QUALITY PROFILE FOR THE AMERICAN HOUSING SURVEY

Rameswar P. Chakrabarty
Bureau of the Census, Washington, DC 20233 *

1. INTRODUCTION

The main objective of The American Housing Survey (AHS) is to provide important information on housing conditions, size and composition of the housing stock, and the characteristics of its occupants to support policy and program decisions by the Department of Housing and Urban Development (HUD).

In this paper we evaluate the quality of AHS data. We briefly describe potential sources of error in AHS data and quality control procedures used in the operation of the survey, and document what is currently known about the magnitude of errors in AHS data. The discussion primarily focuses on nonsampling errors; information on sampling errors are provided in published Current Housing Reports. This paper summarizes a comprehensive document (Chakrabarty, 1992c) on the potential sources of error in the AHS. The document unifies and summarizes available information on data quality from many reports and memoranda developed over the years.

2. DESCRIPTION OF THE SURVEY

The AHS, conducted for the HUD by the Bureau of the Census, is actually two separate data collection efforts. One is a national sample and the other a metropolitan sample (MS). The AHS-National is a biennial survey of occupied and vacant housing units in the United States. The AHS-MS is a quadrennial survey of 44 large metro areas, 11 per year.

2.1 Sample Design, Frames and Undercoverage

The AHS is a stratified multi-stage probability sample of housing units. AHS-National's current sample design is based on the 1980 Census.

The selection of housing units (HUs) within primary sampling units (PSUs) requires five separate non-overlapping sampling frames: (1) address enumeration districts (EDs), (2) area EDs, (3) special places, (4) new construction and (5) coverage improvement. Frame development and sample selection within sample PSUs involve a complex system of automated and manual operations. For the area ED frame, a field operation -- listing of addresses in sample blocks -- is also necessary. All these operations are subject to errors.

- Units constructed without permits in permit-issuing areas may be missed.
- If a permit is issued for a new structure at an existing address, that address may receive a duplicate chance of selection.
- Adequate coverage of mobile homes presents a variety of problems.

The magnitude of these coverage problems is not generally known, but is believed to be small in relation to the universe.

Schwanz (1988a) estimated that undercoverage of mobile homes constructed after 1980 was close to 25 percent in the 1985 AHS-National. Coverage of new mobile home parks in address EDs was very poor.

3. NONRESPONSE ERROR

3.1 Unable-to-Locate Units

The units that cannot be located by interviewers are recorded as unable-to-locate (UTL) units. The UTL rates were less than 0.5% in address segments and exceeded 2% only in area segments in rural areas in the 1985 AHS-National (Schwanz 1988b).

3.2 Noninterviews

The Type A noninterview rates were 4.2%, 3.2%, and 4.2% for the AHS-National in 1985, 1987, and 1989 (Chakrabarty 1992b).

3.3 Item Nonresponse

Item nonresponse rates vary widely from item to item. Weidman (1988) estimated nonresponse rates for a set of 43 items for the 1985 AHS-National. Five items out of 43 had nonresponse rates greater than 10% and 12 had rates greater than 3%. If the questionnaire as a whole meets the minimum requirements for a completed interview, missing data for select items are estimated by imputation. The imputed values are, at best, probable in nature, and thus, subject to error, so potential biases from item nonresponse cannot be completely eliminated.

4. MEASUREMENT ERRORS

4.1 Interview Mode

Telephone interviewing from an interviewer's home has become an acceptable alternative to personal interviewing in the AHS-National as a result of the telephone experiments conducted in 1981 and 1983. In the 1987 AHS-National, one third of the sample was assigned to computer-assisted telephone interviewing (CATI) at a centralized facility and the remaining two thirds continued to be interviewed in the field. (See Leadbetter, et al 1991).

(1) The results indicate that CATI respondents report higher incomes than non-CATI respondents and lower levels of general deficiency items. It is not known which treatment provides better data overall, but it is speculated that CATI income estimates are probably better than non-CATI. This speculation about income estimates is based on the assumption that with computer's assistance CATI tends to ensure that all appropriate income questions are asked.

(2) It is also speculated that the non-CATI

estimates for general deficiency items are probably better than the CATI estimates because it seems unlikely that respondents would over-report items which may have a negative connotation (e.g., water leakage).

4.2 Response Errors

Reinterview data can be used to obtain a statistical measure of discrepancies in responses called 'index of inconsistency'. A summary of such indices computed from reinterview data from 1973 through 1985 has been compiled by Chakrabarty (1992a). Opinion questions like adequacy or inadequacy of recreation facilities, and items that are not easy to remember like the number of electrical blown-outs in the last 90 days, have a high level of inconsistency.

4.3 Response Error in Year Built Data

Stating the year in which the structure was built has always been a problem for respondents in the AHS and other surveys, e.g., CPS, and in the Census as well, particularly when he/she is not the first owner of the housing unit or when the respondent is renting rather than buying.

A content reinterview for the 1980 Census showed that the year built data have considerable response variance and bias (over-reporting and under-reporting). The multiunit structure data displayed higher response variability and bias than the single unit data. Also, the response variability in the year built data in the 1980 Census was at about the same level as in the 1970 Census (see, Bureau of the Census, 1986). Similar reinterview data from the AHS (National or MS) is not available.

The "year built" items was one of two items selected for a record check in the "Tampa AHS Census Match Study" (Tippett, 1988). The overall agreement of responses with the assessor's file was about the same for both Census and AHS respondents. Owners naturally had better information on when the unit was built compared to renters in the Census. The high (14.5%) nonresponse rate for renters in this study for AHS might have biased the result. In

any case, the differences between owners and renters based on a small sample were not statistically significant.

Young (1982) compares year built data for all housing units in the 1980 Census and AHS. Several discrepancies exist between AHS and Census estimates. A difference of 2.7 million units for the 1970-80 cohort is most striking. Young states that, "there are several possible reasons for the 1970-1980 cohort difference of 2.7 million units:

A potential response error problem in the census. We know from past experience (1970 Census evaluation program) that this is a problem.

An excessive number of erroneous inclusions in the census, e.g., duplicates, erroneous enumerations, etc. that were built during the period 1970-1980.

Serious undercoverage problems in the AHS of units built during the period 1970-1980."

4.4 Problems with the Number of Units in Structure Question

The number of units in a structure is a basic housing characteristic. A respondent is asked how many units there are in the structure in which his/her unit resides. A distinction is made between a housing unit, e.g., an apartment, townhouse, condominium, and the structure in which the unit is contained. The structure or building may consist of one or many units. Furthermore, single unit structures are classified as either detached or attached to other structures. This question seems to give respondents a conceptual problem, especially in classifying townhouses, duplexes, and small attached units and in making a distinction between a housing unit and a structure.

Tauber, et al (1983) compared 1980 Census estimates of the totals of the "units in structure" categories with AHS estimates. The differences, except the totals, are greater than those expected from sampling error. Since the census was taken as of April 1, 1980 and the AHS date was

October 1980, the total estimate of housing units is expected to be 800,000 to 1,000,000 units higher in the AHS than in the census due to new construction. This is not the case however; the increase was only 335,000 units. The most notable difference existed in the "5 or more units" category.

Young (1982), who also examines the problem, states that the possible reasons for this discrepancy are:

"Census misclassification error. There has been some concern that census respondents might have incorrectly identified certain types of single (or 2-4 unit structures) as 5+ structures, e.g., attached townhouses or garden apartments.

Serious undercoverage problems may exist in our current surveys for picking up new large multiunit structures."

Finally, we will consider a study described by Abernathy (1987) for the 1987 AHS-MS. The responses from Wave I of the Regional Office Pre-edit were compared to the responses from the last enumeration period for AHS. This is part of the continuing quality control program which checks for and corrects inconsistencies. When the "units in structure" response is found to be inconsistent with the previous answer, the response is flagged.

The two main types of inconsistencies are as follows: "units that were classified as one attached' one year and in a multiunit structure the other year; and units that were classified as in multiunit structures both years, but the number of units in the structure between survey years was inconsistent." Also, part of the quality control process was not only to detect the types of inconsistencies with the previous year, but also to check the corrected responses with the previous year. In other words, once the correction cycle is run on the data that are flagged as "units in structure inconsistent", the responses are again checked with the entries from the previous enumeration period. At this point it has been determined that the majority of the corrected entries are consistent with the prior

year's entries. Abernathy concludes, "it appears that the pre-edit research is doing its job in reducing the classification problems that exist with the current year's data."

4.5 Problems with the Tenure Question

Tenure is important as a basic housing characteristic. The tenure question asks the respondent if he/she owns the unit, rents for cash, or occupies without payment of cash rent. The tenure question presents few conceptual problems for respondents, but the owner occupancy rates are persistently higher in surveys than in the census. This fact is documented by Tauber, Thompson, and Young (1983).

In the Tampa AHS Census Match Study (Tippett, 1988) the occupancy rate for owners in the AHS was 45% compared to 42% in the test census. Out of the 324 respondents who replied to both the test census and the AHS, 304 agreed and 20 gave conflicting responses. Thirteen of those twenty responses were reconciled. During the reconciliation reasons for the discrepancies were discovered and listed in the report as follows: "for two cases, a change of tenure had occurred, so both were correctly enumerated; others resulted from mismarking of the item, different respondents, or a temporary interruption in the rent." These incidental discrepancies are not indicative of any problem that is inherent in the tenure question, and they do not help to explain the problem of the differences in the owner occupancy rates between the census and the AHS.

As an additional note, once the results have been reconciled the tenure item has an L-fold index of inconsistency in the low range, 11.08. This indicates that the respondents are answering the tenure question reasonably well.

4.6 Verification of Reporting of Cooperatives and Condominiums

To evaluate the accuracy of the classification of housing units as cooperatives and condominiums in the AHS-National, part of the reinterview program for 1979 and 1983 focused on verifying responses to the AHS questions on cooperative and condominium status.

The verification followup showed that of the 1,634 units originally reported as condominiums, 62 (3.8%) were not condominium. And out of 196 units reported as cooperative in the original interview 19 (9.7%) were verified to be not cooperatives (Hartnett, 1985).

5. COMPARISON OF AHS WITH OTHER DATA

AHS data have been compared with census data to find differences in year built, units in structure and tenure items in section 4. In this section we provide comparisons of AHS utility costs with data from the Residential Energy Consumption Survey (RECS) and income data with independent estimates.

5.1 Comparison of AHS Utility Costs with RECS

RECS, conducted by the Department of Energy collects utility costs data from utility company records. RECS data are, therefore, more accurate than AHS data provided by household respondents. A comparison of AHS utility costs with RECS data is provided in the code book for AHS (HUD and Bureau of the Census, 1990). The results clearly show that AHS reports higher utility costs than the Residential Energy Consumption Survey. The discrepancy is fairly consistent over time, and also consistent for single-family detached homes. A plausible reason for the higher AHS figures is that households are more concerned about and, therefore, over-emphasize high-cost months when they mentally average their bills for the AHS interviewer.

The estimation of utility costs for AHS-National by regression using monthly utility cost data from the RECS public use file and some common RECS/AHS housing characteristics as independent variables was researched by Sliwa (88a, 88b). Sliwa (1989) provided specifications for deriving annual costs for electricity and natural gas. This method is now used to

improve utility cost estimates for AHS.

5.2 Comparison of AHS Income with Independent Estimates

It is well-known that income statistics derived from household surveys are generally biased due to response errors. Respondents tend to underestimate income. A comparison of AHS income data with independent estimates of income (from GNP accounts, the Social Security Administration, the Veterans Administration, etc.) and CPS is provided in HUD and Bureau of the Census (1990). The results show that the AHS estimates are lower than the independent estimates for total income and for every category other than self-employment income. The CPS estimate is also low but comes closer to the independent estimate. This may be largely due to the differences in income questionnaires and timings of CPS and AHS. More detailed and extensive questions about income sources and amount by source are asked in the CPS than AHS. Also, the CPS March supplement for income coincides with the income tax time when respondents are more aware of nonwage incomes like interest, dividends, etc.

Recently, Williams (1992) provided an extensive comparison of the data on income that were collected in the 1989 AHS-National and the March 1990 CPS. This analysis at least partially supports the hypothesis that the AHS income estimates are lower than CPS largely due to the less detailed AHS income questions.

10. SUMMARY

In this paper we briefly describe potential sources of error in AHS data and quality control procedures used in the operation of the survey, and document what is currently known about the magnitude of errors in AHS data. Coverage error, noninterview and item nonresponse rates, and measurement errors in housing characteristics, shelter cost and income data are discussed. Results of experiments with telephone interviewing are also provided. Data discrepancies between the AHS and the decennial census and/or other surveys seem to exist in some housing characteristics such as year built, number of units-in-structure and tenure, largely due to nonsampling errors.

ACKNOWLEDGEMENTS

REFERENCES

Abernathy, J.F. (1987), "1987 AHS-MS - Results of Research on Wave 1 Pre-edit Rejects for Units in Structure Inconsistent", Internal Census Bureau memorandum to Cannon, November 3.

Bureau of the Census, (1986), "Content Reinterview Study: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview", 1980 Census of Population and Housing, Research and Evaluation Report PHC80-E2, U.S. Government Printing Office, Washington, D.C.

Chakrabarty, R.P. (1992a), "Summary of Nonsampling Errors in the AHS through Reinterview Studies from 1973 through 1985", Internal Census Bureau memorandum for the record, July 15.

_____ (1992b), "A quality profile: American Housing Survey - National Sample" Internal Census Bureau document, December 18.

Hartnett, B. (1985), "Results of the Verification of Cooperative and Condominium Followup for the Annual Housing Survey - 1983 National Sample", Internal Census Bureau memorandum to Montfort, March 22.

HUD and Bureau of the Census (1990). "Codebook for the American Housing Survey, Data Base: 1973-1993", U.S. Government Printing Office, Washington, D.C.

Leadbetter, S.D., Cole, C.G., Meier, A.J., and Huang, H. (1991), "The Impact of Computer Assisted Telephone Interviewing on Data from the American Housing Survey", presented at the Joint Statistical meeting in Atlanta, GA, August 19.

Schwanz, D.J. (1988a), "Mobile Home New Construction for 1985 AHS-National", Internal Census Bureau memorandum to Montfort, February 3.

_____ (1988b), "1985 Type-A Unable-to-Locate Rates for the AHS-National Sample Units", Internal Census Bureau memorandum for the distribution list, November 14.

Silwa, G.E. (1988a), "Regression Analysis of Utility Costs Using the 1984 Residential Energy Consumption Survey (RECS) Public Use File", Internal Census Bureau memorandum for the distribution list, October 12.

_____ (1988b), "Specialized Regression Analysis of Utility Costs Using the 1984 Residential Energy Consumption Survey (RECS)", Internal Census Bureau memorandum for the distribution list, November 4.

_____ (1989), "Specifications for Deriving Annual Costs for Electricity and Natural Gas in the American Housing Survey (AHS)", Internal Census Bureau memorandum to Bartlett, September 29.

Tauber, C., Thompson, J., and Young, A.F. (1983), "1980 Census Data: The Quality of the Data and Some Anomalies", Internal Census Bureau memorandum.

Tippett, J. (1988), "Tampa AHS-Census Match Study", Preliminary Research and Evaluation memorandum, Internal Census Bureau memorandum to Norry, January 13.

Weidman, L. (1988), "Nonresponse Rates for the American Housing Survey", Internal Census Bureau memorandum to Schwanz, April 7.

Williams, B.T. (1992), "Comparison of 1989 AHS and CPS Income Reporting", Internal Census Bureau memorandum for the record, May 18.

Young, A.F. (1982), "Comparison of 1980 Census Provisional Estimates with AHS Estimates for Selected Characteristics", Internal Census Bureau memorandum to Bailar, June 18.