests of individual contrasts, we made no adjustments to compensate for making multiple comparisons. However, because the p value was so low

for all the tests that were significant, undoubtedly the same conclusions would hold even if we made such adjustments.) However there is no evidence of differences between Versions 2 and 3.

TABLE 1. Average Total Persons and Usual Residents Rostered by Version

	Version 1 (n=173)		Version 2 (n=177)		Version 3 (n=159)	
	Total Persons	Usual Residents	Total Persons	Usual Residents	Total Persons	Usual Residents
Total	2.90	2.83	3.94	3.13	3.87	2.99
Black Males	1.12	1.09	1.80	1.45	1.55	1.18
Black Females	1.48	1.45	1.93	1.60	1.98	1.60
Non-Black Males	0.16	0.16	0.04	0.02	0.12	0.06
Non-Black Females	0.13	0.13	0.04	0.02	0.06	0.05

Note: Because of missing data, row entries within a column may not sum to the column total. Figures for Version 1 reflect only persons listed using the standard rostering questions.

3.2 Number of Usual Residents Rostered

All three versions of the questionnaire contain an item to distinguish usual residents from others listed on the roster. In each version, respondents were asked, "Does NAME usually live here?" about each person on the roster. Those for whom the answer was positive are considered usual residents.

There is no evidence of overall differences in the number of usual residents reported in the three versions. However, if we examine the means for Black males--a group with a particularly high level of undercoverage--significant differences do emerge. Version 2 had the highest mean number of Black male usual residents. (See Table 1.) Version 2 yields an increase of 33% in the number of Black males relative to Version 1. A contrast comparing Version 2 (which did not require full names to be used) with the other two versions (both of which did require full names) is also highly significant-- $t(506)=2.72$ ($p < .001$). Thus, relaxing the requirement that full names be given appears to substantially increase the number of Black males reported as usual residents. However, there is no evidence that the experimental roster questions had any effect by themselves.

The differences across versions in the number of usual residents reported did not extend to Black females.

3.3 Complex Definitions for Residences

The "usual resident" question may not be the best way to determine which people belong to a given household for several reasons: 1) We expected that our broad-based coverage questions may pull in people that we would regard as usual residents, but are not so

considered by respondents; 2) we also expect that respondents may be unwilling to acknowledge some people as usual residents but will give us honest answers to other questions by which we can classify them as belonging to the unit, and 3) we also expect the experimental roster questions to net some people who have no usual place of residence and who therefore are not identified as usual residents by the respondent but should be included in the final roster because they are currently staying at the unit.

In this section, we examine one of many possible definitions for who should be included as a housing unit resident. The basic idea is that a person is a resident of the housing unit where he or she stays most often. More specifically, the definition encompasses three major classes of persons:

1) Persons identified by the household respondent as usual residents and who either

- have no other place they frequently stay,
- stayed at the housing unit in question at least as many nights in the past month as any other place,
- moved into the housing unit within the past 30 days, or
- has no single other place that they spent more time at than here during each of the last 6 months.

2) Persons identified by the household respondents as not usual residents here and who either

- have no other place they frequently stay and either:
 - stayed at the housing unit in question the previous night, or
 - stayed at the housing unit in question at least as many nights than at any other place in the past month
- stayed some other place frequently, but has no usual residence according to the respondent and stayed here at least as many nights last month as stayed in any other place
- have a usual residence elsewhere, but stayed here at least 15 nights during each of the past 6 months

3) College students (or other boarding students) living away from school who are regarded as members of the household by the household respondent. (Some surveys do not list college students living away at school at the sampled housing unit.)

The questionnaires did not include all the items needed to classify every person in the sample under this alternative definition. Therefore, it was impossible for us to classify some of the persons rostered as clearly meeting the definition or clearly failing to meet it. Results from the Living Situation Survey (Schwede, 1993) may help us determine whether or not people whose living arrangements we are currently uncertain about would meet the complex definition.

Table 2 shows the mean number of persons per household who met the definition above. The "conservative" column includes persons who definitely met the criteria; the "liberal" column includes persons who met the definition under the conservative criterion and all persons who either were not usual residents of the sample unit and had no other place where they frequently stayed or were usual residents of the sample unit but had another place where they frequently stayed. The "usual residents" column includes persons with an affirmative answer to the question "Does NAME usually live here?" Version 1 could not be classified under the complex definitions because we can't determine if any of the people from the standard question stayed last night.

We expected the complex definitions to classify a larger number of people as belonging to the household than does the "usual resident" definition. However, Table 2 averages differ little from each other. This may indicate that there are few transient-type people in these households that would be missed by the standard definition. But, after looking at the number of people whose residency we are uncertain about, it is not clear. For example, of the 318 Black males rostered in Version 2, 260 met the liberal criterion and 58 did not meet it. Of the 58 who did not meet the criterion, though, there are 31 Black males whose residency could not be determined under the complex definition. This does not necessarily imply the undefined people would be transient-type people who would meet the criterion for residency, but since some may, the averages in Table 2 under the complex definitions may be underestimates even for the liberal criterion.

TABLE 2. Average Number of Residents on the Final Roster by Version and Definition

	Version 2 (n=177)			Version 3 (n=159)		
	Liberal Criterion	Conservative Criterion	Usual Residents	Liberal Criterion	Conservative Criterion	Usual Residents
Total	3.18	3.12	3.13	3.06	2.97	2.99
Black Males	1.47	1.45	1.45	1.23	1.16	1.16
Black Females	1.62	1.59	1.60	1.62	1.60	1.60
Non-Black Males	0.02	0.02	0.02	0.06	0.06	0.06
Non-Black Females	0.02	0.02	0.02	0.05	0.05	0.05

Note: Because of missing data, row entries within a column may not sum to the column total.

4. Performance of the Roster Questions

A major purpose for this analysis is to determine if there are a substantial number of individuals who are added to the initial roster by the later rostering questions.

4.1 Examination of the On-diagonal and Off-diagonal Entries

In Table 3, the rows represent the questions that added persons to the roster and the columns represent the quantitative follow-up questions. The questions in the first four columns correspond to those in the first four rows. (See Section 2.2. The row questions appear in the first set of bullets and the column questions appear in the second set of bullets.)

Note that the first two roster questions (How many people stayed here last night? How many other people stayed here at least one night during the last month?) have been combined in the first row because they correspond to a positive number answer to the Column 1 question (How many nights did NAME stay here during the last month?). The questions listed in Rows 5 and 6 of the table do not correspond to any column question. Consequently, the questions in Section 2.2 have been reordered in the table to put these two questions last. Rows 7 and 8 are for records with roster-question numbers that are meaningless or missing, respectively.

Each individual is tabulated in the row corresponding to the question that brought him/her into the roster. The column corresponds to the first quantitative-information question for which the answer was a positive number or yes. For example, if the answers to the Column 1 and 2 questions were 0 and 5, respectively, for an individual, that individual would be tallied in Column 2, irrespective of their answers to the Column 3 and 4 questions. An individual is listed in Column 5 if the answers to the Column 1-4 questions were all either zero or no. The last column lists all individuals who could not be classified in a preceding column due to missing information.

The top number in each cell is the number of people who met the criteria of that cell. The number below it in parenthesis is the number of people who both met the criteria and are considered usual residents by the respondent.

We first discuss here the top numbers in the cells formed by Rows 1-4 and Columns 1-5. When there is consistency between the roster question and the quantitative information, the entry falls on the diagonal. Inconsistencies fall off the diagonal. Column 5 entries are inconsistent and considered above the diagonal.

Overall, 93 percent (1105/1185) of the entries are on the diagonal. Ninety-eight percent of the people rostered with the staying questions are on the diagonal. The people rostered with the eating and member questions are 54 and 36 percent on the diagonal, respectively.

Few entries (31) fall above the diagonal. The phrasing of the roster questions is slightly different from

the phrasing of the quantitative information questions, and may have prompted different answers from the respondent, resulting in some of the above-diagonal entries (and also some of the below-diagonal entries).

Above the diagonal entries could also be the result of a household respondent who initially admitted the attachment of a person to the household by adding them to the roster and then minimized the attachment as the interview progressed into the quantitative questions. The respondents in these households may have concluded that either further information was not wanted or should not be provided for these people.

Below diagonal entries indicate individuals who should have been added to the roster earlier. That is, answers to the quantitative follow-up questions indicated that the person should have been listed already. There were 49 below-diagonal entries. Of these 49 people, 41 stayed at least one night in the residence. Thus, the eating and household membership questions may have reminded respondents in some cases of other people who had also stayed there. In other cases, these questions may have brought in people that the respondent purposely avoided mentioning when asked who stayed in the house. Then, as the roster-building questions focused on eating or household membership questions, they were willing to add other people to the roster. Whether the respondent inadvertently or

purposely did not initially list these people as staying in the house, this makes asking the later roster questions especially useful.

4.2 Usual Residents

The key purpose in asking a set of questions to build an initial roster is to obtain as complete a final roster as possible. In this subsection we consider the impact of each of the rostering questions on the final roster. For purposes of this discussion, a person is defined to be on the final roster if and only if the person is considered by the respondent to be a usual resident. The discussion on usual residents in this section are at least partially dependent on the order in which the rostering questions were asked.

Across all roster questions, 78 percent (1030/1314) of the people listed are usual residents. The roster questions about people who stayed last month at the dwelling (Row 1) or usually stay (Row 5) in the household bring in 93 percent (962/1030) of the usual residents. The roster questions about people who ate last week (Row 2) or usually eat (Row 6) in the household added only 1 percent (6/1030) of the usual residents, and the household membership questions (Rows 3 and 4) bring in fewer than 1 percent (5/1030).

Five percent (50/1030) of the usual residents cannot be tied to a specific roster question because of missing

Table 3. Roster Questions By Quantitative Follow-Up Questions for Version 2 and Version 3

Roster Questions	Total	Quantitative Follow-Up Questions					
		Nights Stayed > 0	Days Ate > 0	Considered a Member	Considers Self Member	All Entries Zero or No	Problem with Blanks
Total	1314 (1030)	1134 (1013)	96 (1)	29 (3)	3 (1)	24 (0)	28 (12)
Stayed Last Night/Month	1051 (957)	1027 (946)	5 (1)	3 (1)	1 (1)	3 (0)	12 (8)
Ate This Week	127 (5)	32 (5)	68 (0)	4 (0)	0	13 (0)	10 (0)
Member Who May Be Away	28 (5)	9 (4)	7 (0)	10 (1)	1 (0)	1 (0)	0
Considers Self Member	1 (0)	0	0	1 (0)	0	0	0
Usually Stays But Away	11 (5)	3 (2)	0	3 (1)	1 (0)	1 (0)	3 (2)
Usually Eats, Not This Week	34 (1)	6 (1)	16 (0)	7 (0)	0	5 (0)	0
Other	8 (7)	8 (8)	0	0	0	0	0
Missing	54 (50)	49 (48)	0	1 (0)	0	1 (0)	3 (2)

Note: The top number in each cell is the number of people who met the criteria of that cell. The number below it in parenthesis is the number of people who met the criteria and are considered usual residents.

information (Row 8). Forty-eight of these 50 people stayed in the households every night of the month. All 50 have no other place where they frequently stay. Probably all 50 fit into the "Stayed Last Night/Month" category, a reallocation which would raise the percentage of usual residents that are in Rows 1 and 5 from 93 percent to 98 percent.

4.3 Roster Questions that Worked and Didn't Work

The roster questions that listed people who stayed in the household clearly worked. These questions captured at least 91 percent of the usual residents. The people rostered with the eating and member questions captured only six percent of the usual residents. According to the household respondent, almost all of these stayed at least one night in the house. (See Table 3.)

5. Conclusions

Compared to Version 1, Version 2 produced about a 33% increase in the average number of Black males reported as usual residents by the household respondents. This indicates considerable promise for improved coverage of Black males by permitting anonymity in survey interviews. There was no evidence, however, that the broader set of roster questions increased the average number of usual residents reported, though it did increase the total number of people reported with at least casual household associations. Since Version 2 had the broad set of questions as well as the anonymity feature, we cannot be sure that anonymity by itself would achieve significant coverage gains.

There were a number of people in the sample that may have complex living arrangements. This is indicated by a substantial group that did not unequivocally belong to the unit nor obviously belong to some other unit. It is possible that many people of this type who were added through the broad set of roster questions met the criteria of our complex definition of belonging to the unit, but unfortunately we did not ask for sufficient information to classify them. For example, it may be that many of the people who frequently stay elsewhere spend more nights at this unit than at other units, but we could not ascertain if this was true. Or, perhaps people of this type have a usual residence elsewhere, and therefore don't belong to the household. Results from the Living Situation Survey may help determine whether there are a substantial number of people who meet this or other complex definitions of household belonging but are not considered usual residents by the household respondent. This research survey is now being conducted by Research Triangle Institute for the Census Bureau (Schwede, 1993).

We are planning additional testing for some type of anonymity in Census Bureau surveys. There are of course difficulties when full name is not collected, especially in longitudinal surveys like Survey of Income and Program Participation which follows movers. We will have to weigh these difficulties against the potential gains of improved coverage. At this point the potential is high. We would also like to conduct more testing of expanded roster questions. We will focus on asking questions that can classify people with complex living arrangements under complex definitions of household belonging.

6. Acknowledgements

We thank Randall J. Parmer and Michael J. Roebuck and Lynn Weidman of the Census Bureau for their useful comments. All errors and opinions, however, are solely the authors. We would also like to thank Carol Macauley for her typing of this paper.

* This paper reports the general results of research undertaken by Census Bureau Staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

REFERENCES

- Hainer, P., Hines, C., Martin, E., and Shapiro, G. (1988), "Research on Improving Coverage in Household Surveys," in *Proceedings of the Fourth Annual Research Conference*, Washington, DC: U.S. Bureau of the Census, pp 513-539.
- Kearney, A., Tourangeau, R., Shapiro, G., Ernst L. (1993). "Coverage Improvement from Experimental Residence Questions," SRD Research Report Number Census/SRD/RR-93/09, Statistical Research Division Report Series, U.S. Bureau of the Census, Washington, D.C.
- Schwede, L. (1993), "An Experimental Test of Residence Rules: The Living Situation Survey," in *Proceedings of the Survey Research Methods Section of the Annual Meeting of the American Statistical Association* (forthcoming).
- Shapiro, G., Diffendal, G., and Cantor, D. (1993). "Survey Undercoverage: Major Causes and New Estimates of Magnitude," in *Proceedings of the 1993 Annual Research Conference*, Washington, DC: U.S. Bureau of the Census (forthcoming).
- Tourangeau, R. (1993), Final Report: SIPP Roster Questions under contract number MRO92-01. Unpublished report submitted to Gary Shapiro, U.S. Bureau of the Census.