HOUSEHOLD ATTACHMENT AND SURVEY COVERAGE

Elizabeth Martin, Bureau of the Census
Elizabeth Martin, Bureau of the Census, Washington, D. C., 20233

Keywords: roster, decennial census, questionnaire design

Introduction

Errors made by respondents or interviewers in listing persons on household rosters are an important source of coverage errors in censuses and surveys. Within-household omissions account for about one-third of all census omissions, and are higher for males and minorities, and nonrelatives within households (Hogan, 1992; Ellis, 1994; Fay, 1989). Despite the evidence, the household roster has not been approached systematically as a survey measurement problem. Most surveys lack standardized questions and procedures to help interviewers decide whether to list persons whose residence is ambiguous, leaving these determinations to the interviewer's discretion and skill.

Research suggests several reasons why respondents may erroneously omit persons from household rosters. Persons may be concealed due to concerns about how the information is used by Government or others (Hainer et al., 1988; de la Puente, 1993; Tourangeau et al., forthcoming). Complicated living situations, transiency, and tenuous attachments to households make it difficult to determine who should be counted as a household member. Mobility among multiple households contributes to residential ambiguity (Bates and Gerber, 1994). In ambiguous situations, respondents' judgments are influenced by intentions and agreements, financial contributions and permanence of attachment, and other criteria which may conflict with official residency rules (Gerber, 1990; 1994). Arcane terminology and counterintuitive instructions may confuse or mislead respondents (Gerber, 1994; Gerber, Wellens, and Keeley, 1996). Household respondents may lack information about persons in their household, and may assume that part-time residents have a home elsewhere, when they don't. There may be disagreements within households about who belongs there and who doesn't (Hainer, 1987).

Different understandings of roster errors may well imply different strategies for solving them. This paper describes an attempt to improve coverage of tenuously attached persons by expanding the roster questions and probes. The strategy is to cast a broad net in order to identify persons with any attachment to a household, no matter how weak or tenuous. Screening questions are asked to weed out nonresidents, and determine where tenuously attached persons should be counted. This design strategy offers the potential advantage of capturing information about persons in the gray area, who might otherwise be missed entirely, and permits the analyst to evaluate alternative residence criteria and population estimation strategies.

An experimental roster design was implemented in a special pilot survey, the Living Situation Survey. I first briefly describe the survey, then explore two alternative measures of household attachment, and offer some preliminary ideas about how they might be used in developing population estimates.

Background

The Living Situation Survey was designed by Census Bureau researchers and conducted by RTI in 1993. Extensive cues and probes were used to build rosters that included all persons with any attachment to the sample households, including (for example) persons who spent a night in the housing unit during the 2 month reference period, who received mail or messages there, had a key, contributed money for rent or bills, and so on. Cues also targeted undercounted categories, such as live-in employees, boarders, foster children, etc. The probes were developed based on evidence about undercounts, as well as cognitive and anthropological research on how people think about residency issues.

Interviews were conducted in 999 households (representing a 79.5 percent response rate) oversampled from areas with high concentrations of minorities and renters. A total of 3,549 people were listed on household rosters. The weighted mean number of persons listed per housing unit in the LSS (3.62) is significantly greater than the mean of 2.63 persons per occupied housing unit in the 1990 census. The added probes in the LSS were especially effective at identifying more young minority males, who were less likely to be mentioned in response to more standard probes (Sweet, 1994).

The LSS also asked the household respondent questions about the residency status of all persons listed on the roster. The questions and distribution of responses to them are shown in Table 1. (Except where noted, all results are weighted to total national households and to account for oversampling and adjust for nonresponse.) Item 1, usual residence, is the official basis for allocating

individuals to households in the census, so is of special interest. At the aggregate level, the responses to these questions appear very consistent: about three-quarters of the persons rostered using the new, inclusive procedure were residents of the sample households, and one-quarter lived somewhere else.

Table 1. Measures of Household Attachment

1. "Do you consider this address to be your/ NAME's usual residence, that is the place where you/NAME live(s) and sleep(s) most of the time?"

Yes	76.3%
No	23.7
Total	100.0

2. "Do you/Does NAME have a usual residence somewhere else?"

Yes	25.6%
No	74.4
Total	100.0

3. "Do you consider yourself/NAME to be a member of this household?"

Yes	78.9%
No	21.1
Total	100.0

4. "Please think about all the time you/NAME actually spent here since (DATE). Were you/Was NAME living here, staying here, visiting here or something else?" *Not asked about "casual visitors" who had a usual residence elsewhere, and spent 7 or fewer nights in the sample household during the reference period.*

Living here	76.3%
Staying here	1.2
Visiting	1.9
Not asked *(casual visitor)*	20.5
Total	100.0

The mean number of usual residents (identified using item 1) per housing unit in the LSS is higher than the census for all race/ethnicity categories, but is significantly higher only for the total population (2.76) and for Hispanics (Sweet, 1994).

For a small but important group of marginal residents, household respondents' reports were often inconsistent with census rules and with reports of the individuals themselves. Nine percent of the persons rostered in the LSS (excluding casual visitors) had complex living situations, and household respondents' determinations of "usual residence" agreed with census rules for only 69 percent of them (Sweet and Alberti, 1994). The LSS followed up a sample of non-casual visitors rostered in the survey, and conducted individual interviews with them (or with proxies reporting for them). Sweet and Alberti (1994) find that in 95 percent of cases, the household respondent and the individual agree on the individual's usual residence (proxy reports for the individual were excluded from their analysis). The 5 percent who disagreed tended to have complex living situations. Potential omissions due to inconsistent assessments of household membership were significantly higher for young, minority males compared to other groups (Schwede and Ellis, 1994).

To date, research based on the Living Situation Survey points to several conclusions. First, the expanded probing resulted in larger numbers of people listed on household rosters, with evidence of increases in undercounted categories (Hispanics, as well as young, minority males). Compared to the census, there was a 38 percent increase in the number of people rostered per household, but only a 5 percent increase in the number of usual residents per household. Second, household respondent reports of who lives in a household should not be taken as unproblematic. Third, people use different criteria and in many cases make different residency determinations than would be implied by the census residency rules. Fourth, living situations which are ambiguous and fluid are particularly vulnerable to misreporting and unreliable reporting.

Alternative Measurement Strategies

Two-item Measure of Residence Status: In her evaluation of coverage gains achieved using more inclusive roster probes, Sweet (1994) used a single measure (item 1--usual residence in the sample household) to screen for rostered persons who should be included. A refinement of her procedure might include an additional measure of residence: usual residence elsewhere (item 2). The weighted joint distribution of these two variables is in Table 2.

Table 2. Weighted Joint Distribution of Two Measures of Usual Residence, for all Rostered Persons (in millions)

		Do you/NAME have a usual residence somewhere else?		
		Yes	No	Total
Do you consider this address to be your/ NAME's usual residence?	Yes	*a* 3.8 1.2%	*b* 249.4 75.1%	253.2 76.3%
	No	*c* 76.8 23.1%	*d* 2.1 .6%	78.8 23.7%
	Total	80.6 24.3%	251.4 75.7%	332.0 100.0

Variances have not been calculated but are undoubtedly large, so these figures should be taken as illustrative. However, they are interesting. Note that the total of 332 million grossly overestimates the size of the U. S. population. As noted above, the expanded probes resulted in many more persons being listed on household rosters, including many persons who did not reside in the sample units, so we expect a substantial overestimate of the total population if we include all rostered persons. Second, note that the cell a estimate of 3.8 million people with two usual residences, and the cell d estimate of 2.1 million with no usual residence, imply that the common assumption that each person has one and only one usual residence failed for an estimated 1.8 percent of the persons rostered in the LSS.

The figure of 2.1 million represents an estimate of one component of the national homeless population--persons with no usual residence, but with some attachment to a household. This figure probably includes some persons who are in transit or moving between one household and another and are not really homeless. Further research and analysis would be needed to evaluate the composition of this group. Nevertheless, this methodology represents a sample-based method which may yield improvements over many current methods for estimating the homeless population or its components, many of which are not based on probability samples.

It is possible to estimate the total U. S. population by weighting each cell to reflect eligibility for census enumeration in the sample household. Members of cell a are at risk of double-counting because they are reported as having two usual residences, and therefore should be weighted by .5 to reflect their potential enumeration at both places. Cell b is weighted 1.0 since it represents residents of sample households who have no other usual residence. Cell c is weighted 0; these persons are not residents of the sample households and would be enumerated at their usual residence elsewhere. Finally, cell d is weighted 1; census rules specify that persons with no usual residence are eligible for enumeration in the households where they are found. These weights yield an estimate of the total U. S. population of 253,339,000. Demographic analysis is often taken as the "gold standard" for census coverage; Table 3 presents this and other benchmarks for comparison.

Table 3. Some Benchmarks

Method	Total Population	Accuracy	
		% of DA	Sex Ratio
Demog. analysis[a]	253,394,000	100.00	96.9
1990 Census[b]	248,710,000	98.15	95.1
PES[c]	253,978,000	100.23	95.8
LSS	253,339,000	99.98	94.3

[a] Robinson et al. (1993).
[b] Robinson et al. (1993).
[c] Post Enumeration Survey; Bounpane (1991).

The estimated population total based on LSS is closer to the demographic analysis estimate of total population than either the Post Enumeration Survey or the 1990 census. The sex ratio is worse than the other estimates.

Three Item Measure of Household Attachment

A further refinement would replace the two-item measure with a multi-item scale using items 1, 3, and 4 (see Table 1), interpreting them as indirect measures of a latent variable, attachment to the sample household. For item 4, "living" and "staying" are combined, and casual visitors (who were not asked this question) are classified as "visiting." We apply the Rasch measurement model (Rasch, 1960/1980), which posits that the underlying trait entirely accounts for associations between responses to the 3 scale items. The model is tested by applying RASCHPLX (Fay and Turner, 1989) to the cross-

classification of the variables using unweighted data.[1]

Table 4. Observed and Fitted Frequencies for 3 Measures of Household Attachment

Response to Item				Model	
1	3	4	Obs.	2-factor	Rasch
Y	Y	L/S	2634	2628.89	2634.00
N	Y	L/S	46	51.11	47.23
Y	N	L/S	12	17.11	13.58
N	N	L/S	46	40.89	43.19
Y	Y	V	14	19.11	11.19
N	Y	V	34	28.89	35.58
Y	N	V	9	3.89	10.23
N	N	V	719	724.11	719.00
d.f.				1	2
X^2				11.68	1.32
p				.006	>.5

Table 4 compares the observed frequencies for the cross-classification of the 3 household attachment measures with expected frequencies under the Rasch model and under an alternative log-linear model which assumes that each pair of variables is associated. The Rasch model fits very well ($p>.5$) as can be seen by the close correspondence between the fitted and observed frequencies. The two-factor model is rejected ($p<.006$).

The close fit of the Rasch model implies that these data are consistent with the assumption that there is an underlying latent dimension of "household attachment," with the Rasch parameters representing heterogeneity among individuals in strength of household attachment. Once variation among individuals in latent attachment is explicitly accounted for by including the Rasch parameters in the model, the associations between usual residence, household membership, and "live/stay" versus "visit" vanish.

As a stronger test (see Duncan, 1984), the Rasch model was fitted to the same cross classification, stratified by relationship to household respondent (self, related, nonrelated), a variable which is strongly associated with household attachment. The Rasch model still fits well, as it does when an alternative stratifying variable, sample domain, is used (results not shown).

The close fit of the Rasch model supports an interpretation of these 3 items as a scale of household attachment. The number of "attached" responses given by respondents is summed, with values ranging from 0, the weakest level of attachment, to 3, corresponding to "attached" answers to all 3 questions.

Table 5 shows that persons with stronger attachments to sample households are mentioned earlier in the roster process than those with weaker attachments. It took only 1.06 roster probes, on average, to elicit names of persons in level 3, but an average of 4.59 cues were needed before persons at level 0 were mentioned. One may infer that with less probing, many persons at the weaker levels attachments would not have been mentioned, at all. Weighted counts in the second row show that the vast majority of persons are in the most attached category, and are mentioned with little probing. The estimated number in this category, 248.8 million, is very close to the number of persons in the 1990 census. This correspondence lends credence to the supposition that traditional roster methods tend to identify only persons with strong attachments to households. The next three rows show some characteristics of persons at each attachment level. As attachment weakens, the percent of persons who are nonrelatives increases monotonically. Black and Hispanic males aged 18-29, historically hard-to-enumerate, are also overrepresented at weaker levels of attachment, although the pattern is not monotonic. Unrelated persons, and young minority males, may be missed in censuses and surveys partly because they tend to have weaker attachments to households (as hypothesized by Fay, 1989) and are unlikely to be mentioned unless more extensive probing is used. Persons classified as "homeless" because they are said not to be usual residents of the sample unit or any other place make up a disproportionate share of level 2.

[1]The use of unweighted data is justified by the fact that, when the Rasch model fits, the results are independent of the sample of persons (Loevinger, 1965; cited by Duncan, 1984). Duncan (1984:216) expands on this point: "*any* collection of respondents from a population in which the model holds--provided only that it includes respondents giving [all of] the relevant patterns and is not selected on the basis of the responses as such--constitutes a sample for the purpose of estimating [model parameters]. The model is 'sample-free' in this specific sense."

Table 5. Mean Number of Roster Probes, by Level of Household Attachment

	Attachment to this Household			
	3 Strong	2	1	0 Weak/ absent
Mean no. probes to elicit name	1.06	1.22	3.48	4.59
Weighted N (millions)	248.8	8.4	6.9	65.5
1990 Census	248.7			
% nonrel.	2.2	7.8	8.9	21.4
% Black, Hisp. male 18-29	2.3	6.1	1.4	3.3
% homeless	0	22.9	.8	.2
% URE	1.2	36.2	99.1	99.8

The bottom row shows the percent of persons at each level reported to have a usual residence elsewhere (URE). Individuals at levels 0 and 1 are almost uniformly reported to have usual residences elsewhere, while persons at level 3 are nearly uniformly said not to have a URE. Level 2 is more mixed. (Despite the high correlation, URE does not form a scale with the 3 measures of household attachment; fitting a Rasch model to all 4 measures yields a poor fit. Belonging to another household intuitively seems a dimension separate from attachment to this household.)

The consistency between household attachment and the URE measure, and the near homogeneity of categories 0, 1, and 3, suggest the use of household attachment as a criterion for allocating persons to households. A multi-item scale should be more reliable than a determination based on any single item. For example, one might exclude categories 0 and 1 as nonresidents, and include category 3 as residents; category 2 would be decided using on other information, or perhaps a weighted fraction of this category would be included.

Finally, it is of interest to consider the relationship between household attachment and individuals' own claims of residence in the sample household, presented in Table 6. A sample of LSS respondents was followed up for individual interviews, and a subset of these are included in this table. Persons who are clearly residents or clearly nonresidents are excluded (see note b). In other words, this table represents the gray area, of persons whose residence is somewhat uncertain.

Table 6. Percent of Individuals[a] Reporting Usual Residence in Sample Households, by Level of Attachment, for a Subset of Respondents[b]

Attachment to this Household			
3 Strong	2	1	0 Weak/ absent
97.6	90.1	42.4	8.0

[a]Both proxy and self reporters for the individual interview are included.

[b]Excluded are: core household members, who spent 8 or more nights in the unit during the ref. period and are reported as living there, and casual visitors, who spent 0-7 nights in the sample unit and are reported to have a URE.

As would be expected, the fraction of individuals claiming usual residence increases dramatically as household attachment grows stronger. However, individual interviews indicate many more claims of residence in the sample units than would be expected at marginal levels of attachment. Over 40 percent of individuals at level 1, and 8 percent at level 0, claim usual residence in the sample households. These results call into question the information about attachment provided by household respondents for levels 0 and 1, or else suggest that individuals (or their proxies) exaggerate their claims to residence in the sample households. Possibly, household respondent reports are not trustworthy for marginally attached persons.

Conclusions

Censuses and surveys almost never question the reliability of household respondent reports of the residence status of individuals in their households. Results reported here suggest that, for persons with tenuous or ambiguous household attachments, they may not be reliable. Further exploration of the relation between household respondent and self reports of residence status is called for.

The measures of household attachment developed for the LSS may be the basis for alternative strategies for

estimating population and/or determining where individuals should be counted in a more discriminating fashion than has been possible to date. Two items, measuring usual residence in the sample unit and elsewhere, might be explored more rigorously as the basis for population estimates which explicitly account for the possibility that persons may have multiple usual residences, or none. Three items have the interesting property that their cross-classification may be fit by the Rasch measurement model, supporting their interpretation as a scale measuring an underlying dimension of "household attachment." The fit of this model suggests a different way of thinking about residence--as a continuum, and a matter of degree rather than as an all-or-none proposition. For a segment of the population, residence determinations appear more complex and difficult than survey practitioners have tended to assume. The quality of measurement may be improved by using multi-item scales that explicitly allow for persons in the gray area, whose residence is ambiguous. Indeed one may question whether any single item measure--such as usual residence--could be said to represent measurement at all. It is a slim reed on which to rest determinations as critical as where persons should be enumerated in the decennial census or household surveys.

Note

Results in this paper are attributed to the author and do not necessarily represent views of the Census Bureau. Jeff Moore, Eric Schindler, Gary Shapiro, and Elizabeth Sweet provided helpful comments and corrections.

References

Bates, N. A., and Gerber, E., (1994), "A Typology of Temporary Mobility Patterns and Reporting of Usual Residence," Proc. of the Surv. Res. Meth. Sec., Am. Stat. Assoc.

Bounpane, P. A., (1991), Census Bureau Releases Refined Estimates from Post Enumeration Survey of 1990 Census Coverage. June 13, 1991, press release.

de la Puente, M. (1993) "Why Are People Missed or Erroneously Included by the Census: A Summary of Findings From Ethnographic Coverage Reports." Proceedings of the Conference on Undercounted Ethnic Populations. Washington, DC: Department of Commerce.

Duncan, O. D., (1984), "Measurement and Structure," in Turner, C., and Martin, E. (eds.), Surveying Subjective Phenomena, Vol. 1. New York: Russell Sage.

Ellis, Y., (1994), "Categorical Data Analysis of Census Omissions,"

Fay, R. E., (1989), "An Analysis of Within-Household Undercoverage in the Current Population Survey," Proceedings, Annual Research Conference. Washington, DC: Bureau of the Census.

Fay, R. E. and Turner, C. F., (1989) RASCHPLX: Rasch Analysis for Contingency Tables with Complex Sample Designs, unpublished, Washington DC: National Academy of Sciences.

Gerber, E. R., (1990), Calculating Residence: A Cognitive Approach to Household Membership Among Low Income Blacks. Unpublished report, Bureau of the Census.

Gerber, E. (1994) The Language of Residence: Respondent Understandings and Census Rules. Unpublished report, Bureau of the Census.

Gerber, E.R., and Bates, N.A., (1994), "Respondents' Understanding of Residence Terminology in Cognitive Research and in the Living Situation Survey," Proceedings, Annual Research Conference, Bureau of the Census.

Gerber, E. R., Wellens, T., and Keeley, C., (1996), "'Who Lives Here?: The Use of Vignettes in Household Roster Research," Proceedings, Am. Assoc. for Public Opinion Research.

Hainer, P., (1987),A Brief and Qualitative Study Exploring the Reasons for Census coverage Error among Low Income Black Households. Unpublished report, April 1987.

Hogan, H., (1992), "The 1990 Post-Enumeration Survey: Operations and New Estimates," Proc. of the Surv. Res. Meth. Sec., Am. Stat. Assoc.

Loevinger, J., (1965), "Person and Population as Psychometric Concepts," Psychological Review 72:143-155.

Rasch, G., (1960/1980), Probabilistic Models for Some Intelligence and Attainment Tests. Chicago IL: University of Chicago Press.

Robinson, J. G., Ahmed, B., Das Gupta, P., and Woodrow, K., (1993), "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis," Journal of the American Statistical Association 88: 1061-1079.

Schwede, L. and Ellis, Y., (1994), "Exploring Associations between Subjective and Objective Assessments of Household Membership: The Living Situation Survey." Proc. of the Surv. Res. Meth. Sec., Am. Stat. Assoc.

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

Sweet, E. M., and Alberti, N. S., (1994), "Coverage Implications of Using the Term 'Usual Residence,'" DSSD 2000 Census Memorandum Series #H-17. Bureau of the Census.

Tourangeau, R., Shapiro, G., Kearney, A., and Ernst, L., (forthcoming), "Who Lives Here? Survey Undercoverage and Household Roster Questions," Journal of Official Statistics.