Documentation of Changes in the 1997 American Housing Survey

July 2001

For:
Office of Policy Development
and Research
U.S. Department of Housing and Urban
Development

By:
ICF Consulting

ICF
CONSULTING

Table of Contents

1. Acknowledgements

This document was developed by ICF Consulting for HUD's Office of Policy Development and Research under Contract Number R-OPC-21229, Task Order 2.

ICF would like to thank Ron Sepanik and Dav Vandenbroucke at HUD and Jane Kneessi, Paul Harple, Barbara Williams, Anne Jean, and Tom Blatt at the Census Bureau for their valuable contributions.

Staff at ICF who worked on this document include: Gregory Watson, Heather Gazan, and Kevin Blake.

Questions about the document can be directed to:

At ICF Consulting:

Gregory Watson
Project Manager
ICF Consulting
9300 Lee Highway
Fairfax, VA 22031
Email : `gwatson@icfconsulting.com`
Voice: 703-934-3694
Fax: 703-934-3156

At the U.S. Department of Housing and Urban Development:

Ronald Sepanik
Division Director
Housing and Demographic Analysis Division
Policy Development & Research
U.S. Department of Housing and Urban Development
451 Seventh St., S.W.
Washington, DC 20410
Email: `Ronald_J._Sepanik@hud.gov`
Voice: 202-708-1060 x5887
Fax: 202-708-3316

2. Introduction

In 1997, the U.S. Department of Housing and Urban Development (HUD) and the U.S. Bureau of the Census changed the way in which American Housing Survey (AHS) data is collected and processed. Prior to 1997, information from approximately 50,000 households was collected by interviewers visiting or calling survey respondents to ask them questions printed on a paper survey. Interviewers would record all responses on the paper survey form for each respondent and then send these surveys to Census where the data would be coded and entered into a database. Using this process, the interviewer was responsible for following the proper pattern of questions based on the responses received during the interview. For example, if the respondent reported that the household was a renter-occupant of the property, the interviewer would skip the set of questions about mortgage payments, interest rates, and real estate taxes. In 1997, the process of using a paper survey to ask questions and record responses was eliminated in favor of a computerized system called CAPI – Computer Assisted Personal Interviewing. Using CAPI, interviewers were able to ask questions and record responses directly into a laptop computer provided to them. More importantly, CAPI determined for the interviewer which questions should be asked based on past responses, effectively automating much of the decision-making process that was previously left to the interviewer.

As a result of the new CAPI system, the interviewing process became much quicker and the data entry process was greatly reduced. The introduction of CAPI also provided Census and HUD with an opportunity to introduce other changes to the survey – all intended to improve the quality of the data and collect more accurate information. These changes ranged from a refinement of some of the survey questions to a change in the universe of respondents for specific questions. These refinements, while intentional, changed the data in a way that – depending on researchers' ultimate analysis purposes – prohibits some longitudinal analysis without a caveat explaining the changes in the data. Unfortunately, unintentional changes were also made to the survey and, as a result, to the data. Some questions were asked of the wrong universe of respondents in 1997 and the wording of certain questions and the answer choices given to respondents also changed slightly.

As part of the process to ensure a high level of data quality and integrity, HUD contracted with ICF Consulting in the summer of 1998 to work closely with both HUD and Census to replicate 17 of the tables printed in the AHS reports and compare results with Census. Through this process, data entry errors, data mis-codings, changes in the data collection process, and problems in the data collection process that should be addressed in future surveys were discovered.

This document describes the most common and important changes that were made to the AHS in 1997. First, the overall survey issues, particularly the move from paper to CAPI – are described with details about the changes to specific individual variables. Intermediate-level changes are then described. These changes generally refer to issues that are not survey-related, however they do pertain to the data in its entirety. Micro-level issues are then described. These issues speak to specific changes to individual variable. Throughout the document, each change is described, the reason for the change is discussed, and the effect(s) of the particular change are documented.

In addition to describing changes to the data, certain issues that have been a source of confusion to users of the AHS are clarified.

This document brings together some of the key findings and issues for AHS data users identified by ICF Consulting, HUD, and Census during this quality control process. In some cases, the document recommends referring to working documents released by HUD for further detail or other examples. This document is intended to complement – but not substitute for – the other valuable documents about the AHS. These documents are listed in the section of this report entitled "Other Documents and Resources" (page 60).

3. Macro-Level Changes

The modifications of the data collection and processing stages in 1997 affected all subsequent analyses of the AHS. There were several HUD and Census Bureau data collection and processing policies and procedures that changed in 1997, many of which are visible to the users of the AHS. These macro-level changes are discussed in this section.

3.1. Conversion to CAPI

Before 1997, the primary tool for data collection for the AHS was the paper-based survey instrument. Census Field Representatives (interviewers) completed the instrument, either in person or by telephone and data from the completed interviews was then key punched and entered into a computer system. A copy of the paper questionnaire that was used in the 1995 AHS – and is representative of earlier years – can be found in Volume 1 of the AHS codebook.

In 1997, the Census Bureau changed from the existing survey process to a paperless system. This new system – Computer Assisted Personal Interviewing (CAPI) – was implemented to accomplish a number goals. First, the system is intended to make the interviewing process faster and more accurate. The pattern of questions posed to respondents is no longer left up to the interviewer with CAPI, instead the computer makes those determinations. CAPI automatically follows the question skip pattern, thus speeding up the interview and ensuring that respondents are asked the proper questions. Second, the conversion was intended to reduce costs and speed up the release of data to the public. No longer is it necessary to enter data from the paper survey into a database, now the information is entered directly into the system. This process also reduces the introduction of errors that generally occur during the data entry process. Finally, CAPI was also implemented in an effort to reduce the burden on respondents with some questions asked only once and then verified in later years. This process is only used for variables thought to be fairly constant over time. This technique, also known as "dependent interviewing," can lead to significantly faster interviews and additional data reliability in the future. This technique was partially introduced in 1997 and fully implemented in 1999.

The implementation of CAPI led to several unforeseen changes in the data collection process, which in turn led to inadvertent changes in the AHS data. These changes are documented here and in other documents published by HUD and the Census Bureau. HUD and Census are also continuing to remove potential problems or errors in the data for future use.

While there is no published questionnaire available for 1997 and later years of the AHS, the questions and the question interview pattern are available as part of the Field Representative's guide for the AHS released on the Census Bureau Web site. The computer code controlling the interviews, known as "Q-Code" ("Questionnaire Code"), can be found on the HUD USER Web site. The Field Representative's guide is much easier to read than the Q-Code but does not include information on skip patterns or the exact rules for question flow.

3.2. New Data Processing Structure

Along with new data entry procedures, the AHS data are now manipulated and processed using a completely new set of software. In years prior to 1997, Census manipulated the data using custom FORTRAN programs developed internally, running on a UNISYS mainframe. In 1997, Census moved the entire processing of the AHS into SAS running on SUN Microsystems workstations. This change to SAS necessitated redesigning the processing system and writing new programs to produce the tables in the printed report. As part of the conversion, Census designed their programs to take advantage of some of the special features and functions available in SAS, which in turn led to changes in variable codes.

3.3. New Data Structure

Following the changes in data entry and processing, the Census Bureau took advantage of the flexibility of the SAS software to generate the output data files in a new format. Prior to 1997, the public use file (PUF) of the AHS was a single text (ASCII) file with a "flat" data structure. In such a file, each housing unit's information is contained on a single logical record. The file has a set number of characters for each observation with each variable having a fixed location in the file. This file format that had been established many years ago, however, and was not Year 2000 (Y2K) compliant – all variables containing years were allocated only two digits rather than the four digits required. In the process of changing the data structure to make the data file Y2K compliant, Census took advantage of the opportunity to enhance the data format in other ways as well.

The "raw state" of the AHS data is now a collection of 10 files, each containing data in a different format or layout. Each record type corresponds to one or more module(s) in the data collection instrument or processing. This new file format removes the issue of fixed variable lengths and allows for a wider range of values. In addition, the new file format corresponds to the actual data collection process. Additional information on the files can be found in the section on file formats on page 56. The files are:

- HOMIMP – questions about upgrades and remodeling

- HOUSEHLD – household level information (on internal Census data, this file is called "houshld")

- JTW – journey to work or commuting information

- MORTG – mortgage information

- OWNER – questions just for owners of rental properties

- PERSON – individual person level information

- RATIOV – verification of income to cost when the ratio of income to cost is outside of certain tolerances. (Note: This file is present beginning with the 1998 AHS metro sample. This file is not present in 1997.)

- RMOV – recent mover information

- TOPPUF – general unit characteristics and data recodes

- WEIGHT – weighting information

At first, the files were available only in a SAS format. Later, HUD created ASCII (text) versions of the files and made them available to users. The ASCII files are not a Census Bureau product. The ASCII versions have the same multi-file layout as the SAS versions, but differ in the coding of missing values. For a description of missing value codes in both file formats, please see page 11.

The general structure of the collection of files in 1997 is different from the 1995 file structure and earlier versions. The 1997 files are more analogous to a relational database so that it is **not** possible to simply merge the files together to create the flat file. Each file is based on the unit of observation for that file. A file, for example, on the housing unit structure has one observation for each housing unit while the file on people living in the housing unit has one observation for each person in the household (and so if there are multiple persons in the household, there will be multiple observations). As a result, the files have different numbers of observations depending on base unit of observation.

Users can use the raw information in its current state, develop their own programs to read the data in and convert it, or convert the data into a format that is similar to the older format. ICF Consulting has also developed a conversion program that combines the different data files into a flat-file format that is similar to that of the 1995 dataset. The final format is a SAS dataset with a similar structure to the original flat file with each observation equal to one housing unit. That program can be found on the HUD User Web site (`www.huduser.org`) under the links for AHS data. The program is also included in the distribution of the data in SAS format from the HUD User Web site. Currently, the conversion program is written in SAS and works only in SAS. AHS users not using SAS will need to convert the programs to the system they are using or use a program developed by another AHS user. The AHS electronic mailing list exists, in part, to facilitate such exchanges. The Census Bureau and HUD do not officially support the flat file format. Further discussion on this topic can be found in "Discussion of the File Flattening Program" on page 55.

The HUD USER Web site also has a page to distribute user-contributed programs developed using the AHS. These user-contributed programs can be found at `http://www.huduser.org/datasets/ahs/ahsprev.html`. No warrantees about the suitability of these programs are provided by HUD. If you wish to contribute to this library of programs, please send them to: Dav Vandenbroucke at `David_A._Vandenbroucke@hud.gov`.

Census and HUD plan to change the data format for the 2001 AHS. This change will involve combining the WEIGHT, TOPPUF, and HOUSEHLD files into one file. This change will allow users to perform a greater number of household level analyses without having to merge the component data files together.

3.4. Other Access to the AHS Data

Not all users need the entire set of micro-data to perform their analyses. Subsets of the data are available from the Census through the FERRET System on the Census Bureau Web site. This Web-based system allows for the extraction of a list of variables instead of the whole data set and also allows the online creation of simple custom tabulations. The custom tabulations can be weighted or unweighted as requested by the user.

The FERRET System works with data from the years 1997 and later. The FERRET System supercedes the Data Extraction System (DES) that works with the 1993 and 1995 AHS data. Current Web site addresses for each system are located in the Appendix on page 62.

4. Intermediate-Level Changes in the AHS

This section discusses some of the changes that may have resulted from the processing of the data or may have affected the characteristics of the data.

4.1. Missing Values

One of the changes in the 1997 AHS is the way in which missing values are coded. Previously, different types of missing values were coded using all 9's and then counting down by one from there; for example, 9999 = Not applicable, 9998 = Not answered, etc.

The 1997 AHS changed the coding of missing values by taking advantage of the SAS missing value functions. This function allows different types of missing values to be assigned different values. Analogous missing values are also present in the ASCII version of the microdata.

In the SAS data files, the missing values are as follows:

B = Not applicable
D = Don't Know
R = Refused
Blank or **.** = Not reported

In the ASCII text version of the data, the system of coding missing values is more complicated. When the range of valid values for a particular variable contains two or more digits *or* the largest valid value is greater or equal to 6, then the codes are the following:

-9 = Not applicable
-8 = Refused
-7 = Don't Know
-6 or Blank = Not reported

Otherwise, when the valid values for a variable have only one digit *and* the largest valid value is less than six, then the coding for missing variables is:

9 = Not applicable
8 = Refused
7 = Don't Know
6 or Blank = Not reported

This change in the coding of missing values means that programs developed to use previous years of AHS data may need to be revised to run properly and accurately. If programs are not adapted, the program may crash, report errors, or produce erroneous results.

4.2. Disagreement Between Data, Documentation, and Reports

Although extensive efforts were undertaken to avoid the introduction of errors into the 1997 AHS, a few discrepancies exist among the data, the data documentation, and

the printed reports. Generally, errors are more likely to be present in the documentation than in the data or the printed report. Errors exist in the documentation because the data documentation was issued before the quality control process began and has not been changed or updated to reflect changes resulting from the quality control process. Errors that were discovered in the data or printed reports were corrected as a part of the quality control process for the AHS.

4.3. Suppression of Data

There are certain data that HUD and Census collected and originally intended to release publicly, but did not release for a number of reasons. Typically, Census suppresses data because of unacceptable disclosure risk or because the Bureau decides that the data are non-comparable or erroneous.

Examples of variables that have been suppressed include:

- SMSA in extremely small MSAs
- Condo/co-op distinction
- METRO variable (Note: Replaced by METRO3, page 16)
- Flush toilet breakdowns in 1997

Discussion of reliability, computation of sample errors, and certain other statistical issues can be found in Appendix D of *American Housing Survey for the United States: 1999* available in PDF format on the Census Bureau Web site.

4.4. Differences Between the Public Use AHS Files and Internal Census Files

The Public Use Files (PUF) released by the Census Bureau and the files used by the Census Bureau in the development of the tables for the printed reports are not the same. The internal files contain two types of data that are not released to the public: certain variables and non-top-coded data. This is not a change in 1997, but rather a topic where there is sometimes a lack of understanding among users.

Variables that are not released to the public are generally used either in the data collection process or created and used for internal calculations. These internal variables are often used for quality control purposes or to construct other variables that are released to the public. These internal variables can be identified in Volume 3 of the codebook. These variables have a blank value under the SEQ# column – the fourth column from the left. Variables released to the public have a number in the SEQ# column. There are a few cases in which variables were originally intended for release but were later held back and cases where variables were not originally intended for release and were ultimately included in the PUF. These, however, are unusual circumstances

The PUF has certain variables, primarily financial ones, which are top-coded by the Census Bureau. Top-coding is a process where all values above a certain level are set to an artificial ceiling. This editing of data is performed to ensure that respondent confidentiality is protected. In 1997, the top-code was set to be either the highest top-

code value for 1985 and later years or the 97th percentile, whichever value is greatest, with a few exceptions. For 2001 and beyond, the top-coded value for a variable will be equal to the 97th percentile in the previous year or the top value existing in the variable for the 1999 PUF, whichever value is larger. Some variables, mostly in the Alterations and Repairs module, will be calculated at the 99.5th percentile. There are exceptions to this rule, such as certain variables that will be top-coded at the $100,000 level. This top-coding is intended to provide an additional element of confidentiality for individuals at the extreme high or tail end of the distribution, who may be more visible than those individuals in the middle of the distribution.

There are a very few variables, such as Total income (ZINC and ZINC2) and "other income" (VOTHER), that are also "bottom-coded" at negative $10,000.

4.5. Calculations

Some information published in the printed AHS and available in the dataset is based on calculations using data collected from individual questions on the questionnaire. The Census Bureau performs these calculations to generate summary variables that are based on the component data that was reported.

For example, the family income variable (ZINC) is simply created by summing the income of all members of the family. Other information, such as outstanding mortgage principal, is calculated based on the mortgage terms and mortgage payment information provided by respondents.

4.6. Allocation Matrix Changes

While the Census Bureau makes all reasonable efforts to collect valid responses to all survey questions, respondents do not always provide information. Usually non-response occurs for several different reasons, ranging from not knowing the answer to refusing to answer the question. For most variables, the Census Bureau enters a missing value code for that response; however, for certain key variables, Census applies a procedure to allocate or assign values to those missing data. This process, known as allocation at the Census Bureau, is often referred to as imputation of missing values.

The Census Bureau uses a procedure called "hot-decking" in which a missing value is replaced with the last value for that variable for a unit with the same selected characteristics. The collection of possibilities for units with the same characteristics is referred to as an allocation matrix whose dimensions are based on a set of selected variables being allocated. Different matrices are used to create replacements for missing values for different variables or questions.

A more complete description of the allocation matrices and the allocation process can be found in the Appendix to Volume 3 of the codebook, the first year covering 1997. However, during the data quality control process – after Volume 3 of the codebook was released – some of the characteristics used to define the allocation matrices were changed. The variables used in the construction of the matrices were changed to produce what are believed to be more accurate allocations for missing values. As a result, the Appendix to Volume 3 is not current. HUD intends to release the current allocation matrices on the Web.

For exact details of each current allocation matrix definition, please contact the American Housing Survey Branch at the Census Bureau. Their telephone number is 301-457-3235 and email address is `ahsn@census.gov`.

4.7. Medians

One of the common statistics used to describe an aspect of data distribution is the median. The standard way of calculating a median is to sort the values in order and determine which value is the exact middle of the distribution. This is then the median value. In cases where there are an even number of observations, the median is then usually calculated as a point half-way between the two mid-points.

The Census Bureau has traditionally used an approximation to generate medians, especially when using discrete values, rather than calculating medians based directly on the raw data. This policy is unchanged in 1997, but is discussed here to inform users who may calculate medians differently from the Census Bureau. The method of calculating the median used by Census is approximately as follows:

1. Take the categories used in the printed reports.

2. Determine which category contains the median.

3. Determine what percentage of cases in the median category fall below the median, assuming a uniform distribution inside of the band.

4. Apply the percentage in step 3 to the range of the band to determine an estimated median value.

Note that in 1997 and earlier, this process is used for both continuous and discrete variables. Medians reported for discrete variables will be fractional. In 1999 and later, the AHS will not have medians calculated using this method for discrete or single-value categories, such as rooms and bedrooms.

Since this procedure is not the "standard" way to compute medians using microdata, the medians computed by statistical packages will be different from those published in the printed AHS reports.

5. Micro-Level Changes

5.1. Introduction

This section addresses changes to individual variables or small groups of variables on a particular topic. These changes are noted so that users of the data can better understand and use the data. Some of the variable changes described in this section result from intended variable changes, while others are unexpected changes in the values when compared with earlier years of data.

This section is divided into two sections: the first section addresses issues related specifically to the housing unit and the second deals with issues related to the household occupying the unit. Financial issues are included in issues relating to the household. Certain changes that refer to classes or groupings of variables can be found in the section beginning on page 49.

5.2. Housing Unit Characteristics

5.2.1. Metro

Item	Description
1997 and later variable name:	METRO3 – Metropolitan area status
Pre-1997 variable name:	METRO
Reason or Justification for change:	Census is unable to convert 1990-based urbanized areas into the 1980-based geography used by the AHS. This means that housing units added to the survey after 1994 cannot be classified by urbanized area. The METRO variable was simplified to take this into account.
Location of change:	Data file, codebook. This variable is computed by the Census Bureau.

Effect of change:	It is not possible to identify metropolitan area status to the same degree as pre-1997.
	
	The new coding is:
	
	    1 = Central City
	
	    2 = Suburb, urban (old METRO codes 2 and 3)
	
	    3 = Suburb, rural (old METRO code 4)
	
	    4 = Non-metro, urban (old METRO codes 5 and 6)
	
	    5 = Non-metro, rural (old METRO code 7)
	
	Note: The new coding has fewer distinctions than the older coding.

Is this change believed to be an improvement?	No. Slight loss of information in the data.
What should be considered when comparing different years of data?	The pre-1997 data cannot be compared completely with the 1997 and later data.
Is the change permanent?	The METRO3 variable is also present in 1999, and is presumed to be present in later years also.

5.2.2. Housing Quality

Item	Description
1997 and later variable name:	ZADEQ – housing quality
Pre-1997 variable name:	ZADEQ
Reason or Justification for change:	The ZADEQ variable is computed from a collection of other housing quality variables. The value of ZADEQ is dependent on both the component variables and the level of severity for the component variables: • plumbing (both presence and number of breakdowns) • heating (both presence and number of breakdowns) • electric (both presence and number of breakdowns) • upkeep (based on leaks, cracks, holes, rats, etc.) • hallways (based on hallway lights, steps, railings • kitchen Slight changes were introduced into the questionnaire, and to the creation of certain variables such as KITCHEN (see KITCHEN, page 24) and about plumbing breakdowns (see plumbing, page 19). Definitions were changed very slightly for the CRACKS and HOLES variables among others to make the questions clearer to the respondent.
Location of change:	Questionnaire and data processing. This variable is computed by the Census Bureau.
Effect of change:	Certain slight differences in ZADEQ from pre-1995 to later.
Is this change believed to be an improvement?	Yes.
What should be considered when comparing different years of data?	Due to the definitional changes, the data is not directly comparable.
Is the change permanent?	Yes.

5.2.3. Plumbing

Item	Description
1997 and later variable name:	PLUMB
Pre-1997 variable name:	PLUMB
Reason or Justification for change:	Coding of the variable has changed. The concept of exclusive use is no longer tied to each item and is asked only at units reporting one or fewer full bathrooms.
Location of change:	Codebook, data

Effect of change:	New coding. The new coding is:
	
	    1 = With all facilities
	
	    2 = Lacking some or all facilities

Is this change believed to be an improvement?	Not Applicable
What should be considered when comparing different years of data?	The coding across years is different, and so users cannot achieve the same level of detail.
Is the change permanent?	Yes.

5.2.4. Plumbing Breakdowns

Item	Description
1997 and later variable name:	IFTLT – If toilets/plumbing are broken
Pre-1997 variable name:	IFTLT
Reason or Justification for change:	Error in questionnaire. Questionnaire mistakenly asked about times <u>the</u> toilet was unusable, as opposed to times when <u>all</u> toilets were unusable. There also may be changes in the data because the questionnaire changed to ask exclusive use only for units with one or fewer full baths.
Location of change:	Questionnaire
Effect of change:	Data in 1997 about number of toilet breakdowns is erroneously high. In addition, since the housing quality variable (ZADEQ) is partially dependent on this information, more houses may be reported in "poorer" condition. (See Housing Quality, page 17)
Is this change believed to be an improvement?	No, the change is an error.
What should be considered when comparing different years of data?	The 1997 figures are likely to be inaccurately high.
Is the change permanent?	No, the question was corrected in the 1999 AHS.

5.2.5. Sewage Breakdowns

Item	Description
1997 and later variable name:	IFSEW – If sewage system broke down
Pre-1997 variable name:	IFSEW
Reason or Justification for change:	Error in questionnaire skip pattern where households which have public sewers and have reported inside water leaks miss being asked if they had a sewage breakdown.
Location of change:	Questionnaire.
Effect of change:	Published totals may be slightly low.
Is this change believed to be an improvement?	No.
What should be considered when comparing different years of data?	Error in the data collection may cause 1997 numbers to be slightly low.
Is the change permanent?	No.

5.2.6. Rooms

Item	Description
1997 and later variable name:	ROOMS – The number of rooms in the unit
Pre-1997 variable name:	ROOMS
Reason or Justification for change:	In 1997, the data were collected differently. Prior to 1997, the respondent was asked a series of questions about different types of rooms in their unit. (Questions 26a-c in the questionnaire.) In 1997, the respondent was asked to describe what rooms were in the home, going through floor by floor. (INVEN9a in ahsinven.q of the Q-Code.) This change was thought to provide clarity and accuracy in collecting information from respondents.
Location of change:	Questionnaire and data collection
Effect of change:	The data results, even for the same units, were somewhat different than previous years. Based on a review of the data, Census returned the questions to the pre-1997 format in the 1999 AHS.
Is this change believed to be an improvement?	The change was originally believed to be an improvement, but after reviewing the data collected, the processing method returned was changed to the pre-1997 method in 1999.
What should be considered when comparing different years of data?	There may be apparent inconsistencies in the number of rooms in a specific unit when tracking a unit across time. The 1997 data show fewer units with small numbers of rooms, perhaps because of double counting.
Is the change permanent?	No, the 1999 AHS is following the process used prior to 1997. 1997 is the only year with the change in processing method.

5.2.7. Business Rooms

Item	Description
1997 and later variable name:	EXCLUS – Number of rooms used exclusively for business BUSPER – Number of rooms used for business and personal use
Pre-1997 variable name:	OFFICE
Reason or Justification for change:	Error in questionnaire. The specifications for rooms questions mistakenly dropped finding out whether any room in the unit was used both for business space and other purposes. The questionnaire does not capture if rooms are shared business space.
Location of change:	Questionnaire.
Effect of change:	Incorrect data were collected.
Is this change believed to be an improvement?	No.
What should be considered when comparing different years of data?	Not possible to compare 1997 with other years. The 1997 information is suppressed in the report.
Is the change permanent?	No, corrected in 1999.

5.2.8. Heating Equipment

Item	Description
1997 and later variable name:	HEQUIP – Heating equipment
Pre-1997 variable name:	HEQUIP
Reason or Justification for change:	Considerable questionnaire changes to aid respondent identification of equipment type.
Location of change:	Questionnaire
Effect of change:	Differences between pre-1997 and 1997 published data.
Is this change believed to be an improvement?	Yes.
What should be considered when comparing different years of data?	The data are not directly comparable because there are significant questionnaire changes.
Is the change permanent?	Yes.

5.2.9. Kitchen

Item	Description
1997 and later variable name:	KITCHEN – Presence of complete kitchen facilities
Pre-1997 variable name:	KITCHEN
Reason or Justification for change:	Change in questionnaire and component variable. The KITCHEN variable reports whether or not the unit has complete kitchen facilities. Printed versions of the report required that an oven also be present. In 1997, the questions about kitchens changed, including the removal of the instruction for "this household's use only," and a replacement instruction for households in multiunit structures, if the household shared kitchen facilities. In addition, the questions now explicitly state that the kitchen equipment (refrigerator and cooking equipment) must be in working condition. Previously, that requirement was part of the interviewer's instructions but may not have always been told to the respondent. In 1997, the definition of OVEN has changed. (See OVEN, page 25.) As a result of these changes, both the data and printed tables are no longer comparable with pre-1997 results.
Location of change:	Questionnaire and printed reports.
Effect of change:	Data on kitchen are not directly comparable.
Is this change believed to be an improvement?	Unknown.
What should be considered when comparing different years of data?	Data on kitchen are not directly comparable.
Is the change permanent?	Yes.

5.2.10. Oven

Item	Description
1997 and later variable name:	OVEN – Presence of a microwave oven.
Pre-1997 variable name:	OVEN – Presence of a standard oven or a microwave oven.
Reason or Justification for change:	Questionnaire change, believed that asking about microwave ovens rather than traditional ovens was more relevant.
Location of change:	Questionnaire
Effect of change:	Data from pre-1997 and 1997 are not comparable since they refer to very different kitchen equipment.
Is this change believed to be an improvement?	The change is in information collected. The change is neither positive nor negative.
What should be considered when comparing different years of data?	The data is not comparable pre-1997 and 1997. The 1997 data is more focused than the pre-1997 data.
Is the change permanent?	Yes.

5.2.11. Wire

Item	Description
1997 and later variable name:	WIRE
Pre-1997 variable name:	WIRE
Reason or Justification for change:	Change in definition. A code of "1" for WIRE now includes wire enclosed in plastic coverings in addition to wire enclosed in walls or metal coverings.
Location of change:	Questionnaire
Effect of change:	Count in 1997 and later is higher.
Is this change believed to be an improvement?	Yes.
What should be considered when comparing different years of data?	Count in 1997 and later is higher.
Is the change permanent?	Yes.

5.2.12. Structure Type

Item	Description
1997 and later variable name:	TYPE – Structure type
Pre-1997 variable name:	TYPE
Reason or Justification for change:	1997 Codebook is in error. Coding is different from 1995 and earlier. The correct coding is: 1 = House, apartment, flat 2 = Mobile home, no permanent rooms added 3 = Mobile home, permanent room added 4 = Housing unit in nontransient hotel, motel, etc. 5 = Housing unit in transient hotel, motel, etc. 6 = Housing unit in rooming or boarding house 7 = Boat or recreational vehicle 8 = Tent, cave, or railroad car 9 = Housing unit not specified 10 = Unoccupied site for mobile home 11 = Group quarters
Location of change:	Codebook, data
Effect of change:	Codes 10+ are different from earlier years
Is this change believed to be an improvement?	Yes
What should be considered when comparing different years of data?	Coding of unusual values is different from earlier years, which use codes from 1 through 14.
Is the change permanent?	Yes. Codebook will be updated.

5.2.13. Abandoned or vandalized buildings within ½ block

Item	Description
1997 and later variable name:	EABAN – Abandoned or vandalized buildings within ½ block
Pre-1997 variable name:	EBAN
Reason or Justification for change:	1997 Codebook omits coding, which is the same as 1995 and earlier: 1 = Yes, one 2 = Yes, more than one 3 = No 4 = No other buildings within ½ block
Location of change:	Codebook
Effect of change:	None
Is this change believed to be an improvement?	Not applicable.
What should be considered when comparing different years of data?	Data should be comparable.
Is the change permanent?	There was no change. Codebook will be updated.

5.2.14. Age of other buildings within ½ block

Item	Description
1997 and later variable name:	EAGE – Age of other buildings within ½ block
Pre-1997 variable name:	EBAGE
Reason or Justification for change:	1997 Codebook omits coding, which is the same as 1995 and earlier: 1 = Older than sample unit 2 = About the same 3 = Newer than sample unit 4 = Very mixed 5 = No other residential buildings within ½ block
Location of change:	Codebook
Effect of change:	None
Is this change believed to be an improvement?	Not applicable.
What should be considered when comparing different years of data?	Data should be comparable.
Is the change permanent?	There was no change. Codebook will be updated.

5.2.15. Bars on windows of buildings within ½ block

Item	Description
1997 and later variable name:	EBARCL – Bars on windows of buildings within ½ block
Pre-1997 variable name:	EBARCL

Reason or Justification for change:	1997 Codebook omits coding, which is the same as 1995 and earlier:
	
	    1 = Yes, one building with bars
	
	    2 = Yes, more than one
	
	    3 = No

Location of change:	Codebook
Effect of change:	None
Is this change believed to be an improvement?	Not applicable.
What should be considered when comparing different years of data?	Data should be comparable.
Is the change permanent?	There was no change. Codebook will be updated.

5.2.16. Trash or junk in streets or properties within ½ block

Item	Description
1997 and later variable name:	EJUNK – Trash or junk in streets or properties within ½ block
Pre-1997 variable name:	EJUNK
Reason or Justification for change:	1997 Codebook error. Correct coding is the same as 1995 and earlier: 1 = Major accumulation 2 = Minor accumulation 3 = None
Location of change:	Codebook
Effect of change:	None
Is this change believed to be an improvement?	Not applicable.
What should be considered when comparing different years of data?	Data should be comparable.
Is the change permanent?	There was no change. Codebook will be updated.

5.2.17. Hallway lights working

Item	Description
1997 and later variable name:	LTSOK – Hallway lights working
Pre-1997 variable name:	LTSOK
Reason or Justification for change:	The 1997 codebook omits the coding of this variable, which is the same as in previous surveys: 1 = No public halls 2 = All in working order 3 = Some in working order 4 = None in working order 5 = No light fixtures in public halls
Location of change:	Codebook
Effect of change:	None
Is this change believed to be an improvement?	Not applicable.
What should be considered when comparing different years of data?	Data should be comparable.
Is the change permanent?	There was no change. Codebook will be updated.

5.2.18. Railings on common stairs firmly attached

Item	Description
1997 and later variable name:	RAILOK – Status of Railings
Pre-1997 variable name:	RAILOK
Reason or Justification for change:	The 1997 codebook omits the coding of this variable, which is the same as in previous surveys: 1 = No railings 2 = Railings not loose 3 = Railings loose
Location of change:	Codebook
Effect of change:	None
Is this change believed to be an improvement?	Not applicable.
What should be considered when comparing different years of data?	Data should be comparable.
Is the change permanent?	There was no change. Codebook will be updated.

5.2.19. Reason for vacancy

Item	Description
1997 and later variable name:	VACANCY – Reason for vacancy
Pre-1997 variable name:	VACANCY
Reason or Justification for change:	The 1997 codebook omits variable, which is the same as in previous surveys: 1 = For rent 2 = For rent or sale 3 = For sale 4 = Rented, not occupied 5 = Sold, not occupied 6 = Held for occasional use 7 = Other vacant 8 = Seasonal summer 9 = Seasonal winter 10 = Other seasonal 11 = Migratory Usual Residence Elsewhere (URE) cases can have any of the vacancy codes marked.
Location of change:	Codebook
Effect of change:	None
Is this change believed to be an improvement?	Not applicable.
What should be considered when comparing different years of data?	Data should be comparable.
Is the change permanent?	There was no change. Codebook will be updated.

5.2.20. Land Rent Frequency and Amount

Item	Description
1997 and later variable name:	FLRENT – Frequency of Land Rent LRENT – Land Rent amount
Pre-1997 variable name:	FLRENT, LRENT
Reason or Justification for change:	Questionnaire error, where certain groups of people missed being asked the question. Households who should have been asked but were not asked include: owner-occupied units not in mobile homes, condominiums, or cooperatives and owner-occupied units who do have land rent but not homeowners' association fees.
Location of change:	Questionnaire
Effect of change:	Certain groups not asked question, more allocations than typical.
Is this change believed to be an improvement?	No.
What should be considered when comparing different years of data?	More allocations than usual occurred.
Is the change permanent?	No.

5.3. Household characteristics

5.3.1. Income

Item	Description
1997 and later variable name:	ZINC, ZINC2 – family income and household income
Pre-1997 variable name:	ZINC, ZINC2
Reason or Justification for change:	No change intended.
Location of change:	ZINC, ZINC2
Effect of change:	In 1997, more households than expected claimed to have under $5,000 in income, compared with previous years of the AHS and with other data sources such as the Current Population Survey. Number of households at the lower end of the income spectrum is incorrectly high.
Is this change believed to be an improvement?	No.
What should be considered when comparing different years of data?	Remember that there may be a data issue with income on the lowest end of the spectrum and so be cautious about comparisons with income or income-related variables (e.g. ZSMHCP). There was also a universe change in 1997. Related adults who do not provide support financially for the household do not go through the regular family interview questions.
Is the change permanent?	Ideally no. The cause of this change in responses and differences from other surveys is unknown.

5.3.2. Non-Wage Income

Item	Description
1997 and later variable name:	VOTHER – Non-wage income
Pre-1997 variable name:	VOTHER
Reason or Justification for change:	Questionnaire error. VOTHER can be positive or negative. In certain special cases, zero was entered instead of the correct negative values.
Location of change:	Questionnaire, data.
Effect of change:	For an unknown number of cases, VOTHER is higher than it should be. This number of cases is believed to be small.
Is this change believed to be an improvement?	No.
What should be considered when comparing different years of data?	In 1997, the values for VOTHER may be slightly higher than is accurate.
Is the change permanent?	No.

5.3.3. Receive Farm Income in the last 12 months

Item	Description
1997 and later variable name:	CROPSL – Receive farm income in the last 12 months
Pre-1997 variable name:	CROPSL
Reason or Justification for change:	The 1997 codebook incorrectly lists this variable as INCFM.
Location of change:	Codebook
Effect of change:	None
Is this change believed to be an improvement?	Not applicable.
What should be considered when comparing different years of data?	Data should be comparable.
Is the change permanent?	There was no change. Codebook will be updated.

5.3.4. Mortgages

Item	Description
1997 and later variable name:	MG and every other variable in the Mortgage module of the questionnaire and the codebook
Pre-1997 variable name:	MORT and other variables on mortgages.
Reason or Justification for change:	In 1997, enhanced questions about mortgages were added to the AHS questionnaire. Information on up to four (4) mortgages, three (3) home equity loans, and one (1) reverse mortgage were put into the questionnaire. This contrasted with two (2) mortgages in the pre-1997 data. Note also that if a reverse mortgage was reported, other questions on mortgages were not asked. The rationale is that a household can't have both a traditional and a reverse mortgage at the same time.
Location of change:	Questionnaire, data, and codebook. These questions are in the AHSMORT.Q module of the Q-Code.
Effect of change:	Additional information on the financing of the properties collected. Also see the Home Equity Loans (page 40) for additional information on those loan types.
Is this change believed to be an improvement?	Yes.
What should be considered when comparing different years of data?	The data are largely comparable, but there is additional information available in the 1997 and later data.
Is the change permanent?	No. In 2001, additional changes will be made, further enhancing the data collection and data quality.

5.3.5. Home Equity Loans

Item	Description
1997 and later variable name:	HEL, HENUM, HETYP1, HECR1, HELMP1, HEBAL1, HEBAM1, HEINW1, HEINF1, HEPMT1 – variables all dealing with home equity loans. There are analogous variables for home equity loans number 2 and 3.
Pre-1997 variable name:	None.
Reason or Justification for change:	Data improvement, collection of additional information in the questionnaire.
Location of change:	Questionnaire and then the processing. These questions are in the AHSMORT.Q module of the Q-Code.
Effect of change:	Additional information available about the population.
Is this change believed to be an improvement?	Yes.
What should be considered when comparing different years of data?	The information is present only in 1997 and later. It is not present in earlier years.
Is the change permanent?	No. In 2001, additional changes will be made to further enhance the data quality and data collection. These changes will allow better distinction between the types of home equity loans, and loans versus lines of credit. The HEL, HENUM and HETYP1 variables will no longer exist. HELUMP and HELUMN will be new variables referring to home equity, lump-sum loans. HELC and HELCN will be new variables referring to home equity lines of credit.

5.3.6. Housing Costs – No Cash Rent Units

Item	Description
1997 and later variable name:	ZSMHC – Monthly Housing Cost
Pre-1997 variable name:	ZSMHC
Reason or Justification for change:	In 1997 for the first time, units reporting "no cash rent" still had a value calculated for them for total housing costs.
Location of change:	Processing.
Effect of change:	Fewer units reporting having no monthly housing costs, slight increases in the number of units with small monthly housing costs reported.
Is this change believed to be an improvement?	Yes.
What should be considered when comparing different years of data?	There will be slight differences, as described above, in the distribution of the data.
Is the change permanent?	Yes.

5.3.7. Housing Costs – Owner

Item	Description
1997 and later variable name:	ZSMHC – Monthly Housing costs
Pre-1997 variable name:	ZSHMC
Reason or Justification for change:	Addition in computation of housing costs for owners. The 1997 costs include costs on up to four mortgage products, more than before. This is due to the enhancements in data collection about mortgages. (See page 39.) Also see information on recodes for housing costs, page 41.
Location of change:	Printed reports
Effect of change:	Better capturing of housing costs for owners.
Is this change believed to be an improvement?	Yes.
What should be considered when comparing different years of data?	Costs for owners may be very slightly higher.
Is the change permanent?	Yes.

5.3.8. Housing Subsidy

Item	Description
1997 and later variable name:	SUBRNT, VCHER, PROJ – variables dealing with public or assisted housing.
Pre-1997 variable name:	SUB, SUBLOC, PROJ
Reason or Justification for change:	HUD and Census adjusted the housing subsidy questions in an attempt to improve accuracy of collected information. This change was based on previous HUD research that found significant overcounting. HUD made the change following research conducted by the National Opinion Research Center (NORC).
Location of change:	Questionnaire, data, codebook. These variables are in the AHSHCST2.Q module of the Q-code.
Effect of change:	Different questions led to changes in the data. As a result, data cannot be directly compared from year to year.
Is this change believed to be an improvement?	Yes. HUD though is also interested in testing if the change in questions have made an improvement.
What should be considered when comparing different years of data?	Different questions were used, meaning slightly different responses may have been elicited as a result.
Is the change permanent?	As of now, the change is permanent, but HUD is interested in exploring other ways to further improve this series of questions.

5.3.9. Heating Costs

Item	Description
1997 and later variable name:	BUYE, BUYF, BUYG, BUYO, BUYT, and BUYW
Pre-1997 variable name:	BUYE, BUYF, BUYG, BUYO, BUYT, and BUYW
Reason or Justification for change:	1997 Codebook is in error. Correct coding to use is the same as in 1995 and earlier. That coding is: 1 = Not used 2 = Included in rent, condo fee, or other fee 3 = Obtained free
Location of change:	Codebook, data.
Effect of change:	None
Is this change believed to be an improvement?	Not applicable.
What should be considered when comparing different years of data?	Data should be comparable.
Is the change permanent?	There was no change. Codebook will be updated.

5.3.10. Lodgers

Item	Description
1997 and later variable name:	LODSTAT or LODSTA1-LODSTA16 in flattened file format
Pre-1997 variable name:	LODGE1-LODGE8
Reason or Justification for change:	There was a change in the definition of lodgers. In 1997, a lodger could be a relative or a non-relative. Before 1997 and after, a lodger could be only a non-relative. In 1999, the data collection was further changed to ask more people this question. In 1997, the question was not asked of people younger than 16. In 1999, the question was not asked of individuals younger than 14.
Location of change:	Questionnaire skip patterns and processing of data.
Effect of change:	Count of lodgers is higher in 1997.
Is this change believed to be an improvement?	Uncertain.
What should be considered when comparing different years of data?	The number of lodgers in 1997 is larger because of the expanded universe of eligible people. This change is not present in 1999 and later.
Is the change permanent?	No. This change is only present in 1997.

5.3.11. Rent paid by lodgers

Item	Description
1997 and later variable name:	LODRNT (in person file) LODRNT1-LODRNT16 in flattened-file version.
Pre-1997 variable name:	RENT1-RENT16
Reason or Justification for change:	There was a change in computation for reporting rent paid by lodgers in the published report.
	Before 1997 in the published report, the rent paid by lodgers was the rent paid by the first lodger reported inside of the household. In 1997 and later, the rent paid by lodgers is the sum of the rent paid by all lodgers residing in the household.
	This change in computation better captures the total rent received by the household. The change only makes a difference when there are two or more lodgers residing in the household.
Location of change:	Processing and report.
Effect of change:	More accurate capturing in rent paid by lodgers to a household.
Is this change believed to be an improvement?	Yes.
What should be considered when comparing different years of data?	The 1997 and later figures will be higher and more accurate than prior years.
	Note that the figure is the total rent received by the household and not the amount of rent each lodger is paying.
	Note this is only in the published report. The data are unchanged.
Is the change permanent?	Yes.

5.3.12. Relationship codes

Item	Description
1997 and later variable name:	REL, or if using a flattened version of the data: REL1-REL16.
Pre-1997 variable name:	REL, REL2-REL16
Reason or Justification for change:	New coding structure used to describe the relationships.
Location of change:	Data and codebook.
Effect of change:	Different values are now used to describe the relationship possibilities. The coding in the codebook is incorrect. The correct coding is: 1 = Reference person with relatives in household 2 = Reference person without relatives in household 20 = Spouse (husband/wife) 22 = Child of reference person 23 = Grandchild of reference person 24 = Parent of reference person 25 = Sibling (brother/sister) of reference person 26 = Other relative of reference person 31 = Unmarried Partner (with own relatives) 32 = Unmarried Partner (without own relatives) 33 = House/Roommate (with own relatives) 34 = House/Roommate (without own relatives) 35 = Roomer/Boarder (with own relatives) 36 = Roomer/Boarder (without own relatives) 37 = Other Nonrelative (with own relatives) 38 = Other Nonrelative (without own relatives)
Is this change believed to be an improvement?	Yes.
What should be considered when comparing different years of data?	The coding of the variables are not directly comparable.
Is the change permanent?	Yes.

5.3.13. Food Stamps

Item	Description
1997 and later variable name:	QFS1 – Food Stamps
Pre-1997 variable name:	QFS1
Reason or Justification for change:	Data error in collection. Skip patterns in questionnaire were improperly set, and so not everyone was asked questions on food stamps correctly. Some were incorrectly not asked, while others were incorrectly asked.
Location of change:	Questionnaire, data. The error was in AHSINCOM.Q of the Q-Code.
Effect of change:	Data in 1997 is high.
Is this change believed to be an improvement?	No, data error.
What should be considered when comparing different years of data?	The totals reported in the 1997 tables are higher than other years.
Is the change permanent?	No, the problem in the questionnaire was fixed in 1999.

6. Discussion of Other Notes and Considerations for Users

This section discusses some changes in the data that are not due to the conversion of the AHS, but rather are naturally occurring changes in the data itself.

6.1. Changes in Opinion

There is a series of questions that asks the respondent to report their opinion about their neighborhood. Because these are opinions and not based on objective standards, they often change from year to year. Respondents' standards may change over time and their perceptions of the surrounding areas may change. In addition, there were minor changes in the wording and order of questions that may have affected the results.

The questions that ask for the residents' opinions are in the Neighbors module. These variables include:

- STRNA – street noise or heavy street traffic
- CRIMEA – neighborhood crime
- ODORA – smoke, gas, or bad smells
- STRNB – street noise or heavy street traffic bothers respondent
- STRNC – street noise or heavy street traffic bothers respondent enough to want to move
- CRIMEB – neighborhood crime bothers respondent
- CRIMEC – neighborhood crime bothers respondent enough to want to move
- ODORB – odors bothers respondent
- ODORC – odors bothers respondent enough to want to move
- NPROBS – no problems in neighborhood (see note below)
- NOPROB – no problems in neighborhood (see note below)
- NOISE – noise is a problem in the neighborhood
- LITTER – litter or housing deterioration is a problem in the neighborhood
- BADSRV – poor city/county services is a problem in the neighborhood
- BADPRP – undesirable commercial, institutional, or industrial property in neighborhood
- BADPER – people are a problem in the neighborhood
- OTHNHD – other problem in the neighborhood
- NEWTRN – public transportation is present in the neighborhood
- TRN – satisfactory public transportation in the neighborhood
- SATPOL – satisfactory police protection in the neighborhood
- SCH – elementary school in neighborhood is satisfactory

- SCHM – elementary school so unsatisfactory that respondent would consider moving from the neighborhood

Note on NPROBS versus NOPROB. NPROBS is a question that asks: "Is there something about the neighborhood that bothers you?" If yes, then the respondent is asked what bothers, and they have the option of saying "No problems" which is the NOPROB variable. Theoretically, there shouldn't be responses under NOPROB unless the person is being inconsistent.

When using the variables described above, caution must be used when comparing the results across years.

6.2. Normal Changes in the Nation

Certain changes present in the 1997 AHS were due to "normal" changes in the nation. These changes may occur because the country has evolved or because public knowledge has become better.

An example of a change in the nation is the increase in the number of people aware of the safety of their drinking water. Due to the EPA Safe Drinking Water Act as amended in 1996, water providers are required to annually produce a report for their consumers about the safety of their drinking water. The large increases in the numbers of respondents knowing about their drinking water may be due, in part, to the distribution of that report.

More common changes in the nation are not whole-scale changes, but rather shifts in values from one characteristic to another. For example, gradual upward shifts in household income occur because of inflation and real economic growth. To see specific table items that may have changed due to normal changes in the nation, please examine the "comparer results" output. The "comparer results" can be found in the AHS section of the HUD USER Web site.

6.3. Respondent Answers Instead of Interviewer Observation

Starting in 1997, all respondents were asked directly about the neighborhoods surrounding their housing units. Before 1997, Census Field Representatives completed this section when making a visit to conduct the interview or to update the address listings for multi-unit buildings. A visit was required to update the lists for multi-unit buildings. While interviews could be obtained by telephone for both multi-unit structures and single-unit structures, a visit to update listings was not required of the single-unit structures. Thus, single-unit structures were visited only when a phone interview was not possible, and so data were not always collected for single-unit structures. In previous reports, the information published was limited to multi-unit structures.

The change in data collection, as well as the data coverage improvement by collecting information for single-unit structures, led to shifts in the overall data reported. It is not known who is more accurate, the trained interviewer who is being consistent or the respondent who knows the neighborhood much better.

These variables include:

- EABAN – Abandoned or vandalized buildings within a half block
- EAGE – Unit about the same age as nearby buildings
- EAPTBL – Any apartment buildings (note: question new for 1997)
- EBAR – Any windows covered by metal bars
- EBARCL – Buildings with bars on windows within a half block
- EBOARD – Any boarded up windows
- EBROKE – Any broken windows
- ECOM1 – Any businesses or institutions, such as stores, restaurants, schools, or hospitals within half a block (note: question new for 1997)
- ECOM2 – Any factories or other industrial structures (note: question new for 1997)
- ECRUMB – Holes or open cracks or crumbling in the foundation
- EGOOD – No listed conditions seen in/on unit
- EGREEN – Open spaces, parks, woods, farms, or ranches within a half block
- EHIGH – Apartment buildings 7 or more stories tall within a half block
- EHOLER – Holes in roof
- EJUNK – Trash or junk in streets or properties within a half block
- ELOW1 – Single-family town/row houses within a half block
- ELOW2 – Apartment buildings less than 4 stories tall within a half block
- EMID – Apartment buildings 4-6 stories tall within a half block
- EMISSR – Missing roofing materials
- EMISSW – Outside walls have any missing siding, bricks, or other wall materials
- EMOBIL – Any mobile homes
- EPRKG – Any parking lots within a half block
- EPRKGA – Nearby parking lots usable by anyone
- EPRKGR – Nearby parking lots for residents only
- EPRKGS – Nearby parking lots for shoppers or workers
- EROAD – Condition of roads within a half block
- ESAGR – Roof's surface is uneven or sagging
- ESFD – Any single family homes
- ESLOPW – Outside walls slope, lean, buckle, or slant
- ETRANS – Railroads, airports, or highways with at least 4 lanes within a half block

- EVROD – Evidence of rodents in unit
- EWATER – Large bodies of water within a half block

7. Notes for Microdata Users

This section provides tips for potential users of the AHS microdata starting in 1997. This section will assist people who are trying to manipulate or use the AHS data in the distribution provided by the Census Bureau.

7.1. File Format

As discussed earlier (page 8), the structure of the AHS dataset was changed to take advantage of the new processing system and to reduce file sizes by eliminating "filler" variables (such as leaving room for more persons than exist in the given households). However, this change requires data users to refine their existing programs that analyze the AHS data.

A program to convert the 1997 data into a single file, similar to the earlier format, is distributed by HUD on the HUD USER Web site. This program was developed as part of the quality control process conducted by ICF Consulting. Further discussion of how this program works can be found in the Discussion of the File-Flattening Program located on page 55. The program was written for SAS, and as of now, there are no implementations in any other statistical packages. Users may prefer to keep the datasets in their supplied form and combine them as appropriate. The Discussion of the File-Flattening Program includes specific notes on how to combine the files and other cautions (page 57).

Census and HUD plan to add householder characteristics directly to the household dataset beginning with the 2001 survey and so the file structure will change slightly.

7.2. Missing Values in SAS

As mentioned in the Discussion of Intermediate Level Changes in the AHS (page 11), missing values are handled differently beginning in 1997.

As a default, SAS recognizes only one value as a missing value, namely the "**.**" (period) which is referred to as the "system missing" value. However, SAS actually has the ability to distinguish among more than 20 different missing values. The Census Bureau took advantage of this feature in the standards for data coding in 1997 and later years.

However, the other missing values coded in the AHS can cause error messages and warnings, because SAS recognizes them as character values in numeric fields unless SAS been told explicitly to expect them as missing values.

The command for SAS to treat the other values as missing is:

```
missing b d r;
```

That command should be present in DATA steps using the variables in computations. Not using the command will lead to a large series of errors about invalid numeric data.

For more information, please see the MISSING command in the SAS Language guide.

7.3. AHS Electronic Mailing List

One valuable resource about the AHS is the AHS e-mail discussion list. This list is frequented by both users of the AHS and the staff who created the AHS. The list serves as a public place for discussion and answers to questions. The list usually generates a low volume of messages.

Information about the mailing list can be found on the HUD USER Web site.

8. Discussion of the File-Flattening Program

8.1. Introduction

As noted previously, because Census completely replaced the processing system for the AHS in 1997, different file formats were designed and issued for public use by Census. The file-flattening program, however, enables users to convert 1997 AHS data into a format that is more similar to AHS data released between 1984 and 1996. By using the file-flattening program, users will be required to make fewer modifications to their existing programs developed for pre-1997 data. On the other hand, the data is converted into a format that is not in the exact format as described by the documentation for the 1997 and later AHS. In addition, the flattened version of the file has a much larger file size than is perhaps necessary.

This section describes the data format in general terms and how it can be combined. It is not necessary to collapse the file to use the data. However it may be more convenient to do so. This section may be of assistance to others who are considering using different statistical packages to analyze the AHS.

8.2. 1997 and Later Data Format

The 1997 and later AHS data format is analogous to a relational database. There is a collection of different files that can be joined based on different sets of merging variables. The merging may not be one-to-one merging but rather one-to-many. For example, for a household living in a housing unit, the (one) housing unit can be matched to the multiple (many) persons comprising the household.

The greatest challenge to working with a file formats such as this is ensuring that the information is merged or joined correctly. This section will describe the different files of data and what characteristics are necessary for merging properly.

The most important variable for merging is the CONTROL variable. This variable is the control code for the housing unit. CONTROL is a 12-digit variable unique to each housing unit. This variable is present in all of the files.

Repeating the files list from earlier, the files are as follows:

- HOMIMP – questions about upgrades and remodeling

- HOUSEHLD – household level information

- JTW – journey to work or commuting information

- MORTG – mortgage information

- OWNER – questions just for owners of rental properties

- PERSON – individual person level information

- RATIOV – verification of income to cost when the ratio of income to cost is outside of certain tolerances. (Note: This file is present beginning with the 1998 AHS metro sample. This file is not present in 1997.)

- RMOV – recent mover information

- TOPPUF – general unit characteristics and data recodes

- WEIGHT – weighting information

Each file has its own level of observation. The following table shows the file, level of observation, and the variables needed to merge on, and to what file it should be merged. The word "(multiple)" next to the merging variable indicates that there are multiple observations with the same values for the merging variable inside of a single household.

The column listing what file the data can be merged to should be used as guidance only. There are other valid possibilities, this is to merely give an idea. For example, TOPPUF and HOUSEHLD can be thought of interchangeably for merging purposes.

File Matching Key

File	Unit of observation	Merging variables	To what:
HOMIMP	home improvements, upgrades, and alternations	CONTROL (multiple)	TOPPUF
HOUSEHLD	household	CONTROL	TOPPUF
JTW	individual commuting patterns	CONTROL (multiple) PLINE	PERSON
MORTG	mortgages	CONTROL (multiple)	TOPPUF
OWNER	owners of rental properties	CONTROL	TOPPUF
PERSON	persons comprising household	1. CONTROL (multiple) 2. CONTROL (multiple) MOVGRP (multiple) 3. CONTROL (multiple) PLINE	1. TOPPUF, HOUSEHLD 2. RMOV 3. JTW
RATIOV	special cases, household level	CONTROL	TOPPUF
RMOV	Groups of persons inside of a household who are recent movers	CONTROL (multiple) MOVGRP (multiple)	PERSON

TOPPUF	housing unit	CONTROL	all other files
WEIGHT	Household	CONTROL	TOPPUF

8.3. Notes on Combining Data

When combining the data, care should be taken. The basic level of observation is generally the housing unit, and improper merging or improper use of weights will lead to inaccurate results.

8.3.1. Merging Variables

In order to keep all of the responses for a particular housing unit together, users must merge or join files using the variables described in the table above. Failure to use the correct merging variables will lead to incorrect merging and combining of data from different households or housing units together.

8.3.2. Weights

Weights are assigned only at the household level; there are no person level weights present in the AHS.

8.3.3. Householder

The householder is *not* necessarily the first member listed for the household or even the person with the lowest PLINE value. The householder is determined through the REL variable. REL values of 1 or 2 identify the householder.

8.3.4. Number of Observations per Housing Unit for a Module

Only the HOUSEHLD, TOPPUF, and WEIGHT modules have exactly one observation for every housing unit in the AHS. Other modules may have none, one, or more than one observation for each housing unit. Certain modules have fewer total observations, such as the HOMIMP module (because of the relatively few cases doing home improvements), some have more, such as PERSON module (because more than one person often lives in a household).

8.3.5. Notes on Merging or Joining

When combining data from the different modules, users must be careful to retain the data in a valid format. An overriding goal during the creation of the file-flattening program was to allow all the data to be analyzed at a housing unit level while still allowing the flexibility to examine the data at a greater level of detail, such as the person level.

8.3.6. Validating Combining of Files

Users are strongly encouraged to check their merging or joining of data by comparing their results against summary statistics generated with the public use file (PUF) which is available from the HUD USER Web site. Those summary statistics will match exactly. Users may also wish to compare against results published in the printed AHS report, but there is a chance that those will not match exactly due to the Census Bureau's use of internal data.

8.3.7. Merging and Missing Values

When combining data, users should be certain that they understand how the statistical package they are using treats situations where a one-to-one correspondence in observations for the two data sets does not exist. The basic issue is how the program carries values from one observation to the next. Different statistical packages handle this in different ways depending on the options used. Some statistical packages will carry the value through while others will create a missing value based on the merge.

For example, if the HOUSEHLD module has 1 observation and the PERSON module has more than 1 matching observation, does the merging technique apply the values from the HOUSEHLD set to every matching observation in the PERSON set, or just to the first matching person?

9. Discussion of Any Anticipated Changes in the AHS

The structure of the AHS is expected to stay relatively constant for the next few years. There may be minor enhancements to certain questions, but there are no plans for a major redesign like the one that took place with the 1997 AHS.

A new codebook for the 1997 and later AHS is currently being developed by ICF Consulting under contract to HUD. This codebook should be released to the public in early 2002.

10. Other Documents and Resources

The items below are all available from HUD USER. Most can be downloaded from the HUD USER Web site. Note that all URLs are correct as of the writing of this report, but they may change in the future.

HUD USER can be contacted through:

HUD USER
P.O. Box 6091
Rockville, MD 20849
1-800-245-2691
TDD: 1-800-483-2209
Fax: 1-301-519-5767
E-mail: `huduser@aspensys.com`

Web: `http://www.huduser.org/`

AHS Codebook – The basic codebook describing the file layout and data dictionary for the 1993 and earlier editions of the American Housing Survey. (There was a version for 1995 based on the 1993 codebook. The 1995 version was never released to the public however. There are no significant differences from the 1993 version.)

AHS Codebook, Volume 2 – A supplement to the basic codebook. This was published in 1995 and provides information on additional variables present in the 1995 AHS. These variables are primarily present in the remodeling module and a few others.

AHS Codebook, Volume 3 – This is a large-scale revision of the AHS codebook. This is the first codebook to deal with the 1997 and later data. This codebook was created using a computer program that analyzed the "Q-Code." It has not been updated when the Q-Code has been updated nor when the coding of variables has changed.

Appendices for the American Housing Survey – In each year, there are appendices published as a part of the printed American Housing Survey reports. Some of these appendices are augmented each year, while others are completely replaced. Users should refer to the appendices for the particular year they are examining. The published reports are in the H150 series of publications from Census. The 1997 report can be found on the Web at: `http://www.census.gov/prod/99pubs/h150-97.pdf`. The appendices are:

Appendix A – This appendix provides the definitions of items that are present in the AHS.

Appendix B – This appendix provides a description of the sampling and weighting that is used in the AHS.

Appendix C – This appendix provides a listing of historical changes to the AHS as well as some known data errors in previous years. The appendix covers the surveys up to and including the 1999 AHS survey.

Appendix D – This appendix to the AHS describes possible errors in the data and how to compute standard errors.

File-flattener program – A program developed by ICF Consulting to put the 1997 AHS in a similar file format as previous years. It is available as part of the distribution of the AHS on the HUD User Web site.

Comparer Results – The output of a program written by ICF Consulting that performs a line by line comparison of 1993 and 1995 tables with 1997 tables. With each comparison are identifications of the differences if known. The results are available on the HUD User Web site at `http://www.huduser.org/datasets/ahs/ahsdata97.html` under "Comparison with 1993 and 1995 data."

Q-Code – Short for "Questionnaire Code." This is the actual computer source code used to control the interviewer's machine during the interview. This code contains all of the information on the questions and skip patterns in the AHS. It is not easy to read, and there is no public documentation available. It is available from the HUD User Web site at `http://www.huduser.org/datasets/ahs/ahsprev.html`. The Census Bureau has made available a more readable version of the questionnaire process by putting their field manual for interviews up on the Web. This can be found at `http://www.census.gov/hhes/www/housing/ahs/over.html` and is Part D of the Field Representative's manual.

Income Limits – Shortly after the release of each public use file, HUD releases a corresponding dataset with estimates of area median incomes, fair market rents, poverty incomes, and very-low income cut-offs for each record in the AHS. Analysts may find this file useful for producing research consistent with official definitions. This data can be downloaded from HUD User.

11. Appendix: Glossary

AHS – American Housing Survey.

Census – U.S. Census Bureau, Department of Commerce. The main Census Web site is `http://www.census.gov/` The page focusing on the American Housing Survey can be found at `http://www.census.gov/hhes/www/ahs.html`.

Current Population Survey (CPS) – Another survey conducted by the Census Bureau tracking households, population, income, and other issues.

FERRET – A tabulation and data extraction system present on the Census Bureau Web site that allows extraction of variables from the data. This system works with 1997 and later data. 1993 and 1995 AHS data can be extracted using the Data Extract System (DES). FERRET can be found at: `http://ferret.bls.census.gov/cgi-bin/ferret` In addition to the AHS, FERRET works with a series of other data products.

HUD – U.S. Department of Housing and Urban Development. `http://www.hud.gov/`.

HUD USER – The dissemination arm for HUD's Office of Policy Development and Research. HUD USER is also a repository of HUD's public data sources. `http://www.huduser.org/`.

ICF – ICF Consulting. A private consulting firm that assisted HUD and Census with the quality control of the AHS and also developed this document. `http://www.icfconsulting.com/`.

PUF – Public Use File. The data file that the Census Bureau releases for public use. The file omits information that could be used to identify the respondents.

SAS – A statistical analysis software developed by the SAS Institute. `http://www.sas.com/`.

Skip patterns – The rules for what questions to ask and not ask the respondents. Not all questions are relevant to all respondents. For example, the skip patterns would avoid asking renters if they have mortgages.