THE ROLE OF QUESTIONNAIRE DESIGN IN REDUCING CENSUS COVERAGE ERROR[1]

Elizabeth A. Martin and Deborah H. Griffin, U.S. Bureau of the Census
Elizabeth A. Martin, U.S. Bureau of the Census, Washington, DC 20233-9150

Keywords: residence rules, household roster, mail questionnaire

I. THE PROBLEM

It is fair to say that until recently there has been no systematic attempt, by the Census Bureau or other survey organizations, to approach household rostering as a measurement issue. Rather, it is commonly treated as an administrative detail prior to the survey itself. Yet, accurate household rosters are critical to ensure accurate and complete representation of the population in a census or survey. The instructions and questions used to compile a household roster essentially serve as rules to link persons to housing units or other units of the sample universe. Ideally, the residency rules and roster instructions assign each and every member of the population as resident of one and only one housing unit.[2] If persons are left off rosters who should be included, then the coverage of the population is incomplete even if the coverage of housing units is perfect. If persons are included in multiple rosters, then the survey or census is flawed by overcoverage.

Several types of problems may generate errors of omission or erroneous inclusion on household rosters.

- First, the rules themselves may be incomplete or logically inconsistent or vague, so they do not function to assign each person to one and only one household. Respondents may ignore rules which seem counterintuitive and do not match their own understandings of who lives in their household.
- Second, respondents or interviewers may not understand the rules or how to apply them.
- Third, some persons may have living situations which make it difficult to uniquely assign them to a household under any set of rules.
- Finally, there may be reasons why residents do not want to reveal full details about household membership.

In the paper below, we begin by summarizing evidence about sources and magnitude of coverage errors within households. Next, we summarize the history of the residence rules used in the decennial census and critique the roster questions and instructions used in the 1990 census. Finally, we recommend alternative mail questionnaire rostering strategies and methods which should be explored and evaluated experimentally.

II. EVIDENCE ABOUT SOURCES AND MAGNITUDE OF WITHIN-HOUSEHOLD CENSUS COVERAGE ERRORS

Table 1: Rates of Nonmatched Persons by Type

	Nonhispanic White & Other	Black	Hispanic	Asian Pacific Islander	Total
Within household	1.3	4.3	3.3	2.6	1.8
Whole household	1.6	4.7	2.7	3.0	2.0
Housing unit	1.6	2.5	3.0	1.4	1.8
Other	0.2	0.4	1.0	0.5	0.3
Total	4.7	11.9	10.0	7.5	5.9

Source: Hogan (1992)

Table 1 shows estimates of rates of 4 different types of nonmatches (or omissions) in the 1990 census, based on Hogan's (1992) analysis of the Post Enumeration Survey. The first category is a fair estimate of "within household" coverage error as it summarizes nonmatched persons within a household where other persons were matched. About 1.8 percent of the total population were omissions of this type. The estimates are highest for Blacks, Hispanics and Asian and Pacific Islander populations but the relative contribution of the within household component remains about a third for all groups.

The second component of "whole household" nonmatches occurs when an occupied unit is erroneously identified as vacant or the wrong persons are enumerated in place of the correct ones. About 2.0 percent of persons were nonmatches in households where all persons were nonmatches but the housing unit was enumerated. This represents another third of the nonmatched persons.

Housing unit coverage error occurs when all persons were nonmatches because the housing unit was a nonmatch; this also accounted for a third of the

nonmatches. The final category, "other," includes incomplete cases or processing errors.

Table 2 presents results of analyses by Moriarity and Childers (1993) of census omissions in housing units classified as occupied. About 1.3 percent of the population were erroneously left off mail questionnaires, and 3.0 percent of the persons listed on mail return forms were included in error. Both error rates increase when the data are provided by a non-household proxy, rather than a member of the household. Additionally, the presence of nonrelatives in a household significantly increases the likelihood of both erroneous inclusions and exclusions. (Ellis (1994) and Fay (1989) also find that nonrelatives are far more likely than relatives to be omitted.)

Ethnographic studies of causes of minority undercount in the 1990 census provide vivid examples of the types of households which were enumerated incorrectly in the census (see de la Puente, 1993, for a summary of the ethnographers' findings). For example,

Table 2: Errors on Mail Return Questionnaires

	Rate of Omission	Rate of Erroneous Inclusion
Respondent		
• Household member	1.3	3.0
• Proxy	6.9	7.0
Date of Completion		
• 3/18 - 3/24	1.0	2.4
• 3/25 - 3/31	1.3	2.9
• 4/1 - 4/7	1.6	3.2
• 4/8 - 4/14	2.1	5.3
Relation		
• All Related	1.1	3.0
• 1 or more Nonrelatives	4.7	4.2

Source: Moriarity and Childers (1993) Griffin and Moriarity (1992)

Mexican migrant workers in Marion County, Oregon, had a particular type of arrangement, which Montoya (1992) called ad hoc household. Ad Hoc households came together as a practical response to poverty and a lack of affordable housing. Each slot in these households was allocated by money, so relationships were ephemeral. Information was difficult to obtain because persons living in ad hoc households protected, and in some cases did not know, the identities of those living with them. Montoya noted that in these circumstances, coverage was largely determined by the coincidence of who happened to be present during an enumerator's visit. Similar observations regarding unrelated Hispanic males sharing the same housing unit were made in other sample areas with recent immigrants.

A more complex household centered around a nuclear core is described by Romero (1992) in San Francisco. The core was a Salvadoran immigrant couple with their two children. The couple rented a three bedroom apartment and took in nine other Salvadoran immigrants to make the rent. The couple and two children shared one bedroom. Another bedroom was occupied by a woman and her unmarried male partner, their baby, two children from her previous marriage and her partner's brother. Three recent Salvadoran immigrants occupied the third bedroom. Two were unrelated and the third was the father of the wife. Interaction among household members was minimal and impersonal. Each bedroom had a lock and the bathroom and kitchen were shared. The refrigerator was divided into different sections and dry and canned goods were kept in the bedrooms. Of the 13 persons living in this "household" only 6 were enumerated by the census.

Complex households also characterized two Haitian sites in Florida, where there was usually a core family group with other individuals in the periphery who came in and out of the household depending on their life circumstances (Wingerd, 1992). The Haitian households often included persons who were described by other household members as "just passing through," who stayed anywhere from 2 weeks to 4 years. New arrivals from Haiti were referred to as "just comes," the community's term for someone literally just arrived or released from detention. When there was no relative to stay with, a "just come" typically was taken in as a boarder by a friend of a friend for an indefinite stay.

Gerber (1990, 1994) has applied the methods of cognitive anthropology to investigate how people decide questions of where someone lives. When life circumstances are complex and ambiguous, her informants used criteria, such as peoples' intentions and agreements, the location of belongings, and where mail is received, to determine residency. These considerations are not part of the census definition, and may lead respondents to leave off "marginal" people who should be included, or to include them in error. In particular, calculations of residence may lead respondents to leave off nonrelatives. This is shown dramatically in Table 3, which presents estimates of gross omission and erroneous enumeration rates based on the ethnographic sample. These are calculated as dual system estimates, but these data are from sites selected purposively in areas expected to have large undercounts, so the rates not generalizable and are much higher than characterize the census generally. Gross omissions increase dramatically for persons in more peripheral relationships, such as boarders, housemates, and unmarried partners. Erroneous

enumerations also increase, but not as steeply, so the net effect is a sharp increase in net undercount rates for persons in more marginal relationships.

Table 3
Rates of Gross Omissions and Erroneous Enumerations, by Relationship, for Ethnographic Sample Cases

	Gross omission rates		Gross erroneous enumeration rates		Net undercount
	%	N	%	N	Difference
Householder, spouse	14.7	3563	10.5	3756	4.2
Son, daughter	16.6	2796	10.8	2916	5.8
Other relative	27.9	760	15.0	799	12.9
Nonrelative	44.2	529	16.8	487	27.4
Total	18.8	7648	11.4	7958	7.4

Note: Gross census omissions are calculated as the number of Census Day residents counted by the ethnographer but not census, divided by the number of census day residents counted by the ethnographer. Gross census erroneous enumerations are calculated as the number of persons counted in the census who were not census day residents of the site, divided by the number of persons counted in the census. Cases with missing data on relationship, or for whom residency status on Census Day is uncertain, are excluded. Table 3 is based on the 10/92 version of the ethnographic data.

III. HISTORICAL PERSPECTIVE ON CENSUS RESIDENCE RULES

Next we examine the rules and consider how their implementation in the census may give rise to coverage errors. The basic census residency rules were already in place in the 1820 census: each person was to be enumerated at his or her "usual abode" as of the day enumeration commenced; persons who died after that day were to be included and babies born after it were to be excluded; persons with "no settled place of residence" were to be enumerated with the family where they were staying on census day (Bureau of the Census, 1989).

Over the decades, special rules were added specifying where certain groups should be counted, such as seafaring men or those employed in navigation (1850 census), children absent for the purpose of education (1870 census), and so on. By 1990 the number of residency rules had grown to 17.

Up until the 1960 Decennial Census the enumerator was responsible for compiling the household roster. That Census was the first major implementation of self response, thus shifting the task of determining a roster from the enumerator to the household respondent. The basic approach employed in censuses over the past four decades has been to request a roster of "usual residents."

The 1990 questionnaire begins with a statement of the basic principle that persons must be counted at their "usual residence," defined as "the place where the person lives and sleeps most of the time." This definition has no reference period, allowing respondents to interpret "most of the time" over a week, a month, years, a lifetime. In addition to several listing instructions, the questionnaire introduces numerous rules which include or exclude specific categories of persons, such as persons temporarily away, college students, etc. Several rules contradict the general principle that a person is to be counted where he or she lives and sleeps most of the time, and others contradict the commonsense notion that a person should be counted at home.

By the time a respondent is ready to begin the task of listing the roster of household members, he or she in theory has been exposed to one principle, one definition, three instructions, and 15 rules about categories of persons to include or exclude on the list. Various residence terms have been introduced, including usual residence, living, staying, home, and household. Do respondents pay attention to this material? Can they absorb and understand it? Does it make any difference if they do or not? Do their rosters conform to the rules? If not, what rules do they apply, and how do they understand the task of rostering?

We have already seen evidence that respondents commonly deviate from census rules by leaving nonrelatives off the roster when they should be listed. More evidence comes from the Living Situation Survey, conducted by Research Triangle Institute for the Census Bureau (see Schwede, 1993, for a detailed description). Interviews completed in 999 households used extensive probing to compile expanded rosters. For each household, the roster included not only persons who lived there, but those with tenuous and casual attachments as well--persons who kept their belongings or received mail there, or who slept there even one night during the 2-3 month reference period. After compiling the roster, household respondents were asked, for each person reported, "Do you consider this address to be (NAME's) usual residence, that is the place where (NAME) lives and sleeps most of the time?" Thus, the respondents were asked to apply the census usual residence definition, although they were not exposed to the rules, as respondents in the census are. The survey went on to conduct individual interviews with the rostered persons, to determine their own assessments of their usual residence and collect data about their movements in and out of the household during the reference period. Casual visitors who were rostered but

had stayed a week or less and had a home elsewhere, were not followed up and are excluded from all figures cited below.

Most of the rostered persons-- 91 percent of the sample-- had one clearly defined residence and stable living situation, according to analysis by Sweet (1994). For the 9 percent of persons with atypical living situations, household respondents' determinations of "usual residence" were frequently inconsistent with census residence rules, as shown by Sweet and Alberti (1994). Of these persons with atypical living situations, there were four times as many persons who should have been included but weren't, as the reverse: persons who were counted when they should not have been. This result suggests a possible underreporting bias in household respondents' determinations of usual residence for persons with ambiguous or complex living situations.

A similar bias is suggested by comparisons of household respondents' and individuals' own determinations of usual residence, which were inconsistent in about 5 percent of cases. There was an apparent underreporting bias due to persons who claimed a sample household as their usual residence but were not claimed by household respondents. There were three times as many of these individuals as the reverse, persons who said they were not usual residents but household respondents said they were. The reasons for bias are suggested by Gerber's (1994) cognitive study of how people determine residency in complex situations portrayed in vignettes. She finds that persons who provide space to highly mobile persons may lack information and assume the existence of a valid residence elsewhere, based on other aspects of a person's lifestyle.

Inconsistent reporting of usual residence also may result from respondents' confusion about its meaning. Gerber and Bates (1994) report that some respondents thought the term "usual residence" implied there was another residence besides the permanent residence. Others assumed the intent was to ask about permanent or official address, and interpreted the roster instructions as asking for a list of persons with long term or official attachments to the place.

It appears there is a small but important group of people--perhaps 5 to 9 percent--whose residential arrangements are ambiguous or subject to uncertainty or disagreement within the household. The numbers of such persons in the Living Situation Survey are too small to be very certain about their impact on coverage. Nevertheless, Table 4 presents national estimates of potential omission and erroneous enumeration errors due to complex living situations, and to inconsistent household and individual classification of individuals' residency. The standard errors on these estimates are too large to draw conclusions about the relative magnitudes of rates of omission and erroneous enumeration. However, it is suggestive that these data are consistent with the hypothesis of a net underreporting bias on the part of household respondents in complex or ambiguous circumstances. In combination with the evidence of Tables 2 and 3, it appears that household respondents may be too restrictive in deciding whether to include marginal or peripheral persons as household members, and therefore leave off some persons who should be rostered. A larger sample is needed to draw more definite conclusions.

Table 4
Estimates of Potential Coverage Errors
Based on Living Situation Survey

	Potential omissions		Potential erroneous enumerations	
	%	s.e.	%	s.e.
Due to unusual living situations	2.30	1.49	.58	.26
Due to inconsistent individual and household respondent classification of individual's residency	.59	.37	.11	.09

Source: Sweet and Alberti (1994)

As an antidote to this tendency, we have proposed an alternative rostering strategy which casts a broader net by eliciting information about persons with tenuous or casual attachments to a household as well as the more central, core members, as was done in the Living Situation Survey. If it works, such an expanded roster would have the advantage of including persons who are erroneously left off using current methods. A drawback is that an expanded roster is likely to pull in persons who do not live in the household but somewhere else. To avoid duplication, screening questions asked after the roster has been compiled are needed to permit residents of other households to be identified so they can be eliminated from the final count.

Several recent methodological studies tested an expanded roster strategy combined with screening questions. In a mail questionnaire experiment conducted as part of the 1990 census, Bates (1991) found that a simplified and somewhat expanded roster yielded significantly more names on the roster, compared to the standard census form. (However, the experimental panel containing the expanded roster did not obtain significantly different results from those experimental panels which did not, so results were not conclusive.) In another experimental test conducted in

personal interviews in Chicago, Washington DC, and Baltimore, Kearney et al. (1993) found no increase in the number of usual residents enumerated using expanded roster questions, although there was an increase in the number of Black males listed when respondents were not asked to give full names. Sweet (1994) analyzed the coverage gains obtained by the expanded set of roster probes used in the Living Situation Survey, and found that, compared to the census, there were significant increases in the mean number of usual residents per occupied housing unit for Hispanics and for the total population. She also found that the additional probes used in the Living Situation Survey were especially effective at adding to the roster young, minority, males who were not identified by more conventional probes.

In a 1994 questionnaire experiment, Pausche (1994) found that a form with expanded roster probes yielded a significantly larger number of rostered persons than a slightly improved version of the 1990 questionnaire. The number of usual residents per household identified by the two approaches was not significantly different, however. The problem was that the screener questions incorrectly eliminated persons who should have been counted as household residents. (This result may be further evidence of household respondents' tendency to be too strict in making determinations of usual residence for persons living or staying in their household.)

V. CONCLUSIONS

Our review of recent research to improve coverage within households by improving the design of household rosters suggests several specific and several general conclusions.

To begin by summarizing specific results: experimental and other research shows that additional roster probes can be designed to add people to household rosters, and that the persons added are in the categories missed using traditional roster methods (namely, young, minority, males.) The research further indicates that about 9 percent of the population have complicated or unclear living situations which put them at risk of being incorrectly enumerated. For about 5 percent of the population, household respondents and the individuals themselves disagree about the residency status of persons within households.

Evidence suggests that household respondents tend to err on the side of excluding marginal or peripheral household residents. This may be due to flaws in the ways the census roster questions were asked and in their premise that each person can be assigned to one and only one "usual residence." The variety of residence terms used in the traditional census roster may also be a source of underreporting by household respondents. Gerber and Bates (1994) hypothesize that respondents assimilate the various contrary meanings and multiple rules by interpreting the roster as a request for the permanent or official residents of the household. Possibly, the screening questions developed in the more recent tests to determine whether rostered persons really are household residents similarly are flawed. In both the 1994 Coverage Test and (less definitely) in the Living Situation Survey, household respondents appear on balance too strict in determining residency, judged in terms of the census rules.

More generally, recent Census Bureau research demonstrates that the task of compiling a roster of household members is more complex than had been assumed. Research has clearly demonstrated that respondents make mistakes that lead to coverage errors. Past and present research shows that certain categories of persons--highly mobile or transient persons, nonrelatives, college students--are particularly at risk of being mistakenly included or excluded on household rosters.

If these errors can be reduced, census and survey coverage can be improved, and the differential undercount of minority populations can be reduced. Because a majority of persons are enumerated by self-response, even a small reduction in response error could have a major impact on coverage. Although the omission rate for mail return households was estimated to be 1.3 percent, this represents close to 2 million census omissions.

Questionnaire design research represents an obvious answer to this problem. In lieu of expensive coverage improvement programs, we need to design a questionnaire that helps respondents provide accurate and complete household rosters. We need to simplify the task and eliminate as much confusion as possible. To do this, it may be necessary to revise the residency rules themselves. Because the rules are numerous, complicated, counterintuitive, and mutually contradictory, we believe they are either ignored, or they make the roster task more difficult for respondents who try to apply them. Either way, they contribute to errors. We believe that now, 174 years after the basic rules were put in place, is the time to examine them critically, with the aim of simplifying them.

The research reviewed in this paper is just the beginning. It is critical that we continue the cycle of analysis, qualitative exploration, and experimentation, that recently has begun to yield both knowledge and practical improvements in roster methods. As we continue to gain insight into the roster process, potential causes of coverage error emerge. We have a clearer picture than we did even a year ago of the sources of

errors in household rosters, and of potential methodological improvements which may reduce or eliminate many of these errors. One issue we have not yet confronted is the mode of data collection and how it affects rostering using this questionnaire. We hope to report back to you in a year additional improvements in that and other aspects of our roster research.

REFERENCES

Bates, N. 1991. "The 1990 Alternative Questionnaire Experiment: Preliminary Report of the 100-percent Items." Preliminary Research and Evaluation Memorandum No. 108. Washington DC: Bureau of the Census.

Bureau of the Census. 1989. 200 Years of U. S. Census-Taking: Population and Housing Questions, 1790-1990. Washington D.C.: Bureau of the Census.

Childers, D. 1993. "The Impact of Housing Unit Coverage on Person Coverage", Internal Memorandum, Washington, DC: Bureau of the Census.

de la Puente, M. 1993. Why Are People Missed or Erroneously Included by the Census: A Summary of Findings from Ethnographic Coverage Reports. Report prepared for the Advisory Committee for the Design of the Year 2000 Census, March 5, 1993. Washington: Bureau of the Census.

Ellis, Y. 1994. "Categorical Data Analysis of Census Omissions", Internal Memorandum. Washington, DC: Bureau of the Census.

Fay, R. E. 1989. "An analysis of within-household undercoverage in the Current Population Survey." Proceedings, Annual Research Conference. Washington: Bureau of the Census.

Gerber. E. 1990. Calculating Residence: A Cognitive Approach to Household Membership Judgments among Low Income Blacks. Washington DC: Center for Survey Methods Research, Bureau of the Census.

_____. 1994. The Language of Residence: Respondent Understandings and Census Rules. Washington, DC: Center for Survey Methods Research, Bureau of the Census.

Gerber, E. and N. Bates. 1994. "Respondents' Understanding of Residence Terminology in Cognitive Research and in the Living Situation Survey." Proceedings, Annual Research Conference. Washington, DC: Bureau of the Census.

Griffin, D. and C. Moriarity. 1992. "Characteristics of Census Errors", American Statistical Association Proceedings 1992. Alexandria, VA: American Statistical Association.

Hogan, H. 1992. "The 1990 Post-Enumeration Survey: Operations and New Estimates" American Statistical Association Proceedings 1992. Alexandria, VA: American Statistical Association.

Kearney, A., R. Tourangeau, G. Shapiro, and L. Ernst. 1993. "Coverage Improvement from Experimental Residence Questions," paper presented at the annual meeting of the American Statistical Association.

Montoya, M. D. 1992. Ethnographic Evaluation of the Behavioral Causes of Undercount: Woodburn, Oregon. Prepared under JSA with University of Oregon. Washington: Bureau of the Census.

Moriarity, C. and Childers, D. 1993. "Analysis of Census Omissions: Preliminary Results", American Statistical Association Proceedings 1993. Alexandria, VA: American Statistical Association.

Pausche, J. 1994. "1994 National Census Test: Quantitative Comparisons of Coverage", Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Romero, M. 1992. Ethnographic Evaluation of Behavioral Causes of Census Undercount of Undocumented Immigrants and Salvadorans in the Mission District of San Francisco, California. Prepared under JSA with San Francisco State University Foundation. Washington: Bureau of the Census.

Schwede, L. 1993. "An Empirical Exploration of Residence Rules: The Living Situation Survey", American Statistical Association Proceedings 1993. Alexandria VA: American Statistical Association.

Sweet, E. M. 1994. "Roster Research Results from the Living Situation Survey." Proceedings, Annual Research Conference. Washington DC: Bureau of the Census.

Sweet, E. M. and Nicolas S. Alberti. 1994. "Coverage Implications of Using the Term 'Usual Residence'." Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Wingerd, J. 1992. Urban Haitians: Documented/Undocumented in a Mixed Neighborhood. Report prepared under JSA with Community Service Council of Broward County, Inc. Washington: Bureau of the Census.

NOTES

[1] This paper reports the results of research undertaken by Census Bureau staff. The views expressed are the authors' and do not necessarily reflect those of the Census Bureau. We thank Nancy Bates, Dave Desjardins, Eleanor Gerber, and Laurie Schwede for helpful comments, and Elizabeth Sweet for assistance with statistical computations.

[2] In this paper, we ignore the group quarters and institutional populations, but ideally the residence rules in a census should assign each person uniquely to a household or another type of place where enumeration is conducted.