

American Housing Survey: Data Users' Frequently Asked Questions
Maintained by David A. Vandembroucke, U.S. Dept. HUD
Last Update: September 9, 2008

Contact Person:

David A. Vandembroucke
U.S. Department of Housing and Urban Development
451 7th Street SW, Room 8218
Washington, DC 20410

david.a.vandembroucke@hud.gov
Phone: 202-402-5890
Fax: 202-708-3316

Introduction

This is a collection of specific questions asked by data users about the American Housing Survey. As with any such document, it is always a work in progress. Please send your additions or corrections to the contact person listed above.

Many of the items in this FAQ are answers to questions asked on the American Housing Survey mailing list. You can find out how to subscribe to this mailing list by going to the HUD USER web site at <http://www.huduser.org/emaillists/ahslist.html>.

Many of the entries have a date stamp in curly braces, such as {6/6/2008}, which indicates the last time the entry was updated.

At the end of this document is a glossary that lists the descriptive labels of all variables mentioned in this document.

Contributors

Many people have contributed to the information in this document. Prominent among them are Duane McGough, Ron Sepanik, and Carolyn Lynch, of or retired from HUD, Tom Blatt, Paul Harple, Bill Hartnett, Dennis Schwanz, and Barbara Williams of or retired from the Census Bureau, and Greg Watson, formerly of ICF Consulting. David A. Vandembroucke of HUD edited all contributions. He is responsible for any errors. Note that some of the answers below are written in the first person. The "I" or "we" in those may refer to any of the contributors, HUD, the Census Bureau, or the AHS team.

General Questions

Q. How did the AHS get started?

The original motivation for the surveys came from the post-riot commissions in the late 1960s (Kaiser Committee and the Douglas Commission). The organizing meeting for the AHS was held in June 1970 at the Office of Management and Budget in Washington, DC. Participants included representatives of all the major domestic agencies, including all Cabinet Agencies: Labor; Health, Education and Welfare; Agriculture; Commerce (DoC); and of course, Housing and Urban Development (HUD). The proposal for an annual survey of the nation's housing stock, and of at least sixty metropolitan areas, to provide needed housing data on which to base housing and community development policies and programs between the decennial censuses, was a joint proposal from the Housing Division of the Bureau of the Census, DoC, and the Office of Economic and Market Analysis, HUD, each of which had the approval of their respective cabinet secretaries.

Outside, or HUD internal, data collection resources were not seriously considered, because only the Census Bureau had complete access to the confidential records of the decennial censuses, from which the basic national and metropolitan samples were to be selected. In addition, the Bureau controls the confidential records of the building permit, housing starts, and housing completions surveys, which were to be used to update the decennial census sample in permit-issuing areas. Neither other Federal agencies nor outside, private, firms had the national resources nor statistical and data processing capabilities to handle the mass of data expected to be collected.

For years, HUD had engaged the Census Bureau to collect housing production and marketing data to fill gaps in the housing statistical picture. New home sales, begun in 1959 by the Bureau of Labor Statistics for a HUD predecessor agency, HHFA, was transferred to the Census Bureau in 1963 because BLS was unable to carry on the work satisfactorily, and the data series fit nicely with the building permits and housing starts series already carried out by the Census Bureau, but in the Construction Division, not the Housing Division. In 1968, the Housing Completions Survey was added to the New Home Sales Survey funded by HUD and done by Census. In 1969, the Survey of Market Absorption of New Apartments was piggybacked on the existing production and marketing surveys, but was carried out by the Census Housing Division, whetting their appetite for more HUD dollars. In 1972, the Census Construction Division added the Survey of New Mobile Home Placements as a HUD sponsored and funded survey. The AHS, begun in 1973, is the most prominent in a line of HUD-funded and directed surveys conducted by the Census Bureau. Given the firm working relationship between HUD and Census, there was no question but that the AHS would follow that pattern.

The initial organizing meeting at OMB in 1970 unanimously approved the HUD-Census proposal for the AHS, and OMB set about assuring that adequate funds would be provided in the HUD research budget for transfer to Census to implement the survey.

Q. What are the important contact points for AHS information?

Note that the HUD contact is listed at the beginning of this document.

HUD USER

Web site: <http://www.huduser.org/datasets/ahs.html>

Phone: (800) 245 2691

Email: helpdesk@huduser.org

Mail:

P.O. Box 23268

Washington, DC 20026-3268

CENSUS BUREAU

Web site: <http://www.census.gov/hhes/www/ahs.html>

Phone: (301) 763-3235

Email: ahsn@census.gov

Mail:

Bureau of the Census

HHES Division

Washington, DC 20233-8500

Attn: AHS Branch

Q. What is the National Survey?

The National Survey (AHS-N) is conducted in every odd-numbered year. It is a nationally representative survey of the housing stock of the United States. It is a longitudinal survey, with the same housing units being surveyed each time (plus additions for new construction).

Q. What is the Metropolitan Survey?

Note: As of this writing (June 2008), the AHS metropolitan survey schedules are undergoing revision. The description below is based on current plans.

The Metropolitan Survey (AHS-MS, or “metro”) is conducted in odd-numbered years. It cycles through a set of 21 metropolitan areas, surveying each one about once every six years. Like the national survey, the metro survey is longitudinal. However, metro survey samples have been redrawn more often than the national samples, and this reduces the time spans where longitudinality applies.

The metro survey program has changed many times, mostly in response to changes in the AHS budget. The summary above is the current plan, which was implemented in 2007. However, budget cuts forced a change to the 2009 survey, which we hope will be temporary. The 2009 survey includes five metropolitan areas (Chicago, Detroit, New York, Northern New Jersey, and

Philadelphia) as reinforced samples of the national survey, plus one independent metropolitan survey (Seattle).

During the period 1985-2004 the AHS surveyed 41 metropolitan areas. During 1985-1995, there were metro surveys every year, and the time between surveys of the same area was four years. During 1996-2004, the metro surveys were conducted in even numbered years, although there was none in 2000. In addition, during that period the six largest metropolitan areas (the five listed above plus Los Angeles) were surveyed every four years (that is, in 1995, 1999, and 2003) by a special reinforced samples as part of the national survey.

Microdata for the independent metropolitan surveys are published as a separate metropolitan dataset. This dataset includes a few variables that the national dataset does not, most notably ZONE, for submetropolitan areas. Microdata for the metropolitan surveys using the reinforced national samples are included in the national data sets. They do not include the ZONE variable. {6/6/2008}

Q. In what forms are the data published? How can I get them?

The main distinction is between microdata and tabular data.

Microdata include the individual (edited) responses from each interview. The version of the data available to the public is sometimes called the public use file, or PUF. These data are released by the Census Bureau to HUD in SAS datasets. HUD then produces ASCII versions of the datasets, to make them easier to import into other analysis software. Datasets from the 1995 and later surveys are available for download in both formats from the HUD USER web site. All datasets since the beginning of the survey are available on CD-ROM, from HUD USER and the Census Bureau. Datasets beginning with the 1997 survey can also be downloaded using the Census Bureau's FERRET system. Datasets from the 1993 and 1995 national surveys are available through the Census Bureau's Data Extraction System.

Tabular data are formatted tables, aggregated by certain characteristics. They are easier to use but less flexible than the microdata. Tabular data are available in printed reports, downloadable reports, online tables, and CD-ROMs. Printed reports and CD-ROMs can be ordered from HUD USER or the Census Bureau. Online and downloadable tables are available at the Census Bureau web site.

Q. How large are the samples?

Sample sizes depend on the year and the kind of survey (national or metropolitan). National survey samples sizes have been stable, slowly growing over the years. The 2005 survey selected 59,450 units for interview. Of these 2,800 were found to be ineligible, because the unit no longer existed or did not meet the AHS definition of a housing unit. Thus, the sample size was 56,600 units.

The metropolitan survey sample sizes have varied more over the years, in response to budgetary pressures. The 2004 survey selected 68,683 units for interview, of which 2,821 were found to be ineligible, leaving 57,862 units, or about 4,500 per metropolitan area.

Q. How are new units added to the sample?

The Census Bureau samples building permits monthly to select new construction cases for all of the surveys including AHS. There is about a 7-month lag time from when a permit is issued to when the sample is selected and ready to be added to the sample for a survey. For AHS-N, this means we add about 24 months (from 6 months prior to the start of the previous interviewing to 7 months prior to the start of the current interviewing) of new permit sample cases to each AHS-N survey year. For 2011 AHS-N, we will be adding permit sample from permits issued in March 2010 through February 2011.

Q. What are the response rates for the surveys?

In the 2005 national survey, interviews were conducted for approximately 50,500 eligible units, yielding a response rate of about 89 percent.

In the 2004 metropolitan survey, a total of 52,450 interviews were conducted, yielding a response rate of about 91 percent. The current plan is for 3,500 cases per metro area.

Q. What is the smallest unit of geography for the National survey? The Metropolitan survey?

The national survey identifies the four Census regions (REGION). It also identifies metropolitan areas over 100,000 in population (SMSA). Units are also identified as being in central cities, etc. (see the METRO3 variable).

The metropolitan survey identifies metropolitan areas and “zones,” which are groups of Census tracts that comprise at least 100,000 in population and are as homogeneous as we can make them, given the size limit. In most metropolitan areas, zones do not cross central city boundaries. As far as possible, zones do not cross county boundaries or include more than one whole county.

Q. Can you do state estimates?

No. The sample sizes are too small to yield meaningful state estimates.

Q. Which metropolitan areas are identified in the national survey?

See the coding for SMSA in the codebook and the section on specific variables below.

Q. Which areas were covered in each metro survey?

See the list of publications at the Census Bureau web site, <http://www.census.gov/hhes/www/housing/ahs/metropolitandata.html>.

Q. Which Metro Survey panels were used in each survey year?

Download the document “History of AHS-MS Interviewing” from the HUD USER web site, <http://www.huduser.org/datasets/ahs/ahsprev.html>.

Computer Questions

Q. What codebooks do I need?

1. For surveys **up to 1993**, you need the “volume 1 codebook.” This is available in paper from HUD USER or as a scanned PDF document on the HUD USER web site, <http://www.huduser.org/datasets/ahs/ahsprev.html>. Its official title is *Codebook for the American Housing Survey Data Base: 1973 to 1993*.
2. For the **1995 and 1996** surveys, you need the volume 1 codebook and the volume 2 codebook. Volume 2 can be ordered from HUD USER or downloaded in Microsoft Word format from <http://www.huduser.org/datasets/ahs/ahsprev.html>.
3. For the surveys beginning from **1997 onward**, you need the “new codebook,” *Codebook for the American Housing Survey, Public Use File: 1997 and Later*, which can be downloaded in PDF format from <http://www.huduser.org/datasets/ahs/ahsprev.html>.

Q. What software do I need?

AHS datasets contain tens of thousands of records and almost two thousand variables. This is far too much data to fit into a spreadsheet. Most users access AHS data with a statistical analysis package. Neither HUD nor the Census Bureau endorses any specific commercial products for use with the AHS. Statistical packages common among our users include SAS, SPSS, and Stata. A few users employ standard database applications, such as Microsoft Access or Lotus Approach.

Q. How can the SAS datasets be transferred for use with other software?

One program that will do this is called Stat Transfer. There is a web site about it at <http://www.stattransfer.com/>.

Another is called DBMS/Copy. You can read more about it at <http://www.dataflux.com/Product-Services/Products/dbms.asp>

Mention of these programs does not constitute an endorsement by HUD or Census. We have not even tried them ourselves. We would appreciate any comments from AHS users who have used these products, or any other similar products.

Q. How can I use AHS datasets with Microsoft Access?

While we don't have a lot of experience with this, we suggest that you start with the ASCII version of the data files. The files for the newer (1997 and later) surveys are comma-delimited, and Access has a tool to import such files. However, you may encounter difficulties because there are so many variables (also called columns or fields) in the AHS data. Access itself can handle that number of variables, but the import tool chokes at 256 fields. One work-around is

described at <http://www.nber.org/data/access2excel.html> . (Note that this method is not supported by either HUD or the Census Bureau. If you try it, you're on your own.) {3/18/2008}

Q. How do I convert the SAS files for use with SPSS?

The SPSS syntax command line to read the various datasets in the SAS transport file is:

```
get sas data="file name" dset (dataset name)
```

Where dataset names depend on the survey year, as shown in the table below:

<u>1997</u>	<u>1998</u>	<u>1999</u>	<u>2001-2005</u>	<u>2007</u>
houshld	houshld	houshld	newhouse	newhouse
person	person	person	person	person
jtw		jtw	jtw	jtw
homimp		homimp	homimp	homimp
mortg	mortg	mortg	mortg	mortg
owner	owner	owner	owner	owner
rmov	rmov	rmov	rmov	rmov
	ratiov	ratiov	ratiov	
toppuf	toppuf	toppuf		
weight	weight	weight		

Note that if you leave off the dataset name, the command will extract just the first dataset in the transport file, which will probably be houshld or newhouse.

AHS users have also reported that the data conversion program DBMS/COPY successfully converts AHS SAS files into SPSS format—including all value labels and missing value codes. The DBMS/COPY website is at <http://www.dataflux.com/Product-Services/Products/dbms.asp>. Note that this is *not* an endorsement by HUD or the Census Bureau. {6/6/2008}

Q How do I import the comma-delimited files into STATA?

This applies to the ASCII version of the datasets, for survey years 1997 and later. With a comma delimited file, the user should select File>Import>ASCII data created by a spreadsheet. This will generate a dialog box which prompts the user to select the file. At the bottom of this box, the Comma Delimited button should be selected. This should work for Stata versions 8 and 9. This is equivalent to the insheet command.

Q. How do I import the SAS transport files into STATA?

If you have Stata version 8 or later, you can convert the SAS transport files into Stata files. The command is "fdause" and then the path to the file, or you can choose File - Import - SAS XPT from the pull down menus. This is much easier than working with seven ASCII files. The other advantage is it keeps all the variable names and descriptions that come with the SAS download. {3/5/2008}

Q How do I read the data on the pre-1997 CD-ROMs?

The older microdata files are in fixed-column ASCII format. In order to turn them into datasets, you will need to specify the columns of each line that correspond to the AHS variables. You can find the record layouts for the older surveys (through 1993) in the Volume 1 codebook (available in paper from HUD USER, or as a scanned PDF at <http://www.huduser.org/datasets/ahs/ahsprev.html>). The record layouts for the 1995 and 1996 surveys are in the Volume 2 codebook, which can be downloaded from the HUD USER web site. Alternatively, most of the CDs have a record layout file in the document directory. These files go under different names, depending on the CD. The files actually give you information about the distribution of values for each variable, but you do not need all of that just to import the data. All you need to look at is the beginning of the lines, which gives the description, variable name, and starting column number. Since the variables extend until the next variable starts, you can see the column numbers you need to read. For example, here is an edited portion of one such file:

<u>DESCRIPTION</u>	<u>VARIABLE</u>	<u>COLUMN</u>
CONTROL	CONTRO	1
STRATUM NUMB	THIV70	13
1970 SMSA	SMSA	16
1980 SMSA	PMSA	20
SURVEY YEAR	YEAR	24
PANEL	IMONTH	26

This tells you that the variable CONTRO, which is the control number, begins with column 1. The next variable, THIV70, begins in column 13. Thus, CONTRO is in columns 1-12. Similarly, the THIV70 is in 13-15, SMSA is in 16-19, PMSA is in 20-23, and YEAR is in 24-25. IMONTH begins in column 26.

You may want to look at the library of user-written programs, which you can download from the HUD USER web site at <http://www.huduser.org/datasets/ahs/ahsprev.html>. As of this writing, this archive includes programs to convert the files for all the older surveys into SAS datasets, and to convert the metropolitan files from 1984 to 1996 into SPSS. We will update the archive as new contributions come in, and so it is worth checking if you need other years or formats. You may be able to use one of the existing programs as a template to show you how to proceed. Most of the variables are the same. You will have to edit the column numbers to be consistent with the record layout files. It will save you some typing and maybe some head scratching. If you do write new conversion programs, I would very much like to have copies of your code to include in the library.

Q. What are the anomalies on the 1993-1994 Metro Survey CD-ROM?

- 1993: The 1993 data file on the CD-ROM has three sets of ten records that repeat. This is an error whose origin is unknown. Records 51-60, 61-70, and 71-80 are identical. You should delete all but one of these sets.
- 1994: The 1994 data on the CD-ROM comes in eight files, D2120A.DAT, D2120B.DAT, etc., through D2120H.DAT. Each of these files includes the data for a

single metro area. There is one more file, D2120I.DAT, in the 1994 directory. This file is NOT part of the dataset. It has a completely different file layout and includes alphabetic as well as numeric characters. We do not know what this file is for, if anything.

Q. What is the "flat file," and what is the "file flattener program" used for?

Before 1997, AHS datasets were released as a single file. Each record of this file contained all the data for a single housing unit. For those entities (such as persons) which occurred more than once in the same housing unit, this single file included one set of variables for each possible entity. Thus, there were fifteen race variables (RACE1-RACE15), one for each person in the housing unit. Since most housing units did not contain this many persons (and some contained none at all), many of these variables were filled out with missing values. This arrangement of data is referred to as the flat file.

Beginning in 1997, the AHS datasets were reorganized on a relational basis. There is, for example, a separate Person file, containing one record for each person in the survey. These records can be linked to the housing unit file by means of the CONTROL variable. There are similar files for recent mover groups, commuters, and home improvement projects. In addition, the Ratio and mortgage files have only one record per housing unit, but they have records only for those units for which there is relevant data.

Because many long-term AHS users were used to using the flat file format, HUD released a file flattener program to convert the post-1997 set of files into a single flat file. This program has been updated for each survey since, but it is available only for SAS. In order to reduce the need for the file flattener, in 2001 we began adding the householder's characteristics to the housing unit file (now called newhouse). Thus, one can tabulate most housing characteristics by age, race, etc. of householder, using only one file.

Q. How do I weight the data using SAS?

In any SAS procedure step, you simply add the "weight" command and indicate the variable you want to use for weighting. For example, if you want to produce a frequency distribution by TENURE using the standard AHS weight variable (WEIGHT), you would write:

```
proc freq;
  tables TENURE;
  weight WEIGHT;
run;
```

Q. How do I weight the data using SPSS?

Under the "Data" menu, choose "Weight Cases..."; a dialog box appears. Select a numerical variable from the list in the dialog box; click the "Weight cases by" radio button; then click the "OK" button. A "Weight on" indicator will appear in the lower right corner of the SPSS window. Choose "Do not weight cases" to turn the weight off.

These instructions are based on SPSS 12, but they may apply to previous versions. {6/6/2008}

Analysis Questions

Q. How do you handle missing data? Do you impute all or some missing data?

Some variables have values for missing data. In the surveys beginning in 1997, there are separate values for refused, don't know, and blank. Other variables have missing values replaced by imputed values. Each variable with imputed values has a corresponding "allocation variable" which indicates whether it has been edited or imputed. See the appropriate codebook for details.

Q. How do I identify the householder in the person records?

To identify the head of household or the householder, you need to use the REL variable. A REL value of 1 or 2 is the householder.

Note also that the flat version of the AHS data file, created by the file flattener program, organizes the data so that the householder is always the "first" person, the person whose characteristics would be identified with a 1, such as AGE1, SEX1, REL1, etc.

Finally, all of the householder's person-record data are included in the Newhouse file, in variables with the prefix "HH" (such as HHAGE, HHSEX, etc.). Thus, if all you want is to filter or subdivide your data on the basis of the householder's characteristics, you can do that using Newhouse alone, without having to link to Person. {6/6/2008}

Q. How can I tell if the respondent was (or was not) the reference person?

In 1997 and later surveys, to determine if the reference person and the respondent are the same:

- If DLINE1=PLINE and for that person REL=1 or REL=2, then the respondent is the household reference person.
- If DLINE1=PLINE and REL>2 then the respondent is not the household reference person.
- You can identify the relationship of the respondent based on the REL variable.
- If DLINE1=B (so no one living in the unit), then check RESPTYP, and that will report the type of respondent.

In surveys, before 1997, there is no RESPTYP. Instead, DLINE1 contains special codes if the respondent is not the reference person. This will occur only for vacant or URE interviews (ISTATUS not equal to 1). {3/5/2008}

Q. How can I calculate current equity in an owner-occupied unit?

The AHS does not ask respondents about their current equity. It is not possible to calculate more than a rough estimate because on the public use file the variables you would use are topcoded

(see the topic on topcoding below). However, you can download a copy of the SAS “table specification” program the Census Bureau uses to produce publications from HUD USER. Simply go to the web page for the survey year of interest. In it, you will find a recode for something called OTPIN. This is outstanding principle. If you use a program like this to get outstanding principle from the public use file you will get a distribution that is in effect top coded because of the necessity of using variables that are topcoded.

If you take your calculated outstanding principle and subtract it from the value in the variable VALUE (value of the unit) you get an estimate of equity. However, VALUE is top coded on the public use file also.

This is the only way I know to get an equity estimate. However, my recommendation is to not use it. I really do not think that the estimate would be very good, especially if you are trying to get some measure of aggregate equity.

Q. How can I identify first-time buyers?

Use FRSTHO. There are three things you have to be careful about. First, the concept applies to only owner-occupied units. Thus, you must restrict yourself to records where STATUS='1' and TENURE='1'. Second, the name of the variable is misleading. FRSTHO contains the answer to the question, “Did you ever own a home before?” Thus, a “yes” answer indicates a *repeat* buyer, and a “no” answer indicates a *first time* buyer. The third thing to watch is that the question is *not asked* if the respondent indicates that the household used the sale of a previous home as a source of down payment. In this case, FRSTHO='B'. So, to identify a first-time buyer, you need to make sure FRSTHO='2.' Checking to see if FRSTHO ne '1' is not sufficient, since blanks would pass this test. To identify a repeat buyer, the best way is to make sure FRSTHO does *not* equal '2.’”

Q. What is the universe of the recent mover questions? People who have moved in the past 12 months or since the previous survey?

During the interview (and hence for the microdata), we get info on people moving in since the last interview (or last two years, in metro surveys). When we publish tables, the cutoff is the last 12 months. So, the answer is ‘both:’ 2 years in the dataset, but 12 months in the report.

Note that in the metro survey datasets, only people who moved within the past two years are flagged as recent movers—not all people who moved since the previous survey. (The time between surveys of the same metro area is longer than two years.) {3/18/2008}

Q. What are the time periods for the equipment breakdown variables?

For water interruptions (IFDRY, etc) and blown fuses (IFBLOW, etc), it is last 3 months or length of residence, whichever is shorter. The water leak variables refer to the last 12 months or length of residence, whichever is shorter.

Q. What is the income limits file?

For programmatic purposes, HUD estimates certain income-related values for each metropolitan area and non-metropolitan county in the United States. These values include area median income, low income, and very low income. These estimates are used in determining eligibility for housing assistance programs, such as Section 8 vouchers. HUD also estimates the Fair Market Rent for each of these areas, and this is used as a payment standard in rental subsidy programs. Many assistance programs also refer to the official poverty income standard, which is published each year by the Department of Health and Human Services.

Because these values have many uses in housing research, HUD produces a dataset, called the "Income Limits File," which includes estimates for each record in the national AHS dataset. This file can be matched to the AHS microdata in order to classify household incomes relative to area median incomes, or rents relative to FMRs. The details of how these estimates are prepared are beyond the scope of this FAQ. The income limit files for recent survey years are available for download from the HUD USER web site. Note that the income limits files are *not* part of the American Housing Survey per se. They are auxiliary files prepared by HUD using the AHS public use file and HUD administrative data.

Beginning in 2003, the AHS newhouse file contains the income limits. These limits are added to the public use data by the Census Bureau, using HUD administrative data. Note, however, that income limits for units outside identified metropolitan areas are weighted averages based on heating degree days, region, and metropolitan status. This is done in order to protect confidentiality.

Q. Why does the AHS show many fewer efficiency apartments than the 2000 census?

Generally speaking data from Census 2000 and data from the AHS should be compared only with extreme caution. Many items in the two statistical efforts show statistically different results. These differences may be the result of a variety of data collection and processing differences between the AHS and Census 2000 such as the methods of data collection (mailed questionnaires in the census, personal and telephone interviews in the AHS), differences in processing procedures and sample designs, the sampling variability associated with the sample data from both the AHS and the census, and the nonsampling errors associated with the survey estimates and the census data.

There are a variety of possibilities as to what may be causing these differences. The difficulty is, of course, the amount of time and money it would take to investigate and verify any of the possibilities. One possibility is for the case where someone is renting a room or two in the house of another person. If in a large number of these cases in the census these people are counted as living in a separate housing unit but in the AHS they are counted as being part of the household from whom they are renting and living in a room of that household's unit, then the AHS would have much smaller counts of one room units and units with no bedroom.

In addition, there was change in the AHS method of data collection that has significantly affected the AHS counts of rooms and bedrooms. In 1995 and earlier in the AHS the counts of

one room units and units with no bedrooms were at levels roughly twice that observed in the AHS from 1997 and beyond. In 1997 the AHS changed from collecting data using a paper questionnaire to collecting data uses questions programmed into a laptop computer. This did appear to have a significant impact on the data collected for rooms and bedrooms. We have yet to be able to come up with the reason why. {2/29/2008}

Q. What are the weight variables for? Why are there more than one (PWT, Weight, WGT90GEO, WGT00_90)?

The “S” in AHS stands for Survey. Unlike the decennial census, the AHS does not enumerate every housing unit in the country. Instead, a representative sample is drawn, and this sample can be used to estimate the number and characteristics of housing units the nation. In effect, each AHS unit represents a larger or smaller number of housing units in the nation. In order to get a proper count of units of a certain type, you need to count the number represented by each sample unit.

The *pure weight* (PWT) is the simplest representation of this concept. It is the inverse of the probability of selection. Thus, if the sample unit had a 1/2500 probability of being selected for the sample, then it represents 2500 housing units, and PWT=2500. However, PWT does not account for certain difficulties in conducting surveys, particularly nonresponse. If a household refuses to participate in the survey (or can’t be found, etc.), it will fail to represent those housing units. In addition, the Census Bureau uses a system of projecting the total number of housing units in the nation which is believed to be superior to the total that would be computed from AHS data alone. The *final weight* (WEIGHT) adjusts the pure weight to account for bias introduced by nonresponse. It also is adjusted to agree with the control totals taken from Census Bureau projections.

The AHS is a longitudinal program, and so it is important that its weighting system is consistent across time. However, some aspects of housing change over time and are difficult to synchronize with the longitudinal dataset. One of these is the definition of metropolitan areas, which is set by the Office of Management and Budget. The AHS uses 1980 metropolitan area definitions, which were current when the current sample was first drawn for the 1985 national survey. Since then, metropolitan area definitions have changed several times. WEIGHT is kept consistent with the 1980 definitions. However, the Census Bureau uses 1990 definitions in the tabulations for the published reports. The WGT90GEO variable gives users access to the weights used in the report. At the national and regional levels, WEIGHT and WGT90GEO yield the same results, but at the metropolitan level they do not.

The foundation of the control totals mentioned above in the discussion of WEIGHT is the decennial census. When the data from a new census becomes available, these control totals change, and WEIGHT is updated for survey datasets released after that time. The first survey to make use of 2000 census-based control totals was the 2003 AHS. When the 2003 file was released, the Census Bureau also released a special set of weights for the 2001 AHS, based on 2000 census controls, rather than the original 1990 census controls. This is the WGT00_90 variable, which is available as separate download from HUD USER (see <http://www.huduser.org/datasets/ahs/ahsdata01.html>). The “90” in the variable name refers to the fact that it is based on 1990 metropolitan area definitions.

So which weight should you use?

- Use WEIGHT unless you have a specific reason to use a different weight. It is the most robust weight variable and the one most consistent across time.
- USE PWT for longitudinal analysis, when you are trying to develop a consistent set of weights to apply to more than one AHS dataset. (See the weighting discussion for the Components of Inventory Change project at <http://www.huduser.org/datasets/cinch.html> for a more technical discussion of how to do this.) Note that this is a very specialized application which is seldom necessary. If you are simply comparing AHS totals across time, use WEIGHT, not PWT.
- Use WGT90GEO if:
 - The distribution of housing units across metropolitan areas or between metropolitan and non-metropolitan areas is of particular importance in your work, or
 - Matching your tabulations to AHS reports is important (but see elsewhere in this document for a discussion of why this may not be possible).
- Use WGT00_90 if you are comparing 2001 tabulations with later surveys (in which case you should use WGT90GEO for those later tabulations). {3/18/2008}

Q. What are topcoding and bottomcoding? How can I tell if a value has been topcoded or bottomcoded? What rules does the AHS use for this?

The Census Bureau and HUD are dedicated to protecting the confidentiality of survey respondents. Thus, the microdata files contain no personally identifiable information, such as names or addresses. However, respondents could still be identified if they had extreme characteristics, such as very high incomes, very large houses, etc. To prevent this from happening, the Census Bureau edits such extreme values and substitutes a “topcode” (for large values) or “bottomcode” (for small values) in the place of the actual characteristics of the housing unit. In general, such recodes do not affect means or medians. They do affect variances. Users using multivariate statistical procedures, such as regression, should use their judgment about whether to include such records in their working data.

Very few AHS variables are bottomcoded. Only the nonwage income variables that can include business losses are given this treatment, to avoid violating the confidentiality of households that experience large negative incomes.

A number of variables are topcoded. Most follow the practice of masking the top three percent of values, replacing them with the mean of the values so masked. To give a very simple example, if income values over \$200,000 were topcoded, and there were only three such households, with incomes of \$200,001, \$500,000, and \$1,000,000, the income value for all three would be \$566,667. The rules for topcoding can be more complicated than this, depending on the variable. Also, some topcoding rules are themselves kept confidential by the Census Bureau. For a description of the publicly-revealed procedures used in the AHS, see the file, “Confidentiality documentation for public use file data users” in the Technical Supplements section at <http://www.huduser.org/datasets/ahs/ahsprev.html> .

For specific numerical information about the topcodes used in a recent survey, see the HUD USER web page for that survey. Beginning in 2003, HUD has posted a spreadsheet that shows the maximum value, the number of records coded with the maximum value, and the next-lower value for each variable subject to topcoding. This will show you where the topcoding cutoff lies, how many records are topcoded, and what the substituted value is for the topcoded cases. Corresponding information is provided for bottomcoded variables. {4/7/2008}

Survey Changes

Q. How can I tell which surveys years use a certain variable?

Go to the HUD USER web site, <http://www.huduser.org/datasets/ahs/ahsprev.html>, and download the latest “[AHS Variable Name Index](#).” You’ll find it in the “Technical Supplements” section. This index is a comma-delimited file that you can import into most spreadsheet, database, or statistical programs. It cross-tabulates all the variable names ever used in the AHS PUFs with the survey years. {3/5/2008}

Q. What changes were made in the income questions in 2007?

In the 2005 survey, family members were individually asked the collective amount of interest, dividend and rental income. In 2007, the family members were asked about each of the three income types separately. Also in 2007, family members were asked to report alimony and child support separately from their other sources of income. {6/6/2008}

Q. What changes were made in the income questions in 2005?

The income questions were revised considerably in 2005. Before that survey, we collected wage income from all family members, a list of income sources received by *any* family member, and total income from nonfamily members of the household. Beginning in 2005, we began collecting the source and amount of income from each family member. Nonfamily members are asked about wage, self-employment, and other income separately. In addition, the categories of nonwage income were regrouped to more closely follow the questions used in the American Community Survey. The questions on Supplemental Security Income and disability income were split off from the welfare income question, so that disabled persons could be identified more easily. See the document “User Note on New 2005 American Housing Survey Income Questions” on the HUD USER web site (<http://www.huduser.org/datasets/ahs/ahsdata05.html>) for details. {4/3/2008}

Q. What changes were made in the manufactured housing sample in 2005?

A new sample of manufactured/mobile homes was selected from the 2000 census in an attempt to improve coverage of manufactured/mobile homes built between 1980 and 2000. One-half of this sample was included in the 2005 interviewing and, as a result, one-half of the 1980-based manufactured/mobile homes sample was not included. These excluded units are included on the 2005 PUF and assigned a NOINT code = 38(unit eliminated in subsampling).

Q. Why does the survey show that 66 percent of Hispanic householders were white in 2001, but 92 percent in 2003?

Two or three things are going on with the Hispanic counts. First, in accordance with the OMB directives, the order of the Hispanic origin and Race questions was reversed in 2003 (placing Hispanic origin before Race). Second, we went to the 1-person-reporting-multiple-race concept. Third (and most important), the "other" race category has been eliminated. Starting with 2003, we began using the complete set of person edits used by the Current Population Survey. These do not allow "other" entries in RACE; rather they allocate one of the five specified response categories for all "other" persons. The AHS has added a sixth race category but only to cover the persons who reported more than one race. In the past, the "other" race category contained write-ins (such as "human being," "brown," etc.).

Previously, many Hispanic householders (about 30 percent of the total in 2001) stated that they were "other" race--and over three-fourths of the "other" race householders (78 percent) were Hispanic. In 2003, although persons may say they are an "other" race, the edits will not let this stand. So while over 7,000 persons (6,100 of whom were Hispanic) said they were "other" race, once the edits are complete, there are none. The CPS edits assign a race of "white alone" to most of the Hispanics who claimed to be "other" race--92 percent of them. However this is in line with what Hispanics say who actually report one (or more) of the specified races—93 percent say "white alone." {3/5/2008}

Thus, this edit change (and probably the question changes as well) result in many more Hispanics being white without a proportional increase in the number of Hispanic persons.

Q. Why do the control numbers in the 2002 Metro survey have only eleven digits, when the numbers in the corresponding previous surveys have twelve?

This is a mistake. In the 1980s, the National control became 12 digits in order to show the code for the supplemental samples. Since metro did not have the subsamples, the control codes were given an extra trailing zero for the sake of consistency. For unknown reasons, the extra zero was left off CONTROL in the 2002 survey. So, if you want to link 2002 records with previous surveys, either delete the trailing zero from the previous ones or add a zero to 2002 values. {3/25/2008}

Q. What changes were made to the mortgage questions in 2001?

1. The questions on reverse mortgages are moved to the end of the module and asked only if the respondent reports no other type of mortgage.
2. First the respondent is asked, "Not counting home equity loans, is there a mortgage or any loans on this house?"
3. If they say yes to the above question we asked, "How many mortgages or loans are there now on this house?"

4. Next, we asked (regardless of the answer to 2 or 3 above), "Do you have a lump sum home equity loan, that is, a home equity loan that is paid out in a one-time lump-sum amount and that must be repaid over a period of time?"
5. If they say yes to 4 above, we ask, "How many lump sum home equity loans do you have?"
6. Regardless of the answers to 2 thru 5 above we ask, "Do you have a home equity line of credit, that is, a home equity loan that allows you to borrow against it as often as you wish up to a fixed limit?"
7. If they say yes to 6 above, we ask, how many home equity lines of credit do you have?"

In the remaining questions on the module home equity lump sum loans are treated like regular mortgages except for the fact that regular mortgages are treated as primary if the respondent reports having both regular and lump sum loans. The line of credit loans are treated as a special type of loan, and we collect data on them, but it is in less detail than the data collected for regular and lump sum loans.

Q. How do the mortgage variables in 1999 compare with those of 2001?

Trying to match 1999 mortgage data and 2001 mortgage data is not always possible.

MG and REGMOR collect the same information in 1999 and 2001, the existence and number of *regular* mortgages (although REGMOR was not edited in 1999 and is of limited use in that year; see discussion under MCNT below). In 2001 lump-sum home equity loans were collected with the variables HELUMP and HELUMN. HELUMP and HELUMN did not exist in 1999 and do not have an equivalence in 1999. In 2001 home equity line of credit loans were collected with the variables HELC and HELCN. HELC and HELCN did not exist in and do not have an equivalence in 1999.

In 1999 home equity loan data were collected with the following variables which do not exist in 2001. HEL asked if they had a home equity loan (includes both lump sum and lines of credit) and HENUM asks for the number of home equity loans. Also in 1999 HETYP1, HETYP2, and HETYP3 asked if the first, second and third (if they existed) home equity loans were a lump sum or a line of credit.

In 2001 lump sum home equity loans are treated as regular mortgages for the purpose of collecting detail data. In 1999 lump sum home equity loans were treated as home equity loans for the purpose of collecting detailed data. See the discussion under MCNT below.

MRTYP1 is a new variable in 2001. In 2001 lump sum home equity loans are treated as regular mortgages in the collection of detailed data. Priority is given to the regular mortgages. MRTYP1 simply indicated if the first mortgage for which detailed data was collected was a regular mortgage or a lump sum home equity loan. MRTYP2 would give you the same information for the second mortgage for which detailed data was collected. There are not such

variables in 1999, and they would not be relevant in that year, as home equity lump sum loans were treated as home equity loans, not as regular mortgages in 1999 for the purpose of collecting detailed characteristics data.

MCNT is created by the edits. It represents the count of mortgages for which we collected detailed mortgage data. In 1999 and earlier we collected data only for regular mortgages, and so in 1999 if there is a numeric value in MCNT, MG will equal one. It is also worth noting that in 1999 and earlier, we did not edit REGMOR, and so REGMOR is sometimes in disagreement with MCNT in 1999. REGMOR in 1999 and earlier is not a valuable variable for most tabulations.

In 2001 we decided to treat home-equity lump sum mortgages the same as regular mortgages for the purpose of collecting detailed mortgage data. Home-equity lump sum mortgages can be identified by a '1' in helump (yes they have one) and a 1-10 in HELUMN (count of the number of lump sums). IN 2001, MCNT equals the sum of REGMOR + HELUMN. REGMOR is an edited variable in 2001 and is usable to determine the number of regular mortgages. In 2001 REGMOR is never larger than MCNT (although it could be only if MCNT were a 4 and REGMOR = a 5, 6 or 7), but it is smaller and in some cases blank when MCNT has a numeric entry. For example, if you cross MCNT and MG in 2001 you find that there are 648 cases where MCNT is a '1' but MG is a '2' and REGMOR is a 'B'. This is because HELUMN is a '1'.

MG is the yes/no answer for the existence of *regular mortgages*. In 1999 we collected detailed data for regular mortgages only. In 2001 we collected the detailed data for regular mortgages and home equity lump-sum mortgages. REGMOR is the count of regular mortgages. Unedited in 1999, it is of little use in that year and in 1997. Edited in 2001, it provides the correct count of 2001 regular mortgages (MCNT did this in 1999---MCNT counts the total of regular and home equity lump sums in 2001).

HELUMP is a new variable in 2001. It is the yes/no answer for the existence of lump-sum home-equity loans.

HELUMN is a new variable in 2001. It is the count of lump-sum home-equity loans.

Q. What was the reason for the August 2002 revision of the 2001 dataset?

Based on the situation reported by a data user, we re-examined the way that the alterations and remodeling data were processed. The problem was that the summary recode (RAN) for the number of jobs completed by a household sometimes did not match the number of individual job records for the same household.

This occurred because slightly different criteria were used to tally a job in the counter -- that is, the variable, RAN -- than were used to output a job record (which would include the variables RAS, RAD and RAH). The differences were limited to a few job types -- those related to carpeting, other inside and other outside jobs, and disaster-related work. Sometimes more job records were generated than tallied in RAN and sometimes the reverse occurred. However, the number of inconsistencies was not large.

The programs have been rewritten to make the value in RAN agree with the number of job records. This change will affect RAN (total number of jobs done) and RAC (total amount spent). In addition, the number of RAS codes of 1 (disaster loss), 51 (carpeting added over subflooring), 52 (carpeting added over finished flooring), 64 (other major inside improvements) and 70 (other major outside improvements) may change along with the related variables of RAD (cost of individual job) and RAH (who did the work for the individual job). {2/5/2008}

Q. What was the reason for the November 2002 revision of the 1999 dataset?

The number of Alterations and Remodeling/Repairs/Replacement records changed. We had an overall increase of 268. They breakdown to 144 added for carpeting installed/replaced over subflooring, 135 added for carpeting installed/replaced over finished floor, 9 lost from missed inside jobs and 2 lost from missed outside jobs. The missed inside and outside jobs were jobs that had only an N in the field which means there are no other missed jobs. {2/5/2008}

Q. Why are the frequencies for PAPHLT and LD MAYB so much larger in 1999 than they are in 1997?

PAPHLT and LD MAYB both concern whether a household was notified that the housing unit may have lead paint before they moved in. The 1999 automated instrument had a programming error in it. The universe for these items (PAPHLT and LD MAYB) was supposed to be households (in old units) where the respondent moved to the unit in September of the previous year or later. In 1997, this was correct. However in 1999, the order of a couple of lines of coding changed, and instead of asking only for recently moved in households, the question was asked of all households in old units.

So, the 1999 frequencies are too big. You can obtain the correct universe by checking the move month and year of the respondent and selecting cases where MOVE = 1999 or (MOVE = 1998 and MOVM >= 9). {3/12/2008}

Q. What was the data error in the 1998 Metropolitan Survey of Baltimore?

The 1998 Baltimore Central City estimates of vacant units (especially vacant for rent and other vacant) are larger than anticipated, and the estimates of renter-occupied units are smaller than anticipated, because of the misclassification of such units during data collection. Although this misclassification occurred in the central city only, it also affects estimates for the metropolitan area as a whole.

Units that should have been classified as "renter occupied" were coded as either "vacant for rent" or "other vacant," with most probably being classified as "vacant for rent." Because the problem occurred during data collection, we cannot correct it, as would be possible if this were a processing error.

The actual size of the misclassification is impossible to measure. We believe it to be large because the 1998 AHS rental vacancy rate for the Baltimore Central City was 29.3 percent

compared with a 1998 Housing Vacancy Survey (HVS) rate of 3.8 percent and a Census 2000 rate of 7.6 percent.

The collection of the characteristics data for both occupied and vacant units was most likely done correctly. The estimated levels of each characteristic are believed to be too high for vacant units (particularly in the central city) and too low for occupied units. The error is not in the collection of the characteristics data, but only in the classification of the unit as occupied or vacant.

Although we do not know the correct estimate, we believe the relative underestimate of renter-occupied units in the central city of Baltimore is small. All units classified as "renter occupied" were probably classified correctly. As a result, most analytical conclusions made using renter occupied or total occupied data for the Baltimore metropolitan area and/or the central city of Baltimore should be reasonably sound. The owner occupied data were not affected.

Although we do not know the correct estimate, the relative overestimate of vacant-for-rent and other-vacant units is probably large. A significant number of these vacant units were actually renter occupied. As vacant interviews were conducted at these units and characteristics data were collected, the characteristics data collected are reasonably correct. However, the resulting estimated levels of any characteristic are probably too large and should not be used. Any use of data for vacant units in the 1998 AHS for the Baltimore metropolitan area, especially the central city of Baltimore, should be made with caution. If data for occupied units and vacant units are added together, however, the resulting analyses are presumably reasonable.

Q. Why don't the Control numbers in the 1995, 1996, and 1998 Metro Surveys match previous years?

These surveys were new samples. They are not longitudinal with respect to previous metro surveys.

Q. How did housing adequacy (ZADEQ) change in the 1997 redesign?

Severe upkeep problem is a recode of six data items: outside water leaks, inside water leaks, holes in floors, cracks in walls, peeling paint and signs of rats. For a unit to qualify as having a severe upkeep problem, five of the six conditions have to be present. This is, therefore, not a simple comparison.

There were major changes in the question wording for all of the six feeder questions.

- The inside and outside water leaks and signs-of-rats questions each were split into subquestions. This was to limit the number of concepts that respondent had to process at one time. Thus, the approach is to find out if the problem has ever existed at the unit, and then ask if the problem has happened recently.
- For the cracks, holes, and paint questions, the 1997 versions of these items have definitions as part of the question. Thus, a hole must be "big enough to catch their foot on," and a crack

must be “wider than the edge of a dime.” The peeling paint uses the 1995 definition of extent, but the question is split into two parts.

- Perhaps as a result, the counts of the upkeep components have changed:

<u>Item</u>	<u>Frequency</u>	<u>Frequency</u>
	<u>1997</u>	<u>1995</u>
Outside water leaks	13,181	15,999
Inside water leaks	9,647	11,411
Holes in floors	1,611	1,074
Cracks in walls	6,936	4,527
Peeling paint	3,625	3,673
Signs of rats	923	1,332

- There may be some significance to the fact that the questions where the time reference has been emphasized have all decreased in magnitude. The questions where definitions were added have all increased (or held stable) in reporting.

The data collection and processing do not seem to be the villains in this change. The bridge had to merge data for a couple of items to produce the eventual out-variables for the different types of water leaks but this operation looks all right. Cracks, holes, peeling paint and signs of rats were allocated when missing but the number of cases involved was less than 2 percent. In addition, with one exception, the ratio of allocated problems to allocated non-problems was slightly higher than the respondent-reported ratio of problems to non-problems. The exception was signs of rats. For that item, the household was 10 times more likely to be imputed rats than to report them (6.7 percent versus 0.6 percent). Still the number of 'ratty' cases involved was very small.

There was some error in the ICF tally for severe upkeep problems. I believe the count should be 120,400 while the ICF table showed 90,000. My breakout was (in thousands) -

- 21.1 with LEAK, ILEAK, HOLES, CRACKS and BIGP (RATS is missing)
- 29.9 with LEAK, ILEAK, HOLES, CRACKS and BIGP (RATS is blank)
- 20.2 with LEAK, ILEAK, HOLES, CRACKS, BIGP and RATS
- 7.4 with LEAK, ILEAK, HOLES, CRACKS, and RATS
- 12.9 with LEAK, ILEAK, HOLES, BIGP and RATS
- 7.7 with LEAK, ILEAK, CRACKS, BIGP and RATS
- 11.8 with LEAK, HOLES, CRACKS, BIGP and RATS
- 9.4 with ILEAK, HOLES, CRACKS, BIGP and RATS

This does not bring the total up to the previous years' count, but it is closer. In the long run, the main source of the change seems to be the new questions, not the processing.

Another change is that vacant and URE units are all coded 'B' beginning in 1997. Many of the items that go into ZADEQ were observation item before 1997. From 1997 on they are asked of the respondent, and the information is no longer obtained for vacant and URE units.

Report Questions

Q. Why doesn't my tabulation of [something] match the publication?

One reason may be topcoding. The public use file has a number of topcoded variables. The publications are all done from the Census Bureau's internal files. For more complicated tabulations, you might want to look at the table recodes file, a SAS program which shows how the Census Bureau recodes variables for tabulation. The 1999 version is available for download from HUD USER at http://www.huduser.org/datasets/ahs/1999table_recoded.txt.

One specific example is units in structure. Units in Structure tabulations in the publication cannot be duplicated using the public use file (PUF) even when the 1990 geography based weights are used because the topcoding for the variable NUNITS (number of units in structure) is done at a very detailed geographic level. As a result there are areas of the country where NUNITS on the national file are actually topcoded at the 2 or more unit level. Therefore no tabulation for multiunit structures from the PUF will match the publication.

Another reason may be weighting. Beginning in 2001, the AHS reports uses weights based on the 1990 metropolitan area definitions. The WEIGHT variable in the dataset uses 1980 metropolitan area definitions, to be consistent with previous years. While both sets of weights should give the same national totals, there may be slight differences for variables that are affected by the differences in metropolitan classification between 1980 and 1990. The variable WGT90GEO contains the weights using 1990 metropolitan geography. {2/5/2008}

Q. Why don't the sample sizes in the national public use files and the reports match exactly?

There are a very small number of cases (fewer than 100) with such unusual geography that we cannot include them on the public use file without compromising confidentiality. And that's all we can say about it.

Q. Why don't PUF tabulations of the "big six" metro areas match their reports?

In 1995, 1999, and 2003, the metropolitan surveys of the six largest metropolitan areas (Chicago, Detroit, Los Angeles, New York, Northern New Jersey, and Philadelphia) were conducted as supplementary samples of the national survey. Thus, the microdata for these metro areas are part of the national files for these years. The weights on the public use file (PUF) are adjusted to national control totals. The Census Bureau reweights the data using local control totals for the tabulations in the metro reports. Thus, you will not be able to match the metro publications using the PUF.

The SMSA codes included in each of the "big 6" metro areas are:

- Chicago = 1600, 0620, 3960, 3965, 9991
- Detroit = 2160

- Los Angeles = 4480
- New York = 5600, 5380, 5660, 9992
- Northern New Jersey = 5640, 0875, 3640, 5015, 5190, 8480, 9993
- Philadelphia = 6160

Q. What is the rationale behind using the incomes of families and primary individuals as opposed to household income in calculating monthly housing costs as percent of current income?

It is a trade off. Neither way is perfect. However, currently we are doing what has been done historically. If you were to use monthly housing costs as a percent of household income the argument is that you would be including the income of a significant number of people who are not directly contributing to housing expenses. An example of this is a lodger who pays the householder rent (which is recorded as part of the householder's income). The argument in the other direction is that by using monthly housing costs as a percent the income of families and primary individuals you are losing the income of a significant number of people who do contribute to housing expenses. An example is a roommate who does pay part of the rent or utility costs.

Q. Why does the "Year Round" column of the report not match a tabulation of YRRND from the PUF?

Despite the similarity of their names, the report column and the PUF variable are based on different concepts. In the report, "Seasonal Use" is based on how the unit is *used*, not how it is equipped. Essentially, seasonal units are those for which STATUS = 2 or 3 and VACANCY = 8. Year-round units are all the rest. In the PUF, YRRND is determined by the unit's physical equipment—particularly whether it has heating equipment. Even if a unit is not suitable for year-round use, it may be used that way, particularly in warm parts of the country. Conversely, a unit may be suitable for year-round use but only used seasonally, as in a well-equipped vacation home in Maine. {3/12/2008}

Q. Why are vacancy rates calculated from the microdata differently from the report?

The publication and microdata treat UREs differently (mostly for historical reason). First, URE units (STATUS = 2) are simply URE vacants in the publication. We do *not* consider their vacancy status in any of the tabulations. This is consistent with what is done in the Housing Vacancy Survey (HVS). In the microdata, we have vacancy status for UREs. Second, the variable YRRND is not relevant. It has to do with the suitability of the unit for year-round use, not the unit's actual status. That is, a unit may be classified as year-round even though it is not suitable for year round use.

What we are actually doing in the report is:

(STATUS = 3 and VACANCY = 1-2) /
((STATUS = 3 and (VACANCY = 1-2 or VACANCY= 4)) and (STATUS = 1 and
TENURE = 2-3))

Or in English:

Vacant for Rent (STATUS = 3 and VACANCY = 1-2)
DIVIDED BY
THE SUM OF

- 1) Vacant for Rent (STATUS = 3 and VACANCY = 1-2)
- 2) Rented not yet occupied (STATUS = 3 and VACANCY = 4)
- 3) Renter Occupied (STATUS = 1 and TENURE = 2-3)

The sum of 1+2+3 is known as the rental inventory in both HVS and in the AHS publications.

In 1999, we have:

- Vacant for Rent = 2,718.850
- Renter not yet occupied = 291,730
- Renter Occupied = 34,007,000

This yields a rental vacancy rate of 7.34 which is consistent with the 7.3 percent in the publication and in accordance with the definition in appendix A:

$$2718850 / (2718850 + 291730 + 34007000) = 7.34 \text{ percent}$$

If you exclude the 291,730 from calculation, you get a vacancy rate of 7.40 (higher than what is in the publication)

Q. Why does the "with 2 or more living rooms, or recreation rooms, etc." value in Table 1A-6 differ from the same row in Table 2-7, in the 1997-2003 reports?

An error was made in the table specifications from 1997 to 2003. The line "With 2 or more living rooms or recreation rooms, etc." was intended to include living rooms (LIVING), recreation rooms (RECRM), dens (DENS), and family rooms (FAMRM). In the 1997 through 2003 AHS total inventory tables (1A-6, 1B-6, 1C-6, and 1D-6) for the item Selected Amenities the line "With 2 or more living rooms or recreation rooms, etc." does NOT include family rooms. What you have in these tables for these years are a sum of living rooms (LIVING), recreation rooms (RECRM), and dens (DENS).

This error was not made in the occupied chapters (tables 2-7, 3-7, 4-7, 5-7, 6-7, and 7-7). These data are correct in the years 1997 through 2003 in that the data are tallied as intended.

The error was corrected for the 2005 AHS publication. In 2005 data in the total inventory tables are tallied as intended and agree with the occupied chapters. {3/25/2008}

Coding Questions

Q. What are the differences among B, D, R, '.', and just blank?

The SAS system allows for different kinds of missing values. In the AHS data, we use this capability to indicate the reason why a value is missing. The easy ones to explain are B, D, and R:

- B: Not applicable. This data item was never collected for this case. A simple example is that rent is not collected for owner-occupied units. Note, however, that the AHS instrument has skip patterns that do not ask certain questions in less obvious situations. Pay attention to the “out of universe conditions” listed in the codebook to understand when a B code is appropriate.
- D: Don’t know. The respondent was asked the question but said he or she didn’t know the answer.
- R: Refused. The respondent was asked the question but refused to give the answer.

Note that the distinction between D and R is sometimes ambiguous, and to some extent the coding is up to the field representative (FR).

A blank or a SAS missing, (a blank in a character variable or a . in a numeric variable) means not reported, except where a “mark all that apply” variable is involved (an explanation of this type of variable is below). There are two possible ways to get a not reported. First, the specific item accepts a non-response meaning the FR can press ENTER to proceed without having to enter an answer. Second, the interview was cut short but the respondent was in universe for a question he/she never received. For example, the respondent had to leave three quarters through the interview to pick-up her son. The items the respondent was in-universe to be asked will be set to blank unless set or imputed in the edit.

The AHS instrument includes some “mark all that apply” questions in which the FR reads a list of responses and codes an “X” if the response is applicable, leaving the result blank if not. For those questions, a blank code essentially means “no.”

Some blank results are added by the editing process. For type “C” noninterviews (permanent removals from the housing stock) many variables are coded blank that could arguably be coded B, as they don’t apply to no-longer existent structures. However, such inconsistencies are usually easy to spot if one pays attention to the value of STATUS (STATUS='4' for noninterview cases).

All of the SAS missing value codes are translated to negative numbers in the ASCII version of the dataset. See the codebook section on missing values for details.

Q. Why are some categorical variables numeric, while most are character?

During the 1997 AHS redesign process, we decided to treat any variable that had two or more digits as numeric. That way when a program does a comparison to the variable for the value '1' for example, you can just look for a '1'. When you deal with two or more digits, you can get each of the following which could also mean '1' - '1', '01', '1 '. In the last example, one can easily argue that it isn't a '1'; it's a two digit number between 10 and 19 where the second digit was missed. The possibilities when you get to three digits are worse than this.

Note that this answer applies to the 1997 and later surveys. Before 1997, all of the variables in the dataset, except for CONTROL, were numeric. The single-digit categorical variables were converted to character in the 1997 redesign in order to make the computer files smaller.
{3/5/2008}

Q. Why don't I see topcoded values of 99998 (or similar)?

For many numeric variables, the codebook shows the highest value as all nines, except for the last digit, which is 8. For example, the highest value for UNITSF (square footage) is shown as "99998," with the label "99,998 square feet or more." While this value is the largest this variable could ever take, the Census Bureau's topcoding system will usually preclude your ever seeing it. See the notes section of each codebook entry for the topcoding rule that applies. For example, the UNITSF entry is "Topcoded at the 97th percentile." In addition, for recent surveys you will find a spreadsheet on the appropriate HUD USER web page that shows empirical topcoding information, including: the highest value found in the file; the number of cases with this value; and the second-highest value found in the file. Thus, you can see the topcoded value, the number of topcoded cases, and the highest non-topcoded value for that variable in that year.
{2/5/2008}

Q. What do 8 and 9 codes mean in pre-1997 survey data?

In 1985 and earlier years 8 is always 'not reported' and 9 is 'out of universe.' The same goes for 98 and 99, 998 and 999, etc., for variables with more digits than one.

Q. The codebook lists out-of-universe conditions for variables, but some of the variables used in them are not listed in the codebook. What do these mean?

These are internal Census variables that are not part of the public use file. Here is what they mean:

- GQINST -- defines units as being in group quarters or not, and whether the group quarters are institutional (jails) or non-institutional (dorms) units
- FPHEQP -- determines if a fireplace is considered to be heating equipment or not
- HHMEM -- identifies if a current occupant is a household member or not
- INCSAM -- identifies whether a serial number is in sample for the first time, or is a returning case
- OWNLOT1 -- determines if mobile home occupants own the land where the unit is sited

- OWNLOT2 -- determines if a vacant or URE-occupied mobile home is on owned or rented land
- PEELRM -- determines which room has the most peeling paint (in the lead paint module)
- PHAFLG -- identifies units that are part of a public housing authority (this is not actually used)
- REPLHH -- identifies replacement households (used to create the SAMEHH variable)
- REMOVE -- This appears to be typo for RMOV, the recent mover group (which is on the PUF). It will be corrected the next time the codebook is updated.
- RMOVM -- month of move for recent movers
- ROOM -- variable for room heaters as secondary equipment (it is split into FLIN and FLOT)
- SPOUS -- line number of spouse
- SPOUS_FLG -- identifies if spouse present in household
- SURVYEAR -- survey year
- TOTINC -- recode at time of interview for total family income (uses unedited data) {3/12/2008}

Q. How is AMTX coded?

AMTX is described on page 8 of the Volume 3 codebook: "Property tax per year. Truncate to the next lower \$50, and then add \$25. If over \$1000, truncate to next lower \$100, then add \$50." This is *incorrect*.

The actual code is:

```
if amtx >= 1 then
do;
  amtx = amtx - 50;
  amtx = round(amtx,100);
  amtx+50;
end;
```

Thus, it is truncated to the next lower \$50, rounded to the nearest \$100, and then \$50 is added.

Q. What is the universe for ACCESSC and ACCESSB?

“I notice that the variable ACCESSC, which inquires whether an entry system is necessary to enter the community, is flagged out of universe for 95 percent of the records. What is the screener being used for this question? ACCESSB (entry to building) similarly has a high share of out of universe. NORC (Majority of neighbors 55+) and AGERES (Age restricted development) have relatively high shares of out of universe, but I infer these are only asked when there is one person age 55 or older in the household.”

ACCESSC is asked only where GATED = yes, don't know, or refused; ACCESSB is asked for multiunit buildings regardless of the answer to GATED (Walls/fences surrounding community). You are correct about the restrictions on NORC and AGERES.

Q. What is the relationship among BURNER, COOK, AND OVEN?

These variables are included in the survey in order to determine whether a housing unit has complete kitchen equipment (summarized in the variable KITCHEN). If a unit has a cookstove with range and oven (COOK='1'), then there is no need to ask about burners or oven separately. Hence, both BURNER and OVEN will be set to B (not applicable). Note also that after the 1997 redesign, OVEN refers to *microwave* oven only. {6/6/2008}

Q. Is there an error in the way BUYG is coded?

In 1999, there was an error, which miscoded some responses for BUYG. In some cases, BUYG was blanked when it should have been set to 1 (for "not used"). Users can correct for this by looking at the value of AMTG. If BUYG is properly blanked, AMTG should have a value. If AMTG is blank, then treat blank values of BUYG as "1."

Q. How is CONTROL coded in the metropolitan surveys?

In the metro surveys, CONTROL values are unique only within a metro area. Thus, you may have records with the same CONTROL values but different SMSA values. In order to match records between datasets, you should sort on both SMSA and CONTROL.

Q. What is the meaning of the "blank" values in DISTJ and TIMEJ?

Those are persons whose place of work is not in the United States.

Q. What do DLINE, PER, PERSON, and PLINE represent?

Before answering the question, an explanation of how the data is set up and organized is in order. As part of the data entry, each person living in the housing unit is assigned a "person number." This person number does NOT necessarily correspond to who the head of the household (or householder) is.

So, first answering the question about the different variables:

- DLINE-person number of the respondent. The respondent is not necessarily the head of the household. This is the primary person answering the survey.
- PER-number of people in the household. This is a total count and does not refer to any individuals.
- PLINE-the person number of the particular individual in the household. Each person in the household is assigned a unique number. These numbers are not necessarily continuous and are assigned based on how the people have been listed during the survey, and NOT following any rules.
- PERSON-same as PLINE. This is the equivalent variable present in the journey to work module (JTW). It is used to identify in that module.

Q. What do we know about the coding of the "e" neighborhood variables?

Here is a partial answer. We still really do not know what is going on completely.

As an example, we looked at ESFD. Coming off the questionnaire, ESFD would have a '0' in it *if it were not checked*. It would have a '1' in it *if it were checked* by the interviewer. ESFD would have a '9' in it if an edit were run to put a '9' in it.

We are confused in that we have two people at the Census Bureau who have run this data and they have gotten different results in 1987. For single-family interviews, one person is getting all 9s and the other person is getting all 0s and 1s at this point. One is running an ASCII file and the other a SAS file, but except for the formatting, the files should be the same. We are still trying to resolve this inconsistency but it will take some time.

In the meantime, all I can tell you is that between 1987 and 1995, we had telephone interviewing and therefore, most single-family homes were not visited. If not visited, ESFD would not have been asked of the interviewer. Coming out of the questionnaire, it would have a '0' in it. A '0' in 1987 through 1995 therefore, could mean that this is a telephone interview (I believe that this is what most '0s' mean) or it could mean that the unit was visited (only new units or replacement households were visited if they were single family structures) and ESFD is 'no'. A '1' would mean that the unit was visited and the answer is 'yes'.

In short, for single-family structures in the years 1987 through 1995 assume the following for now:

- 1) That you cannot get reliable totals for the E variables for single-family structures.
- 2) That codes of '1' are valid yeses but they do not represent all the yeses we would have gotten if we had visited all single-family homes.
- 3) That many of the '0s' simply mean that the unit was not visited. Only some of the '0s' mean that the answer is no.
- 4) That any '9s' were put there using some edit and that in most cases the '9s' were '0s' in units not visited coming out of the questionnaire.

We are not at all sure as of now why the 1987 data appears to be so different from the 1989 through 1995.

Q. What is the relationship among FAMNUM, FAMREL, and REL?

The key point is that there is a difference between "reference person" and "reference person of family."

- There is one reference person per household.
- There is one "reference person of the family" for each separate family in the household.
- If there is at least one family in the household, the reference person is always also the "reference person of family" for FAMNUM = 1.

- If there is more than one family in the household, the reference person of the family for FAMNUM > 1 may or may not be a relative of the reference person.

Specifically,

- If FAMNUM = 0, then FAMREL = 0 for everyone. The REL values for everyone show them as either the reference person or a nonrelative.
- If FAMNUM = 1, FAMREL shows everyone as being either the reference person of the family or a relative of the reference person of the family. REL shows everyone to be either the reference person or a relative of the reference person. Note that the reference person is always the "reference person of family."
- If FAMNUM > 1, FAMREL shows everyone as the "reference person of the family" or a relative of the ref person of the family. REL shows that none of the members of these families are the reference person (REL > 2 always). They may or may not be relatives of the reference person. {2/13/2008}

Q. Under what circumstances will FRSTHO be blank?

There are two circumstances where FRSTHO is specifically mentioned in the blanking:

1. If the unit is not occupied or not an owner unit (if STATUS ne '1' or TENURE ne '1')
2. If the source of down payment was sale of previous home (if DWNPAY eq '1').

Q. How are HO1, HI1, HP1, HV1, and HH1 coded?

- 0 = no
- 1 = yes
- 8 = not reported
- 9 = not applicable.

For HMod11 (raised lettering or Braille) we did not ask, "which limitation required the use of this aid?" (hi, ho etc).

Q. What is the relationship between HHMOVE and SAMEHH?

SAMEHH is created in the "bridge" program that moves data from the survey instrument into the Census computer system. It simply reverses the coding of the instrument variable rephh. A code of '1' (yes) simply means that at least one (not necessarily all) of the current household members was living in the unit during the previous interview period.

HHMOVE gives the information for MOVE (year move) for the householder. Whether or not anyone in the household was present during the previous interview period (SAMEHH) often has little relationship to the year the householder moved (HHMOVE). {2/29/2008}

Q. How are HRamp, HElev, Hrail, Hwide, HHndl, Hpush, HSckt, HSink, HBath, HKit, HRais, HPhon, HFlas, HOthr, HHelp, HCane, HChair, HCart, HDEV coded?

1 = yes

2 = no

8 = not reported

9 = not applicable

Q. What happened to INCFM?

INCFM is listed in the volume 3 codebook as indicating whether a household received farm income. In the dataset, this variable is called CROPSL.

Q. Is the IPOV variable contained as part of the income limits file the same that is used for the 'Below Poverty Level' in the published AHS volume? That is, should I be able to duplicate those results using IPOV?

IPOV is computed by HUD. We use the official poverty level definitions, of course. However, we do not explicitly coordinate with Census to make sure that tabulations based on IPOV will match the reports. The AHS reports also use the official poverty level cutoffs. A possible source of difference with the IPOV value would be that the official cut-offs are issued monthly, and the appropriate figure is applied to each AHS case based on month of interview. IPOV is an annualized figure, and so there will be some cases that fall differently in AHS report than in a classification by IPOV.

Q. What is the difference between ISTATUS and STATUS?

In the surveys from 1997 on, ISTATUS and STATUS are identical. When the coding was revised for the 1997 survey, STATUS was used all through the various programs. At the end of this, we noticed that previous codebooks had used ISTATUS ("interview status") for this variable. Since changing the name would require extensive rewriting of the supporting programs, we just added ISTATUS "to avoid confusion." "Computers are going to save us a lot of work someday."

Q. What is the difference between KITCHEN and KITCH?

KITCHEN (sometimes KITCH in older files) is a recode that indicates whether a unit has complete kitchen facilities. It is a yes/no variable. KITCH is the number of kitchens in the unit—a count of rooms.

Q. Why are the values of LMED (area median income) considerably higher than the median value of ZINC2 (household income)?

There are several reasons for this. LMED is based on HUD estimates of area median incomes for individual metro areas and nonmetropolitan counties. Although the people who create the estimates (our Economic Market Analysis Division) do use data from the individual AHS metro

surveys where available, they use many other data sources, such as decennial census and random-digit dialing surveys. Thus, there is no specific reconciliation between LMED and ZINC2.

ZINC2 is known to underestimate income, mainly through missing nonwage income. Also, as was explained in the documentation for the income limits, the LMED estimate for cases where we don't have a valid SMSA value is an average over REGION, METRO3, and DEGREE.

LMED and the other income limits variables are intended to be used to match HUD program eligibility requirements and affordability measures. They are probably the best measures we have of variability in income and cost at the local level. But they aren't designed to add up to national measures based on ZINC2.

Another reason is that HUD estimates median *family* income, not *household* income. HUD income limits are based on "median income for family households." Thus, it more closely corresponds to ZINC in the AHS dataset, although the previous discussion still applies. The difference is that the HUD numbers exclude non-families (e.g., single people, roommates).
{3/18/2008}

Q. Under what circumstances will LPRICE and MORTIN be blank?

I have been doing some tabulations of the characteristics of recent buyers using the 1999 national survey, and I have found a lot of missing values for LPRICE, the purchase price. I can understand "D" and "R" values. I am also getting "B" and "." values. These are all for households where MG='1,' and so they have mortgages. Under what circumstances will "B" and "." show up for LPRICE?

Also, there are 120 cases (among these recent buyers) where MG='1' but MORTIN (type of mortgage insurance) is blank--not "B" but just blank. What's going on there?

You really need to use both LPRICE and CPRICE. From 1997 on, we collect CPRICE (LPRICE will be 'B') if people constructed or had their unit constructed and LPRI CE if they simply bought a unit already constructed. There should never be a value in both variables. Also, always assume in variables that have 'Ds' and 'Rs' that missing can also occur and they usually do. Anytime an interview stops before being completed, a missing can be generated for the remainder of the items that were not asked. Treat these missings as the same as Ds and Rs.

Remember that MG is allocated and MORTIN is not. It is quite possible that we allocated the value in MG and MORTIN would have remained as a missing.

Q. What is the difference between MCNT and REGMOR?

REGMOR is a respondent question. We ask how many regular mortgages they have, and record what they say.

MCNT is computed in the edits. After asking the above question, we gather data on each individual loan. MCNT is the count of how many loans we gathered information on.

In principle, there should be no difference between the two, but we do not reconcile them. In practice, they are different for a small number of cases. MCNT is probably a better variable to use, since it counts the loans that we know exist (in that we got more information about them). In addition, MG and MCNT are reconciled, while MG and REGMOR are not.

In 2001, the plan is to edit REGMOR and MCNT so that they are consistent. However, MCNT will be the count of both regular mortgages and lump-sum home equity loans. Thus, MCNT will always be greater than or equal to REGMOR.

Q. What is the difference between MOVE and MOVYR?

MOVE is a 4-digit variable and is edited. MOVYR is 2-digit and unedited. I have no explanation why the latter is in the public use file at all, but there it is.

Q. What are the differences among MVG, MOVGRP, and RMOV?

RMOV and MOVGRP are supposed to mean the same thing. RMOV is the secondary key in the recent mover data sets. Coming out of the instrument, we had both because CASES (the programming language for the survey instrument) creates a roster index (which is RMOV), so we have been carrying them along. In 1999, 584 cases have a missing RMOV and a good MOVGRP, for reasons we do not understand yet. We suggest you ignore RMOV for now and use just MOVGRP.

MVG comes from the person file. MOVGRP comes from the recent mover file. The file-flattening program merges recent mover information when combining based on control and the mover group number. The program renames one of them to allow the merge to be based on it.

Another oddity you will notice is that In the PERSON file, MVG goes from 1 to 4, but in the RMOV file, MOVGRP goes from 1 to 3. This is because we ask the recent mover questions for only 3 groups of movers per household. There are 4 mover groups indicated on the Person record because all the people (who are movers) that aren't part of the first 3 groups get lumped together into group 4 (so there could be any number of actual mover groups but we don't--and can't--differentiate after the first 4).

Q. What is the difference between NOPROB and NPROBS?

NPROBS provides a yes/no response to the question "Is there <anything/anything else> about the neighborhood that bothers you?" It is asked after the respondent was asked about heavy traffic, neighborhood crime, smoke/gas or bad smells, and street noise.

If the person answers 'yes' to NPROBS they have an opportunity to indicate all that apply from the following list, which includes the variable NOPROB:

- No problemNOPROB
- NoiseNOISE

Litter or housing deterioration	LITTER
Poor city/county services	BADSRV
Undesirable commercial, institutional, or industrial property	BADPRP
People.....	BADPER
Other	OTHNHD

So if someone says, "Yes, there is something about the neighborhood that bothers me," he *can* then say, "no problem."

Q. What is the difference between OWNPHA and PROJ?

The variable PROJ on the file is a recode of PROJ1 (question 12d_OWNPFA -- universe, renter-occupied units) and PROJ2 (question 14a_PROJ universe, selected vacant units). In other words, the PROJ variable is originally collected when it is not a regular interview. The OWNPHA variable is originally collected when there is a regular interview with the occupant. This can be seen in module AHSHCST2.Q of the "Q-code" which is the code for the interviewing framework for the computer assisted telephone interviewing. Both of these were then combined into the PROJ variable on the public use file.

Q. When should I use PRENT instead of RENT?

RENT is contract rent, but PRENT is what the household actually pays. The difference occurs when a unit is subsidized in some way (and thus the government or some other entity pays part of the contract rent). If you are interested in household expenditures, use PRENT if it has an entry. If you are interested in the cost of a unit independent of who lives there, use RENT. {3/18/2008}

Q. What is PVALUE for? Why is it mostly missing?

PVALUE is the "extra" value question asked at non-standard units. Units in multifamily structures, with a business on the property, or with more than 10 acres are asked PVALUE (value of entire property), and well as VALUE (value of unit/value of unit and lot). So, PVALUE should mostly be missing. {2/13/2008}

Q. What is the difference between RAC and RAD?

RAD (old RACOST) is the amount reported for an individual job. This variable is on the HOMIMP dataset. Each eligible household may have several RAD variables. RAC is the summed amount for all jobs done by the household. In the 1997-1999 surveys, this variable is with the recodes (TOPPUF). After that, it is in the household file (NEWHOUSE).

Q. How is RAS coded? (1997 version)

- 1 = disaster required repairs
- 2 = created finished bathroom from unfinished space
- 3 = created finished bedroom from unfinished space
- 4 = created finished kitchen from unfinished space
- 5 = created finished recreation room from unfinished space

6 = created other finished inside room from unfinished space
7 = added bathroom onto home
8 = added kitchen onto home
9 = added bedroom onto home
10 = added other inside room onto home
11 = added/replaced garage
12 = added/replaced porch
13 = added/replaced deck
14 = added/replaced carport
15 = added/replaced other outside structure
16 = moved walls in bathroom
17 = added/replaced cabinets in bathroom
18 = added/replaced flooring in bathroom
19 = added/replaced counter tops in bathroom
20 = added/replaced toilet in bathroom
21 = added/replaced tub/shower in bathroom
22 = added/replaced sink in bathroom
23 = added/replaced lighting fixtures in bathroom
24 = added/replaced other electrical items in bathroom
25 = painted, papered, or wall tiled bathroom
26 = moved walls in kitchen
27 = added/replaced cabinets in kitchen
28 = added/replaced flooring in kitchen
29 = added/replaced counter tops in kitchen
30 = added/replaced other built-in appliances in kitchen
31 = added/replaced sink in kitchen
32 = added/replaced lighting fixtures in kitchen
33 = added/replaced other electrical items in kitchen
34 = painted, papered, or wall tiled kitchen
35 = bedroom created through structural changes
36 = other room created through structural changes
37 = added/replaced roof over entire home
38 = installed/added siding to home
39 = replaced/covered siding on home
40 = added internal water pipes to home
41 = replaced internal water pipes in home
42 = added electrical wiring to home
43 = completed rewired the electrical wiring in the home
44 = added/replaced fuse boxes or breaker switches
45 = added doors or windows to home
46 = replaced doors or windows in home
47 = added plumbing fixtures to home
48 = replaced plumbing fixtures in home
49 = added insulation to home
50 = replaced insulation in home
51 = added wall-to-wall carpeting over bare subflooring

- 52 = added wall-to-wall carpeting over a finished floor
- 53 = added other types of flooring over bare subflooring
- 54 = replaced finished flooring with same/different type of flooring
- 55 = installed new paneling or ceiling tiles
- 56 = replaced existing paneling or ceiling tiles
- 57 = installed/replaced central air conditioning
- 58 = replaced built-in heating equipment
- 59 = installed new built-in heating equipment
- 60 = added/replaced septic tank
- 61 = added/replaced water heater
- 62 = added/replaced dishwasher
- 63 = added/replaced garbage disposal
- 64 = other major improvements or repairs inside home (up to three could be reported)
- 65 = added/replaced driveways or walkways
- 66 = added/replaced fencing or walls
- 67 = added/replaced patio, terrace, or detached deck
- 68 = added/replaced swimming pool, tennis court, or other recreational structure
- 69 = added/replaced shed, detached garage, or other building
- 70 = other major improvements or repairs to lot or yard (up to three could be reported)

Q. How is RAS coded? (1999 version)

- 1 = disaster required repairs
- 2 = created finished bathroom from unfinished space
- 3 = created finished bedroom from unfinished space
- 5 = created finished recreation room from unfinished space
- 6 = created other finished inside room from unfinished space
- 7 = added bathroom onto home
- 8 = added kitchen onto home
- 9 = added bedroom onto home
- 10 = added other inside room onto home
- 11 = added attached garage onto home
- 12 = added porch onto home
- 13 = added deck onto home
- 14 = added carport onto home
- 15 = added other outside structure onto home
- 35 = bedroom created through structural changes
- 36 = other room created through structural changes
- 37 = added/replaced roof over entire home
- 38 = added/replaced siding on home
- 40 = added/replaced internal water pipes in home
- 42 = added/replaced electrical wiring, fuse boxes, or breaker switches in home
- 45 = added/replaced doors or windows in home
- 47 = added/replaced plumbing fixtures in home
- 49 = added/replaced insulation in home
- 51 = added wall-to-wall carpeting over subflooring
- 52 = added wall-to-wall carpeting over finished flooring

- 53 = added other types of flooring such as wood, tile, marble, or vinyl
- 55 = installed paneling or ceiling tiles
- 57 = added/replaced central air conditioning
- 58 = added/replaced built-in heating equipment
- 60 = added/replaced septic tank
- 61 = added/replaced water heater
- 62 = added/replaced built-in dishwasher
- 63 = added/replaced garbage disposal
- 64 = other major improvements or repairs inside home (up to three could be reported)
- 65 = added/replaced driveways or walkways
- 66 = added/replaced fencing or walls
- 67 = added/replaced patio, terrace, or detached deck
- 68 = added/replaced swimming pool, tennis court, or other recreational structure
- 69 = added/replaced shed, detached garage, or other building
- 70 = other major improvements or repairs to lot or yard (up to three could be reported)
- 71 = remodeled bathroom
- 72 = remodeled kitchen
- 73 = bathroom created through structural changes
- 74 = added/replaced security system in home

Q. When does RENT=1 have a special meaning?

1. The use of '1' as a code applies to any URE or vacant unit for which RENT is collected. In those cases, it means, "Depends on income of occupant, such as public housing and some military housing."
2. However, for occupied units, '1' is a dollar amount, the same as any other dollar amount.
3. The practice of using a code for 'amount of rent depends on the occupant' began in 1985 National and in 1986 for Metro for vacant units; the code was not adopted for URE units until 1995 for both National and Metro.

Q. What, exactly, does a "no" answer for SAMEDU mean?

If you get a 'no' in the question for SAMEDU, it could mean that:

- the unit is the result of a conversion or merger since the previous survey
- the interviewer went to the wrong place last survey
- the current unit is a replacement mobile home (or, much less frequently, a replacement structure)
- the unit is a vacant mobile home site that was occupied in the previous survey
- the address identifies a location that is now a type C noninterview {2/29/2008}

Q. Why are SAMEHH and MOVE not always consistent?

We generate the variable SAMEHH from two questions in the demographics module of the instrument:

STILIV1: "I have listed {READ NAMES} Are all of these persons still living or staying here?"

STILIV2: "Do ANY of these people still live at this address?"

STILIV2 is asked only if the answer to STILIV1 is 'no'. If both STILIV1 and STILIV2 are answered 'no' the unit is considered to contain a replacement household. Neither variable is on the public use file since their information is coded directly into the variable SAMEHH.

The reason you are seeing conflicts between the MOVE dates and the SAMEHH variable is because there are no consistency edits between the two concepts. MOVE is edited to be consistent with the date of interview (can't move in later); the year the unit was built (can't move in before, except for mobile homes--it's a long story) and age (can't move in before person is born).

So in doing any analysis comparing MOVE and SAMEHH, you may first want to check to see if one or the other variable is allocated. If you find both data items are the household's response, you would have to decide which is more likely to be correct.

Q. What edits are performed on SAMEHH?

The 1999 version of the SAMEHH (Same household members live in unit) variable is improved from the 1997 version, but not perfect. In 1997, the questionnaire instrument initialized the value of SAMEHH as 'zero' and then changed the value to either '1' (same household) or '2' (new household) as data was collected to determine the situation. Unfortunately, the instrument did not manage to identify all the possible locations that should have reset the value of SAMEHH, and so many households were left with the original default value of 'zero'. In fact in 1997, about two-thirds of the interviews had a value of 'zero' in SAMEHH. Research showed that these cases included both new and returning households, and there was no way for users to distinguish between the two.

In 1999, there are no values of 'zero' in SAMEHH. This was easily accomplished by initializing the variable as '1' rather than 'zero.' However, besides this 'improvement', staff also reviewed the survey instrument to plug the holes in coding SAMEHH. A review of the data (comparing prior year and current year names) revealed that in 1999, the great majority of households seem to have the appropriate value in the variable SAMEHH. Due to the complexity of the instrument section that collects the demographic information, there are still some cases where SAMEHH may be incorrect. For example, fewer than 500 persons (out of 102,000 plus) are in households that may be mislabeled as in new households (although they also lived at the unit the previous interview period) and no more than 1,000 persons are in households that may be mislabeled as returning units. These figures are overstatements of the problem, both because SAMEHH is a household, not a person-level variable and because the data for some of these cases is so self-contradictory that the true situation cannot be determined short of re-contacting the sample household.

Q. What is the relationship between SAMEHH and SAMEHH2?

In the instrument:

1. All cases are initially set as REPLHH = 2 (not a replacement hh) at page 20 of QCODE; for vacant and noninterview units, they skip past all the succeeding points in the instrument that change the value of REPLHH, so theoretically all vacant and noninterviews should have SAMEHH = 1
2. At CC_BEGIN (page 134) which is reached by occupied units, cases that were prior year URE, vacant or noninterviews are directed to SET_REPLACE (page 139) where REPLHH is reset as 1 (replacement hh). Cases with INCSAM = 0 (returning units) and are not prior year URE, vacant or noninterviews are sent to STILIV1, and the remainder are sent to SET_REPLACE (page 139) where REPLHH is reset as 1 (replacement hh)
3. At STILIV1 (page 134), if all persons from prior year are still residents (or answer is DK or Ref) then REPLHH reset as 2 (same household); answers of "no" got STILIV2.
4. At STILIV2 (page 136), if any person from prior year is still a resident or answer is DK or Ref then REPLHH is left as 2 (same household); answers of "no" got CHECK_REP
5. At CHECK_REP (page 136), FR asked to confirm that this is a replacement household, if "P" (for "proceed") then REPLHH is reset to 1 (replacement household)
6. There are 2 places in the instrument CK_LEFT3 (page 138) and CK2_LEFT3 (page 165) which compare the number of persons deleted from the household roster with the total number of persons on the roster; and if all persons are deleted then REPLHH is reset to 1 (replacement household), otherwise it is reset to 2 (same household).
7. Finally at SET_REPLACE (page 139), cases routed from CC_BEGIN above, are set to REPLHH equals 1 (replacement household) In the processing: then the PUF version of the variable, called SAMEHH, is created in the bridge; this code sets SAMEHH = 3 when STILIV2 = DK or REPLHH = D, else SAMEHH = REPLHH, unless REPLHH = 1 then SAMEHH = 2, or REPLHH = 2, then SAMEHH = 1.
8. There are no edits that change the value of SAMEHH.

DISCUSSION:

The code above should produce an expected outcome for SAMEHH. However it does not. Theoretically all cases that are not interviews should have a value of '1' in SAMEHH but this is not the case. (Whether that should be the appropriate value is a different question.) There are a little over 1,100 cases in 2005 that are noninterviews, vacants or URE interviews but have a value in SAMEHH of '2'. As far as it can be determined, some of these cases changed interview status during or after the interview but the entry in SAMEHH did not change. For other cases, the reasons are unknown.

For occupied units, SAMEHH appears to generate correctly. An overview comparing SAMEHH by current and previous occupants' names showed the appropriate entry for the situations.

The new variable SAMEHH2 should be better than SAMEHH. It is generated as a recode after the edits are done and is based on current and prior status and well as SAMEHH. So it would have the correct parts of SAMEHH (the information regarding actual changes in the people) and the correct parts of the changes in interview type. SAMEHH2 is also an improvement since it provides more information on the reason the unit is not the same household. {2/13/2008}

Q. How is SMSA coded in the national surveys?

See the entry for this variable in the codebook.

Three pseudo MSA codes, 9991, 9992, and 9993, were added in 1995 to preserve confidentiality for part of the supplemental sample that was added for the 6 cities that year. (Only these 3 MSAs needed the extra protection.) There are parts of these smaller MSAs that cannot be identified on the public use file because they do not meet the 100,000-population rule. Normally, they would have shown up on the public use file as an unidentified MSA. However, since they were added to beef up the sample in these metropolitan areas, we wanted to show that they were indeed part of that larger metro area but at the same time not reveal exactly which part of the metro area. To use 9992 as an example, parts of 5380 cannot be identified by themselves. Therefore, they have been combined with parts of 5600 to form 9992. That way, a user can not know for sure if the case is from 5380 or 5600, and we have preserved the confidentiality of the respondent. The user does need to include these pseudo MSA codes in his tabulations if he is tabulating Chicago, New York, or Northern New Jersey.

Q. In pre-1997 surveys, what is the difference between the variables NOISE and STRN?

The STRNx variables come from the section of the questionnaire that specifically asks about street noise. By contrast, NOISE comes from the section a little further down which asks people to list the things about the neighborhood that bother them. Surprisingly, there isn't a strong correlation between NOISE and any of the STRNx variables.

Q. In 1997 and later surveys, what is the relationship among the street noise variables (STRNA, STRNB, STRNC)? And does the variable NUTRAF still exist?

The old variable STRN has been split into three variables:

- STRNA Neighborhood has heavy street noise/traffic
- STRNB Neighborhood street noise/traffic bothersome
- STRNC Street noise/traffic so bad you want to move

These replace NUTRAF.

The difference among the variables can be seen in tracking through the original q-code. Respondents are first asked if the neighborhood has heavy street noise or traffic. This information is stored in the STRNA variable. Only if they answer yes do they get the question of if it is bothersome (STRNB variable). If they answer yes to that question, they are asked if it is bothersome enough to move (STRNC variable). So, from a reading of the Q-Code that drives the interviewing process, STRNC is a subset of STRNB, which is a subset of STRNA.

That being said, the NOISE variable is collected independently of the STRNx variables. That is a separate set of questions asking, "Is there anything about the neighborhood that bothers you?" If yes, then the respondent is asked for a list of items and can independently respond "Noise." These questions and skip patterns can be found in the neighborhood quality module of the q-code.

Q. How is TENURE different from TEN and HHTEN? Why is HHTEN always 'X'?

TENURE is a unit-level variable that tells you whether the *unit* is owned, rented, etc. TEN is a person-level variable that tells you whether a particular *person* is an owner (named on the title) or renter (named on the lease) of the unit. TEN is coded as 'X' if the person is an owner or renter and is missing otherwise. Note that you would have to check TENURE to tell whether the 'X' means owner or means renter in a particular case. While the reference person will always have TEN='X,' other members of the household may or may not have this value.

HHTEN is one of the "HH..." variables on the NEWHOUSE file. These contain the characteristics of the householder. We include them on NEWHOUSE so that users can run tabulations by characteristic of the householder without having to link to the PERSON file. By the logic of how TEN is coded, HHTEN will always be 'X.' However, we keep it on NEWHOUSE for the sake of completeness.

Q. How is TENURE allocated for "Type A" noninterviews?

1. Tenure is collected early in the interview. We may get an answer to the question on tenure even though we do not collect enough information on the household for the interview to be considered complete (i.e. it ended up being a type A). The answer we get for tenure in one year may or may not be the same as it was in the previous year.
2. If we do not get tenure for a noninterview, but we do have a value from the previous survey that was not the result of any editing or allocation (i.e. it was an actual answer), this tenure is copied into current survey. In this case, the two years will be the same tenure.
3. If we do not get tenure for a noninterview and we either do not have a previous answer or the previous answer is the result of editing/allocating, the 1999 tenure was cold decked (3 times owner and 2 times renter). This answer has a high probability of being different from any answer cold decked in the previous survey.

Units that are type As in one year are likely to be problem units. There is a good chance that they did not have a legitimate tenure answer in the previous year either. Thus, it would not be surprising if these units show different tenures in different years.

Q. How is TPARK coded?

1 thru 20 = number of mobile homes in group
21 = 21 mobile homes or more

Q. How do the utility cost variables interact?

Using gas as an example:

1. Does the unit use gas at all?

BUYG NE 1

2. Does the unit pay for gas?

AMTG >= 1

3. Does the unit pay for gas separately?

AMTG >= 1 AND BILLG = B

a. How much?

value in AMTG

4. Is the gas payment combined with other utilities?

BILLG = 2 (note that AMTG = B in this case)

a. Which ones?

whichever of BILLG_x has an X (see part b, it will be BILLGE)

b. What is the total?

amount in AMT_x of the first of the group listed. Since only electricity precedes gas in question order, this works out to AMTE

5. Is this thing called the gas payment really a combined payment?

GBILL = 2 (yes, a combined payment);

a. For what other utilities?

whichever of BILLOG, BILLFG, BILLTG or BILLWG has an X (note: there is not a "BILLEG" variable due to the priority rule, e.g. the amount goes with the 1st listed of combined items, so if gas was with electricity the conditions in point 4 would apply)

6. Does this amount for utility X include gas in it as well?

xBILL = 2 AND BILLG_x = X

Q. Does ZINC2 not include all the income of all the household members?

In 1999, we stopped collecting income for household members 14 and 15 years of age, but that is the only change. In 1999 ZINC2 should have the income of all household members 16 years of age and older.

Q. How are TLINE1, TFAM1, TDORM1, THOSP1, TJAIL1, TSHEL1, TSRO1, TBUS1, and TPARK1 coded?

These variables are in the alternate housing supplement. There is one set for each person in the household (labeled 1, 2, etc.).

TLINE1 PLine number on roster
TFAM1 Days spent with friends or family
TDORM1 Days spent in dormitory or barracks
THOSP1 Days spent in a care facility or other institution, such as a hospital
TJAIL1 Days spent in jail or prison
TSHEL1 Days spent in a homeless shelter or other shelter
TSRO1 Days spent in a SRO
TBUS1 Days spent in a bus or train terminal.
TPARK1 Days spent on a street, in a park.

Q. Why does WEIGHT take such extreme values (such as from 19 to 25,713 in 2005)?

These values are generally artifacts of unusual circumstances in drawing the sample.

Small weights can come from a unit picked up when the Census Bureau was conducting Components of Inventory Change procedures and was identifying and interviewing all changes in a structure, and not just those that involved AHS-N sample cases. The probability of selection for such cases would be the probability that any unit in the structure was selected for AHS-N. The bigger the structure, the larger the probability of selection, and the smaller the weight (which is the inverse of the probability of selection). Cases in large structures thus have very small weights.

The large weight is for a unit that was selected from a frame of housing units missed in the 1980 Census. This frame was used by several surveys, and so we had to share these missed units. Thus, we had to live with a small sample size of such units, which meant they had to be assigned larger weights.

Q. Why is YRMOR sometimes less than WHNGET?

A big part of the problem is probably the fact that WHNGET is edited to bring it into agreement with MOVE and BUILT. YRMOR is not edited to bring it into agreement with WHNGET. So it is quite possible that for some cases YRMOR has a value that seems to be inconsistent with WHNGET because of the edited values in WHNGET. It is also possible that edits did not cause

the problem but that the respondent gave answers to WHNGET in the RET module that are inconsistent with the answers in the MORTGAGE module.

In order to change this in the future we would have to add some edits for the mortgage module. We would have to be careful if we did this however as it would also change the data for MATBUY and NEWMOR as MATBUY, NEWMOR, and YRMOR are all brought into agreement in the edits. {2/29/2008}

Q. Is the variable ZSPEC still available in the 1999 survey? If not, can it be reproduced, especially the three categories of ownership?

ZSPEC is not available. Here is how I would reproduce it:

```
/* Vacant, URE, Noninterview */
IF Status ^= '1' THEN
    ZSpec = 9;

/* Owned, one unit, < 10 acres, no business */
ELSE IF Tenure = '1'
    AND '1' <= NUnit2 <= '2'
    AND Lot <= 439999
    AND DRShop ^= '1'
    AND Condo = '3' THEN
    ZSpec = 1;

/* cooperative or condominium */
ELSE IF Condo = '1' THEN
    ZSpec = 2;

/* other owner-occupied */
ELSE IF Tenure = '1' THEN
    ZSpec = 3;

/* renters, excluding single unit on 10+ acres */
ELSE IF (Tenure = '2' OR Tenure = '3')
    AND NOT (NUnits = 1 AND Lot >= 440000) THEN
    ZSpec = 4;

/* other renters */
ELSE IF (Tenure = '2' OR Tenure = '3') THEN
    ZSpec = 5;
```

Q. How were ZONES defined in the metropolitan PUFs?

In the early 1980s, the city of Chicago wrote to HUD asking if the areas they defined as "neighborhoods" could be broken out in some way on a file so that they could use the AHS to do more detailed studies. HUD and Census put tracts together into zones that would satisfy the

confidentiality requirements and closely approximate Chicago's "neighborhood" definitions. Having done that, HUD thought this was a good enough idea to apply it to the other metro areas. So HUD people (primarily Kathy Nelson) sat down again and defined "neighborhoods" for the other cities and Census people put the tracts together to satisfy those definitions as well.

Several difficult-to-find papers describe the process of defining zones in more detail. These include:

- Nelson, Kathryn P. 1981. Defining "small" areas within Annual Housing Survey SMSAs. Paper delivered at Southern Regional Demographic Group, Little Rock, AK, October 16.
- Turner, Margery and Edwards, John. 1993. "Affordable Rental Housing in Rental Neighborhoods." In Kingsley and Turner. *Housing Markets and Residential Mobility*, Urban Institute Press.
- Nelson, Kathryn P. and Edwards, John. 1993. "Intra-urban Mobility and Location Choice in the 1980s." In Kingsley and Turner. *Housing Markets and Residential Mobility*, Urban Institute Press.

A paper copy of Nelson 1981 is in the files at HUD. See the FAQ contact about obtaining a copy.

These papers deal with redefining zones after the 2000 revision of the metropolitan area definitions by the Office of Management and budget:

- Vandenbroucke, David A. 2005. "Sausage making at the AHS: Metro Survey Zones." Paper delivered to the Mid-Year Meetings of the American Real Estate and Urban Economics Association, May 31.
<http://www.areuea.org/conferences/papers/download.phtml?id=693>
- Vandenbroucke, David A. 2008. "The Improved AHS Sausage Machine, or American Housing Survey Metropolitan Zones." Paper delivered to the Southern Regional Science Association, March 28. [Electronic copy available from author.] {3/5/2008}

Variable Name Glossary

This appendix lists all the variable names mentioned in this document, along with their descriptive labels. Note that some of the variables are no longer used in current surveys, some were used only in certain supplements, and a few are Census Bureau internal variables that are not publically available. This list is intended only as a way of checking on variables while reading the FAQ. For a comprehensive list of variables, consult the codebook for the appropriate survey year.

<u>NAME</u>	<u>LABEL</u>
ACCESSB	Entry system required to access building
ACCESSC	Entry system required to access communit
AGE	Age of person
AGERES	Age restricted development
AMTE	Average monthly cost of electricity
AMTG	Average monthly cost of gas
AMTX	Annual real estate tax payment
BADPER	People in neighborhood are bothersome
BADPRP	Undesirable nghborhd property bothersome
BADSRV	Poor city/county services are bothersome
BIGP	Area of peeling paint larger than 8 x 11
BILLFG	Other fuels billed with gas
BILLG	Type of billing for gas
BILLGE	Gas billed with electricity
BILLOG	Fuel oil billed with gas
BILLTG	Garbage/trash billed with gas
BILLWG	Water/sewage billed with gas
BUYG	Pay for gas separately
CONTROL	Control number
CPRICE	Cost of construction plus value of land
CRACKS	Open cracks wider than dime
CROPSL	Receive farm income
DEGREE	Average heating/cooling degree days
DENS	# of dens/libraries/tv rooms in unit
DISTJ	# of miles traveled to work
DLINE	Line number of respondent
ESFD	Single family homes within 1/2 block
FAMNUM	Family number
FAMREL	Family relationship
FRSTHO	Ever owned home before
GATED	Walls/fences surrounding community
HBATH	Bathroom designed for wheelchair present
HCANE	Someone has a cane, walker, or crutches
HCART	Someone has motorized or electric cart
HCHAIR	Someone has a wheelchair

<u>NAME</u>	<u>LABEL</u>
HDEV	Someone has any other device
HELC	Has a home equity line of credit
HELCN	# of home equity lines of credit
HELEV	Elevator present
HELUMN	# of lump sum home equity loans
HELUMP	Has a lump sum home equity loan
HETYP1	1st home equity loan type
HETYP2	2nd home equity loan type
HETYP3	3rd home equity loan type
HFLAS	Flashing lights present
HH1	Hearing problems require the use of the first feature
HHELP	Someone has the help of another person with their limitation
HHMOVE	Year householder moved in
HHNDL	Door handle instead of knobs present
HHTEN	Householder is owner/renter of unit
HI1	Getting around inside the house requires the use of the first feature
HKIT	Home has kitchens designed for easier accessibility
HO1	Entering and exiting this residence requires the use of the first feature
HOLES	Holes in floor
HOTHR	Other special features present
HP1	Personal activities require the use of the first feature
HPHON	Home has specially equipped telephone
HPUSH	Push bars present
HRAIL	Extra handrails or grabbers present
HRAIS	Home has raised lettering or braille
HRAMP	Ramps present
HSCKT	Special wall sockets or light switches present
HSINK	Special sink, faucets, or sandwiches present
HV1	Vision problems require the use of the first feature
HWIDE	Extra wide doors present
IFBLOW	Fuses blown or circuit breakers tripped
IFDRY	Unit completely without running water
ILEAK	Any inside water leaks in last 12 months
INUSYR	Year came to US
IPOV	Poverty income
ISTATUS	Interview status
KITCH	# of kitchens in unit
KITCHEN	Unit has complete kitchen equipment
LDMAYB	Seller/agent said lead paint possible
LEAK	Any outside water leaks in last 12 mnths
LITTER	Litter in neighborhood bothersome
LIVING	# of living rooms in unit
LMED	Area median income (average)

<u>NAME</u>	<u>LABEL</u>
LPRICE	Purchase price of unit and land
MATBUY	Got 1st mortgage in same yr bought unit
MCNT	# of regular mortgages
METRO3	Central city / suburban status
MG	Any mortgages on this property
MORTIN	Type of 1st mortgage
MOVE	Year person moved in
MOVGRP	Recent mover group
MOVM	Month person moved in
MOVYR	Year person moved in
MRTYP1	Type of 1st mortgage
MRTYP2	Type of 2nd mortgage
MVG	Recent mover group
NEWMOR	1st mortgage new or assumed
NOINT	Noninterview reason
NOISE	Noise in neighborhood is bothersome
NOPROB	Nothing bothersome in neighborhood
NORC	Majority of neighbors 55+
NPROBS	Anything bothersome in neighborhood
NUNITS	# of units in building
NUTRAF	Traffic in neighborhood is bothersome
OTHNHD	Other problems bothersome in neighborhd
OWNPHA	Unit owned by public housing agency (Census internal variable)
PAPHLT	Received pamphlet about lead paint
PER	# of persons in household
PERSON	Line number of person
PLINE	Line number of person
PRENT	Amount of rent actually paid
PROJ	Bldg owned by public housing authority
PVALUE	Current value of unit
PWT	Pure weight - inverse of prob of selectn
RAC	Cost of replacements/additions to unit
RACE	Race of person
RACOST	
RAD	Cost of alteration/repair
RAH	Hhld member performed alteration/repair
RAN	# of replacements/additions made to unit
RAS	Type of alteration/repair - 1999 version
RATS	Rats seen in unit recently
RECRM	# of recreation rooms in unit
REGION	Census region
REGMOR	# of regular mortgages on unit
REL	Relationship of person to reference person
RENT	Monthly rent

<u>NAME</u>	<u>LABEL</u>
RESPTYP	Type of respondent
RMOV	Mover group number
SAMEDU	Same HU as last enumeration
SAMEHH	Same hhld members live in unit
SAMEHH2	Same hhld members live in unit(recode)
SEX	Sex of person
SMSA	1980 design PMSA code
STATUS	Interview status
STRN	Street noise present
STRNA	Neighbrhd has heavy street noise/traffic
STRNB	Nghbrhd street noise/traffic bothersome
STRNC	St. noise/traffc so bad you want to move
TBUS1	Days spent in a bus or train terminal
TDORM1	Days spent in dormitory or barracks
TEN	This person is owner/renter of unit
TENURE	Owner/renter status of unit
TFAM1	Days spent with friends or family
THOSP1	Days spent in a care facility or other institution, such as a hospital
TIMEJ	Length of trip to work
TJAIL1	Days spent in jail or prison
TLINE1	PLine number on roster
TPARK	# of mobile homes in group
TPARK1	Days spent on a street, in a park
TSHEL1	Days spent in a homeless shelter or other shelter
TSRO1	Days spent in a SRO
UNITSF	Square footage of unit
VACANCY	Vacancy status
VALUE	Current market value of unit
VOTHER	Total non-wage income for family members
VOTHER2	Total non-wage income for household
WEIGHT	Final weight using 1980 geography
WGT00_90	Final weight using 1990 geography and census 2000 controls
WGT90GEO	Final weight using 1990 geography
WHNGET	Year unit bought/obtained/received
YRMOR	Year 1st mortgage obtained
YRRND	Unit suitable for year round use
ZADEQ	Recoded adequacy of housing
ZINC2	Household Income
ZONE	Metropolitan survey zone number
ZSPEC	Recoded renter/owner code